Non-canonical Nucleosides and Proto-Urea-RNA at the Chemical Origins of Life

Dissertation der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

Antony Crisp

München, 2021

Diese Dissertation wurde angefertigt
unter der Leitung von Prof. Dr. Thomas Carell
im Bereich von der Fakultät für Chemie und Pharmazie
an der Ludwig-Maximilians-Universität München

Erstgutachter/in: Prof. Dr. Kirsten Jung (Doktormutter)

Zweitgutachter/in: Prof. Dr. Thomas Carell (Betreuer der Arbeit)

Tag der Abgabe: 25.02.2021

Tag der Mündliche Prüfung: 26.05.2021

TABLE OF CONTENTS

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

Die vorliegende Dissertation wurde weder ganz, noch teilweise bei einer anderen Prüfungskommission vorgelegt.

Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen.

München, den 17.06.2021 Antony Crisp

Antony Crisp

PUBLICATIONS LIST

Publications submitted towards this cumulative dissertation:

11. Nainyté, M., Müller, F., Ganazzoli, G., Chan, C., Crisp, A., Globisch, D., Carell, T. – Amino Acid Modified RNA Bases as Building Blocks of an Early Earth RNA-Peptide World. ***Chem. Eur. J.***, 26: pp. 14856-14860 (2020).

*(highlighted in **Nature News**)*

10. Becker, S., Feldman, J., Wiedemann, S., Okamura, H., Schneider, C., Iwan, K., Crisp, A., Rossa, M., Amatov, T., Carell, T. – Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. ***Science,*** 366: pp. 76-82 (2019)

*("Hot Paper"; highlighted in **ChemistryViews** Magazine)*

9. Okamura, H.*, Crisp, A.*, Hübner, S., Becker, S., Rovó, P., Carell, T. – proto-Urea-RNA (Wöhler RNA) Containing Unusually Stable Urea Nucleosides. ***Angew. Chem. Int. Ed.***, 58: pp. 18691-18696 (2019).
**authors contributed equally*

8. Schneider, C., Becker, S., Okamura, H., Crisp, A., Amatov, T., Stadlmeier, M., Carell, T. – Noncanonical RNA Nucleosides as Molecular Fossils of an Early Earth—Generation by Prebiotic Methylations and Carbamoylations. ***Angew. Chem. Int. Ed.***, 57: pp. 5943-5946 (2018).

7. Becker, S., Schneider, C., Okamura, H., Crisp, A., Amatov, T., Dejmek, M., Carell, T. – Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. ***Nature. Commun.***, 9: 163 (2018).

Other publications resulting from this doctoral work:

6. Lauria, T., Slator, C., McKee, V., Müller, M., Stazzoni, S., Crisp, A., Carell, T., Kellett, A. – A Click Chemistry Approach to Developing Molecularly Target DNA Scissors. ***Chem. Eur. J.***, 26: pp. 16782-16792 (2020).

5. Crisp, A. – Purines and Pyrimidines: Structure out of Soup. ***Nature Chemistry Community, Behind The Paper*** (2019).

4. Becker, S., Schneider, C., Crisp, A., Carell, T. – Non-canonical nucleosides and chemistry of the emergence of life. ***Nature. Commun.***, 9: 5174 (2018).

3. Crisp, A., Carell, T. – In vitro evolution: Rethinking the tools of the RNA World. ***eLife.*** 7:e38297 (2018).

Other publications:

(featured on the front cover)

2. Crisp, A. L., Noble, B., Schwartz, B. D., Willis, A. C., Coote, M. L., Banwell, M. G. – The Synthesis, Structural Characterisation and Chemical Manipulation of the [6+3]Cycloadduct Derived from α-Tropolone O-Methyl Ether and Trimethylenemethane. ***Asian J. Org. Chem.***, 8: pp. 1458-1467 (2019).

1. Crisp, A. L., Li, J., Nugent, J., Matoušová, E., and Banwell, M. G. – The Palladium-Catalysed Intramolecular Alder-ene (IMAE) Reactions of Certain Heteroatom-linked 1,6-Enynes: The Formation of Hexahydro-indoles and -benzofurans. ***Aust. J. Chem.***, 68: pp. 1183-1189 (2015).

Awards and Public Presentations during this PhD:

2019 ***1st prize*** *– Best research lecture. Life Science Munich Retreat, Studienzentrum Josefstal. "Proto-urea-RNA and the chemical origins of life".*

2019 *Poster Presentation. Science of Early Life conference, Kloster Seeon, "Proto-urea-RNA and the chemical origins of life".*

2018 ***1st prize*** *– Best poster presentation. Life Science Munich Retreat, Haus Bergsee. "Studies towards the chemical origins of RNA and polypeptides"*

2017 *Poster presentation. SFB 749 Workshop, Kloster Irsee, "Prebiotic origins of non-canonical RNA bases and polypeptides".*

ACKNOWLEDGEMENTS

In fulfilling the requirements of this doctoral dissertation, a number of individuals have continually provided me with their support and guidance, whom I would like to sincerely thank. First and foremost, I wish to thank my fantastic supervisor Prof Dr Thomas Carell for his guidance, motivation, and for offering me a place within his group. His enthusiasm in all aspects of chemistry have continually inspired me have undoubtedly been what enabled me to have a rewarding and successful PhD experience. I would also like to thank Prof Dr Kirsten Jung for taking me on as an LSM PhD candidate and for providing me with helpful feedback throughout my research endeavours. I am also very grateful to Prof Dr Oliver Trapp, Prof Dr Hoffmann-Röder, and Dr Sabine Schneider for attending my thesis advisory committee meetings and for their valuable contributions toward this research.

Next, I wish to express my gratitude to Dr Markus Müller for his support throughout my degree and for his tireless work managing many aspects of the lab, while helping to develop my projects. Markus displays an extraordinary commitment to the group, and I am tremendously grateful for his help. I also owe my gratitude to Dr Hidenori Okamura, Dr Nobuhiro Tago, Dr. Iacovos Michaelides and Dr Sidney Becker, who trained me in a number of aspects within the lab and have furthered my professional development. Next, I would like to thank Dr. Petra Rovó for lending her expertise in NMR and for the fantastic research collaboration. I would like to thank Sarah Hübner, Marc Nieveler, Kristof Möller, Felix Schäfer, Adriana Neves Vieira and Edmire Shala, all of who were excellent students and who were a pleasure to work together with. I also wish to thank Jonas Feldman for providing me with helpful feedback on the written content of my dissertation. My gratitude is further extended to the past and current members of the Carell group for providing me with their help and continued friendship.

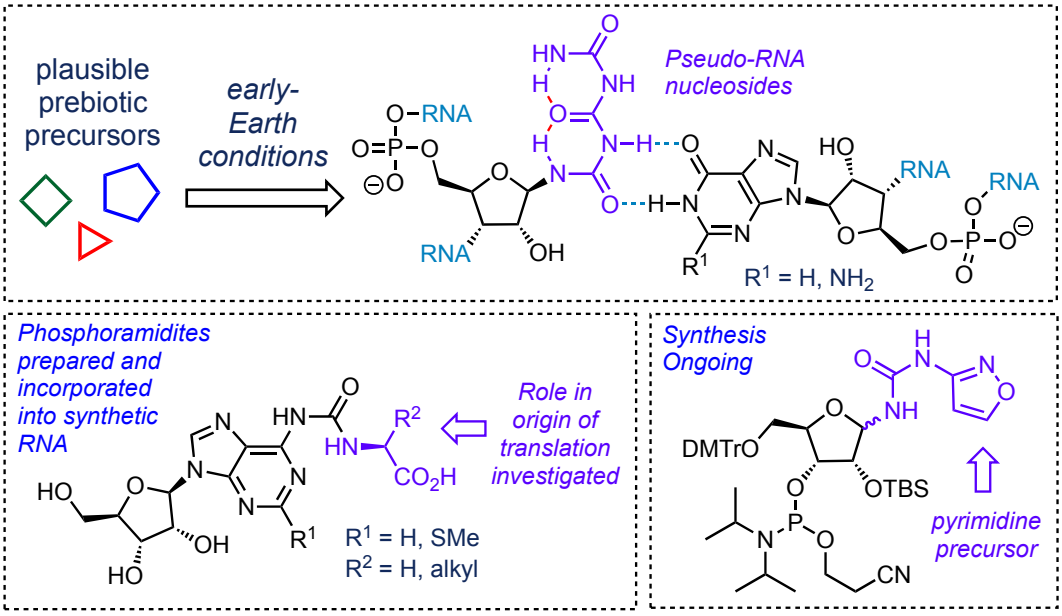
I wish to thank the LMU Department of Chemistry professional staff for their time and use of their facilities during the preparation of my dissertation. In particular, I must express my gratitude to Carrie Louis for her assistance with NMR and for her friendly encouragement. I am also very grateful to Francisca Rosa Mende and Nadine Hamze for supporting my academic progress and providing me with guidance during my studies. I would also like to thank my colleagues Martin, Annika, René, Edris, Giacomo and Philipp from lab F4.001b with whom I worked together daily and who made this PhD a joyful experience. Finally, I wish to thank Mum, Dad, Pete and Clara for their continued care, encouragement and support during my time at the LMU.

ABBREVIATIONS

%	percentage yield	M^+	molecular ion
μ	micro	MALDI	Matrix-assisted laser desorption/ionization
δ	chemical shift		
ν_{max}	absorption maxima (IR)	Me	methyl
A	adenine	mg	milligram/s
Ac	acetyl	MHz	megaHertz
aq.	aqueous	min	minute/s
C	cytosine	mol	mole
COSY	correlated spectroscopy	mmol	millimole
d	doublet/s	m.p.	melting point
DTBS	di-*tert*-butylsilylene	MS	mass spectroscopy
Et	ethyl	*m/z*	mass to charge ratio
ESI	electrospray ionisation	NMR	nuclear magnetic resonance
FaPy	formamidopyrimidine	NOE	nuclear Overhauser effect
g	gram/s	NOESY	nuclear Overhauser and exchange spectroscopy
G	guanine		
Gya	billion years ago	NPE	*p*-nitrophenethyl
g^6A	N^6-glycinylcarbamoyl adenosine	pK_a	acid dissociation constant
		Ph	phenyl
HPLC	high-performance liquid chromatography	ppm	parts per million
		PRPP	5'-phosphoribosyl-1'-pyrophosphate
hr	hour/s		
HRMS	high resolution mass spectrometry	q	quartet
		R_f	retardation factor
HSQC	heteronuclear single quantum coherence	sat.	saturated
		TBSCl	*tert*-butyldimethylsilyl chloride
Hz	Hertz		
IR	infrared	TLC	thin layer chromatography
J	coupling constant	t^6A	N^6-threonylcarbamoyl adenosine
L	litre/s		
m	multiplet	U	uracil
M	molar		

ABSTRACT

The work presented in this dissertation was aimed towards furthering our understanding of the chemical processes that gave rise to the emergence of life on Earth. Specifically, we provide new evidence to support the RNA World Hypothesis, which suggests that the first organisms were preceded by an era in which RNA molecules catalysed their own autonomous replication. To this end, the syntheses of certain prebiotically plausible pseudo-RNA nucleosides derived from Biuret and Triuret are reported. Remarkably, in addition to being hydrolytically stable, these planar H-bonded pseudobases exhibit unique base-pairing properties that mirror that of a UG wobble-base pair when coupled with guanine or inosine within an RNA duplex. Possible mechanisms for an evolutionary transition from such pseudobases to the canonical pyrimidines are discussed. In pursuit of further proto-RNA structures, prebiotically plausible routes to a number of extant methylated nucleobases, as well as the carbamoylated adenosine nucleosides N^6-threonylcarbamoyl adenosine (t^6A) and N^6-glycinylcarbamoyl adenosine (g^6A) are described. These nucleosides are universally conserved and present within the genome of the last universal common ancestor (LUCA), which together with their ease of synthesis, suggests that they might endure today as molecular fossils from our RNA-based progenitors. Preparation of the first nucleoside phosphoramidites for g^6A and related structures were also established, thus allowing their incorporation into RNA strands and investigation of their theorised role in the origin of translation. Finally, a novel prebiotically plausible route to pyrimidine nucleosides is reported, as well as encouraging preliminary results towards the synthesis of a phosphoramidite building block of the 1-(isoxazol-3-yl)-3-ribosylurea precursor from that pathway.



1 INTRODUCTION

1.1 MINIMAL LIFE ON EARTH

The origin of life remains undoubtedly one of the greatest unsolved mysteries of our time.[1] Extant life is observable in nearly every conceivable niche, from deep sea vent-dwelling chemotropes[2] and simple cells buried as deep as 11km beneath the central west Pacific,[3] to hyperthermophiles capable of withstanding extreme acidities, temperatures, and pressures.[4] Though it may be impossible to delineate exactly when or how life began, putative fossilised microorganisms have been described that could be as old as 3.8 billion years.[5] Trapped inside these deposits are the tiniest specks of carbon bearing the unmistakable atomic and isotopic footprint of life. If this evidence were to be true, then life has surely endured for most of Earth's geological history (**Figure 1**).[6] The Earth itself is thought to have formed approximately 4.56 Gya (billion years ago) through a process known as runaway accretion, in which tiny particles slowly accumulated until they eventually formed a terrestrial body.[7] During its first, and appropriately-named *Hadean* Eon (4.54 – 4.00 Gya), the Earth appeared uninhabitable, blanketed by molten rock[8] and likely also disturbed by an enormous impact giving rise to the Moon.[9] Based upon what we currently know of the requirements needed to sustain life, factors such as a suitable atmosphere to retain heat, availability of liquid water, and moderate surface temperatures would first need to be met.[10] One may therefore ask, what exactly was it about the early Earth that breathed life into inorganic matter?[11]

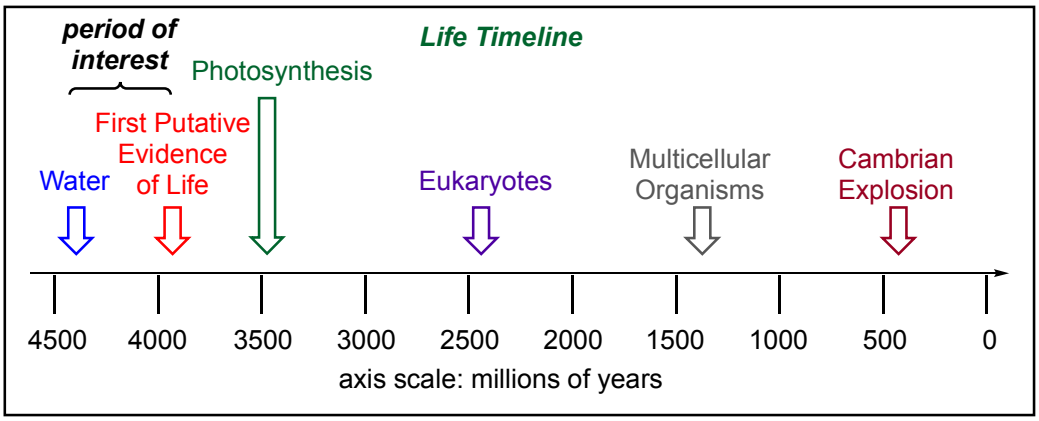


Figure 1 *A simplified timeline of the evolutionary history of life on Earth, outlining certain major developmental events that occurred in the last 4.5 billion years.*[12–14]

If life indeed formed on the Archean earth (4.03 - 2.40 billion years ago), we can assume that it did so in a chemically active, magmatic environment, with the frequent exhalations of water vapour, methane, carbon dioxide, and sulphur.[10] Though it is hotly debated, current evidence suggests that

the early Earth's atmosphere was not oxidising, but rather, weakly reducing.[15] It is through this turbulent geochemical lens that the organic chemist must envisage plausible origin-of-life scenarios. One should note that we can never really discern *exactly* how life began on our planet. The goal of the prebiotic chemist is therefore not to delineate the exact pathway through which natural life emerged, but instead to more broadly demonstrate that life's fundamental aspects can arise through the organisation of organic matter.[16] Thanks to decades of such advances, we now know of simple chemical reactions that make many of life's 'ingredients'.[1] The conditions may have needed to be just right, but as I hope to convince the reader of this dissertation, the early Earth was the perfect stage for diverse chemistry to happen. In the words of Nick Lane, "*if the probability of life in the universe is one in a million billion, then in a million billion planets there is a chance approaching 'one' that life should emerge"*.[6] Evidently, we live on such a planet.

1.2 GIVING 'LIFE' A DEFINITION

The immense challenge in addressing life's origin is compounded by the fact that there is currently no clear consensus on what is or isn't '*living*'.[17,18] In 1994, an expert panel organised by NASA proposed a consensus definition for life: a "*self-sustaining chemical system capable of Darwinian Evolution"*.[19] While this definition has become increasingly accepted among origins-of-life scientists, it has received ample criticism towards the usage of '*Darwinian evolution'*. There are, for example, sterile living organisms that are incapable of reproducing (nor of Darwinian evolution) and would not fulfil the criteria of NASA's definition.[18] It is nonetheless evident that a grey area exists between 'dead' chemicals and 'living' biology. Some experts, such as Nobel Laureate Jack Szostak, have argued that we should not give life a definition, as it distracts us from more pressing questions such as how organic matter transitioned to give biology.[20] Albert Eschenmoser noted that researchers have tended to fall into one of three groups: those who emphasise genetics and heritability; cellular compartmentalisation; or metabolism, depending on which aspect they consider to be most important in their definition of life.[16] Were a researcher to take up the challenge of designing an experimental model for life's origin, then clearly all three aspects would need to be taken into consideration. For the sake of unambiguity, all references to 'life' throughout this dissertation will correspond specifically to the definition given by NASA in 1994.

1.3 THE 'RNA WORLD' HYPOTHESIS

If we accept 'life' to be capable of replication and evolution, then the first lifeforms must have had some means of encoding information. In modern cells, the storage, function and proliferation of genetic information is orchestrated through a web of highly coordinated and optimised chemistry

performed by DNA, RNA, and protein molecules (**Figure 2a**). This concept has become embedded in modern science as the '*central dogma of molecular biology*', and was characterised by Francis Crick in response to the discovery that RNA can be reverse-transcribed.[21] After half a century of progress, the assumption that information flows irreversibly from RNA to proteins still holds up.[22] When it comes to studying the origins of life, however, we begin to encounter certain complications. DNA functions to store our genetic information, a job that it does exceptionally well. That is, however, about *all* that DNA can do. Proteins, on the other hand are needed to keep cells alive by acting as enzymes and catalysing chemical reactions. Without the genetic code, proteins would not exist, and without proteins, DNA could not survive. This poses a chicken/egg problem of what came first, proteins or DNA? The solution is perhaps very simple: That early life used RNA as both genetic material and as a catalyst (**Figure 2b**). As it turns out, RNA can catalyse a plethora of chemical reactions by varying mechanisms.[23] One can therefore imagine that if RNA molecules were once able to self-replicate, they could have even undergone (upon exposure to a section pressure) Darwinian evolution. This concept, now referred to as the *'RNA World Hypothesis'*,[24] was first explored in 1968 by Crick & Orgel[25,26] and is at present the most widely accepted hypothesis for the origins of biological evolution. The *RNA World Hypothesis* became further cemented in the early 1980s when Altman & Cech independently discovered the first catalytic RNA molecules, otherwise known as ribozymes.[27,28]

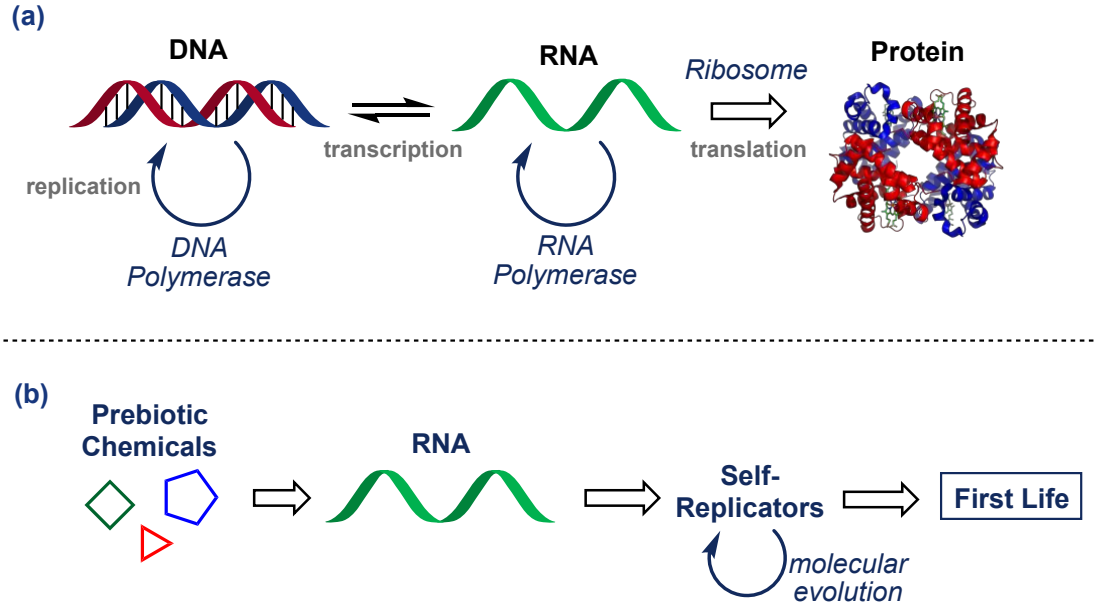
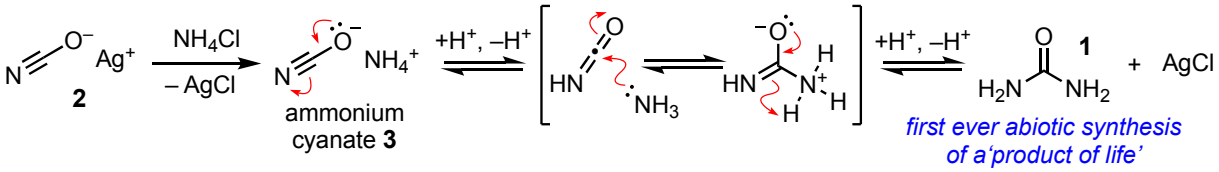


Figure 2 (a) *Schematic representation of the 'central dogma of molecular biology', describing the three-step process (replication, transcription, and translation) through which information is copied flows from genes to proteins.* **(b)** *Depiction of the 'RNA World Hypothesis', a hypothetical stage of evolutionary history in which RNA molecules proliferated before genetic information came to be stored as DNA. Image created using existing artwork*[29] *with adaptations (CC BY-SA 3.0).*

The best evidence for the *RNA World* can be observed in the various roles of RNA in modern cells. The ribosome, the large molecular machine that drives protein synthesis, is for example, in fact a ribozyme. Although no ancient ribozyme with intact RNA copying machinery has yet been discovered, evidence for such machinery can be found within the small ribosomal subunit.[30] The 23S component of ribosomal RNA appears to be directly involved in peptidyl transferase activity, strongly implying primitive peptide-bond catalysis by RNA during the earliest stages of evolution.[31] It has also been long-speculated that the cyclic nitrogenous bases found in modern-day enzyme-coenzyme complexes might be vestiges of an early genetic system.[32–36] Despite this and other strong evidence, many aspects of the RNA World remain uncertain. It is also unclear exactly how RNA itself could have spontaneously arisen on the early Earth. For this to be possible, a transition must have occurred from small inorganic molecules, to complex and highly-ordered organic scaffolds. The main body of work in this dissertation aims to experimentally demonstrate how certain aspects of such a transition could have occurred.

1.4 ORGANIC FROM INORGANIC

As recently as 250 years ago, chemists held the somewhat paradoxical view that the discipline ought to be strictly divided into segments that involved either living organisms or lifeless-matter. Along this line of thinking, Jacob Berzelius combined plant- and animal chemistry and named the joint discipline "organic chemistry", which he distinguished from "inorganic chemistry".[37] Berzelius then formulated what became known as his *Central Dogma of Chemistry*, where he stated that the generation of organic compounds from inorganic ones was physically impossible. Louis Pasteur meanwhile provided definitive proof that "life comes only from life" - at least based upon the prevailing biological conditions on Earth. As is often the case however, dogmatic views tend to be superseded. Thus, the idea that life could be built out of something inorganic first became widely accepted in 1828, when chemist Friedrich Wöhler presented his synthesis of urea **1** from silver cyanate **2** in the presence of ammonium chloride (**Scheme 1**).[38] To Wöhler's astonishment, the mixture gave rise to crystalline urea **1** rather than the expected ionic compound **3**. Although the mechanistic details of this synthesis are not fully understood, the observed reactivity suggests an initial rearrangement of ammonium cyanate **3** *via* proton transfer, followed by nucleophilic attack at the alpha-carbon and subsequent tautomerisation.[39]



Scheme 1 *Wöhler's revolutionary synthesis of urea **1** from the elements of ammonia and cyanate.*

For the first time ever, a product of animal metabolism (urea **1**) had been synthesised in a lab entirely without the help of biology. Wöhler's legacy was immense and kindled the birth of modern organic chemistry. The discovery was also a stark contradiction to the 'vitalist' doctrine at the time, which stated that 'products of life' could only be made by living organisms. In an instant, these doctrines were shattered. The feat of mimicking nature was truly exciting, as Wöhler expressed in a letter to Berzelius: "*I can no longer, so to speak, hold my chemical water and must tell you that I can make urea without needing a kidney, whether of man or dog; the ammonium salt of cyanic acid is urea*".[40] The field of prebiotic chemistry was partly birthed from the pioneering viewpoints of Alexander Oparin, who in 1924, noted that '*there is no fundamental difference between a living organism and lifeless matter*'.[41] Oparin asserted that Earth likely possessed a reducing atmosphere composed of water, methane, ammonia, and hydrogen, a combination which he argued were the raw materials needed to kick-start life. Oparin also emphasised that Darwin's theory of natural selection should apply equally to inanimate molecules as it does to living things. Together with the ideas of John Haldane, these concepts became known as the Oparin-Haldane hypothesis for chemical evolution.[42]

Perhaps the best-known example of prebiotic chemistry and a direct test of Oparin's hypothesis, came in the form of the Stanley Miller and Harold Urey's pioneering electrical discharge experiments, which demonstrated the synthesis of amino acids from a crude mixture of gases.[43] Although chemists had certainly explored prebiotic syntheses before, what distinguished the Miller-Urey experiments, were that they were designed to test a specific hypothesis about the origins of life. In the experimental design, Miller and Urey gave careful consideration to what would constitute prebiotically-plausible atmospheric conditions and temperatures. The experiment itself consisted of a glass apparatus with two connected flasks – one containing a mixture of gases representing the primordial atmosphere, and another containing an aqueous solution representing the primitive ocean (**Figure 3**). Electrodes supplying a high-energy electrical input were applied to the mixture of gasses, which were circulated through the system *via* the refluxing solution. After several days, Miller observed the formation of a yellow organic residue, which was extracted and subjected to paper chromatography, revealing the presence of 5-different amino acids. Recent re-analyses (in 2008) of the mixtures obtained by Miller revealed the formation of many more compounds, including almost all of the proteinogenic amino acids.[44,45] The Miller-Urey experiment was also momentous given how it inspired the next generation of prebiotic chemists. Encouraged by Miller and Urey, others have since carried out similar experiments demonstrating, for example, that hydrogen cyanide **4**,[46] cyanoacetylene **5**,[47] and cyanogen[48] can be produced from an atmosphere composed of methane, nitrogen and carbon monoxide.

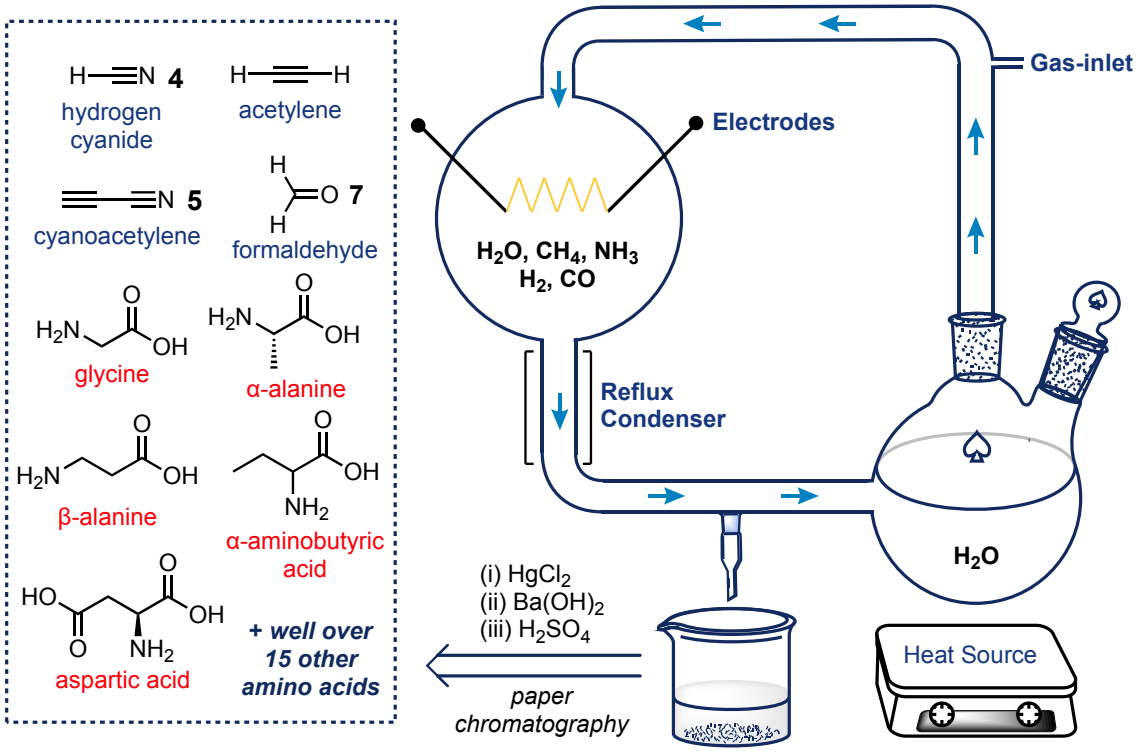
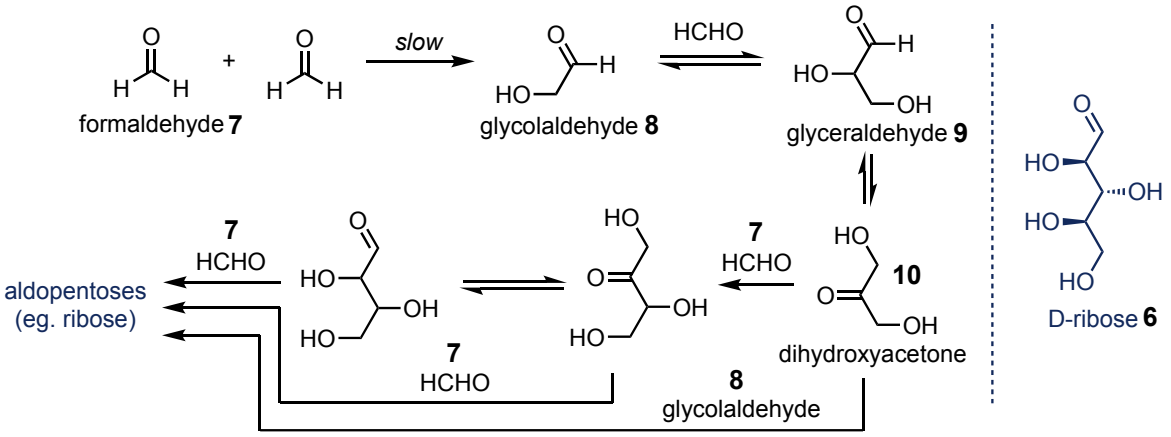


Figure 3 *The spark-discharge apparatus used in the 1952 Miller-Urey experiment (or Miller experiment), which aimed to produce organic molecules by replicating primitive atmospheric conditions of the early Earth.*[43] *Compounds labelled in red were determined by Miller and those labelled in blue were discovered or inferred later.*

Beyond the synthesis of small molecules like urea **1** or amino acids, 'RNA World' hypothesis poses a greater challenge for chemists to elucidate plausible abiotic pathways to the components that make up RNA. Among the most daunting aspects of such an endeavour would be the synthesis of the sugar D-ribose **6**. Aside from the apparent regiochemical issues associated with synthesis of carbohydrates, ribose **6** is pentose sugar with four linear-, open-chain configurational isomers, each with their own set of two enantiomers. This fact alone suggests that the probability of *neat*-ribose emerging spontaneously on the early earth must have been incredibly low. Extraordinarily, Aleksandr Butlerov had already begun to address these obstacles as early as the mid-19th century with his discovery of the formose reaction (**Scheme 2**).[49] In its original version, the formose reaction involved the polymerisation of formaldehyde in the presence of calcium hydroxide to give a complex mixture of different carbohydrates. Major by-products from this process include formic acid and methanol, which are produced *via* the Cannizaro reaction.[50]

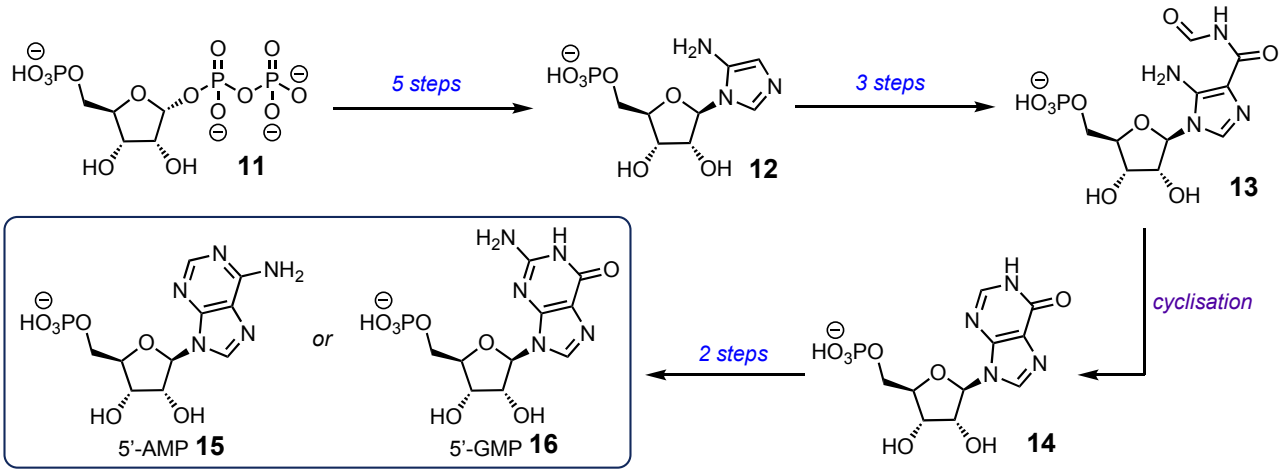


Scheme 2 *The formose reaction (discovered in 1861 by Aleksandr Butlerov) generates glycolaldehyde **8** via the coupling of two molecules of formaldehyde 7, which can subsequently provide sugars through a series of intermediary Aldol, reverse Aldol, and Lobry de Bruyn–Alberda–van Ekenstein reactions*.[49,51]

The exact mechanism of the first step of the formose reaction is still not fully understood.[52] The most widely accepted model involves an initial mechanistically unknown slow dimerization of formaldehyde **7** to afford glycolaldehyde **8**, which can subsequently undergo successive Aldol reactions to give carbohydrates such as glyceraldehyde **9**.[53] Glyceraldehyde **9** can then isomerise to give dihydroxyacetone **10** and subsequently, different tetroses. Larger, linear and branched sugars including ribose are then produced *via* further Aldol reactions and rearrangements involving **7**, **8** and **9**. After nearly two centuries, the formose reaction is remarkably still considered to be the main prebiotically plausible route to ribose **6** and other significant carbohydrates on the early earth.[53] Very recently, for example, Trapp and co-workers were able to show that pentose monosaccharides can be generated under dry, mechanochemical conditions in the presence of a multitude of different mineral catalysts.[54] The group even demonstrated that when coupled with canonical nucleobases as nucleophilic catalysts, the reaction of acetaldehyde with glyceraldehyde **9** can selectively provide the full set of DNA nucleosides at very mild temperatures, thus further exhibiting the synthetic utility of the formose process.[55] The challenge of ribonucleoside synthesis is, however, twofold; abiotic routes are required not only for ribose **6**, but also for sufficiently nucleophilic nucleobases or their precursors. In this work, I will focus mainly on the second point, as well as the roles that nucleosides could have played in the RNA World.

1.5 PREBIOTIC AND DE NOVO SYNTHESES OF RIBONUCLEOSIDES

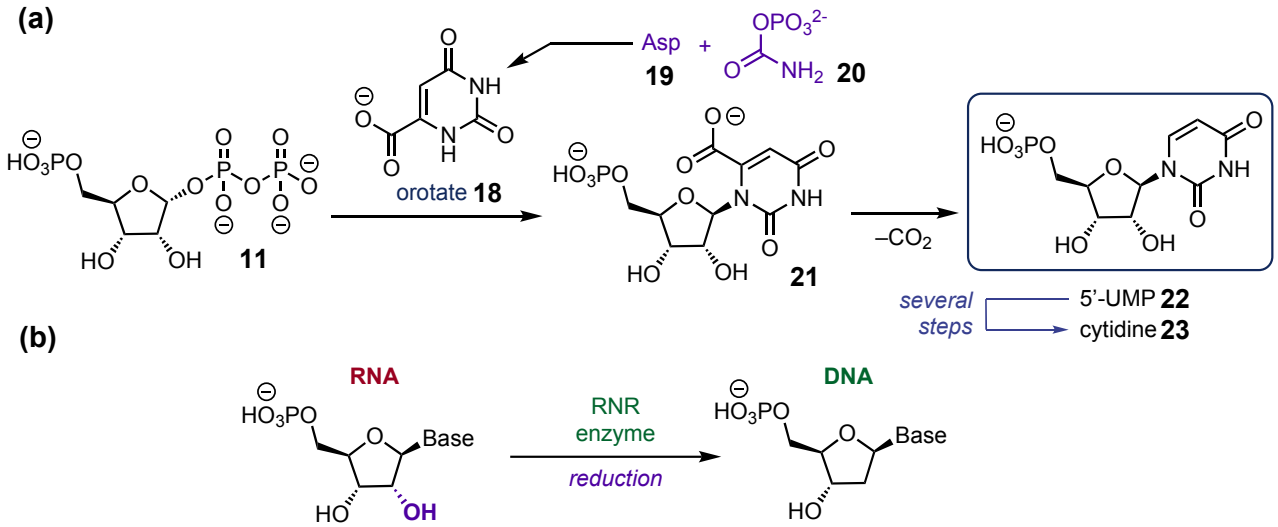
Ribonucleosides are the key building blocks used by all organisms to make the essential biopolymers responsible for life's biochemistry, DNA and RNA. Thanks to their association with the RNA world, ribonucleosides now occupy a central focus in the endeavours of scientists to understand the origins of life. When considering the *de novo* origins of such molecules from a prebiotic perspective, however, we are confronted by several contradictions. Why, for example, does extant biology require so many discrete steps to synthesise ribonucleotides if they were indeed among life's earliest ingredients?[53] In the *de novo* synthesis of purines **(Scheme 3)**, the nucleobases are assembled nearly one atom at-a-time on the pre-attached ribose moiety. This pathway firstly involves amination of 5'-phosphoribosyl-1'-pyrophosphate **11** in the presence of glutamine to give a ribosyl amine, which is then elaborated in five-steps to give **12**. Further oxidation, amination, and formylation of **12** gives **13**, which finally undergoes cyclisation to give the purine inosine-5'-monophosphate **14**. Even more steps are then required to afford 5'-AMP **15** and 5'-GMP **16**. At first glance, such a lengthy and complex process seems outwardly impossible without the aid of enzymatic catalysis, or otherwise complex multistep synthesis in a laboratory.



Scheme 3 *Biosynthetic pathways leading to the canonical purine RNA ribonucleotides. The construction of purines involves the stepwise assembly of a heterocyclic moiety already attached to the sugar (ribose-5'-monophosphate).*[53]

Further puzzling is the stark contrast between the lengthy *de novo* purine biosynthesis and the convergent pyrimidine biosynthetic pathway. For the synthesis of the pyrimidine nucleoside uridine **17 (Scheme 4a)**, the nucleobase orotate **18** is first assembled from the elements of aspartate **19** and carbamoyl phosphate **20**. 5′-phosphoribosyl-1′-pyrophosphate **11** is then substituted with **18** at the 1'-position to give orotidine monophosphate **21**, which is enzymatically decarboxylated to yield uridine-5′-phosphate **22** (UMP). **22** is finally converted into cytidine **23** by a series of functional-group

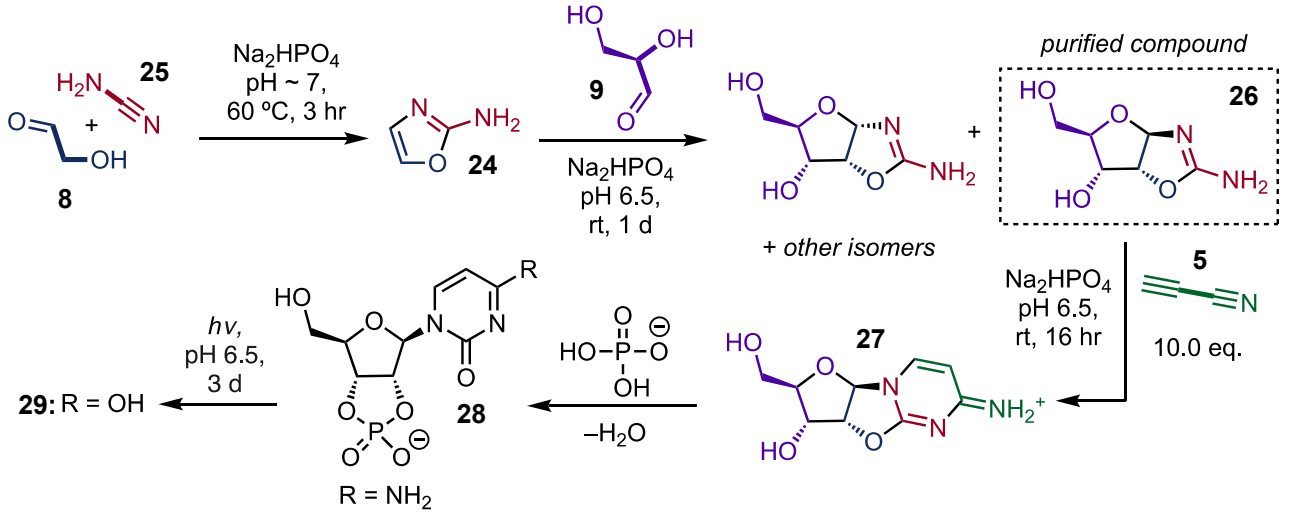
interconversions. Recovered nucleobases (from the catabolic degradation of nucleotides) can later be recycled for nucleotide biosynthesis *via* the salvage pathway, which involves their direct reaction with PRPP **11**.[53] Finally, the conversion of ribonucleotides into deoxyribonucleotides is catalysed by the ribonucleotide reductase (RNR) enzyme, further suggesting that RNA predates DNA from a chemical perspective **(Scheme 4b)**.[56]



Scheme 4 *Biosynthetic pathways leading to the canonical RNA ribonucleotides and DNA deoxyribonucleotides.* ***(a)*** *Nature makes pyrimidine ribonucleotides via the direct glycosylation of orotic acid* ***18****, followed by decarboxylation.*[53] ***(b)*** *Deoxyribonucleotides are biosynthesised via the RNR-catalysed reduction of the 2'-hydroxyl group of an RNA ribonucleotide.*[56]

Given the inherently complex biosynthetic pathways leading to the pyrimidine and purine ribonucleosides, one may ponder how such chemistry could have possibly occurred on the early earth without the aid of enzymes? Many have argued that the only way to address such questions is by validating various chemical scenarios experimentally.[1,57,58] As it turns out, the first plausible prebiotic syntheses of nucleobases were already being explored more than half a century ago[59], with numerous reports providing both the purine[60–63] and pyrimidine[64–66] bases and nucleosides. There was however, at the time of this thesis, no single published synthesis that provided both the purine and pyrimidine nucleosides *via* a unified chemical pathway, starting from a limited set of prebiotically plausible small molecules. Among the published works, the most widely supported approach to the pyrimidines was demonstrated in 2009 by Powner and Sutherland **(Scheme 5)**.[67] Building upon the pioneering work of Orgel,[66] the key feature associated with the Orgel/Powner/Sutherland pyrimidine synthesis, was that it did not require ribose as a starting material, as was previously assumed only to have been available *via* the formose reaction.[49] The synthesis instead involved the condensation of glyceraldehyde **9** with aminooxazoline **24**, which is itself accessible from cyanamide **25** and

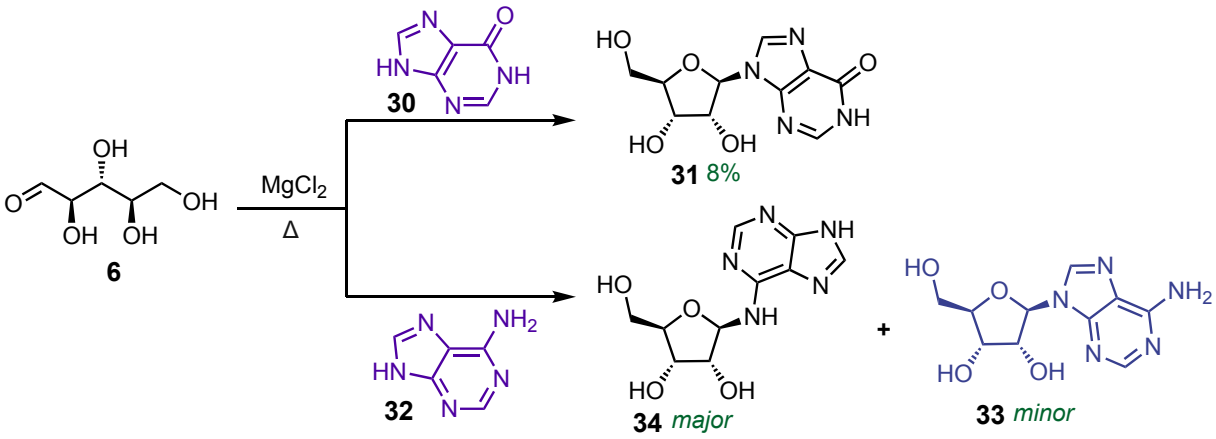
glycolaldehyde **8**. Subsequent reaction of the pentose adduct **26** with cyanoacetylene **5** gave the anhydroarabinonucleoside **27** (as was originally shown by Orgel),[66] which upon nucleophilic ring-opening with phosphate gave cytidine 2',3'-cyclic phosphate **28**. Conversion into the uridine derivative **29** was also possible by photochemical irradiation (254 nm) in buffered solution (pH 6.5).



Scheme 5 *The synthesis of pyrimidine nucleotides from the elements of cyanamide and glycolaldehyde as demonstrated by Powner and Sutherland.*[67]

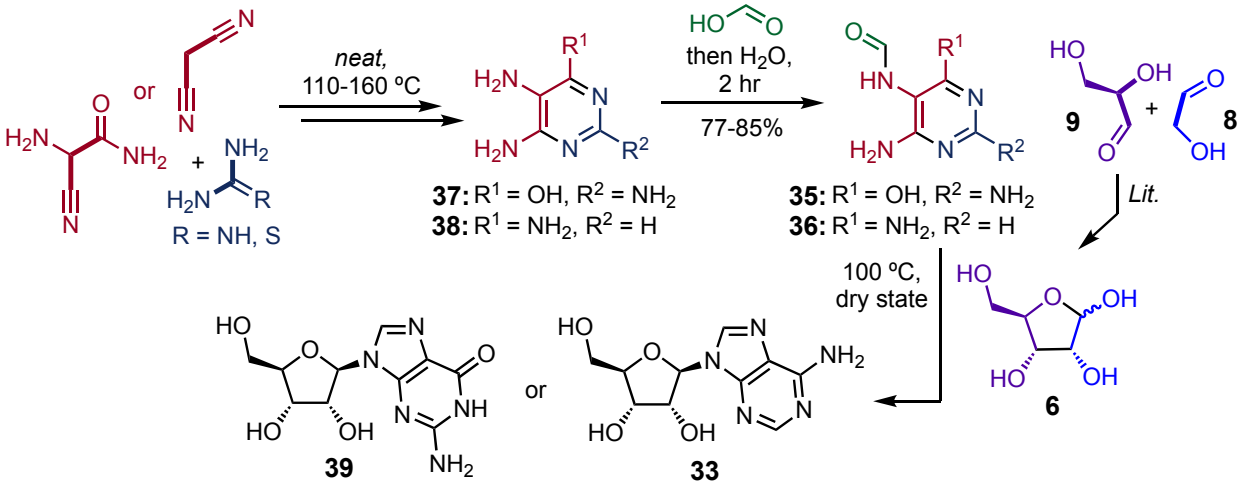
Although the Powner/Sutherland synthesis was the first of its kind to provide the pyrimidine ribonucleotides in high yields, certain issues were raised about the plausibility of the processes involved.[53] The condensation of **9** with **24** for example, provided a complex isomeric mixture similar to that of the formose reaction, where **9** is condensed with **8**. Although selective crystallisation of the ribose amino-oxazoline **26** gave a partially enriched mixture, chromatographic separation of **26** was still necessary to obtain the purified product for reaction with cyanoacetylene, thus bringing into question the pathway's true advantages over an independent prebiotic synthesis of ribose. The pathway was also curiously dependant on the presence of phosphate buffer. Preparation of the critical 2-amino-oxazole intermediate **24** for example, required a 1M solution of phosphate buffer at pH 7, whereas only a small amount of **24** was formed in unbuffered solution. The Sutherland pathway also did not account for how the purine nucleosides could have been generated on the early earth. Although Powner and co-workers have recently demonstrated the formation of 8-oxo purine nucleosides *via* a related but divergent synthetic approach,[68] it remained unclear how such a strategy should provide the canonical set purine ribonucleosides. Through pioneering experiments by Orgel in 1970, it had already been shown that when ribose **6** is heated in the presence of hypoxanthine **30** and either magnesium chloride or certain other inorganic salts, the beta-isomer of furanosylinosine **31** (the hydrolysis product of adenosine) can be formed in up to 8% yield **(Scheme 6)**.[69] When the same

reaction was attempted with adenine **32**, however, the major reaction product was not adenosine **33**, but instead, the undesirable exocyclic (N-6) amine-conjugated isomer **34**.



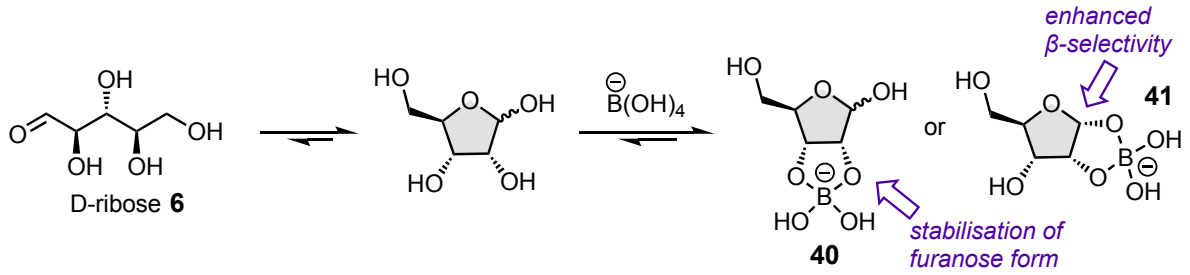
Scheme 6 *Orgel's synthesis of purine nucleosides from ribose **6** regioselectively provides inosine **31**, but instead gives the non-natural N-6 regioisomer of adenosine **34** in the case where adenine **32** is employed as the nucleophile.*[69]

It was not until 2016 that the Carell lab published a complete a complete- and high-yielding synthesis of the purine nucleosides **(Scheme 7)**.[70] The Carell synthesis, in comparison to previous work from Orgel and others, exploited the enhanced nucleophilicity of formamidopyrimidines (FaPys) compared with fully constructed purine nucleobases. The synthesis also drew inspiration from the seminal work from Trinks and Eschenmoser, who in 1987 showed that purine nucleobases can be assembled from FaPys.[71]



Scheme 7 *The synthesis of purine nucleosides from formamidopyrimidines (FaPys) **35/36** and ribose **6**, as demonstrated by Carell and co-workers.*[70] *Ribose **6** can be, among other ways, generated by the condensation of glyceraldehyde **9** with glycolaldehyde **8** in the presence of inorganic catalysts.*[72]

The FaPys **35** and **36**, which were themselves derived from the pyrimidines **37** and **38**, reacted with ribose **6** regioselectively due to the presence of their two relatively nucleophilic 4- (or 6-) amino groups, as well as their C_{2V} symmetrical nature, to give either guanosine **39** or adenosine **33** as well as the expected alpha-furanoside and pyranoside isomers (not pictured). The Carell synthesis also exploited the use of borates, which have been known to stabilise ribofuranosides such as **40 (Scheme 8)** as well as enhancing the yields of pentoses in the formose reaction, as was first shown by Prieur[73] and later investigated further by Benner and co-workers.[74] Borate may also enhance the β–selectivity in glycosylation reactions due to the α–complexation as in **41**, although this has yet to be thoroughly investigated from a mechanistic perspective.[70]



Scheme 8 *The proposed stabilising effects of borate on ribofuranose **40** and its proposed influence on substitution reactions.*

As noted throughout this section, there has been a continued effort among prebiotic chemists to discover plausible syntheses of ribonucleotides from small molecules that could have plausibly emerged the early earth. It should also be apparent that prebiotic synthesis pathways need not necessarily mirror *de novo* biochemistry. One may in fact argue that a certain level of deviation should be expected, given that modern biochemistry is the product of millions of years of evolution. Along those lines, it could perhaps be possible that the molecules which currently make up our genetic systems are distantly removed from those of our ancient progenitors. In the following sections, this topic is further discussed as well as existing evidence for how certain non-canonical nucleosides might be relics of a more ancient, alternative genetic system.

1.6 NON-CANONICAL NUCLEOSIDES AS MOLECULAR FOSSILS

Among the different types of RNA, ribosomal RNAs (rRNA), and in particular, transfer RNAs (tRNA), possess the largest number of post-translational modifications.[75,76] As depicted in **Figure 4**, tRNA and rRNA can be enzymatically modified in a number of ways to provide non-canonical bases and sugars. To this day, more than 600 unique tRNAs have been discovered, and the locations and structures of their many naturally-occurring modified-nucleosides precisely elucidated.[75] The scope of these chemical modifications includes methylations; sulfur- or selenisations; aminations;

thioalkylations; and even aminoacylations. Even more remarkably, an enormous number of non-canonical RNA nucleosides have been identified in all three kingdoms of life.[76] This raises the question of whether it is a coincidence that non-canonical ribonucleosides are so particularly common in rRNA and tRNA, which are themselves thought to be rudimental to the ancient machinery of translation in the RNA world.[30,77] Instead perhaps, life emerged from a more chemically-diverse set of nucleobases that was eventually restricted as a result of chemical evolution.

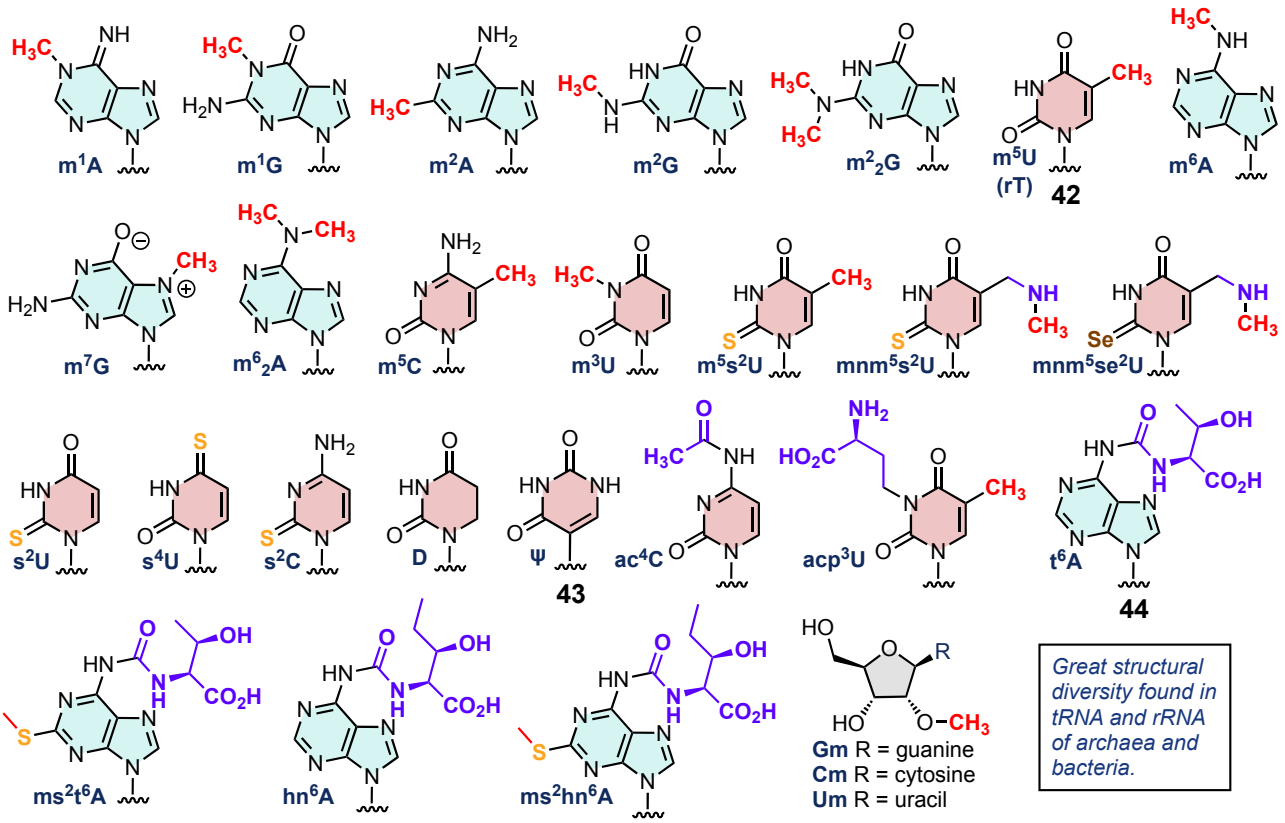


Figure 4 *Selected structures of modified nucleosides found in the tRNA and rRNA of archaea and bacteria. Modifications are indicated in bold and with colour (red = methyl; yellow = sulfur; brown = selenium; purple = other). All of the above structures have been associated with the genome of LUCA through phylogenetic analyses.*[76]

If this were to be the case, then the modified nucleosides found in tRNA and rRNA might in fact be signatures of the genetic code's origins.[76] Strong evidence for this scenario was provided by the Martin group in 2016, where phylogenetic analyses identified that a plethora of extant tRNA and rRNA modifications were present in the genome of the last universal common ancestor (LUCA).[76] Furthermore, many of the non-canonical nucleosides studied by Martin and co-workers are observable all three kingdoms of life. The modifications m^5U (ribo-T) **42** and Ψ (Psi) **43** for example are ubiquitous at positions 54 and 55 of the so-called TΨ loop of tRNA (**Figure 5a**).[75] Of particular

interest to us, are the tRNA modifications whose nucleobases are charged with amino acid moieties, such as in the case of t^6A **44**. Covalent adducts of nucleotides with amino acids such as tRNA esters or aminoacyl adenylates, are indispensable in biochemistry and constitute the biochemical tools upon which translation is built.[77]

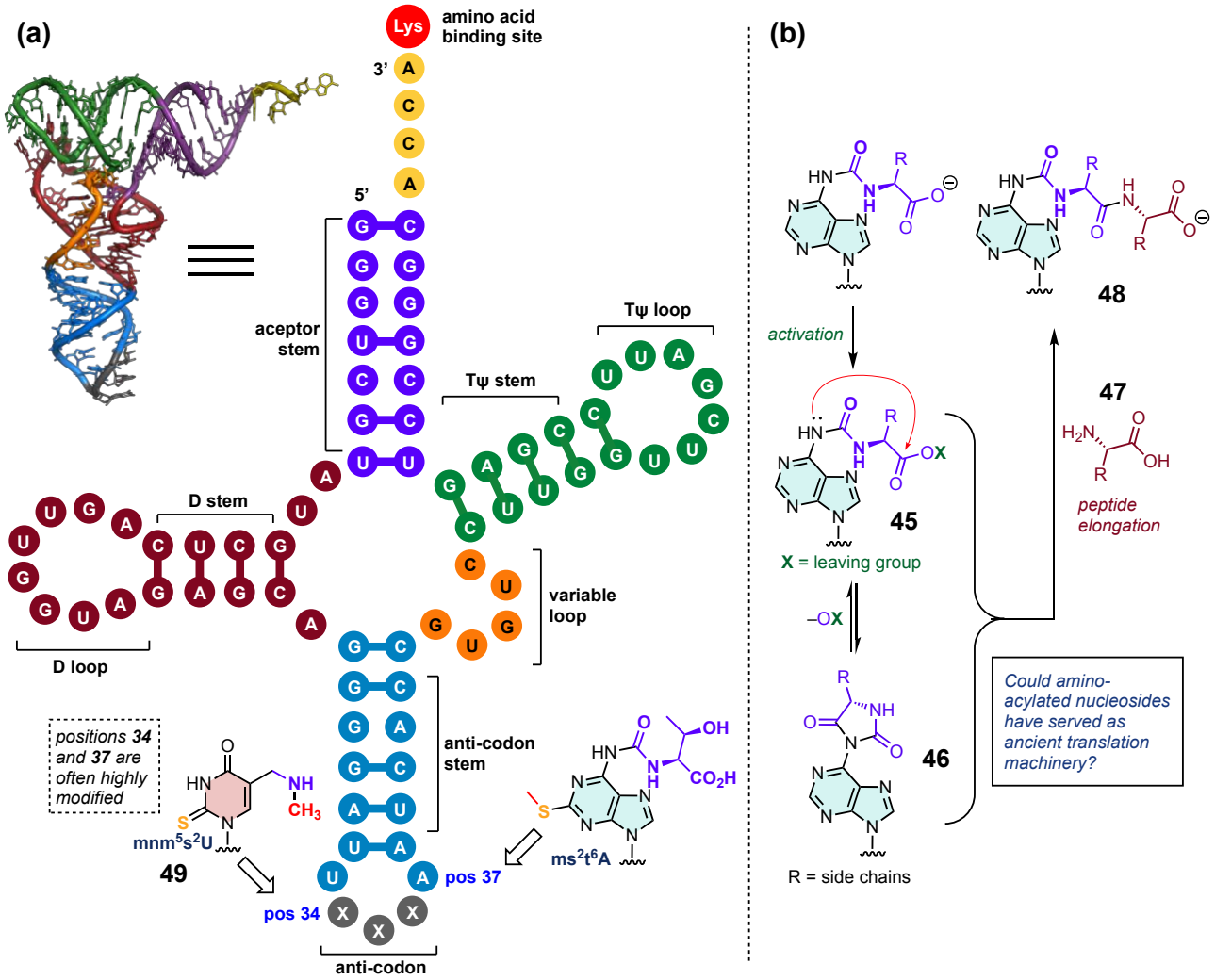


Figure 5 (a) *The structure of an example tRNA with various modifications occurring at position 34 and 37 of the anti-codon stem loop.*[78] **(b)** *A possible mechanism by which information-encoded polypeptides could have originally been generated in the RNA World. Image created using existing artwork*[79] *with adaptations (CC BY-SA 3.0).*

tRNAs are naturally aminoacylated through a cardinal process that occurs in all forms of life – a process that dictates faithful protein synthesis since mischarging leads to sequence errors. In today's cells, however, protein synthesis is carried out using tRNAs charged with amino acids at the 3'-hydroxyl group of the terminal nucleoside located at the 3'- end of the acceptor stem. If translation was once mediated on the prebiotic earth *via* the same type of process, there would be little to protect the resultant ester-bond from hydrolysis (aminoacyl-tRNA esters are readily hydrolysed under either

acidic or alkaline conditions).[80,81] The emergence of translation may therefore instead lie in the primitive chemistry of RNA containing modifications such as **44**, to which amino acids are covalently linked *via* stable urea functional groups. Many such modifications are found at tRNA-positions 34 and 37, directly adjacent to the anti-codon stem loop, further suggesting that the natural ancient progenitor to tRNAs relied on non-canonical nucleosides to facilitate translation. One possible alternative mechanism that could account for the emergence and development of primordial translation was therefore proposed by the Carell group (**Figure 5b**). First, upon activation of the C-terminus, a urea-linked amino acid **45** could cyclise to give a semi-stable hydantoin **46**. Next, the hydantoin ring is reopened *via* the nucleophilic attack of a donor amino acid **47**, thus elongating the peptide chain and giving structures of the form **48**. Further propagation and elongation of polypeptides could be facilitated by cleavage of the N-linked urea in **48**. Successive further additions of amino acids might then be possible if the growing peptide chain was anchored to another RNA at the C-terminus. This could in principle be facilitated by an amine-bearing nucleotide modification such as mnm^{5}s^{2}U **49**, often which are appropriately found at position 34 of the anti-codon stem loop. Importantly, tRNA modifications containing hydantoin moieties such as **46** have previously been identified and characterised in bacterial tRNA.[82] With these points in mind, a mechanistic exploration into whether translation could occur *via* alternative aminoacylation sites within proto-RNAs is currently lacking. It would also be valuable to demonstrate that modifications such as t^{6}A **44** could plausibly have arisen on the early earth, thus prompting the need for prebiotic syntheses. Beyond the genome of LUCA, life presumably underwent many stages of chemical evolution before we arrived at the genetic system found in biology today. This begs the questions of how RNA might have looked even before LUCA. In the next chapter, I discuss plausible alternative genetic systems and recognition units based on chemistry that can no longer be observed in extant life.

1.7 PLAUSIBLE ALTERNATIVE ANCESTRAL RNAS

Among prebiotic chemists, two distinct views have prevailed regarding the prebiotic origins of RNA.[83] In the first instance, RNA is thought to have emerged as a direct result of geochemical processes that specifically favoured its formation on the early earth.[66,67,84] One shortcoming in this model, is that the accumulation of sufficient quantities of complex RNA mononucleotides on the early earth would surely have been a very rare occurrence, even if stemming from highly-favourable synthesis conditions.[85] In the second view, RNA is instead thought to be a product of chemical evolution, and therefore a molecular descendent of another 'proto-RNA'[85,86] In this thesis, I aim to explore both possibilities – that RNA could have itself existed on the early earth, and that it may alternatively have been the product of a multistage evolutionary process.

The first type of adaptation that may have been present in ancestral proto-RNA involves simplification of the ribose moiety and phosphodiester backbone. Joyce and Orgel proposed for example, that RNA might have been preceded by a polymer of "simple, flexible, possibly prochiral nucleotide analogues" that were more accessible through prebiotic synthesis on the early earth.[85] The two chemists further suggested that the prochiral propylene glycol-based polymer **GNA (Figure 6)** could have arisen, given the expected presence of glycerol on the early earth. Meggers and co-workers further validated this claim by showing that chemically-synthesised **GNA** can in fact base-pair with itself and complementary RNA.[87] Inspired by this work, Eschenmoser later demonstrated an α-threofuranosyl-based system (**TNA**), containing vicinally connected (3'→2') phosphodiester bridges.[88] The advantage of **TNA** was that while being constructed from a sugar with only four carbon atoms, the polymer would maintain the same poly-furanosyl backbone repeat unit found in other natural nucleic acids. Despite their prebiotic potential, the proto-RNAs **GNA** and **TNA** have still sometimes been met with doubt.[89] An arguably grander proposal for a proto-RNA involves the potential existence of peptide nucleic acid **PNA**, where a peptide constructed from repeating glutamic- and aspartic acid residues would serve as the nucleic acid backbone, rather than phosphodiester linkages.[90] Unlike RNA, **PNA** might have been accessible under mild non-enzymatic conditions, given that amino acids can be activated in a variety of ways.

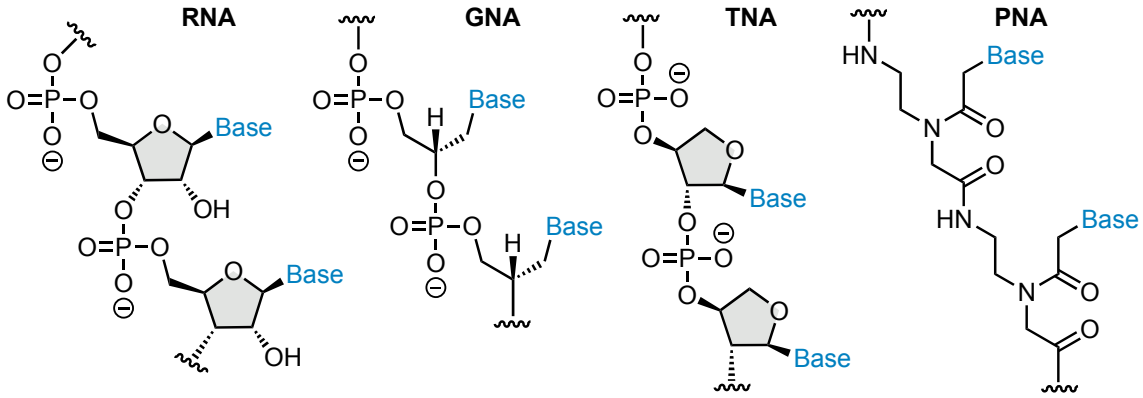


Figure 6 *Examples of non-canonical genetic polymers with hypothesised roles in chemical evolution and the origin of life.*

A second class of modifications that may have been present in proto-RNA involves the simplification or derivatisation of the base-pairing recognition units. Crick and Orgel were among the first to suggest that primitive biology might have taken advantage of a reduced set of heterocyclic bases.[25] They proposed that nucleic acids might have evolved from a purine-only base pairing system. Such systems were later demonstrated experimentally by Eschenmoser and others.[91,92] Numerous studies have since also confirmed the propensity of various heterocyclic systems to form stable base-pairs, each

with their own merits and advantages in origins of life scenarios **(Figure 7)**.[93–97] We argue, however, that while many of these alternative base-pairing systems may indeed be prebiotically plausible, their constituents are in most cases barely simpler than those that make up canonical Watson-Crick RNA. We therefore became interested in whether an information biopolymer could possibly have arisen, containing nucleobases constructed from linear, acyclic, repeating units.

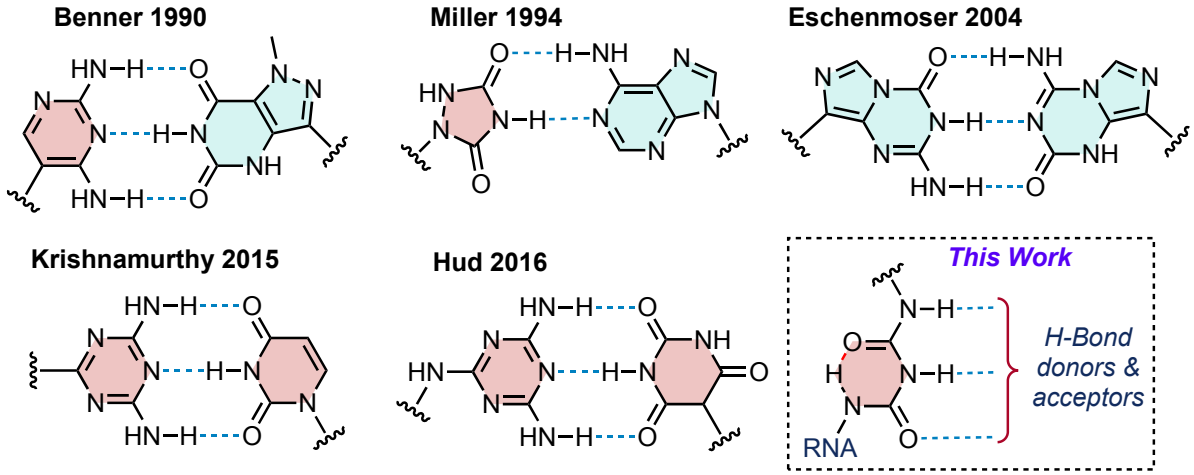
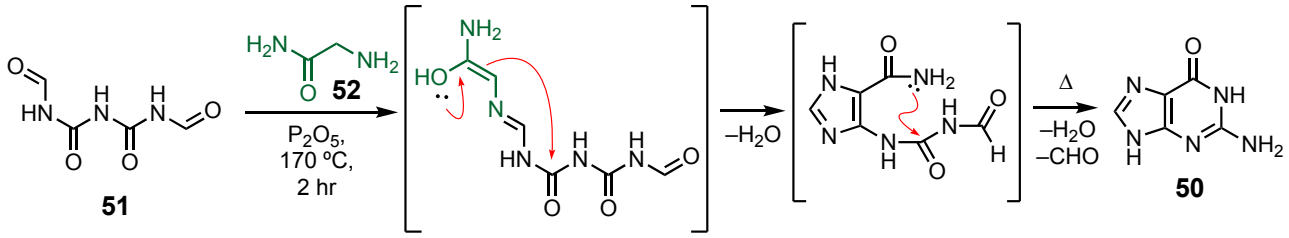


Figure 7 *Examples of alternative base-pairing systems that have been implicated with the chemical origins of life.*[91–97]

Inspired by the pioneering work of Wöhler,[38] we asked whether it would be possible to generate a base-pairing system entirely from polycarbamoylureas with the general repeating structure H_2N-CO-(NH-CO)$_n$-NH_2. In general, such molecules are known to adopt reversible planar-, hydrogen-bonded confirmations, thus potentially allowing them to participate in base-pairing.[98] We were also inspired by recent work from Lagoja, who demonstrated that guanine **50** can be synthesised from the acyclic precursors diformyl biuret **51** and glycinamide **52** under thermally-promoted conditions (**Scheme 9**).[99]



Scheme 9 *Synthesis of guanine **50** from diformyl biuret **51** and glycinamide **52**, shown by Lagoja.*[99]

Based upon these results, we anticipated certain conditions similar to those used by Lagoja and co-workers (for example with activation of a carbonyl moiety, or by formylation of the N-1 amino group) might even allow us to convert an acyclic proto-base such as that shown in **Figure 7** into one that more closely resembles the canonical pyrimidines. Another goal of this thesis was thus, to investigate

whether acyclic systems containing polycarbamoylureas are capable of base-pairing, and whether they could have plausibly existed on the early earth.

1.8 AIMS OF THIS THESIS

As depicted, the primary goal of this thesis was to investigate whether a proto-RNA recognition system based upon acyclic polycarbamoylureas could have plausibly preceded canonical RNA on the early earth (**Figure 8**). To such an end, prebiotic syntheses of ribonucleosides containing compounds such as biuret **53** and triuret **54** as the nucleobase moiety would need to be developed. The nucleosides would then be synthetically incorporated into RNA and their physicochemical and base-pairing properties thus investigated. The secondary goal of the work presented here was to investigate the speculated role of aminoacylcarbomoyl adenosine nucleosides in the origin of translation and polypetides. Specifically, we aimed to verify whether compounds with the general structure **55** are accessible under prebiotically plausible conditions, and to experimentally demonstrate their propensity to facilitate polypeptide synthesis under early-earth conditions. Finally, harnessing the information gained through our other developing goals, we aimed to investigate whether certain hypothesised proto-RNA nucleosides could have plausibly acted as precursors to the canonical Watson-Crick pyrimidines uridine **17** and cytidine **23**. This goal would be achieved by their incorporation into RNA strands, followed by an assessment of their physicochemical properties. More broadly, we aimed consolidate this knowledge to develop a broader picture for how the first life might have emerged from RNA on the early earth.

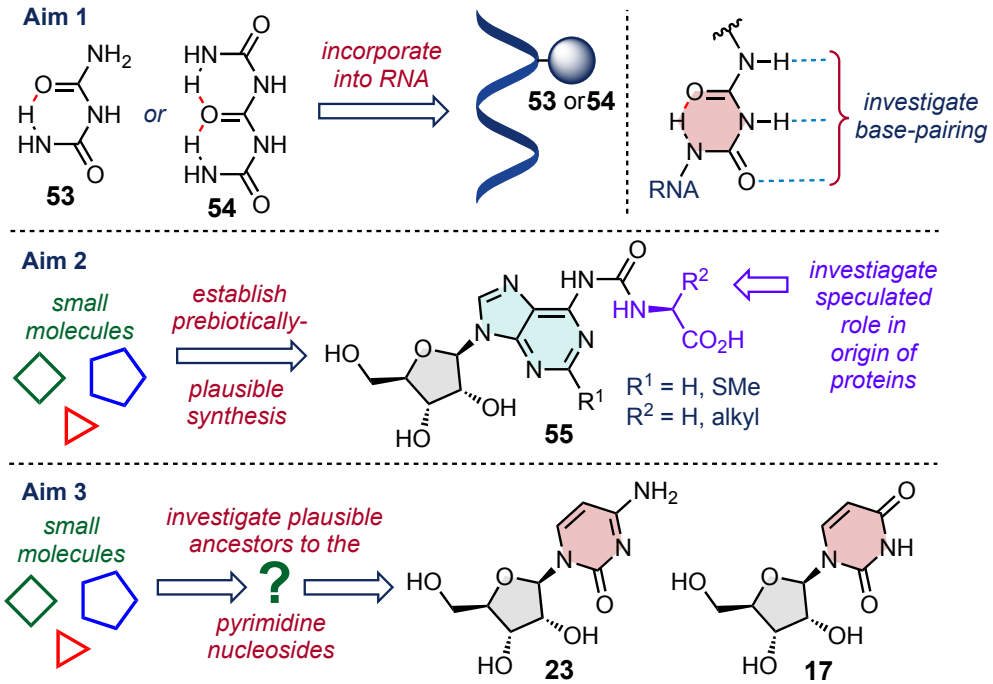


Figure 8 *Aims of this PhD thesis.*

2 PUBLISHED WORKS

2.1 FIRST PUBLICATION

TITLE: proto-Urea-RNA (Wöhler RNA) containing unusually stable urea nucleosides

AUTHORS: H. Okamura[+], A. Crisp[+], S. Hübner, S. Becker, P. Rovó*, T. Carell*

*+ authors contributed equally (shared 1st authorship). * corresponding authors.*

JOURNAL, YEAR, VOLUME, AND PAGE NO.:

Angew. Chem. Int. Ed., **2019**, *58*, 18691-18696

SUMMARY:

It has been proposed that the RNA-world was preceded by an era in which genetic information was encoded by proto-RNAs constructed from molecules that were more likely to have been present on the early Earth. Here, we demonstrate that the prebiotically plausible starting materials biuret and triuret react with ribose to generate (among other products), stable nucleosides capable of base pairing. Synthesis of phosphoramidite building blocks allowed us to incorporate the nucleosides into RNA, and thus determine their base-pairing properties, namely that triuret forms a pair with G, which closely resembles the U:G wobble base pair.

PERSONAL CONTRIBUTIONS:

- Prepared compounds 1-15, as depicted in the publication (syntheses were originally designed and carried out by H. Okamura), and characterised their chemical properties.
- Incorporated nucleoside phosphoramidites into RNA and synthesised homo-oligomers.
- Prepared RNA samples for high-resolution NMR spectroscopy experiments under supervision from H. Okamura.
- Prepared synthetic standards and designed calibration curves in order to quantify prebiotic reactions.
- Conducted prebiotic synthesis experiments (with design and supervision from H. Okamura) and collected/presented the data associated with their results.
- Measured melting curves for inosine-X (X=Bi/Tri/AUCG) pairs, and for certain mismatch pairs.
- Composed and prepared the figures (with the exception of figure 6) together with H. Okamura and T. Carell.
- Composed, edited and revised the manuscript and supporting information together with H. Okamura, S. Hübner, S. Becker, P. Rovó and T. Carell.
- Supervised S. Hübner, who supported the synthesis of phosphoramidite material.

2.1.1 Authorship Declaration

Erklärung über den substantiellen Beitrag zu einer Publikation die für eine kumulative Dissertation eingereicht wird

Ich erkläre hiermit, dass für die Publikation mit dem Titel "proto-Urea-RNA (Wöhler RNA) containing unusually stable urea nucleosides", beide 'geteilter Erstautoren' einen wesentlichen und gleichwertigen Beitrag geleistet haben, der die experimentelle Planung, Ausführung und Vorbereitung des veröffentlichten Manuskripts umfasst. Spezifische Beiträge sind auf der vorhergehenden Seite aufgeführt.

Declaration on the substantial contribution to a publication submitted as part of a cumulative dissertation

I hereby declare that for the publication titled "proto-Urea-RNA (Wöhler RNA) containing unusually stable urea nucleosides", both 'shared-first' authors made a substantial and equal contribution, involving experimental design, execution and preparation of the published manuscript. Specific contributions are listed on the previous page.

München, 06.12.2020 .....

Ort und Datum
Place and date

.....

Unterschrift
Signature

Sendai, 07.12.2020 .....

Ort und Datum
Place and date

.....

Unterschrift
Signature

PERMISION AND LICENSING

Prebiotic Chemistry Hot Paper

International Edition: DOI: 10.1002/anie.201911746
German Edition: DOI: 10.1002/ange.201911746

Proto-Urea-RNA (Wöhler RNA) Containing Unusually Stable Urea Nucleosides

Hidenori Okamura[+], Antony Crisp[+], Sarah Hübner, Sidney Becker, Petra Rovó, and Thomas Carell**

Abstract: *The RNA world hypothesis assumes that life on Earth began with nucleotides that formed information-carrying RNA oligomers able to self-replicate. Prebiotic reactions leading to the contemporary nucleosides are now known, but their execution often requires specific starting materials and lengthy reaction sequences. It was therefore proposed that the RNA world was likely proceeded by a proto-RNA world constructed from molecules that were likely present on the early Earth in greater abundance. Herein, we show that the prebiotic starting molecules bis-urea (biuret) and tris-urea (triuret) are able to directly react with ribose. The urea-ribosides are remarkably stable because they are held together by a network of intramolecular, bifurcated hydrogen bonds. This even allowed the synthesis of phosphoramidite building blocks and incorporation of the units into RNA. Investigations of the nucleotides' base-pairing potential showed that triuret:G RNA base pairs closely resemble U:G wobble base pairs. Based on the probable abundance of urea on the early Earth, we postulate that urea-containing RNA bases are good candidates for a proto-RNA world.*

Introduction

Urea, the bisamide of carbonic acid, is widely distributed in the biosphere and plays a fundamentally important role in the biosynthesis of proteins and the entire N-cycle of organisms in general.[1,2] It is also believed to have formed early on the prebiotic Earth and before the process of chemical evolution that gave the centrally important molecules of life.[3] Urea is a key starting molecule for many prebiotic chemical reactions,[4–11] and was the first organic compound synthesized from inorganic matter (ammonium cyanate) by the chemist Friedrich Wöhler in 1828.[12] Wöhler's synthesis was the starting point of the field of organic chemistry and was, among others, essential to defeating the mainstream ideology of "vitalism", which stated that organic matter contained a special vital force.[13] Current theories about the origin of life are built upon the RNA world hypothesis, which predicts the early formation of information-encoding RNA that was able to self-replicate and that featured properties leading to their survival under early Earth conditions.[14–17] It is assumed that based on the processes of chemical evolution, more and more complex RNA and RNA–peptide structures were created that finally led to the emergence of life.[18,19] RNA and the constituting nucleosides that are needed to establish faithful replication of "genetic" information are, however, rather complex chemical structures. The problem of finding prebiotically plausible pathways to the canonical nucleosides (Figure 1a; known as the nucleoside problem)[20] led to the idea that RNA was potentially proceeded by a proto-RNA that could more easily arise from prebiotically privileged starting materials.[21] As a result, emerging discussions about the origin of life have often emphasized the significance of informational polymers that are simpler than RNA. A revolutionary study from Eschenmoser's group, for example, demonstrated that α-threofuranosyl nucleic acid (TNA) is capable of forming antiparallel duplexes and can even pair with cDNA or RNA.[22] TNA was later simplified to an acyclic polymer known as glycol nucleic acid (GNA),[23–25] and various other XNA backbones have since been investigated.[26] We know that formaldehyde and $Ca(OH)_2$ can give sugars by the

[*] Dr. H. Okamura,[+] A. Crisp,[+] S. Hübner, Dr. S. Becker, Dr. P. Rovó, Prof. Dr. T. Carell
Center for Integrated Protein Science (CiPS[M]) at the Department of Chemistry, LMU München
Butenandtstr. 5–13, 81377 München (Germany)
E-mail: petra.rovo@lmu.de
thomas.carell@lmu.de
Homepage: http://www.carellgroup.de
[+] These authors contributed equally to this work.
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.201911746.

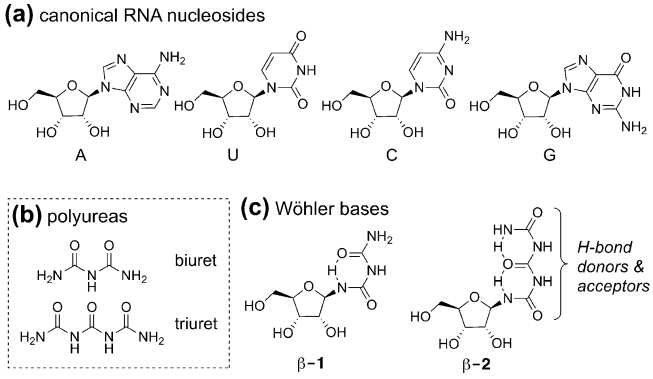


Figure 1. a) The chemical structures of the canonical RNA nucleosides. b) The chemical structures of biuret and triuret. c) Depiction of the urea-based nucleosides with potentially stabilizing hydrogen bonds.

formose reaction,[27–29] as first described by Butlerov.[30] Although it remains unclear how exactly such a process could lead to the substantial accumulation of ribose on the early Earth, recent developments have led to significant improvements in the prebiotic synthesis of ribose from simpler aldoses.[31–33] We also know that urea was one of the most likely nitrogen-containing molecules present on the early Earth and that it can activate mineral phosphate to achieve phosphorylations.[4,5] As urea itself has even been shown to react directly with ribose under mildly acidic conditions,[34] we asked whether the construction of informational base-pairing systems based on the pyrolysis products of urea, biuret and triuret (Figure 1 b),[35] would be possible.

The fundamental chemical problem associated with this notion is that such nucleosides (Figure 1 c) should be highly prone to hydrolysis in an aqueous environment. Investigating the structures of potential urea bases, however, led us to discover that those containing biuret (β-**1**) and triuret (β-**2**) are highly stabile even in water, probably because of the intramolecular H-bonds (Figure 1 c). The question of the possible prebiotic existence of urea (Wöhler) RNA is therefore directly associated with the question as to which extent these non-covalent interactions protect from hydrolysis.

Results and Discussion

To investigate the formation of urea nucleosides under plausible prebiotic conditions, we mixed aqueous solutions of ribose with either biuret or triuret in the presence of boric acid and heated the mixture at 95 °C for 18 h, allowing the mixture to slowly dry down. Boric acid, which forms complexes with vicinal 1,2-diols because of its high electron deficiency, was included for its known stabilizing[36–38] and directing[39–41] effects on ribose. The resulting solid was then taken up in dilute (100 mM) sodium carbonate buffer (pH 9.5) and heated again at 95 °C for 1 h (Figure 2).

The mixture was subsequently analyzed by reverse-phase HPLC-MS. Gratifyingly, we noted significant formation of the corresponding nucleosides with both biuret and triuret. These wet–dry conditions mimic the intermittently concentrating environments that might be found on drying beaches or lagoons (or even today in Death Valley),[42] as was recently discussed in relation to the prebiotic synthesis of canonical and non-canonical purine nucleosides.[39] For the biuret reaction, we detected formation of four major nucleoside products. Of those, two are the α- and β-anomers of the ribofuranosides (α-**1** and β-**1**). These structures were confirmed by independent synthesis of the α- and β-anomers (see below) followed by co-injection studies. The absolute stereochemical configurations of the α- and β-anomers were confirmed by NOESY-NMR spectroscopy (see the Supporting Information). The other two compounds detected in the HPLC-MS experiment are likely the pyranosidic species (not further investigated). Similar but not identical data were obtained for triuret. Here too, we detected four reaction products of which two are the α- and β-anomers of the ribofuranosides (α-**2** and β-**2**), with the remaining compounds likely being the pyranosides again. We focused our initial

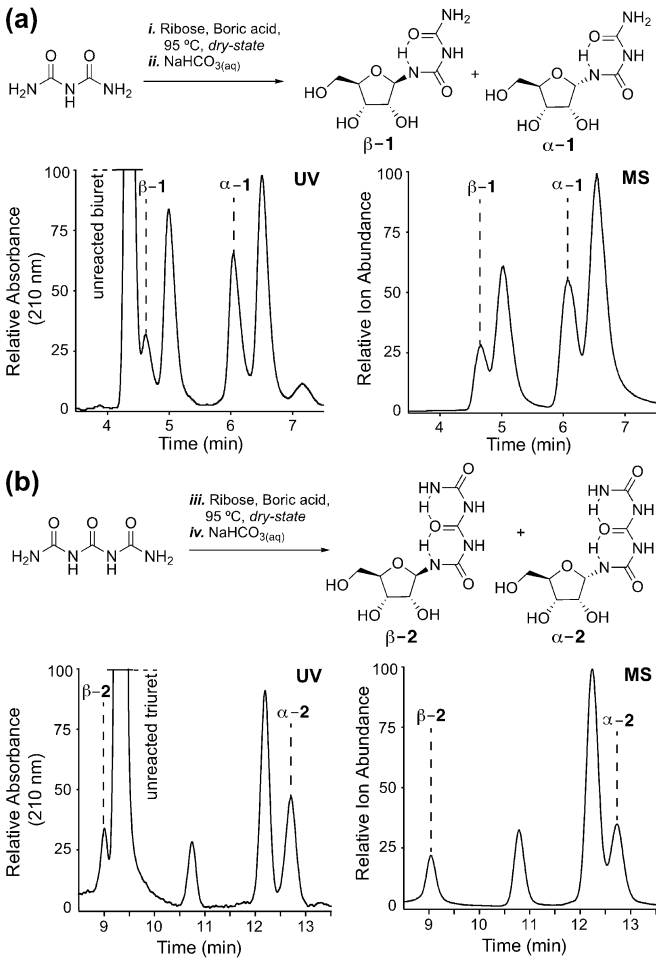
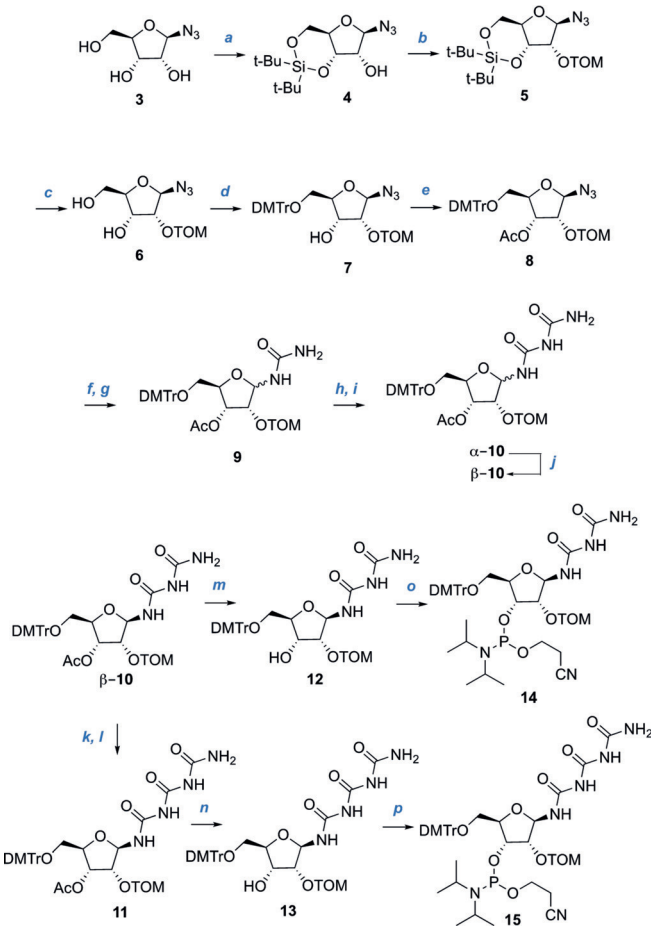


Figure 2. Reaction of a) biuret and b) triuret with ribose and analysis of the reaction mixture by HPLC-MS, showing the successful formation of urea-based nucleosides.

studies on the ribofuranosides and noted to our surprise that both the biuret and the triuret species are quite stable. Both compounds can be kept for prolonged periods of time in an aqueous solution at neutral pH without signs of anomerization. This unusual stability of the biuret and triuret nucleosides prompted us to study if one could generate phosphoramidites and insert them into RNA. This would require the Wöhler nucleosides to survive even the nucleophilic reaction and deprotection conditions needed for solid-phase RNA synthesis.

The synthesis of the phosphoramidite building blocks is shown in Scheme 1. We began with the 3′,5′-silyl protection of 1-azidoribose **3** to obtain **4**, followed by TOM protection of the 2′-OH group to give **5**. Desilylation of the 3′,5′-positions (affording compound **6**) and subsequent DMTr protection of the 5′-OH group furnished compound **7**. We then protected the 3′-OH group with an acetyl group to give **8** and reduced the azide by catalytic hydrogenation followed by reaction of the amine with trimethylsilylisocyanate. This provides the urea riboside **9**. A second reaction with trichloroacetylisocyanate in pyridine followed by cleavage of the trichloroacetate group with basic alumina in methanol gave the biuret riboside **10** as a mixture of the α- and β-isomers (β-**10** and



Scheme 1. Phosphoramidite building block synthesis of the biuret and triuret nucleosides. Reagents and conditions: a) *t*-$Bu_2Si(OTf)_2$, DMF, 0 °C, 1 h; b) *i*-PrSiO(CH_2)Cl, NaH, THF, 0 °C, overnight, 55 % over 2 steps; c) HF-pyridine, pyridine, CH_2Cl_2, room temperature, 1 h, 61 %; d) DMTrCl, pyridine, room temperature, overnight, 77 %; e) Ac_2O, DMAP, pyridine, room temperature, 2 h, 91 %; f) 10 % Pd/C, H_2, THF, room temperature, 2 h, then g) TMS-isocyanate, THF, room temperature, overnight, 76 % (mixture of diastereomers); h) trichloroacetylisocyanate, pyridine, THF, room temperature, 1 h, then i) Al_2O_3, MeOH, room temperature, 1 h, 90 % (d.r. α/β = 1.6:1); j) DBU, THF, 50 °C, overnight, 28 % (96 % based on recovered starting material); k) trichloroacetylisocyanate, pyridine, THF, room temperature, 1 h, then l) Al_2O_3, MeOH, room temperature, 1 h, 58 %; m) NH_3, MeOH, room temperature, 4 h, 74 %; n) NH_3, MeOH, room temperature, 4 h, 76 %; o) bis(2-cyanoethyl)-*N*,*N*-diisopropylphosphoramidite, diisopropylamine-tetrazole, CH_3CN, room temperature, overnight, 70 %; p) bis(2-cyanoethyl)-*N*,*N*-diisopropylphosphoramidite, diisopropylamine-tetrazole, CH_3CN, room temperature, overnight, 54 %.

α-**10**), which we separated by column chromatography (SiO_2, CH_2Cl_2/CH_3OH 100:0→98:2). The unwanted α-isomer (α-**10**) isomerized upon heating in the presence of DBU, which gave a mixture of β-**10** and α-**10**. Iterative anomerization allowed us to increase the total isolated yield of the β-anomer (β-**10**) for subsequent reactions. To obtain the triuret nucleoside **11**, we repeated the trichloroacetylisocyanate reaction followed by basic alumina treatment using the pure β-anomer of the biuret nucleoside (β-**10**). Interestingly, we observed very little anomerization during these reactions (as monitored by TLC). The β-triuret nucleoside **11** was therefore obtained in anomerically pure form and in 58 % yield. Both the β-biuret (β-**10**) and the β-triuret (**11**) nucleosides were next converted into the corresponding phosphoramidites. To this end, we cleaved the 3′-acetyl group with ammonia (even these conditions do not lead to hydrolysis of the glycosidic bond), and then treated the ensuing compounds **12** and **13** with bis(2-cyanoethyl)-*N*,*N*-diisopropylphosphoramidite in the presence of diisopropylamine and tetrazole. This final step provided the phosphoramidite building blocks **14** and **15** in good yields of 70 % and 54 %, respectively.

For the solid-phase oligonucleotide synthesis, we used a standard ultra-mild RNA synthesis protocol. The urea bases were coupled once for 20 min. After full assembly of the RNA strands using Pac chemistry conditions, we deprotected the RNA strands and cleaved them from the solid support with NH_3 in methanol at room temperature. The silyl protecting groups were finally removed using HF-TEA in DMSO at 65 °C. Figure 3 shows the crude HPLC chromatogram of the triuret-containing RNA strand as an example, as well as the MALDI-TOF mass spectrum obtained from purification of the major species. It is clearly evident that the RNA strands are efficiently produced and can be cleanly purified. Similar data were obtained for the biuret-containing strands (Figure S7). It is remarkable that the urea bases survive the RNA synthesis conditions to give RNA strands in excellent purity.

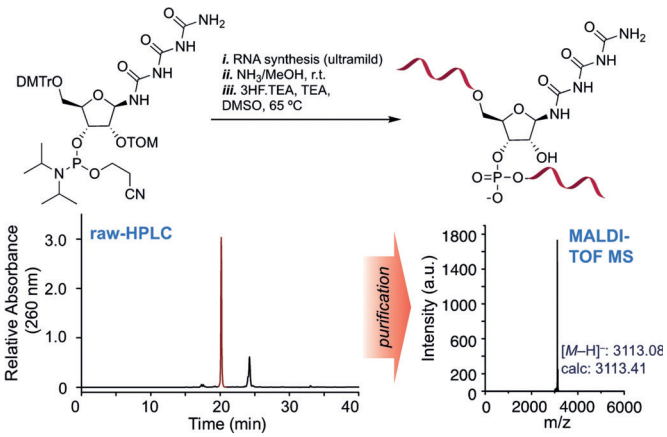


Figure 3. Preparatory HPLC and MALDI-TOF mass data for an exemplary Tri-containing RNA strand with the sequence 5′-CUUACTriCUGA-3′.

This stability allowed us to next investigate the pairing properties of the biuret and triuret bases using thermal melting curve studies (10 mM sodium phosphate, 150 mM NaCl, pH 7.0). If a candidate proto-RNA were to have existed before the emergence of modern RNA, then it stands to reason that its bases needed to pair with the canonical nucleosides to allow a smooth evolutionary transition from proto-RNA to RNA. For the measurements we prepared RNA strands with a C:G or a U:A base pair in a central position. We then exchanged the pyrimidine base C or U with the biuret (Bi) or triuret (Tri) base. The data are compiled in Figure 4b. The unmodified RNA duplexes feature, as expected, rather high melting temperatures of 55 °C (C:G) and

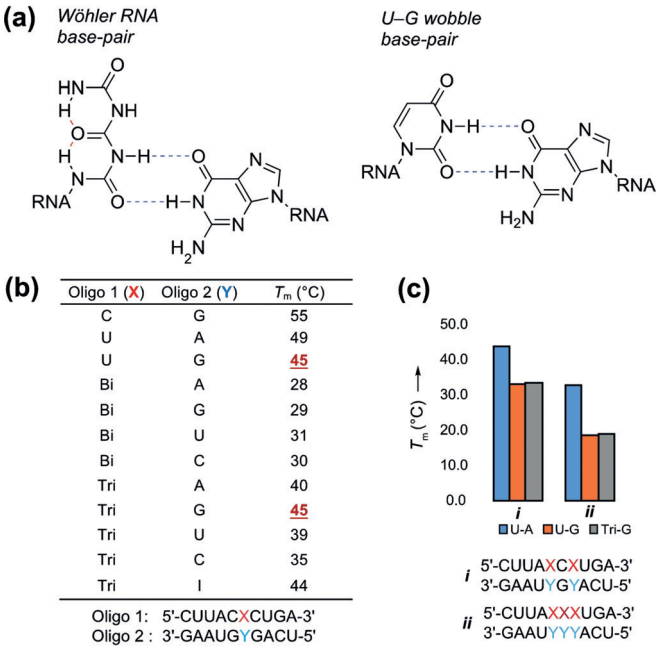


Figure 4. a) Chemical structure and base-pairing properties of triuret and similarity between the triuret-G base pair and a U-G wobble base pair. b) Summary of T_m analyses for oligonucleotides containing biuret and triuret. c) Summary of T_m analyses for oligonucleotides containing more than one modified base or U-G wobble base pair. Solutions were buffered with 10 mM sodium phosphate (pH 7) and 150 mM NaCl.

49 °C (U:A). For all Bi:A/G/U/C base pairs, we measured much lower melting temperatures, which were in addition quite similar irrespective of the counterbase (between 28 °C and 31 °C). This shows that the biuret base does not prefer a particular counterbase and that base pairing is in general weak, if it occurs at all. For the larger triuret base, we noted significantly higher melting temperatures between 35 °C and 45 °C. In addition, we observed a clear base-pairing preference with G. The melting temperature for the Tri:G base pair is around 5 °C higher than those for the others, which clearly points to a significant pairing selectivity. Analysis of the base pairing potential of the triuret base shows that it is in principle possible to form a wobble-type base pair with G. We therefore tested an RNA duplex with a central U:G wobble base pair and observed indeed the same melting temperature (45 °C). Our hypothesis was further validated by the observation that triuret also forms a stable base pair with the structurally related base inosine (I), giving an almost identical melting temperature of 44 °C. To exclude that this is pure chance, we next prepared RNA strands with either two or even three consecutive triuret bases and paired this strand with a counter-strand containing either two or three central G bases. Comparative melting point studies showed exactly the same behavior between the U:G wobble and the Tri:G base pairs.

To gain deeper insight into the origin of the stability of the Tri:G base pairing, we prepared 8 mer palindromic RNA strands that were designed to form canonical C:G, wobble U:G, or Tri:G pairs at the central position (GGUXGACC, where X = C, U, or Tri) and analyzed their 2D ^{1}H–^{1}H NOESY spectra.[43] The high chemical shift similarity in the fingerprint region of the spectra (Figure S17) confirms the same overall structure for the three oligonucleotides, namely formation of an A-form double-stranded RNA (Figure 5c).

The imino region of the NOESY spectra provides direct information on the hydrogen bond interactions between base pairs (Figure 5a and Figure S16). The Tri base has five amide protons (H1, H3, H5, H71, and H72) and three carbonyl oxygen atoms that could potentially be involved in base pairing. Out of the five protons, H3 and H5 are partially or fully solvent-exposed and thus they show strong exchange

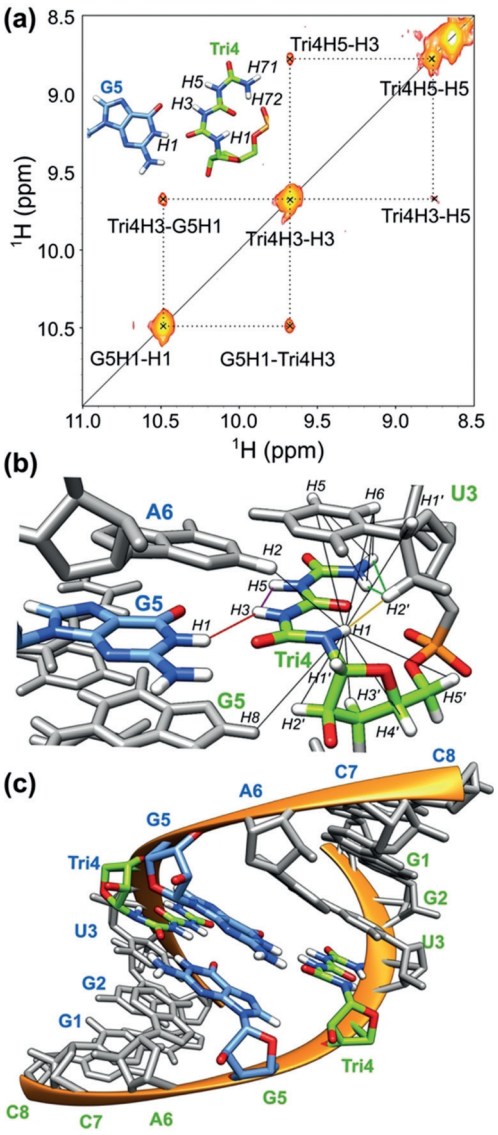


Figure 5. NMR analysis of the triuret:G base pairing. a) Excerpt from the ^{1}H–^{1}H NOESY spectrum (t_{mix} = 40 ms) of the dsRNA (GGUXGACC, where X = Tri) showing the inter-strand cross-peaks between the H1 imino proton of G5 and the H3 amide proton of Tri4. The inset shows a model for the triuret:G base paring. b) NOE contacts of the triuret base amide protons. Essential structure-defining NOE contacts are highlighted for Tri4H1–U3H2′ (yellow), Tri4H71/H72–U3H2′ (green), G5H1–Tri4H3 (red), and Tri4H3–Tri4H5 (purple). Other observed NOE contacts are shown as black lines. c) Structural model of the GGUXGACC oligonucleotide showing the non-canonical base pairing between G5 (blue) and Tri4 (green) bases.

cross-peaks at the water resonance (Figure S19). The number and intensity of cross-peaks between the H1 proton of Tri and the sugar/base protons of the previous uridine base (U3) indicate that the H1 proton is located in a similar arrangement to the H5 proton of a U or a C or as the H8 proton of an A and a G base in canonical dsRNA structures; for example, we observed a strong NOE cross-peak between Tri4H1 and U3H2′ (yellow line in Figure 5b). Similarly, the terminal NH_2 group of Tri shows strong cross-peaks with the H2′ and H5 protons of U3 (green lines in Figure 5c). Thus, the NOESY spectra suggest that both H1 and the NH_2 group point towards the phosphate backbone. Direct evidence for Tri:G base pairing was obtained from a NOESY spectrum recorded with a relatively short (40 ms) mixing time (Figure 5a). This experiment provides a clear cross-peak between the H1 imino proton of G5 and the H3 amide proton of Tri of the opposite strand (red line in Figure 5b). In addition, the TriH3 proton shows an intrabase cross-peak with H5 (purple line in Figure 5c). These, as well as a handful of other NOE cross-peaks between the triuret base protons and the surrounding protons, support the structural model depicted in Figure 4a (all other NOE distance restraints are depicted with black lines in Figure 5b), with the triuret moiety clearly forming bifurcated hydrogen bonds.

To obtain a three-dimensional model for the Tri:G base pair, we performed molecular modeling using the online software ROSIE[44,45] followed by NOE-based structure calculations using the software CNSsolve.[46,47] During the structure calculations, only the conformation of the triuret base was altered while the phosphate backbone and all other bases were kept at their fixed position. Figure 5c displays the obtained low-energy structure model for the double-stranded 8 mer RNA. It is clearly evident that the two Tri:G base pairs are well accommodated in the structure and that the extended network of hydrogen bonds within the Tri structure and between Tri and the opposite G establish the measured stability.

All of these results confirm that triuret, which is itself a condensation product of urea, is able to form a folded pseudobase that pairs with guanine. Finally, in order to further demonstrate the prebiotic plausibility of Wöhler RNA, we synthesized a homo-RNA oligomer containing exclusively triuret bases (Figure 6a). The only additional structural modification was the inclusion of a dye at the 5′-end (Cy3), which was necessary to allow UV detection at 548 nm and therefore purification by HPLC. Remarkably, despite having five, in principle hydrolysable triuret bases in a row, the homo-strand was bench-stable both at room temperature as well as when subjected to the harsh conditions necessary for RNA deprotection and cleavage from the solid support. Figure 6 shows the crude HPLC chromatogram obtained from the material directly after its synthesis and the correct MALDI-TOF mass spectrum, confirming the integrity of the material.

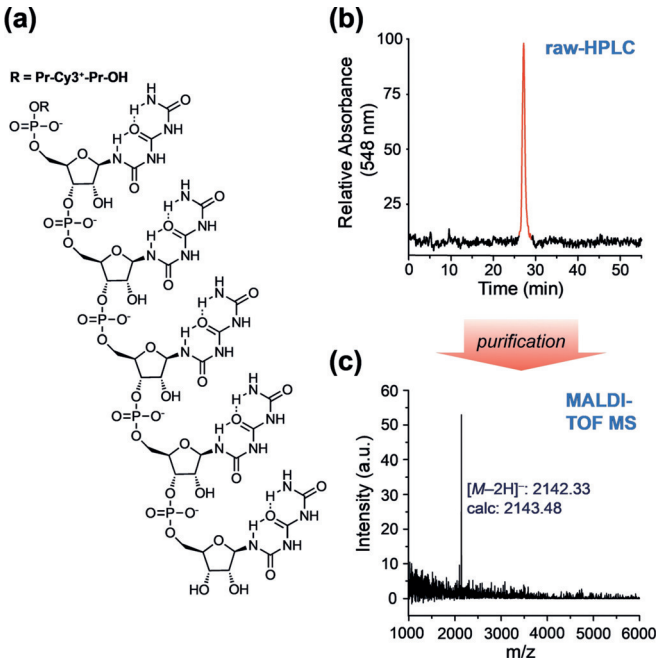


Figure 6. a) Depiction of the 5 mer triuret oligomer containing a Cy3-fluorophore at the 5′-end for better detectability. b) Crude HPLC chromatogram of the triuret oligomer and c) MALDI-TOF mass spectrum of the triuret oligomer.

Conclusion

In summary, we have shown that biuret and triuret are able to condense directly with sugars (here ribose) to form stable bis- and tris-urea nucleosides. Within an RNA strand, triuret is able to form stable wobble-type base pairs with G as well as with the prebiotically relevant base inosine.[48] As discussed, biuret and triuret are obtained upon pyrolysis of urea, one of the most likely building blocks available on the early Earth. Given that various tri-, tetra-, and pentose sugars are prebiotically accessible from glycoaldehyde,[31,32] which is itself accessible from either formaldehyde[30] or HCN by ultraviolet irradiation,[49] our discovery creates the prebiotically attractive possibility of generating information-encoding oligomers whose key building blocks are derived of simple one-carbon units. The chemistry described here now needs to be explored with sugars simpler than ribose. Discussed examples are threose-[22,50] and glycol-based[23,25] backbones.

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for financial support via SFB1309, SPP1784, and CA275/11-1. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. EPiR 741912) and through a H2020 Marie Skłodowska-Curie Action (LightDyNAmics, 765866). H.O. acknowledges support through a Marie Skłodowska-Curie Individual Fellowship (752420) from the European Commission. P.R. acknowl-

edges support from the Center for NanoScience and from the Fonds der Chemischen Industrie.

Conflict of interest

The authors declare no conflict of interest.

Keywords: base pairing · origin of life · prebiotic chemistry · proto-RNA · urea

How to cite: *Angew. Chem. Int. Ed.* **2019**, *58*, 18691 – 18696
Angew. Chem. **2019**, *131*, 18864 – 18869

[1] G. Shambaugh III, *Am. J. Clin. Nutr.* **1977**, *30*, 2083 – 2087.
[2] J. W. Campbell, *Nature* **1965**, *208*, 1299 – 1301.
[3] G. Schlesinger, S. L. Miller, *J. Mol. Evol.* **1983**, *19*, 383 – 390.
[4] R. Lohrmann, L. Orgel, *Science* **1971**, *171*, 490 – 494.
[5] R. Osterberg, L. Orgel, *J. Mol. Evol.* **1972**, *1*, 241 – 248.
[6] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239 – 242.
[7] M. P. Robertson, S. L. Miller, *Nature* **1995**, *375*, 772 – 774.
[8] R. Shapiro, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4396 – 4401.
[9] C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier, T. Carell, *Angew. Chem. Int. Ed.* **2018**, *57*, 5943 – 5946; *Angew. Chem.* **2018**, *130*, 6050 – 6054.
[10] S. Becker, C. Schneider, H. Okamura, A. Crisp, T. Amatov, M. Dejmek, T. Carell, *Nat. Commun.* **2018**, *9*, 163.
[11] N. V. Hud, *Nat. Commun.* **2018**, *9*, 5171.
[12] F. Wöhler, *Ann. Phys.* **1828**, *88*, 253 – 256.
[13] P. J. Ramberg, *Ambix* **2000**, *47*, 170 – 195.
[14] L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99 – 123.
[15] F. H. C. Crick, *J. Mol. Biol.* **1968**, *38*, 367 – 379.
[16] L. E. Orgel, *J. Mol. Biol.* **1968**, *38*, 381 – 393.
[17] W. Gilbert, *Nature* **1986**, *319*, 618.
[18] G. F. Joyce, *Nature* **1989**, *338*, 217 – 224.
[19] G. F. Joyce, *PLOS Biol.* **2012**, *10*, e1001323.
[20] N. V. Hud, B. J. Cafferty, R. Krishnamurthy, L. D. Williams, *Chem. Biol.* **2013**, *20*, 466 – 474.
[21] A. E. Engelhart, N. V. Hud, *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, a002196 – a002196.
[22] K.-U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser, *Science* **2000**, *290*, 1347 – 1351.
[23] L. Zhang, A. Peritz, E. Meggers, *J. Am. Chem. Soc.* **2005**, *127*, 4174 – 4175.
[24] P. Karri, V. Punna, K. Kim, R. Krishnamurthy, *Angew. Chem. Int. Ed.* **2013**, *52*, 5840 – 5844; *Angew. Chem.* **2013**, *125*, 5952 – 5956.
[25] S. Zhang, J. C. Chaput, *Curr. Protoc. Nucleic Acid Chem.* **2010**, *42*, 4.40.1 – 4.40.18.
[26] I. Anosova, E. A. Kowal, M. R. Dunn, J. C. Chaput, W. D. Van Horn, M. Egli, *Nucleic Acids Res.* **2016**, *44*, 1007 – 1021.
[27] K. Usami, A. Okamoto, *Org. Biomol. Chem.* **2017**, *15*, 8888 – 8893.
[28] S. Nitta, Y. Furukawa, T. Kakegawa, *Origins Life Evol. Biospheres* **2016**, *46*, 189 – 202.
[29] V. N. Sagi, V. Punna, F. Hu, G. Meher, R. Krishnamurthy, *J. Am. Chem. Soc.* **2012**, *134*, 3577 – 3589.
[30] A. Butlerow, *C. R. Hebd. Seances Acad. Sci.* **1861**, *53*, 145 – 147.
[31] S. Lamour, S. Pallmann, M. Haas, O. Trapp, *Life* **2019**, *9*, 52.
[32] J. Kofoed, J.-L. Reymond, T. Darbre, *Org. Biomol. Chem.* **2005**, *3*, 1850 – 1855.
[33] M. Oberhuber, G. F. Joyce, *Angew. Chem. Int. Ed.* **2005**, *44*, 7580 – 7583; *Angew. Chem.* **2005**, *117*, 7752 – 7755.
[34] O. Björnberg, M. Vodnala, V. Domkin, A. Hofer, A. Rasmussen, G. Andersen, J. Piškur, *Nucleosides Nucleotides Nucleic Acids* **2010**, *29*, 433 – 437.
[35] F. Chemat, M. Poux, *Tetrahedron Lett.* **2001**, *42*, 3693 – 3695.
[36] A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, *Science* **2004**, *303*, 196 – 196.
[37] R. Scorei, V. M. Cimpoiaşu, *Origins Life Evol. Biospheres* **2006**, *36*, 1 – 11.
[38] Y. Furukawa, M. Horiuchi, T. Kakegawa, *Origins Life Evol. Biospheres* **2013**, *43*, 353 – 361.
[39] S. Becker, I. Thoma, A. Deutsch, T. Gehrke, P. Mayer, H. Zipse, T. Carell, *Science* **2016**, *352*, 833 – 836.
[40] H.-J. Kim, Y. Furukawa, T. Kakegawa, A. Bita, R. Scorei, S. A. Benner, *Angew. Chem. Int. Ed.* **2016**, *55*, 15816 – 15820; *Angew. Chem.* **2016**, *128*, 16048 – 16052.
[41] H.-J. Kim, S. A. Benner, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 11315 – 11320.
[42] P. Kubecka, *Weather* **2001**, *56*, 218 – 221.
[43] J. A. McDowell, D. H. Turner, *Biochemistry* **1996**, *35*, 14077 – 14089.
[44] R. Das, J. Karanicolas, D. Baker, *Nat. Methods* **2010**, *7*, 291 – 294.
[45] S. Lyskov, F.-C. Chou, S. Ó. Conchúir, B. S. Der, K. Drew, D. Kuroda, J. Xu, B. D. Weitzner, P. D. Renfrew, P. Sripakdeevong, B. Borgo, J. J. Havranek, B. Kuhlman, T. Kortemme, R. Bonneau, J. J. Gray, R. Das, *PLoS One* **2013**, *8*, e63906.
[46] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, *Acta Crystallogr. Sect. D* **1998**, *54*, 905 – 921.
[47] A. T. Brunger, *Nat. Protoc.* **2007**, *2*, 2728 – 2733.
[48] S. C. Kim, D. K. O'Flaherty, L. Zhou, V. S. Lelyveld, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 13318 – 13323.
[49] D. Ritson, J. D. Sutherland, *Nat. Chem.* **2012**, *4*, 895 – 899.
[50] H. Yu, S. Zhang, J. C. Chaput, *Nat. Chem.* **2012**, *4*, 183 – 187.

Manuscript received: September 16, 2019
Accepted manuscript online: October 1, 2019
Version of record online: October 30, 2019

2.2 SECOND PUBLICATION

TITLE: Amino Acid Modified RNA Bases as Building Blocks of an Early Earth RNA-Peptide World

AUTHORS:

M. Nainyté, F. Müller, G. Ganazzoli, C. Chan, A. Crisp, D. Globisch, T. Carell*

** corresponding author.*

JOURNAL, YEAR, VOLUME, AND PAGE NO.:

Chem. Eur. J., **2020**, *26,* 14856-14860.

SUMMARY:

The RNA world hypothesis differs from certain other contemporary origin of life theories, in that catalytic peptides are thought to have emerged much later in the chemical-evolutionary timescale. Although it is generally accepted, limited few models have been proposed that describe how RNA-based primordial replicators might have transitioned into those that were able to synthesise peptides. In this work, we report the preparation of RNA molecules containing amino acid-modified adenosine bases by use of novel nucleoside phosphoramidite chemistry. The nucleosides themselves contain structures analogous to the ubiquitous adenosine modification t^6A, which is found in all three kingdoms of life and the genome of LUCA, suggesting that they might be molecular fossils of an earlier genetic system. We show that upon charging of an amino acid, these nucleosides lose their ability to base-pair, thus making them available for potential peptide formation, independent of sequence context.

PERSONAL CONTRIBUTIONS:

- Designed, developed, and carried out the original syntheses of compounds 49, 41, and 43 (aminoacyl nucleoside phosphoramidites).
- Developed and optimised chemistry for the solid-phase incorporation of compound 49 into RNA oligonucleotides.
- Contributed to the overall design of the study through discussions with T. Carell and co-authors.
- Edited and revised the manuscript together with all co-authors.

PERMISION AND LICENSING

Oligonucleotides

Amino Acid Modified RNA Bases as Building Blocks of an Early Earth RNA-Peptide World

Milda Nainytė,[a] Felix Müller,[a] Giacomo Ganazzoli,[a] Chun-Yin Chan,[a] Antony Crisp,[a] Daniel Globisch,[b] and Thomas Carell*[a]

Abstract: Fossils of extinct species allow us to reconstruct the process of Darwinian evolution that led to the species diversity we see on Earth today. The origin of the first functional molecules able to undergo molecular evolution and thus eventually able to create life, are largely unknown. The most prominent idea in the field posits that biology was preceded by an era of molecular evolution, in which RNA molecules encoded information and catalysed their own replication. This RNA world concept stands against other hypotheses, that argue for example that life may have begun with catalytic peptides and primitive metabolic cycles. The question whether RNA or peptides were first is addressed by the RNA-peptide world concept, which postulates a parallel existence of both molecular species. A plausible experimental model of how such an RNA-peptide world may have looked like, however, is absent. Here we report the synthesis and physicochemical evaluation of amino acid containing adenosine bases, which are closely related to molecules that are found today in the anticodon stem-loop of tRNAs from all three kingdoms of life. We show that these adenosines lose their base pairing properties, which allow them to equip RNA with amino acids independent of the sequence context. As such we may consider them to be living molecular fossils of an extinct molecular RNA-peptide world.

The RNA-peptide co-evolution hypothesis describes the emergence of self-replicating molecules that contained amino acids and RNA.[1] At the macromolecular level, this tight coexistence of peptides and RNA is established in the ribosome, where encoding and catalytic RNA is supported by proteins.[2] Although we cannot delineate how such an early RNA-peptide world may have looked like, it seems not too implausible to assume that some of the molecular components may have survived until today as vestiges of this extinct world.[3] tRNAs derived from all three kingdoms of life contain a large number of modified bases,[4] and some of them are indeed modified with amino acids.[3] The most wide spread amino acid modified bases are adenosine nucleosides, in which the amino acid is linked via urea connector to the N^6-amino group of the heterocycle as depicted in Figure 1 a. Particularly ubiquitous are adenosine modifications containing the amino acids threonine (t^6A)[5–7] and glycine (g^6A),[8] together with hn^6A.[9,10] Based upon recent phylogenetic analyses and the fact that t^6A is found in all three kingdoms of life, it has been suggested that such amino acid modified bases were already present in the last universal common ancestor (LUCA), from which all life

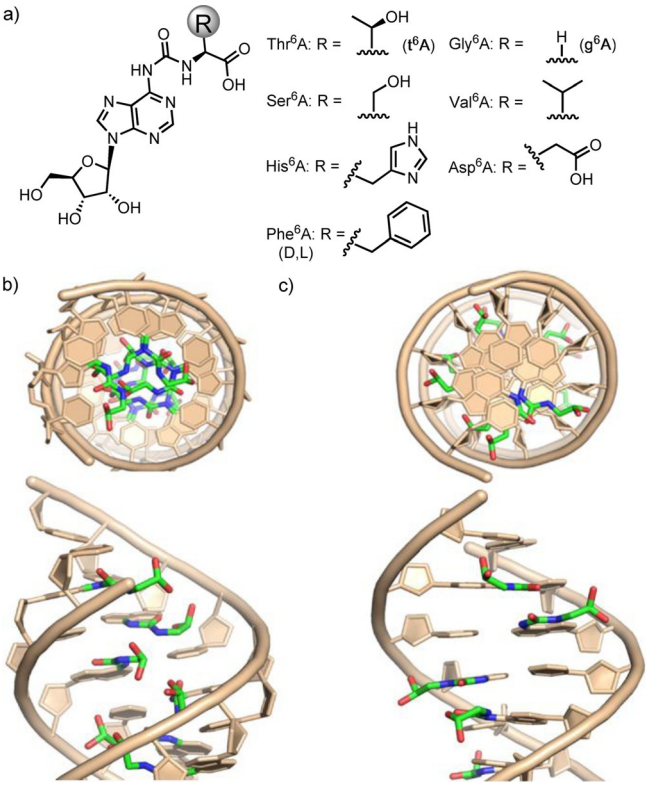


Figure 1. (a) Depiction of the amino acid modified A-bases (aa^{6}A) together with computer visualizations that show how such bases may reside in an (b) A-form RNA duplex and a (c) B-form DNA duplex. The sequence used for the visualization is: 5′-C<u>A</u>U<u>A</u>U<u>A</u>U<u>A</u>UG-3′ with <u>A</u> = g^6A.

[a] M. Nainytė, F. Müller, G. Ganazzoli, C.-Y. Chan, A. Crisp, Prof. Dr. T. Carell
Department of Chemistry, LMU München
Butenandtstr. 5–13, 81377 München (Germany)
E-mail: thomas.carell@lmu.de

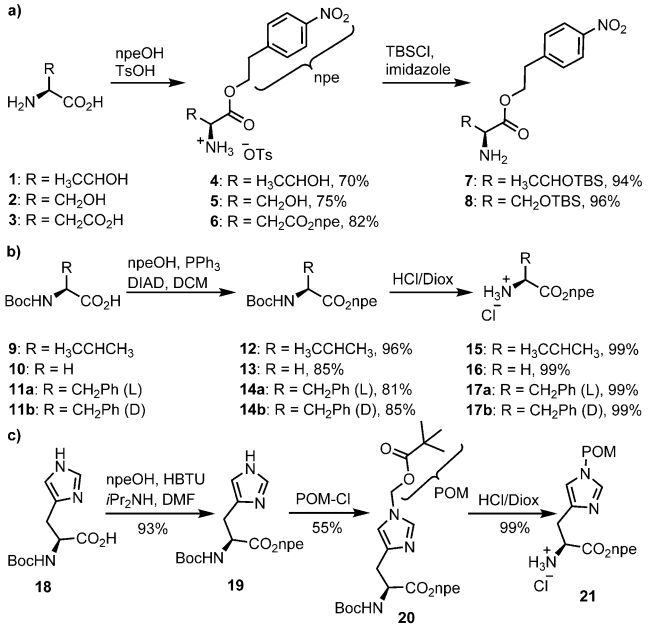
[b] Dr. D. Globisch
Department of Medicinal Chemistry, Uppsala University
Husargatan 3, 75123 Uppsala (Sweden)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/chem.202002929.

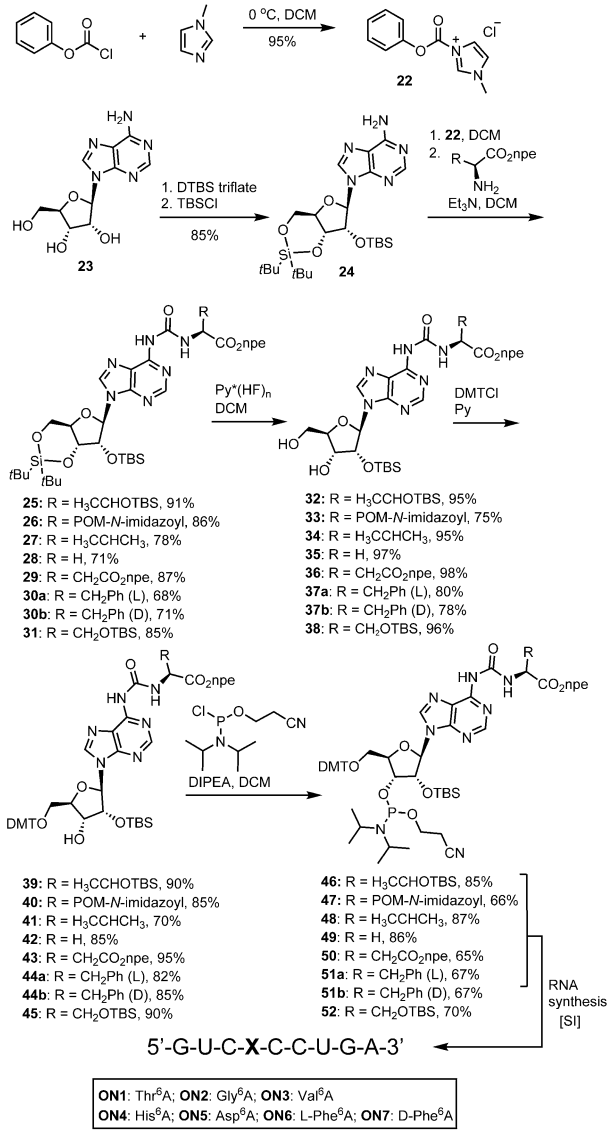
forms descended.[11–14] t^{6}A is for example today found in nearly all ANN decoding tRNAs.[15] We recently reported a plausible prebiotic route to some of these amino acid modified A-bases, which strengthens the idea that they could indeed be living chemical fossils of the extinct RNA-peptide world.[16] Despite the interesting philosophical genotype-phenotype dualism that characterizes these structures and their contemporary importance for the faithful decoding of genetic information, a general synthesis of aa^{6}A modified bases (Figure 1 a) and a systematic study of their properties is lacking.

Here we report the synthesis of a variety of aa^{6}A nucleosides with canonical amino acids (aa = Asp, Gly, His, Phe, Thr,[17] Ser, Val), their incorporation into DNA and RNA and an investigation of how they influence the physicochemical properties of oligonucleotides. We were particularly interested to study how they might affect the stability of RNA and DNA. The computer visualization shows that in A-form RNA (Figure 1 b), the amino acid part of the aa^{6}A base would need to reside inside the helix, shielded from the outside. In the B-form DNA one could imagine a decoration of the major groove with the amino acid side chains as depicted in Figure 1 c.

In the Schemes 1 and 2 we show the synthesis of the different urea linked amino acid A-derivatives (aa^{6}A). We first prepared the amino acid components for the coupling to the A-nucleoside (Scheme 1). Our starting points for Thr^{6}A, Ser^{6}A and Asp^{6}A were the free amino acids **1**–**3**, in which we first transformed all carboxylic acids into the *p*-nitrophenylethyl esters (npe, **4**–**6**).[17] The hydroxy groups of the Thr and Ser compounds were finally protected as TBS-ethers to give the final products **7** and **8** (Scheme 1 a). For Val, Gly and Phe we started with the Boc-protected amino acids **9**–**11**, which we also converted into the npe-esters **12**–**14** using Mitsunobu type chemistry[18] followed by acidic (4 M HCl in dioxane) Boc-deprotection to give the amino acid products **15**–**17** (Scheme 1 b).[19] For His^{6}A, we again started with the Boc-protected amino acid **18** (Scheme 1 c) and used HBTU activation to generate the npe ester **19**. Protection of the imidazole N^τ with POM-chloride followed again by Boc-deprotection furnished the ready to couple amino acid **21**.



Scheme 1. Synthesis of the amino acid building blocks as needed for the coupling to the nucleoside A to give Thr^{6}A, Ser^{6}A, Asp^{6}A, Val^{6}A, Gly^{6}A, Phe^{6}A and His^{6}A.

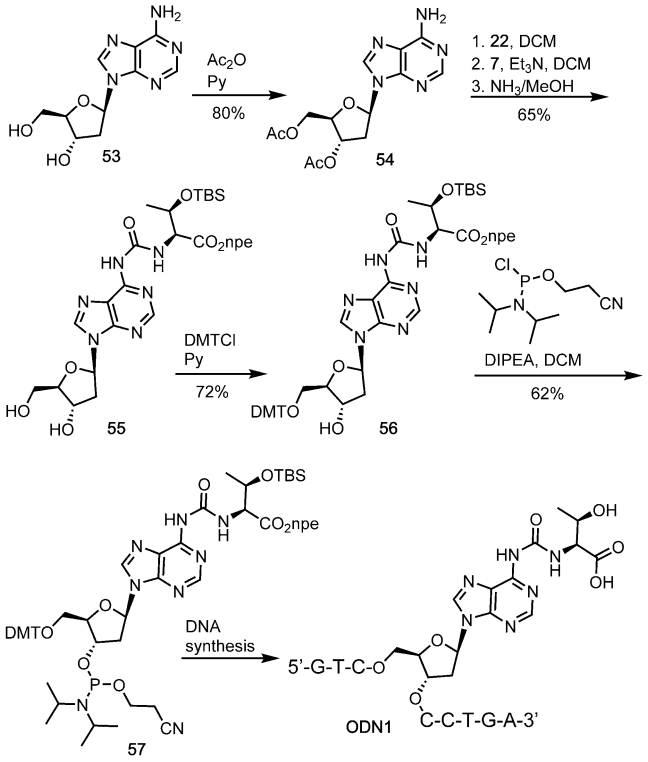


Scheme 2. Synthesis of phosphoramidite building blocks of Thr^{6}A, Ser^{6}A, Asp^{6}A, Val^{6}A, Gly^{6}A, Phe^{6}A and His^{6}A and their incorporation into RNA.

The connection of the amino acid with the A-nucleoside via the urea moiety was next carried out as depicted in Scheme 2. We first treated phenyl chloroformate with *N*-methylimidazole to obtain the 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (**22**).[20] Adenosine was converted in parallel into the cyclic 3′,5′-silyl protected nucleoside, followed by conversion of the 2′-OH group into the TBS-ether.[21] The reaction of compound **24** with the activated carbonate and the corresponding amino acid, provided in all cases the amino acid coupled products **25**–**31** in good to excellent yields. Subsequent cleavage of the

cyclic silylether with HF·pyridine complex,[22,23] protection of the 5′-OH group with dimethoxytritylchloride (DMTCl)[24] allowed the final conversion of the compounds into the corresponding phosphoramidites **46**–**52**. Standard solid phase RNA chemistry[25–31] was subsequently employed to prepare RNA strands containing the individual aa^{6}A nucleosides stably embedded. The standard RNA synthesis protocol did not require any adjustment. In all cases we observed fair coupling of the aa^{6}A phosphoramidites and no decomposition during deprotection. Deprotection required three steps. First, with DBU in THF at r.t. for 2 h we cleaved the npe-protecting group. Second, we deprotected the bases and cleaved from the solid support with aqueous NH_3/$MeNH_2$. Finally, we removed the 2′-silyl group with HF in NEt_3.

In order to investigate how aa^{6}A bases would affect the stability of DNA duplexes we also prepared as a representative molecule t^{6}dA as depicted in Scheme 3. To this end we first acetyl-protected dA **53**,[32] performed the coupling of the protected threonine with the activated carbonate **22**, cleaved the acetyl groups and converted the nucleoside subsequently into the 5′-DMT protected phosphoramidite **57**. The purification of compound **57** was quite difficult due to its high polarity. We needed to use rather polar mixture of EtOAc/Hex (2/1) as the mobile phase for the chromatographic separation. This provided the phosphoramidite **57**, however the material had a lower purity in comparison to the RNA phosphoramidites. Nevertheless, solid phase DNA synthesis and deprotection of the DNA strand **ODN1** proceeded again smoothly and in high yields.



Scheme 3. Synthesis of t^{6}dA phosphoramidite and its incorporation into DNA.

Figure 2a shows as an example the raw HPL-chromatograms of **ON1** (RNA strand with embedded t^{6}A) and the corresponding chromatogram after purification (inset) together with the obtained MALDI-TOF mass spectrum (Figure 2b). The chromatograms of the raw material show a good quality of the obtained RNA material. The analytical chromatogram after purification and the MALDI-TOF data prove the purity of the finally obtained RNA oligonucleotide and the integrity of the t^{6}A-containing RNA strand.

Figure 2c and 2d show the same data set for the t^{6}dA containing DNA oligonucleotide (**ODN1**), proving again the successful synthesis of t^{6}dA containing oligonucleotide. The aa^6(d)A nucleosides can exist in two different conformations.[33] The first, s-*trans*, maintains the Watson–Crick hydrogen bonding capabilities with the urea amino acid oriented towards the imidazole ring system (Figure 3a). This allows formation of a Hoogsteen type 7-membered ring H-bond with the N^7. In the corresponding s-*cis*-conformation, the urea amino acid orients towards the Watson–Crick side thereby establishing a typically strong intramolecular 6-membered H-bond with N^1 (Figure 3b). In order to investigate if the embedding of the amino acid would enforce s-*trans*-conformation and hence Watson–Crick H-bonding, we measured melting points of all aa^{6}A containing RNA strands and of the t^{6}dA containing DNA strand hybridized to the corresponding counter strands (Figure 3). In the RNA:RNA situation we noted for all aa^{6}A strands that we investigated, a single clear melting point, showing that only one conformer of the aa^{6}A base likely exists in the RNA:RNA duplexes. In situation where the aa^{6}A base exists in two different stable conformations, one would expect a more complex melting behaviour. In all cases we saw that the melting point is strongly reduced by 10–15 °C. When we embedded two aa^{6}A building blocks into a short RNA strand no duplex formation

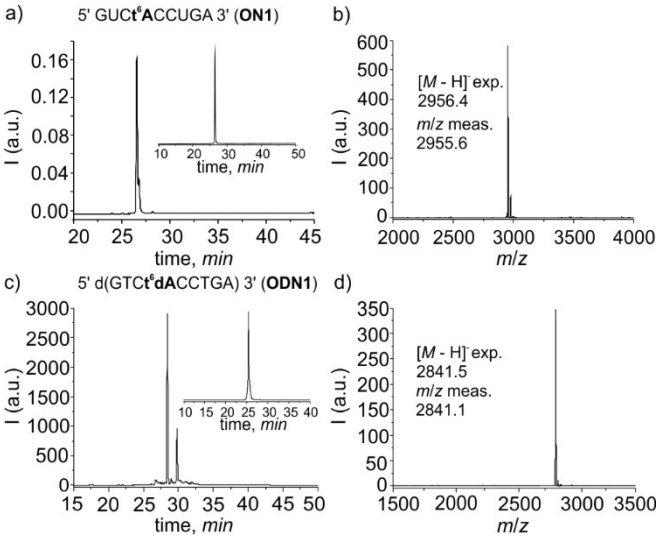
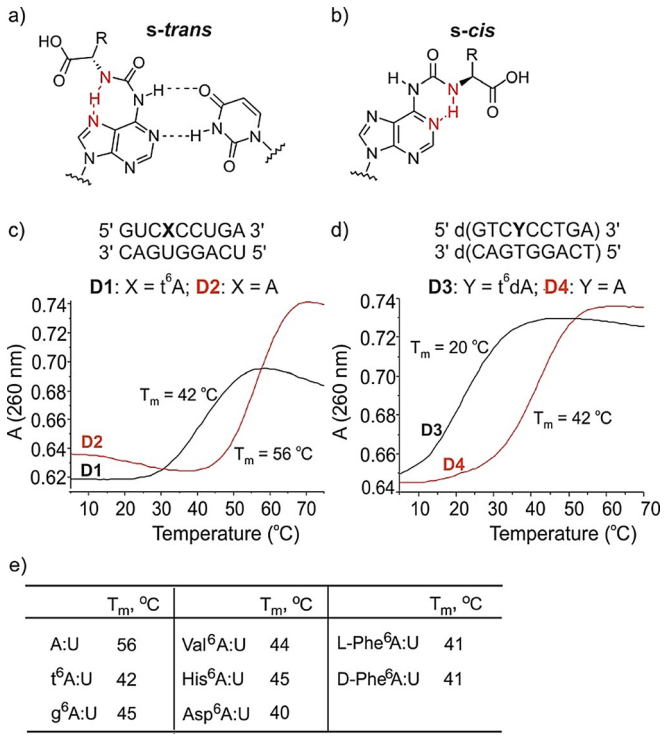


Figure 2. (a) Raw-HPL chromatogram of **ON1**, with the inset showing the HPL-chromatogram of purified **ON1**; (b) MALDI-TOF mass spectrum of **ON1** after purification; (c) raw-HPL chromatogram of **ODN1**, with the inset showing the HPL chromatogram of purified **ODN1**; (d) MALDI-TOF mass spectrum of **ODN1** after purification.



	T_m, °C		T_m, °C		T_m, °C
A:U	56	Val^{6}A:U	44	L-Phe^{6}A:U	41
t^{6}A:U	42	His^{6}A:U	45	D-Phe^{6}A:U	41
g^{6}A:U	45	Asp^{6}A:U	40		

Figure 3. (a, b) Possible conformation, base pairing and intramolecular H-bond of aa^{6}A; (c, d) melting curves measured for t^{6}A containing RNA:RNA duplexes and of a t^{6}dA containing DNA:DNA duplex in comparison with the duplexes containing canonical (d)A:(d)T base pairs; (e) table of the determined melting points.

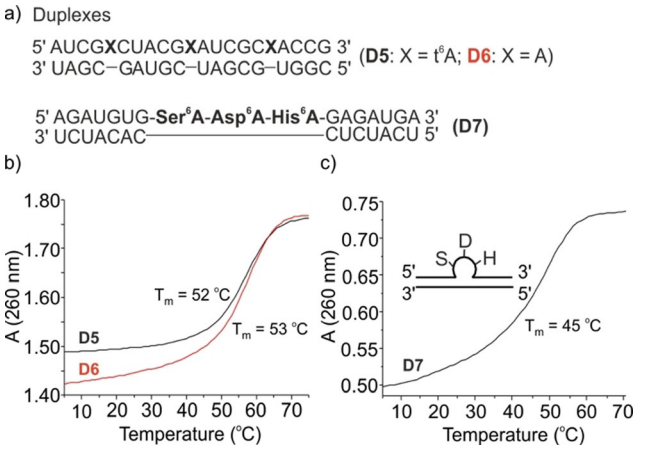


Figure 4. (a) Depiction of the RNA structures containing aa^{6}A nucleobases in extrahelical positions forming either three little bulges or assembling a Ser-Asp-His triad known as the catalytic triad in serine proteases; (b, c) depiction of melting curves of duplexes **D5**, **D6**, **D7**; S: serine, D: aspartate, H: histidine.

was obtained. Even stronger reduction of the melting point was observed for the DNA duplex containing one t^{6}dA. Here, we also saw just one sharp melting point and a reduction of the T_m by over 20 °C. These data show that the aa^{6}A bases and among them t^{6}A and g^{6}A are unable to base pair. Although we have no direct proof of the structure the data argue for a preferred s-*cis*-conformation (Figure 3 b) in agreement with the literature.[34]

This conclusion is also supported by the observation that irrespective of the chirality of the attached amino acid (L- versus D-Phe^{6}A), we measured the same melting temperature. This would not be expected if the s-*trans*-conformation and base pairing would be possible. These data suggest that aa^{6}A nucleosides within RNA position a given amino acid outside the A-form helix in an unpaired situation and hence independent from the counterbase. As such, multiple aa^{6}A containing RNA strands would be structures in which the RNA part is decorated by the amino acid side chains. In order to show that RNA-structures containing multiple amino acids as representatives of an RNA-peptide world can stably form, we prepared two RNA duplexes (Figure 4). In the first (**D5**), we placed three t^{6}A bases as extra bases in an otherwise undisturbed RNA duplex. Indeed, now the stability of this duplex was indistinguishable from the same construct containing just canonical bases (**D6**). Finally, we prepared an RNA duplex **D7**, in which we placed the amino acids Ser-Asp-His directly next to each other to simulate what is known in the peptide world as the catalytic triad present in serine peptidases.[35] Again in this case a stable duplex structure forms with the three aa^{6}A bases creating a loop. Although we do not show any catalytic activity here, we believe that it is easily imaginable that if these amino acids are properly positioned in a stably folded RNA the structure could gain catalytic properties.

The melting data show, that aa^{6}A bases alone are unable to establish base pairing, which hinder them to encode sequence information. On the other side, these bases allow the incorporation of amino acids into RNA structures irrespective of the counterbase. Because RNAs are mostly stably folded structures in which many bases are not involved in any base pairing or establish no Watson–Crick interactions the amino acid adenosine nucleosides allow the stable incorporation of amino acid functionality into RNA.

In summary, here we investigated the synthesis and properties of aa^{6}A nucleoside-amino acid conjugates, some of which (t^{6}A, g^{6}A, hn^{6}A) are today found as key components in the tRNAs of many species. In these tRNAs the aa^{6}A nucleosides reside at the general purine position 37 adjacent to the anticodon loop. They are not involved in base pairing but fine tune the codon-anticodon interaction to enable faithful translation of information into a peptide sequence.[36] Here we show that these bases are indeed unable to base pair. They have to be placed outside the pairing regime that is needed for RNA folding. As such they function as anchors that allow the connection of amino acid to RNA structures independent of the counterbase. The side chains are then available to equip RNA with additional functions that might have been beneficial in an early RNA-peptide world. The fact that aa^{6}A nucleosides are stable structures and until today broadly found in today's RNA make them prime candidates to develop idea about the chemical constitution of the vanished RNA-peptide world.

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for financial support via SFB1309 (325871075) and SPP1784 (255344185). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement n° EPiR 741912) and through a H2020 Marie Skłodowska-Curie Action (LightDyNAmics, 765866). We thank Dr. Tynchtyk Amatov for the initial synthesis of the t^{6}A-phosphoramidite. We thank Dr. Markus Müller for helpful discussions and preparing Figure 1. Open access funding enabled and organized by Projekt DEAL.

Conflict of interest

The authors declare no conflict of interest.

Keywords: amino acid nucleosides · origin of life · prebiotic chemistry · RNA world · RNA-peptide world

[1] P. T. S. van der Gulik, D. Speijer, *Life* **2015**, *5*, 230 – 246.
[2] D. L. J. Lafontaine, D. Tollervey, *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 514 – 520.
[3] H. Grosjean, E. Westhof, *Nucleic Acids Res.* **2016**, *44*, 8020 – 8040.
[4] M. A. Machnicka, K. Milanowska, O. O. Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Howkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, *Nucleic Acids Res.* **2013**, *41*, D262 – D267.
[5] G. B. Chheda, R. H. Hall, J. Mozejko, D. I. Magrath, M. P. Schweizer, L. Stasiuk, P. R. Taylor, *Biochemistry* **1969**, *8*, 3278 – 3282.
[6] S. Takemura, M. Murakami, M. Miyazaki, *J. Biochem.* **1969**, *65*, 489 – 491.
[7] D. M. Powers, A. Peterkofsky, *Biochem. Biophys. Res. Commun.* **1972**, *46*, 831 – 838.
[8] M. P. Schweizer, K. McGrath, L. Bacznyskyj, *Biochem. Biophys. Res. Commun.* **1970**, *40*, 1046 – 1052.
[9] H. Grosjean in *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution* (Ed.: H. Grosjean), Landes Bioscience, Austin, Texas, pp. 1 – 18.
[10] D. M. Reddy, P. F. Crain, C. G. Edmonds, R. Gupta, T. Hashizume, K. O. Stetter, F. Widdel, J. A. McCloskey, *Nucleic Acids Res.* **1992**, *20*, 5607 – 5615.
[11] M. Di Giulio, *J. Theor. Biol.* **1998**, *191*, 191 – 196.
[12] H. Grosjean, V. de Crécy-Lagard, G. R. Björk, *Trends Biochem. Sci.* **2004**, *29*, 519 – 522.
[13] E. Szathmáry, *Trends Genet.* **1999**, *15*, 223 – 229.
[14] L. Perrochia, E. Crozat, A. Hecker, W. Zhang, J. Bareille, B. Collinet, H. van Tilbeurgh, P. Forterre, T. Basta, *Nucleic Acids Res.* **2013**, *41*, 1953 – 1964.
[15] P. C. Thiaville, D. Iwata-Reuyl, V. de Crécy-Lagard, *RNA Biol.* **2014**, *11*, 1529 – 1539.
[16] C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier, T. Carell, *Angew. Chem. Int. Ed.* **2018**, *57*, 5943 – 5946; *Angew. Chem.* **2018**, *130*, 6050 – 6054.
[17] V. Boudou, J. Langridge, A. van Aerschot, C. Hendrix, A. Millar, P. Weiss, P. Herdewijn, *Helv. Chim. Acta* **2000**, *83*, 152 – 161.
[18] O. Mitsunobu, M. Yamada, *Bull. Chem. Soc. Jpn.* **1967**, *40*, 2380 – 2382.
[19] G. Leszczynska, P. Leonczak, A. Dziergowska, A. Malkiewicz, *Nucleosides Nucleotides Nucleic Acids* **2013**, *32*, 599 – 616.
[20] F. Himmelsbach, B. S. Schultz, T. Trichtinger, R. Charubala, W. Pfleiderer, *Tetrahedron* **1984**, *40*, 59 – 72.
[21] V. Serebryany, L. Beigelman, *Tetrahedron Lett.* **2002**, *43*, 1983 – 1985.
[22] B. M. Trost, C. G. Caldwell, *Tetrahedron Lett.* **1981**, *22*, 4999 – 5002.
[23] M. Sekine, S. Limura, K. Furusawa, *J. Org. Chem.* **1993**, *58*, 3204 – 3208.
[24] H. Schaller, G. Weinmann, B. Lerch, H. G. Khorana, *J. Am. Chem. Soc.* **1963**, *85*, 3821 – 3827.
[25] H. Seliger, D. Zeh, G. Azuru, J. B. Chattopadhyaya, *Chem. Scr.* **1983**, *22*, 95 – 102.
[26] M. H. Caruthers, D. Dellinger, K. Prosser, A. D. Brone, J. W. Dubendorff, R. Kierzek, M. Rosendahl, *Chem. Scr.* **1986**, *26*, 25 – 30.
[27] R. Kierzek, M. H. Caruthers, C. E. Longfellow, D. Swinton, D. H. Turner, S. M. Freier, *Biochemistry* **1986**, *25*, 7840 – 7846.
[28] I. Hirao, M. Ishikawa, K. Miura, *Nucleic Acids Symp. Ser.* **1985**, *16*, 173 – 176.
[29] H. Tanimura, T. Fukuzawa, M. Sekine, T. Hata, J. W. Efcavitch, G. Zon, *Tetrahedron Lett.* **1988**, *29*, 577 – 578.
[30] H. Tanimura, M. Sekine, T. Hata, *Nucleosides Nucleotides* **1986**, *5*, 363 – 383.
[31] H. Tanimura, T. Imada, *Chem. Lett.* **1990**, *19*, 1715 – 1718.
[32] E. M. van der Wenden, J. K. von Frijtag Drabbe Künzel, R. A. A. Mathot, M. Danhof, A. P. IJzerman, W. Soudijn, *J. Med. Chem.* **1995**, *38*, 4000 – 4006.
[33] R. Parthasarathy, J. M. Ohrt, G. B. Chheda, *Biochem. Biophys. Res. Commun.* **1974**, *60*, 211 – 218.
[34] F. V. Murphy IV, V. Ramakrishnan, A. Malkiewicz, P. F. Agris, *Nat. Str. Mol. Biol.* **2004**, *11*, 1186 – 1191.
[35] L. Polgár, *Cell. Mol. Life Sci.* **2005**, *62*, 2161 – 2172.
[36] P. C. Thiaville, R. Legendre, D. Rojas-Benítez, A. Baudin-Baillieu, I. Hatin, G. Chalancon, A. Glavic, O. Namy, V. de Crécy-Lagard, *Microb. Cell.* **2016**, *3*, 29 – 45.

Manuscript received: June 18, 2020
Accepted manuscript online: June 23, 2020
Version of record online: October 14, 2020

2.3 THIRD PUBLICATION

TITLE: Noncanonical RNA Nucleosides as Molecular Fossils of an Early Earth—Generation by Prebiotic Methylations and Carbamoylations

AUTHORS:

C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier, T. Carell*

** corresponding author.*

JOURNAL, YEAR, VOLUME, AND PAGE NO.:

Angew. Chem. Int. Ed., **2018**, *57*, 5943-5946

SUMMARY:

According to the RNA World hypothesis, life on earth is thought to have been preceded by self-replicating molecules constructed from the four canonical RNA nucleotides (A, C, G, and U). It has, however, also been speculated that certain modified nucleosides may have emerged in parallel to the canonical bases which would have provided a greater degree of functional diversity on the early earth. Here, we demonstrate the synthesis of a number of methyl-, and carbamoyl-modified RNA nucleosides under conditions compatible with the prebiotic earth. Specifically, we show that *N'*-substituted *N*-methylureas react regioselectively with sodium nitrite to provide *N*-nitrosated species capable of affecting various methylations and carbamoylations. Our experiments thus provide a plausible scenario for the prebiotic origins of certain extant, non-canonical nucleobases.

PERSONAL CONTRIBUTIONS:

- Carried out experiments demonstrating the reactivity of *N*-nitrosated methylureas with amine nucleophiles in the presence of acid and base, with supervision from S. Becker and H. Okamura.
- Contributed to the design and synthesis of an alternative pathway to N^6-aminoacylated nucleosides, involving formamido pyrimidines (FaPys) as reactive intermediates.
- Synthesis of, and towards certain N^6-carbamoylated adenosine nucleosides and their ms^2-modified versions, which were used to assess the outcome of prebiotic reactions in the study.
- Contributed to the overall design of the study though discussions with C. Schneider, S. Becker, H. Okamura and T. Carell.
- Edited and revised the manuscript together with all co-authors.

PERMISION AND LICENSING

Prebiotic Chemistry

International Edition: DOI: 10.1002/anie.201801919
German Edition: DOI: 10.1002/ange.201801919

Noncanonical RNA Nucleosides as Molecular Fossils of an Early Earth—Generation by Prebiotic Methylations and Carbamoylations

*Christina Schneider, Sidney Becker, Hidenori Okamura, Antony Crisp, Tynchtyk Amatov, Michael Stadlmeier, and Thomas Carell**

Abstract: *The RNA-world hypothesis assumes that life on Earth started with small RNA molecules that catalyzed their own formation. Vital to this hypothesis is the need for prebiotic routes towards RNA. Contemporary RNA, however, is not only constructed from the four canonical nucleobases (A, C, G, and U), it also contains many chemically modified (non-canonical) bases. A still open question is whether these noncanonical bases were formed in parallel to the canonical bases (chemical origin) or later, when life demanded higher functional diversity (biological origin). Here we show that isocyanates in combination with sodium nitrite establish methylating and carbamoylating reactivity compatible with early Earth conditions. These reactions lead to the formation of methylated and amino acid modified nucleosides that are still extant. Our data provide a plausible scenario for the chemical origin of certain noncanonical bases, which suggests that they are fossils of an early Earth.*

More than 120 modified bases have been identified in RNA that are important for correct folding into complex three-dimensional structures and for fine-tuning RNA/RNA and RNA/protein interactions.[1–3] Modified nucleosides are, for example, found in proximity to the anticodon stem loop in tRNA, where they are involved in translation of the genetic code.[4,5] Methylated nucleosides such as m^6A are involved in regulating mRNA stability,[6] splicing,[7,8] translation,[9–11] and X-chromosome inactivation.[12] Another methylated nucleoside, m^7G, is part of the 5′-cap structure of eukaryotic mRNA.[13]

The RNA-world hypothesis postulates that life started with self-replicating RNA molecules that were amenable to the processing of chemical evolution through replication, randomization, and selection.[14] Since RNA is able to store genetic information and perform catalytic processes, the hypothesis further posits that an early replicating cell could proliferate and maintain a primitive metabolism in the absence of coded proteins. Noncoded polypeptides[15,16] and simple anabolic pathways[17,18] may have supported an early RNA-based metabolism.

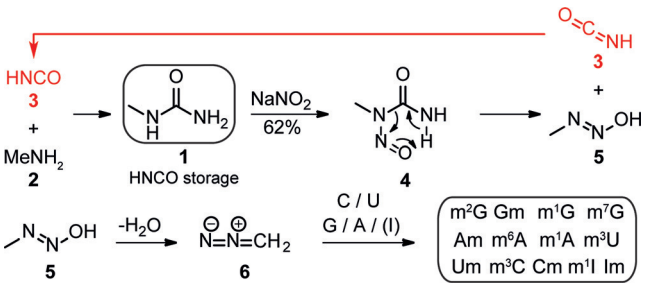
[*] C. Schneider, S. Becker, Dr. H. Okamura, A. Crisp, Dr. T. Amatov, M. Stadlmeier, Prof. Dr. T. Carell
Center for Integrated Protein Science (CiPS^M) at the Department of Chemistry, LMU München
Butenandtstrasse 5–13, 81377 München (Germany)
E-mail: Thomas.Carell@lmu.de
Homepage: http://www.carellgroup.de
Supporting information and the ORCID identification numbers for the authors of this article can be found under: https://doi.org/10.1002/anie.201801919.

This hypothesis requires the presence of the key building blocks of life, such as nucleosides and amino acids, or of primitive anabolic processes that led to their formation.[19] This raises the question of whether life began with only the four canonical bases (A, C, G, and U)[20,21] or if an early pre-RNA was chemically more diverse[22] and contained non-canonical nucleosides.[22] Those noncanonical bases that are still found in RNA might be considered fossils of this early phase of chemical evolution.[23–25] Finding evidence for this idea requires simple chemical synthetic routes compatible with geochemical models of early Earth that generate these noncanonical bases. Here we show that the majority of methylated nucleosides, which play important roles in RNAs of all three domains of life, can be prebiotically generated by the reaction of canonical nucleosides with nitrosylated *N*-methylurea (**1**; Scheme 1).[26]

NO_2^- was potentially available on the early Earth from NO and NO_2,[27] which are formed during lightning in an N_2 atmosphere.[28] Alternatively, NO can form by the reaction of N_2 with CO_2 in hot impact plumes.[29]

Besides the methylated RNA bases, we also find nucleosides modified with amino acids among the many contemporary noncanonical RNA bases.[30,31] They are directly involved in decoding the genetic information.[32,33] We show that our NO_2^--based reactions also provide these modified bases, which suggests an early intimate contact between nucleobases and amino acids that might have formed the basis for the co-evolution of RNA and proteins and the establishment of primitive protometabolic pathways.

The synthetic route starts with methylurea (**1**), which is one of the molecules that was likely present on the early Earth.[34] Methylurea (**1**) is, for example, formed by the reaction of ammonia with methyl isocyanate, which was detected on comet 67P/Churyumov-Gerasimenko.[35,36] Methylurea (**1**) was also shown to form directly in the



Scheme 1. Reactions that lead to the formation of methylated derivatives of canonical nucleobases that are today found in RNA in all three domains of life. Methylurea functions as a storage molecule for reactive isocyanic acid.

Urey–Miller experiment[37] and it is available in high yields from the reaction of methylamine (**2**) with HNCO (**3**) in water (90%).[38] HNCO in turn was detected in interstellar gases[39] and, likewise, on comet 67P/Churyumov-Gerasimenko.[35,40] Urea itself is also known to decompose into ammonium isocyanate.[41,42] Despite the potential formation of **3**, however, it is difficult to conceive the accumulation of HNCO (**3**) because of its high reactivity. If, however, a small amount of methylurea (**1**) is present, it can readily react with NO^+ (Scheme 1). Methylurea (**3**) is easily nitrosylated, which gives *N*-methyl-*N*-nitrosourea (**4**)[26] in a yield of 62%. This compound physically separates as a foam from the aqueous phase, which could potentially allow **4** to accumulate, so that it may have been locally available at high concentrations. Under slightly basic conditions, for example in the presence of borax (reported to be important for ribose-forming reactions),[43] **4** quickly decomposes to furnish 1-hydroxy-2-methyldiazene (**5**) under liberation of HNCO (**3**). As such, only small amounts of HNCO are required to help convert $MeNH_2$ and $NaNO_2$ into 1-hydroxy-2-methyldiazene (**5**). 1-Hydroxy-2-methyldiazene (**5**) in turn eliminates water and decomposes to diazomethane (**6**), which is a common methylating agent.[44] Since all starting materials are likely components of the organic matter on the early Earth, it is, therefore, plausible that diazomethane was an accessible component. The controlled release of **6** from the stable precursor methylurea (**1**) could have made it available for chemical transformations, despite its high reactivity and consequently short half-life time on the early Earth.

When we performed this base-catalyzed formation of diazomethane (**6**) in the presence of the canonical nucleobases, we obtained a large set of methylated compounds (Figure 1). For the experiment we dissolved the nucleosides in a 1:1 mixture of borate buffer and formamide. Formamide is accessible under early Earth conditions through the reaction of HCN with H_2O.[45] *N*-Methyl-*N*-nitrosourea (**4**) was then added to the nucleoside mixture in one portion. After one hour at 70 °C, samples were taken and analyzed by LC-MS and tandem mass spectrometry (Figure 1). To correctly assign the resulting methylated nucleosides, co-injections with synthetic reference compounds were performed (see the Supporting Information). The products were further elucidated by analysis of the fragmentation patterns in $LC\text{-}MS^2$ experiments. When we performed the reaction in the presence of adenosine, we obtained m1A, Am, and m6A, together with the 3′- and 5′-methylated derivatives (marked as mxA; see the Supporting Information). When guanosine was methylated under the same conditions, we detected m7G (7%) as well as Gm, m1G, and m2G, all of which are known noncanonical bases. In the presence of cytidine, the bases m3C and Cm were generated. Furthermore, the reaction of uridine furnished the methylated compounds Um and m3U. m3U was formed in a high yield of 11%. We also investigated the methylation reaction with inosine (I), the hydrolysis product of A.[46,47] When inosine was subjected to the same conditions, we detected the formation of Im and m1I (see the Supporting Information). Importantly, nearly all the methylated nucleosides that we observed are today found in RNAs of all three domains of life.[2,48]

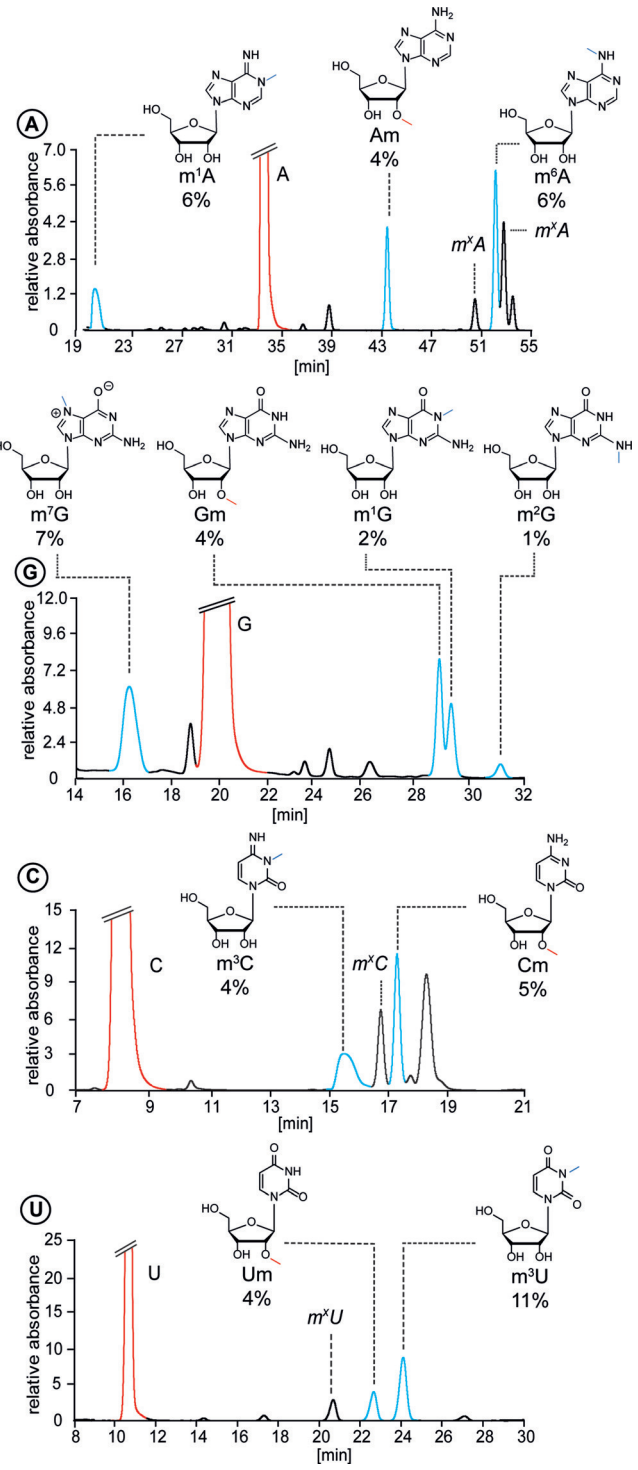


Figure 1. HPLC traces of the reaction mixtures obtained in the reaction of *N*-methyl-*N*-nitrosourea (**4**) in the presence of the canonical nucleobases A, G, C, and U. The modified nucleosides are shown in blue, the canonical ones in red. Peaks labeled with "mˣ" were identified as sugar-modified nucleosides based on data from fragmentation studies (see the Supporting Information).

We next asked the question of whether the simple reactions could be used to enable the attachment of larger chemical moieties, such as amino acids, to the canonical nucleobases to give RNA modifications such as t^6A and g^6A.

This was indeed possible (Figure 2) when we replaced HNCO by methyl isocyanate (CH_3NCO, **7**). CH_3NCO (**7**) can be generated under prebiotic conditions by UV irradiation of CH_4 and HNCO. In an aqueous environment, we observed that **7** reacts rapidly with amino acids such as glycine (**8a**) and threonine (**8b**) to give the corresponding methylurea derivatives **9** (Figure 2) in nearly quantitative yields. Compounds **9a** and **b** can be nitrosylated[49] under the same conditions as **1** to form the nitroso compounds **10a** and **10b** in high yields of 82–95%. A pH change to slightly basic conditions through the use of either phosphate or borate buffer converts the intermediate nitroso compounds **10a**,**b** into the isocyanates of the corresponding amino acids **11a**,**b**. Upon treatment with adenosine, these intermediates react to give the corresponding N^6-derivatives g^6A and t^6A. Since the reaction takes place under basic conditions, not only N^6 but also the 2′-, 3′-, and 5′-hydroxy groups can react with the isocyanate derivative of the amino acids (Figure 2). Interestingly, the selectivity of the reaction can be controlled to favor the N^6 position by the addition of a Ni^{2+} salt, which is generated during prebiotic formation of nucleoside.[22] At the same time, CH_2N_2 (**6**) is formed, which can facilitate subsequent methylations (see the Supporting Information). Interestingly, the amino acid modified nucleosides that are formed as described here, are present today in all three domains of life.[2,48]

Recently, a comparative phylogenetic analysis[50] has suggested that noncanonical bases were likely already present in the ancient parent of all life on Earth, known conventionally as LUCA (the last universal common ancestor). An overlay of the nucleosides accessed in this study with those derived from the genetic analysis shows surprising consensus

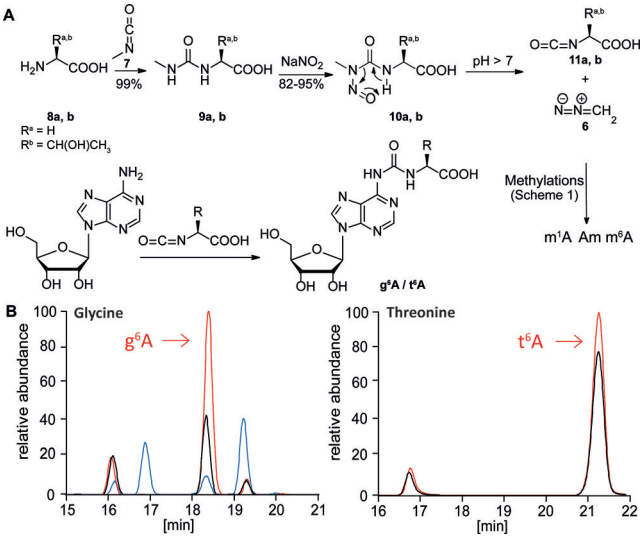


Figure 2. A) Plausible reaction scheme for prebiotic access to t^6A and g^6A. B) MS chromatograms of the reactions of *N*-methylurea derivatives **9a**,**b** with the canonical nucleoside A under formation of isocyanates of the corresponding amino acids. The chromatogram in blue shows the reaction without the addition of Ni^{2+} salts and the ones in black represents the reaction in the presence of Ni^{2+} salts. Co-injections with synthetic standards are shown in red. The additional peaks arise from the reaction of the sugars with the amino acid isocyanate. The selectivity can be increased by the addition of $[Ni(ClO_4)_2]$.

(Figure 3). Most of the simple modifications that were present in LUCA could also be formed by the reactions presented here.

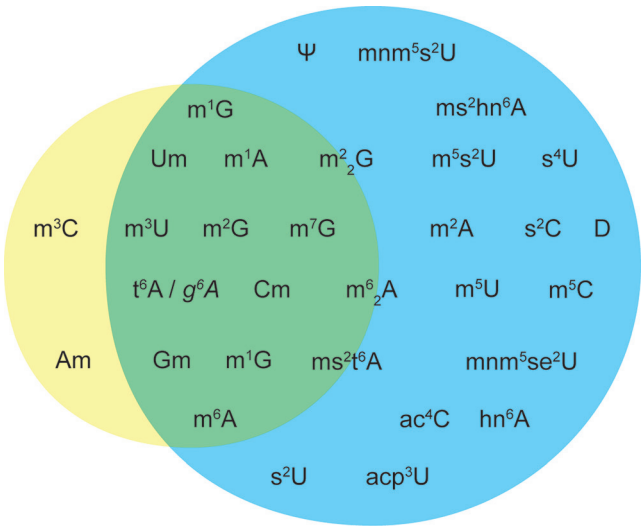


Figure 3. Noncanonical nucleosides generated in this study (yellow area) and noncanonical nucleosides that were found based on phylogenetic analysis to be likely early nucleobases bases in the biosphere (blue area).[50] Modified nucleosides that were found in both studies are shown in the green area. ms^2t^6A, m^2_2G, and m^6_2A were placed at the border since they could also be generated using the here-described reactions starting from ms^2A, m^2G, and m^6A.

In summary, we report a simple cascade reaction that starts from isocyanic acid, methylisocyanate, methylamine, ammonia, and sodium nitrite. In this cascade the unstable molecule isocyanic acid (**3**) is captured by methylamine and stored in the form of methylurea. It can be released under basic conditions from *N*-methyl-*N*-nitrosourea (**4**), which is produced by nitrosylation of methylurea (**1**). These reactions allow us to convert the canonical pyrimidine and purine bases, for which prebiotically plausible formation processes were recently described,[20,21,51,52] into noncanonical nucleosides. As such, our results provide chemical evidence that the canonical and many noncanonical ribonucleosides can form spontaneously under plausible prebiotic conditions. The here described reactions can be linked to the nitrosylation reactions that were recently reported to enable the parallel formation of canonical and noncanonical bases.[22] The noncanonical bases, particularly the amino acid modified purines, potentially increase the chemical diversity of RNA to broaden its folding and catalytic capabilities. This complements ideas that noncanonical base pairs might have existed in pre-RNA.[22,53]

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for financial support through SFB1032 (TP-A5), SPP1784, CA275/11-1, and the Excellence Cluster CiPS^M. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation

programme (grant agreement no. EPiR 741912). We thank Dr. Markus Müller for comments on this manuscript.

Conflict of interest

The authors declare no conflict of interest.

Keywords: methylation · nucleosides · nucleoside modification · origin of life · prebiotic chemistry

How to cite: *Angew. Chem. Int. Ed.* **2018**, *57*, 5943 – 5946
Angew. Chem. **2018**, *130*, 6050 – 6054

[1] T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110 – 7131; *Angew. Chem.* **2012**, *124*, 7220 – 7242.
[2] M. A. Machnicka, K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, *Nucleic Acids Res.* **2012**, *41*, D262 – D267.
[3] A. Czerwoniec, S. Dunin-Horkawicz, E. Purta, K. H. Kaminska, J. M. Kasprzak, J. M. Bujnicki, H. Grosjean, K. Rother, *Nucleic Acids Res.* **2009**, *37*, D118 – D121.
[4] P. C. Thiaville, B. El Yacoubi, C. Köhrer, J. J. Thiaville, C. Deutsch, D. Iwata-Reuyl, J. M. Bacusmo, J. Armengaud, Y. Bessho, C. Wetzel, *Mol. Microbiol.* **2015**, *98*, 1199 – 1221.
[5] J. A. Kowalak, J. J. Dalluge, J. A. McCloskey, K. O. Stetter, *Biochemistry* **1994**, *33*, 7869 – 7876.
[6] X. Wang, Z. Lu, A. Gomez, G. C. Hon, Y. Yue, D. Han, Y. Fu, M. Parisien, Q. Dai, G. Jia, *Nature* **2014**, *505*, 117.
[7] X. Zhao, Y. Yang, B.-F. Sun, Y. Shi, X. Yang, W. Xiao, Y.-J. Hao, X.-L. Ping, Y.-S. Chen, W.-J. Wang, *Cell Res.* **2014**, *24*, 1403.
[8] W. Xiao, S. Adhikari, U. Dahal, Y.-S. Chen, Y.-J. Hao, B.-F. Sun, H.-Y. Sun, A. Li, X.-L. Ping, W.-Y. Lai, *Mol. Cell* **2016**, *61*, 507 – 519.
[9] K. D. Meyer, D. P. Patil, J. Zhou, A. Zinoviev, M. A. Skabkin, O. Elemento, T. V. Pestova, S.-B. Qian, S. R. Jaffrey, *Cell* **2015**, *163*, 999 – 1010.
[10] J. Zhou, J. Wan, X. Gao, X. Zhang, S. R. Jaffrey, S.-B. Qian, *Nature* **2015**, *526*, 591.
[11] X. Wang, B. S. Zhao, I. A. Roundtree, Z. Lu, D. Han, H. Ma, X. Weng, K. Chen, H. Shi, C. He, *Cell* **2015**, *161*, 1388 – 1399.
[12] D. P. Patil, C.-K. Chen, B. F. Pickering, A. Chow, C. Jackson, M. Guttman, S. R. Jaffrey, *Nature* **2016**, *537*, 369.
[13] N. Sonenberg, M. A. Morgan, W. C. Merrick, A. J. Shatkin, *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 4843 – 4847.
[14] A. C. Ferretti, G. F. Joyce, *Biochemistry* **2013**, *52*, 1227 – 1235.
[15] S. Tagami, J. Attwater, P. Holliger, *Nat. Chem.* **2017**, *9*, 325.
[16] J. G. Forsythe, S. S. Yu, I. Mamajanov, M. A. Grover, R. Krishnamurthy, F. M. Fernández, N. V. Hud, *Angew. Chem. Int. Ed.* **2015**, *54*, 9871 – 9875; *Angew. Chem.* **2015**, *127*, 10009 – 10013.
[17] K. B. Muchowska, S. J. Varma, E. Chevallot-Beroux, L. Lethuillier-Karl, G. Li, J. Moran, *Nat. Ecol. Evol.* **2017**, *1*, 1716.
[18] A. J. Coggins, M. W. Powner, *Nat. Chem.* **2017**, *9*, 310 – 317.
[19] E. O. Leslie, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99 – 123.
[20] S. Becker, I. Thoma, A. Deutsch, T. Gehrke, P. Mayer, H. Zipse, T. Carell, *Science* **2016**, *352*, 833 – 836.
[21] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239 – 242.
[22] S. Becker, C. Schneider, H. Okamura, A. Crisp, T. Amatov, M. Dejmek, T. Carell, *Nat. Commun.* **2018**, *9*, 163.
[23] M. P. Robertson, G. F. Joyce, *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a003608.
[24] B. J. Cafferty, D. M. Fialho, J. Khanam, R. Krishnamurthy, N. V. Hud, *Nat. Commun.* **2016**, *7*, 11328.
[25] A. C. Rios, Y. Tor, *Isr. J. Chem.* **2013**, *53*, 469 – 483.
[26] A. Tsubura, Y.-C. Lai, H. Miki, T. Sasaki, N. Uehara, T. Yuri, K. Yoshizawa, *in vivo* **2011**, *25*, 11 – 22.
[27] Y. L. Yung, M. B. Mcelroy, *Science* **1979**, *203*, 1002 – 1004.
[28] W. L. Chameides, J. C. G. Walker, *Origins life* **1981**, *11*, 291 – 302.
[29] J. F. Kasting, *Origins Life Evol. Biospheres* **1990**, *20*, 199 – 231.
[30] G. B. Chheda, *Life Sci.* **1969**, *8*, 979 – 987.
[31] C. Deutsch, B. El Yacoubi, V. de Crécy-Lagard, D. Iwata-Reuyl, *J. Biol. Chem.* **2012**, *287*, 13666 – 13673.
[32] S. E. Kolitz, J. R. Lorsch, *FEBS Lett.* **2010**, *584*, 396 – 404.
[33] F. V. Murphy IV, V. Ramakrishnan, A. Malkiewicz, P. F. Agris, *Nat. Struct. Mol. Biol.* **2004**, *11*, 1186.
[34] T. G. Waddell, L. L. Eilders, B. P. Patel, M. Sims, *Origins Life Evol. Biospheres* **2000**, *30*, 539 – 548.
[35] F. Goesmann, H. Rosenbauer, J. H. Bredehöft, M. Cabane, P. Ehrenfreund, T. Gautier, C. Giri, H. Krüger, L. Le Roy, A. J. MacDermott, *Science* **2015**, *349*, aab0689.
[36] N. F. W. Ligterink, A. Coutens, V. Kofman, H. S. P. Müller, R. T. Garrod, H. Calcutt, S. F. Wampfler, J. K. Jørgensen, H. Linnartz, E. F. van Dishoeck, *Mon. Not. R. Astron. Soc.* **2017**, *469*, 2219 – 2229.
[37] S. L. Miller, H. C. Urey, *Science* **1959**, *130*, 245 – 251.
[38] B. T. Golding, C. Bleasdale, J. McGinnis, S. Müller, H. T. Rees, N. H. Rees, P. B. Farmer, W. P. Watson, *Tetrahedron* **1997**, *53*, 4063 – 4082.
[39] K. Altwegg, H. Balsiger, A. Bar-Nun, J.-J. Berthelier, A. Bieler, P. Bochsler, C. Briois, U. Calmonte, M. R. Combi, H. Cottin, *Sci. Adv.* **2016**, *2*, e1600285.
[40] P. Ehrenfreund, S. B. Charnley, *Annu. Rev. Astron. Astrophys.* **2000**, *38*, 427 – 483.
[41] W. H. Shaw, J. J. Bordeaux, *J. Am. Chem. Soc.* **1955**, *77*, 4729 – 4733.
[42] B. P. Callahan, Y. Yuan, R. Wolfenden, *J. Am. Chem. Soc.* **2005**, *127*, 10828 – 10829.
[43] H.-J. Kim, A. Ricardo, H. I. Illangkoon, M. J. Kim, M. A. Carrigan, F. Frye, S. A. Benner, *J. Am. Chem. Soc.* **2011**, *133*, 9457 – 9468.
[44] K. K. Park, J. S. Wishnok, M. C. Archer, *Chem.-Biol. Interact.* **1977**, *18*, 349 – 354.
[45] R. Saladino, C. Crestini, S. Pino, G. Costanzo, E. Di Mauro, *Phys. Life Rev.* **2012**, *9*, 84 – 104.
[46] R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1970**, *47*, 531 – 543.
[47] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1972**, *67*, 25 – 33.
[48] M. A. Machnicka, K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, *Nucleic Acids Res.* **2013**, *41*, D262 – D267.
[49] T. Machinami, T. Suami, *Bull. Chem. Soc. Jpn.* **1975**, *48*, 1333 – 1334.
[50] M. C. Weiss, F. L. Sousa, N. Mrnjavac, S. Neukirchen, M. Roettger, S. Nelson-Sathi, W. F. Martin, *Nat. Microbiol.* **2016**, *1*, 16116.
[51] H.-J. Kim, S. A. Benner, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 11315 – 11320.
[52] R. Saladino, E. Carota, G. Botta, M. Kapralov, G. N. Timoshenko, A. Y. Rozanov, E. Krasavin, E. Di Mauro, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E2746 – E2755.
[53] C. B. Winiger, M.-J. Kim, S. Hoshika, R. W. Shaw, J. D. Moses, M. F. Matsuura, D. L. Gerloff, S. A. Benner, *Biochemistry* **2016**, *55*, 3847 – 3850.

Manuscript received: February 12, 2018
Accepted manuscript online: March 13, 2018
Version of record online: April 17, 2018

2.4 FOURTH PUBLICATION

TITLE: Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis

AUTHORS:

S. Becker, C. Schneider, H. Okamura, A. Crisp, T. Amatov, M. Dejmek, T. Carell*

** corresponding author.*

JOURNAL, YEAR, VOLUME, AND ARTICLE NO.:

Nat. Commun., **2018**, *9*, 163

SUMMARY:

The molecular building blocks that gave rise to the first life on earth must have emerged through processes that were governed by the composition of the Earth's crust and atmosphere at that time. Although syntheses of RNA nucleosides have previously been reported, no rationale had yet been given as to how those building blocks could have been generated *via* a continuous physicochemical process in high yields. Here we report that geochemical conditions compatible with the early earth could have generated fluctuating wet-dry conditions that allowed for the continuous synthesis of both the canonical- and certain ubiquitous non-canonical RNA nucleosides. Our results provide further evidence that certain non-canonical RNA nucleosides found in nature could in fact be vestiges of the early Earth.

PERSONAL CONTRIBUTIONS:

- Prepared certain nitroso-pyrimidines, and obtained crystallographic data supporting certain structures, under supervision from S. Becker.
- Carried out the synthesis of alternative carbamoyl-substituted nitroso- and formamidopyrimidines as precursors to adenosine nucleosides that were not included in the final study.
- Contributed to the overall design of the study though discussions with S. Becker, C. Schneider, H. Okamura and T. Carell.
- Contributed to the writing and preparation of the original manuscript.
- Edited and revised the manuscript together with all co-authors.

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-017-02639-1 OPEN

Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis

Sidney Becker [1], Christina Schneider [1], Hidenori Okamura[1], Antony Crisp [1], Tynchtyk Amatov [1], Milan Dejmek[2] & Thomas Carell [1]

The molecules of life were created by a continuous physicochemical process on an early Earth. In this hadean environment, chemical transformations were driven by fluctuations of the naturally given physical parameters established for example by wet-dry cycles. These conditions might have allowed for the formation of (self)-replicating RNA as the fundamental biopolymer during chemical evolution. The question of how a complex multistep chemical synthesis of RNA building blocks was possible in such an environment remains unanswered. Here we report that geothermal fields could provide the right setup for establishing wet-dry cycles that allow for the synthesis of RNA nucleosides by continuous synthesis. Our model provides both the canonical and many ubiquitous non-canonical purine nucleosides in parallel by simple changes of physical parameters such as temperature, pH and concentration. The data show that modified nucleosides were potentially formed as competitor molecules. They could in this sense be considered as molecular fossils.

[1] Center for Integrated Protein Science Munich CiPSM at the Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany. [2] Institute of Organic Chemistry and Biochemistry ASCR, 16610 Prague 6, Czech Republic. Correspondence and requests for materials should be addressed to T.C. (email: thomas.carell@lmu.de)

The molecules of life originated around 4 billion years ago under conditions governed by the composition of the Earth's crust and atmosphere at that time[1, 2]. Molecules such as nucleic acids and amino acids must have formed by a continuous physicochemical process, in which greater structural complexity was generated based on fluctuations of the naturally given physical parameters[3]. Geothermal fields, for example, could have established such fluctuations by wet–dry cycles that may have driven chemical transformations, which ultimately allowed the emergence of life[4–8]. The appearance of (self)-replicating RNA was certainly of central importance for the transition from an abiotic world to biology[9–11]. We need to consider, however, that an early genetic polymer might have been structurally different from contemporary RNA. This involves differences regarding the sugar configuration (e.g., pyranosyl RNA) or the presence of other nucleobases[12, 13]. Selection pressure led in this scenario to the chemical evolution of RNA. Contemporary RNA molecules contain four canonical nucleosides (A, G, C, U), which establish the sequence information. In addition, >120 non-canonical nucleosides are present, which govern a diverse set of properties such as correct folding, e.g., to enable catalysis[14]. In fact, the genetic system of all known life is dependent on modified nucleosides. Many of these non-canonical nucleosides are found today in all three domains of life, which indicates that they were present early on during the development of life. For the ubiquitous non-canonical nucleosides we may assume that they were already formed as competitors in parallel with the canonical ones on the early Earth[15]. So far, however, a geochemical scenario that would allow for the parallel formation of canonical and non-canonical RNA building blocks by a continuous process is not known. All reported multistep chemical models so far rely on tightly controlled laboratory conditions and the isolation and purification of central reaction intermediates by sophisticated methods[1, 16–19].

Herein, we report a robust synthetic pathway, which is purely based on fluctuations of physicochemical parameters such as pH, concentration, and temperature, driven by wet–dry cycles. These fluctuations enable the direct enrichment or purification of all reaction intermediates that are directly used for the next synthetic steps. As such, a continuous synthesis is established. Our results show that RNA building blocks can indeed be formed in a prebiotically plausible geochemical environment without sophisticated isolation and purification procedures. The chemical scenario presented here supports the hypothesis that life may have originated in a hydrothermal milieu on land rather than in a deep sea environment. The key assembly step in our pathway is the formation of variously substituted 5-nitroso-pyrimidines (nitrosoPys) that can be converted into formamidopyrimidines (FaPys) in the presence of formic acid and elementary metals (Ni or Fe). When combined with ribose, the FaPy compounds react to give a set of purine nucleosides. This chemical pathway delivers not only the canonical purine nucleosides but concomitantly many of the ubiquitously present non-canonical relatives, suggesting their origin as prebiotic competitor nucleosides (A, ms^2A, m^2A, DA, G, m^2G, $m^2{}_2G$, m^1G). Since chemical evolution depended on those molecules that were available on early Earth, these non-canonical nucleosides may be considered to be molecular fossils, which maintained their essential life-supporting character until the present day.

Results

Selective crystallization of an organic salt. The chemical scenario that leads to a continuous synthesis of RNA building blocks by just fluctuations of physical parameters is shown in Figs. 1 and 2a. The scenario starts with an aqueous solution of malononitrile **1** and different amidinium salts **2a-d** (HCl or H_2SO_4 salts, 400 mM), both recognized prebiotic compounds[18]. In addition,

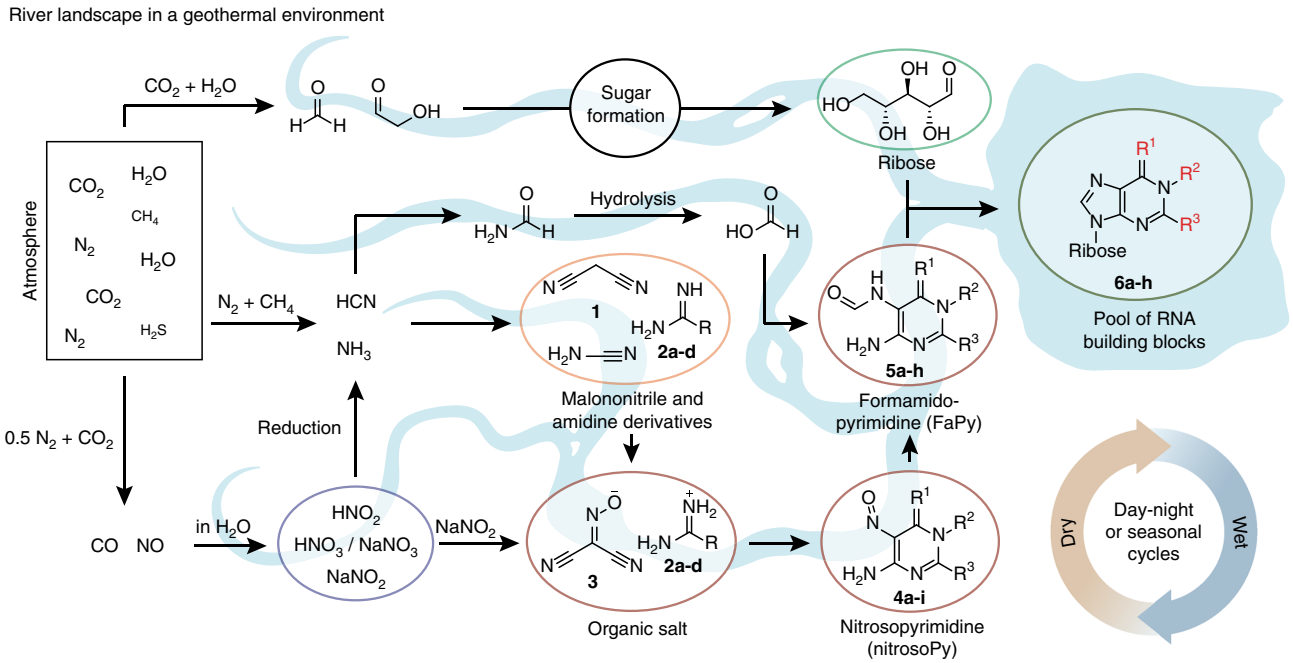


Fig. 1 RNA nucleoside formation pathway. A geothermal environment provides the right set up for the depicted transformations by establishing wet–dry cycles. The prebiotic starting materials are produced from a prebiotic atmosphere and washed into an aqueous environment (e.g. by rain). Major atmospheric components are written in larger letters, whereas minor components are written in smaller letters. Transformations are taking place in different environments, illustrated by various rivers (in light blue). Each environment provides the right setup for different chemistries, leading to several different chemical transformations. This geochemical setup leads to a set of canonical and non-canonical RNA building blocks by continuous synthesis (**6a**, m^1G: $R^1 = O$, $R^2 = Me$, $R^3 = NH_2$; **6b**, ms^2A: $R^1 = NH$, $R^2 = H$, $R^3 = SMe$; **6c**, A: $R^1 = NH$, $R^2 = H$, $R^3 = H$; **6d**, m^2G: $R^1 = O$, $R^2 = H$, $R^3 = NHMe$; **6e**, $m^2{}_2G$: $R^1 = O$, $R^2 = H$, $R^3 = N(Me)_2$; **6f**, G: $R^1 = O$, $R^2 = H$, $R^3 = NH_2$; **6g**, DA: $R^1 = NH$, $R^2 = H$, $R^3 = NH_2$; **6h**, m^2A: $R^1 = NH$, $R^2 = H$, $R^3 = Me$)

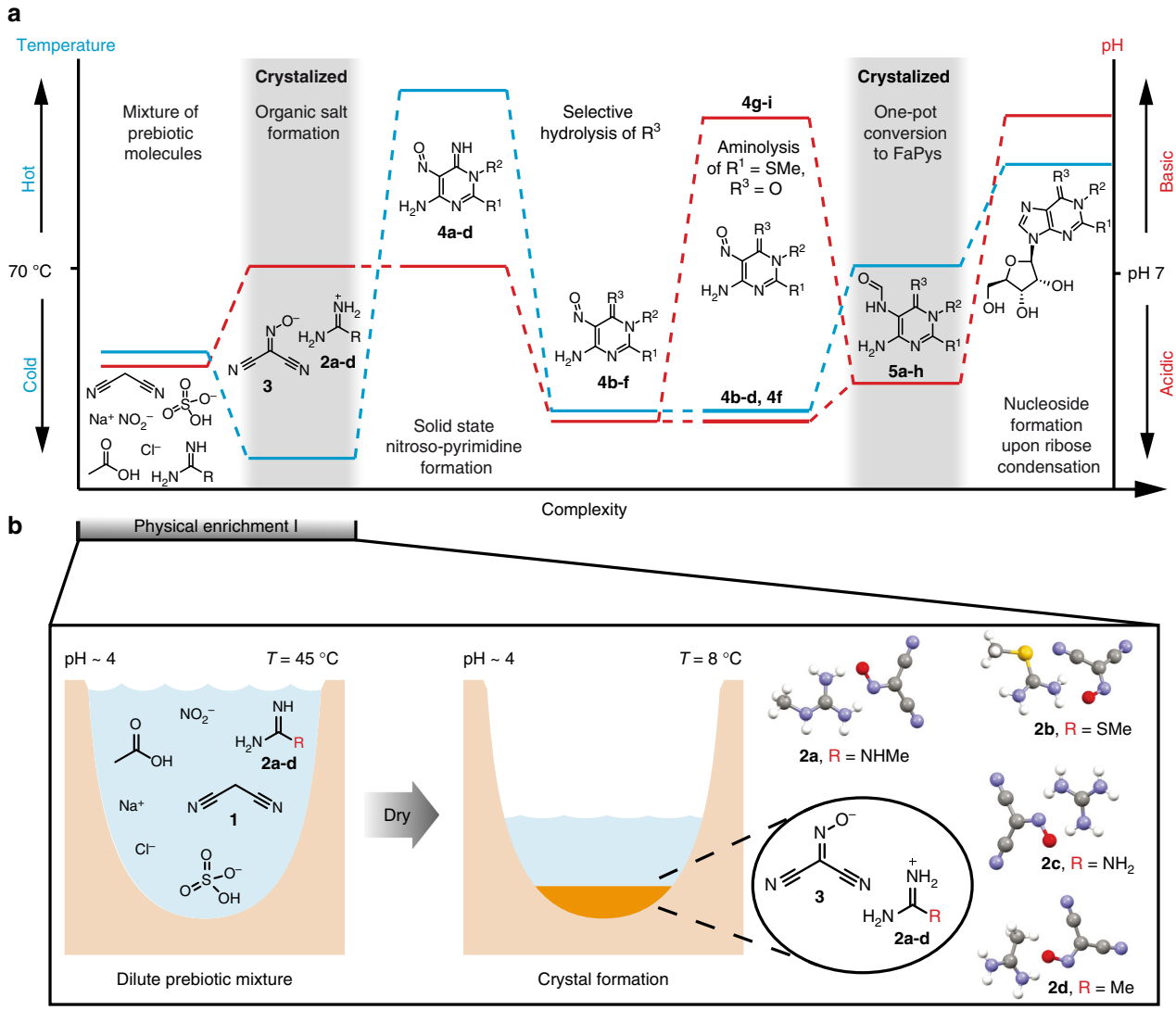


Fig. 2 Chemical complexity created by physical fluctuations. **a** Relative changes of temperature (in blue) and pH (in red) are shown for each synthetic step for the continuous synthesis of purine RNA building blocks from small organic and inorganic molecules. Several wet-dry cycles establish fluctuations of the depicted physical parameters that enable the physical enrichment of intermediates. Gray backgrounds denote compounds that are enriched by crystallization from an aqueous solution. **b** Formation of an organic salt consisting of amidine derivatives **2a-d** and (hydroxyimino)malononitrile **3**. The salt is selectively crystalized by concentrating a dilute mixture of organic and inorganic compounds by slow evaporation. The crystal structures of the four crystalized organic salts are depicted (Supplementary Tables 1–4)

sodium nitrite and acetic acid are present to establish a slightly acidic pH of around 4. Under these conditions the amidine molecules (**2a-d**) are protonated, which leads to their chemical deactivation. This allows selective nitrosation of malononitrile **1** to give (hydroxyimino)malononitrile **3** in situ. Slow evaporation of water under ambient conditions, followed by gentle cooling to 8–10 °C resulted in crystallization of a salt from the ca. 1 M amidinium solution. This crystallization is very robust and resembles naturally occurring concentration processes. The resultant crystals had excellent quality for X-ray analysis, which showed that the salts are formed from the amidinium cations **2a-d** and the (hydroxyimino)malononitrile anion of **3** (Fig. 2b). Interesting is the distance between the negatively charged oxygen in **3** and the positively charged H-bond donor centre of the amidininium units **2a-d**. We determined distances between 1.85–1.95 Å, which is long for a salt bridge but right in the regime for a typical hydrogen bond. This is important because it is supposedly the reason for the comparably low melting temperatures of the salts, which we determined between 110 and 160 °C. The robustness and ease of crystallization establishes a first physical enrichment step that finishes the initial wet–dry phase with the deposition of these salt materials (Fig. 2b).

Nitroso-pyrimidine formation. When the obtained salts containing **2a-d** and **3** are subsequently heated to their respective melting temperatures, transformation into the corresponding nitroso-pyrimidines (**4a-d**, Fig. 3a) occurs. The required temperatures between 110 and 160 °C could have been readily accessible under early Earth conditions, due to, for example, volcanic activity in geothermal fields or sunlight shining on dark surfaces. In order to investigate whether the nitroso-compounds **4a-d** would form in parallel despite their varying structures and different melting points, the different salts were combined in a reaction flask and a temperature gradient (1 °C/5 min, from 100–160 °C) was applied to simulate soil that would slowly heat up. Subsequent ^{1}H-NMR analysis indicated successful formation of the anticipated nitroso-pyrimidines **4a-d** (Supplementary Fig. 1).

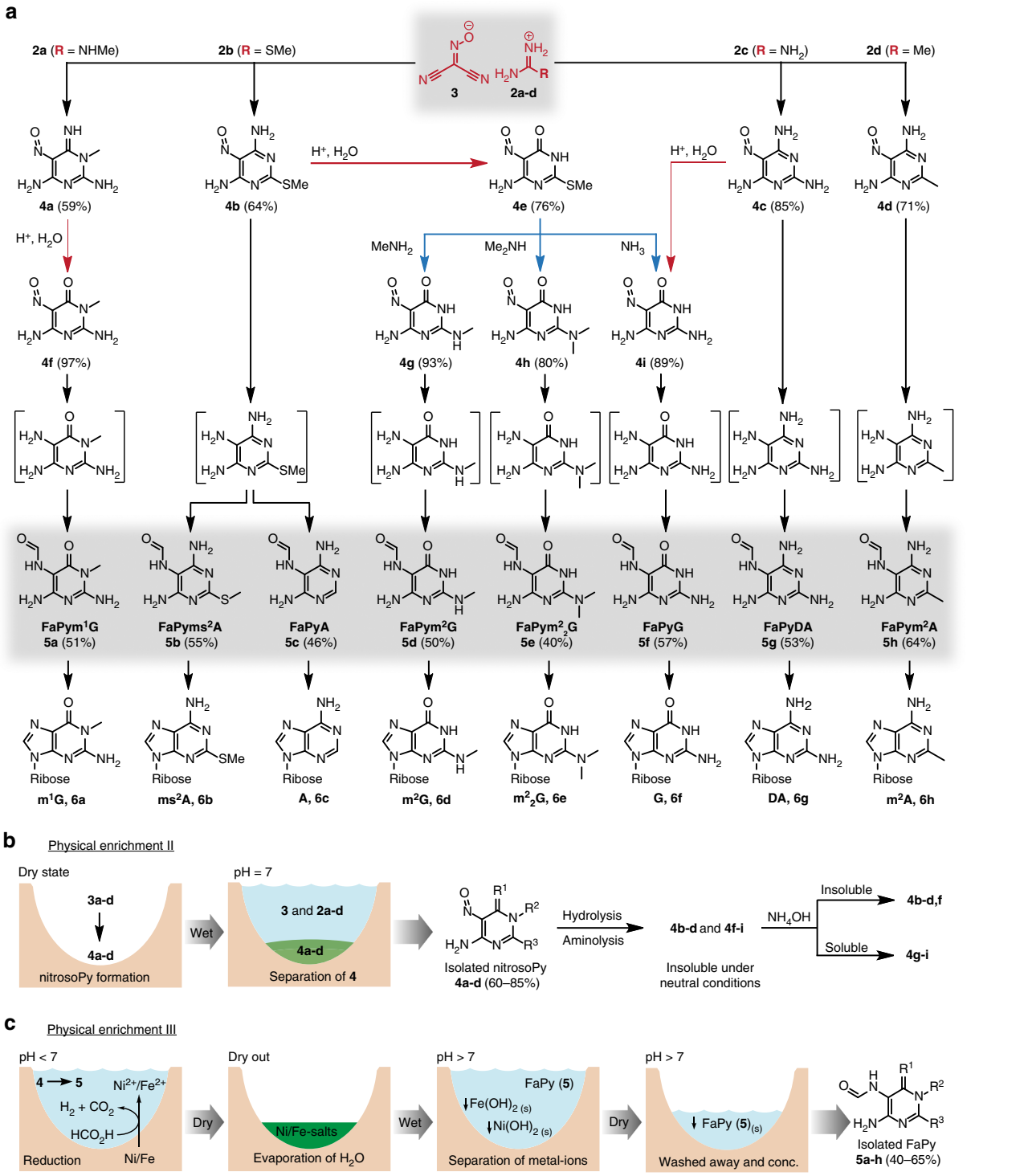


Fig. 3 Reaction scheme and physical enrichment of intermediates. **a** Dry-state reactions of salts containing **2a-d** and **3** provide nitroso-pyrimidines **4a-d**, which can be further diversified by hydrolysis (red arrows) or aminolysis (blue arrows) to give a set of nitroso-pyrimidines (nitrosoPys) **4a-i**. In the presence of elementary Fe and Ni and dilute formic acid, formation of the formamidopyrimidines (FaPys) **5a-h** as direct purine base precursors takes place. In square brackets: non-isolated reaction intermediates. **b** Second physical enrichment of the nitroso-pyrimidines isolated in high purity and yield. **c** Third physical enrichment of the formed FaPys **5a-h** as nucleoside precursors from nitroso-pyrimidines

The resultant nitroso-pyrimidines are stable compounds with melting points typically >250 °C without decomposition. In addition we noted that the nitroso-pyrimidines are rather insoluble in water, which offers the possibility for a second physical enrichment step. Addition of water to the reaction mixture dissolves unreacted starting materials, leaving the nitroso-pyrimidines in basically NMR-pure form behind (Supplementary Fig. 2). In this model, one wet–dry cycle and two physical enrichment steps with a final rain shower or flooding would be sufficient to deposit a mixture of stable nitroso-pyrimidines (**4a-d**) in excellent purities and good chemical yields between 60 and 85% (Fig. 3a).

Diversification by hydrolysis and aminolysis. Depending on the composition and pH of the aqueous environment, which may or may not contain different amines, the nitroso-pyrimidines could undergo further hydrolysis and aminolysis reactions (Fig. 3a). Because these reactions are very slow under neutral conditions, we used dilute HCl to accelerate the processes for investigation. Importantly, we noted a high regioselectivity. Upon treatment overnight at room temperature with 0.5 M HCl, compounds **4a** and **4c** for example are hydrolyzed to afford the oxo-nitroso-pyrimidines **4f** and **4i** in near quantitative yields. Hydrolysis of **4b** to product **4e** was comparitively slower, and under our accelerated conditions a mixture of **4b** and **4e** was obtained. This inefficient conversion would be advantageous in a prebiotic context given that from **4b** the canonical nucleoside adenosine (A) and its 2-thiomethyl derivative (ms^2A) are derived later, whereas **4e** gives rise to guanosine derivatives (G, m^2G, $m^2{}_2G$, Fig. 3a). This allows for the simultaneous formation of canonical and non-canonical bases from the same precursor. In contrast to the 2-amino (**4a,c**) or 2-methyl (**4d**) substituted nitroso-pyrimidines, we noted that the 2-thiomethyl functionality in **4b** and **4e** was prone to undergo selective nucleophilic substitution. Reaction of **4e** with different amines leads to efficient formation of the nitroso-pyrimidines **4g-i** with the concomitant release of methanethiol. Due to its insolubility under basic conditions, nucleophilic substitutions of **4b** are very inefficient. To confirm this, we partially hydrolyzed **4b** to **4e** in the presence of methylamine (300 mM) and dimethylamine (100 mM). The pH was carefully adjusted with Na_2CO_3 to about pH 10. Compound **4b** precipitated, while **4e** stayed in solution, consequently protecting **4b** from further reactions. It is in this context interesting that nucleosides that would form via aminolysis of **4b** have not yet been found in nature. In contrast, **4e** reacts efficiently and after 3–4 days at room temperature **4e** is almost completely converted into **4g** and **4h**, which are direct precursors to the ubiquitous non-canonical RNA bases m^2G and $m^2{}_2G$ (Fig. 3a, Supplementary Fig. 3).

Thus, a few simple chemoselective and regioselective hydrolysis and aminolysis reactions affords a diverse mixture of differently substituted nitroso-pyrimidines (**4b-d, f-i**), all of which possess the right substitution pattern for the synthesis of naturally occuring canonical and non-canonical RNA nucleosides. Because all the formed nitroso-pyrimidines are poorly soluble in water at neutral pH, neutralizing the solutions leads to their efficient precipitation, providing a naturally occurring purification step (Fig. 3b). Importantly, all nitroso-pyrimidines that later give adenosine-derived nucleosides (**4b-d**) are soluble in water under acidic conditions, while the nitroso-compounds that are converted into guanosine-derived nucleosides (**4g-i**, except for **4f**) are soluble under basic pH conditions. These properties allow for potentially divergent chemical pathways leading to A-derived and G-derived nucleosides (Fig. 3b, Supplementary Fig. 4).

Formamidopyrimidine formation as nucleobase precursor. The next wet–dry cycles allow for the formation and isolation of formamidopyrimidines (FaPys) **5a-h**, from nitroso-pyrimidines **4** that are after their precipitation exposed to acidic conditions like dilute formic acid in the presence of elementary Fe or Ni, which are components of the Earth's crust. This leads to reduction of the nitroso-pyrimidines **4** to aminopyrimidines as non-isolated reaction intermediates (Fig. 3a, in square brackets), which are immediately formylated to give the water soluble formamidopyrimidines (FaPys) **5a-h** in a one-pot reaction. During the wet phase, Ni^0 and Fe^0 are converted into the biologically relevant Ni^{2+}/Fe^{2+} ions, while formic acid decomposes into CO_2 and H_2 (Fig. 3c). In the reaction formic acid has a dual function. It provides the H-atoms needed for the reduction and it subsequently reacts with the formed aminopyrimidines to give FaPy compounds that were already shown to be prebiotically valid precursors to purine nucleosides[18]. The Ni/Fe/formic acid environment converts quantitatively all nitroso-compounds **4b-d, f-i** into the corresponding FaPy compounds **5a-h** (Fig. 3a). The water soluble FaPy compounds (under dilute basic conditions) can now be separated from unreacted Ni^0/Fe^0 and from the formed Ni^{2+}/Fe^{2+} byproducts. Under slightly basic conditions (pH ≈ 9–10) the latter compounds precipitate as insoluble carbonate or hydroxide salts. The FaPys **5a-h** are thus washed away, while the transition metal compounds sediment out. Final evaporation of water concentrates the reaction mixture, leading to the crystallization of the FaPy molecules. This third physical enrichment step, involving a wet–dry cycle, leads to the NMR-clean formation of FaPy-derivatives **5a-h** (Fig. 3c).

The 2-(methylthio)-5-nitrosopyrimidine-4,6-diamine (**4b**) gives after treatment with formic acid and elementary Ni two different FaPy products depending on the reaction conditions. One of the products (**5b**) contains a thiomethyl group, while the other (**5c**) is desulfurated. The desulfurization reaction is simply controlled by time and can be promoted when H_2 is bubbled through the solution prior to reaction. Compound **5c** is always generated in a stepwise reaction cascade via compound **5b** which was confirmed by reacting **4b** for 2 h and isolating the only product formed (**5b**, Fig. 3c). The isolated product was immediately subjected to the same conditions, which provided **5c** after 7 days in pure form. This pathway via nitroso-pyrimidines thus affords **5c**, the precursor for the canonical base A under plausible prebiotic conditions[20]. These conditions also lead to the parallel formation of the precursor to the ubiquitous 2-thiomethyl modification (ms^2A), which is today found in all three domains of life.

Formation of canonical and non-canonical nucleosides. All of the prepared FaPy compounds undergo rapid and regioselective condensations with ribose when they are present in the same dry-state environment (Fig. 4). We do not assume that ribose was formed at the same location together with the FaPy compounds since the required carbohydrate chemistry may be incompatible. Several models are available, however, that show ribose formation in different physical environments[21–24]. Even though ribose and FaPys might have formed separately, the water solubility of the FaPys and of ribose allows them to be washed into the same environment by rain or flooding. Evaporation of water in the last wet–dry cycle would enable a condensation reaction under dry-state conditions. Indeed, the physically enriched FaPy compounds (**5a-h**) engage in a rapid reaction with ribose to give the corresponding FaPy-ribosides. Upon dissolution in water and subsequent heating under basic conditions, all four expected purine α/β-ribofuranosides (α/β–f) and α/β-pyranosides (α/β–p) are obtained (**6a-h**, Fig. 4a), completing the last wet–dry cycle. The LC-MS traces of the reactions using both UV- and MS-detection are shown in Fig. 4b. To ensure correct structural assignment we chemically synthesized some of the expected products and performed co-injection studies (Supplementary Methods). These experiments show that the major isomers are the naturally occurring β-configured pyranosides and furanosides. Pyranosides are building blocks for pyranosyl-RNA, which was suggested to be a potential RNA predecessor[12]. Therefore, our scenario delivers the building blocks for this pre-RNA and for RNA. As such it provides the basis for the chemical transition from one genetic polymer to the other directed by selection pressure. Importantly, our continuous synthetic pathway provides next to the canonical bases A and G also the non-canonical β-furanosyl-nucleosides (β–f) m^2G, $m^2{}_2G$, m^1G, ms^2A and m^2A (in red, Fig. 4b), arguing

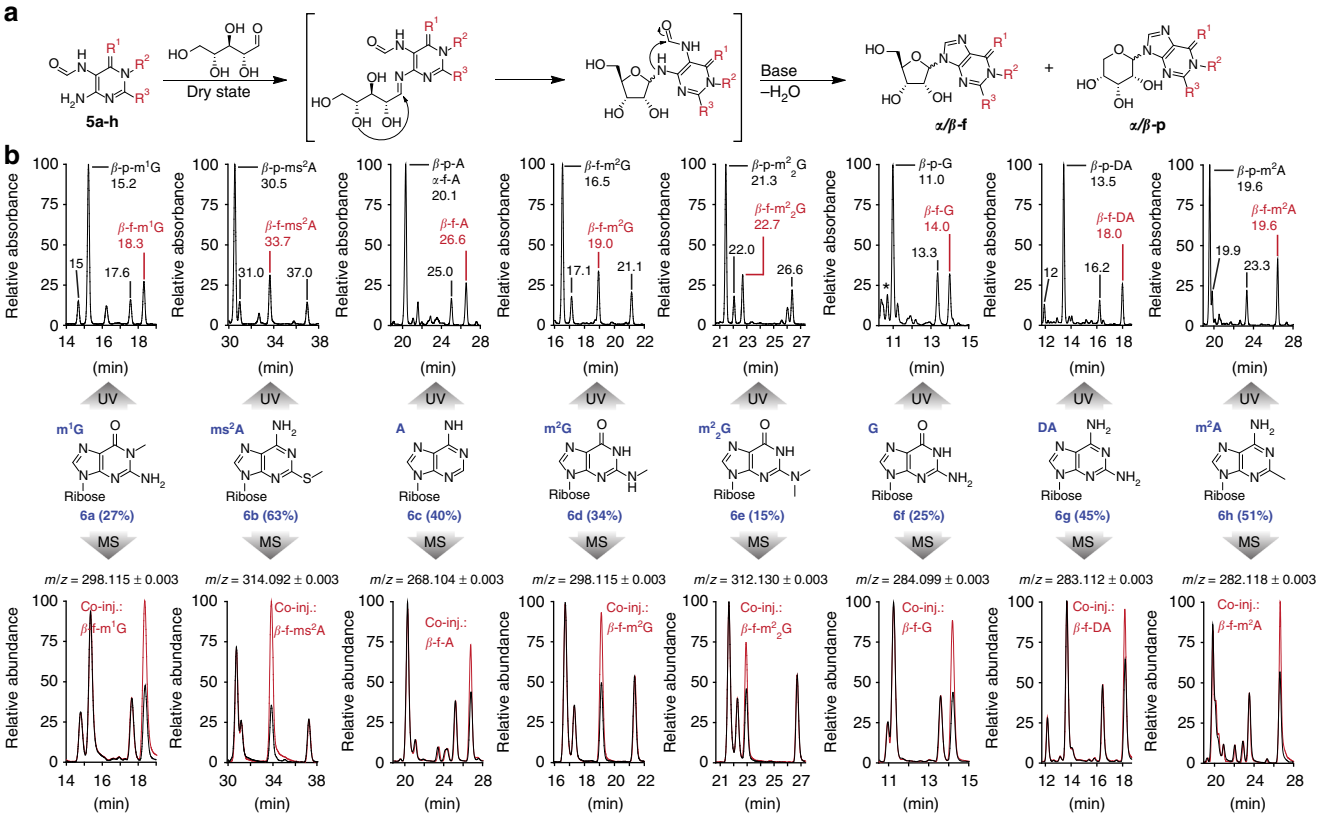


Fig. 4 Formation of RNA nucleosides from nucleobase precursors. **a** Reaction mechanism for the formation of canonical and non-canonical RNA nucleosides **6a-h** from formamidopyrimidines (FaPys) **5a-h** and ribose. The reaction provides the four expected isomers α/β ribopyranosides (α/β–p) and α/β–ribofuranosides (α/β–f). **b** LC-MS analysis of the reaction products from **5a-h** and ribose. Compounds identified by MS detection are labeled (β–p or α/β–f). The UV- and MS-chromatograms show all four expected isomers for each compound (labeled with the retention time or asterisk (*) in the UV-chromatogram). The structural assignment was assisted by co-injection (Co-inj.) studies. The MS traces show in red the co-injection signal obtained with the naturally occurring isomer (β–f)

that the early RNA polymer was structurally already more complex regarding the nucleobases. The ribosylation of the FaPys leading to non-canonical nucleosides is equally efficient to the formation of A and G, with yields between 15 and 60%. Interestingly, we noted that for some A derivatives (m^2A, DA and A) other regioisomers were found as well. These isomers were not formed when pure FaPy starting materials were used that were not derived from our continuous synthesis. We believe that these isomers might be the N3-connected nucleosides, previously proposed by Wächtershäuser for homo-purine RNA[25]. Despite the presence of these side products, we observe efficient N9-nucleoside formation with remarkable yields of up to 60% for the canonical and the non-canonical nucleosides. This work demonstrates that the non-canonical compounds could plausibly have formed as companion and potential competitor compounds in parallel to the canonical nucleosides.

Discussion

Life on earth certainly did not start in a chemists' laboratory, where the relevant compounds are assembled in a step-by-step process from pure starting materials under tightly controlled conditions. Even if the individual reaction steps are performed under plausibly prebiotic conditions, the controlled assembly over many steps with sophisticated isolation and purification procedures of reaction intermediates is an unlikely scenario for chemical synthesis under early Earth conditions. For the process of chemical evolution on the early Earth, we may rather envision a more continuous synthesis, in which small organic molecules, initially formed by volcanic action or lightning, reacted to give increasingly more complex structures (Fig. 1). Here, chemical transformations may have been driven by physical fluctuations, established for example by day-night, seasonal or wet–dry cycles. Such fluctuating parameters might include temperature, concentration, and pH, which have triggered selective precipitation and crystallization to purify and concentrate reaction intermediates (Fig. 2a).

Regarding the central nucleoside building blocks of life, we believe that the four canonical nucleosides were finally selected from a more diverse prebiotic nucleoside pool. These canonical bases today establish the sequence information. The synthesis of the canonical purine (A, G)[18] and pyrimidine (U, C)[16] RNA building blocks has been previously demonstrated in aqueous environments. It is questionable, however, if these multistep synthesis pathways are able to provide all four canonical bases at the same time, which fuels the development of new prebiotically plausible nucleoside formation reactions[17, 26, 27]. Recently, all four canonical nucleosides (A, G, U, C) were accessed in low yields via a one-pot procedure from pure formamide[28]. However, in order to establish a functional genetic system a number of non-canonical nucleosides is required as well that provide other functions related to folding and catalysis[29–34]. Since many of these non-canonical bases are present in all three domains of life, it is likely that they have been early on part of the abiotic chemical selection process[15]. We report here the discovery of a continuous synthesis pathway that enables the efficient production of

canonical and non-canonical purine bases in parallel. Our data show that formation of the many nucleosides needed to establish a functional genetic system is in fact an unavoidable event if we assume the presence of simple starting materials such as formic acid, acetic acid, sodium nitrite, malononitrile (**1**), amidinium compounds (**2a-d**), as well as transition metals like Ni or Fe. These simple compounds react in several successive wet–dry phases, leading to physical enrichment (I, II, and III) of reaction intermediates to finally give RNA building blocks. Wet–dry cycles have already been shown to be a plausible geological scenario especially for polymerization reactions[35, 36]. Our reported chemical scenario shows now that such a geological setup can also result in a diverse set of purine nucleosides by continuous synthesis (**6a-h**, Fig. 4). These nucleosides can be converted into the phosphorylated nucleotides based on recent advances in prebiotic phosphorylation reactions[37, 38]. So far, however, we are not yet able to include this step into our continuous synthesis.

Importantly, all here-reported non-canonical bases are known to exist in the three domains of life. Many are postulated components of the early genetic system of the last universal common ancestor (LUCA), suggesting that they were indeed present already at the onset of biological evolution[39, 40]. Based on the continuous synthesis pathway reported here, we hypothesize that the canonical and at least these non-canonical nucleosides could have formed side by side, dating the formation of the first non-canonical nucleosides back to the origin of chemical evolution around 4 billion years ago. As such, they could have been part of the chemical evolution process that established the putative RNA world[41]. Our chemistry invokes that methylated and thiomethylated nucleosides could particularly have been integral components of the first instructional (pre)-RNA molecules, likely to stabilize folded structures in order to accelerate catalytic processes[29–34]. We therefore propose that these nucleosides could be vestiges and molecular fossils of an early Earth, as it was suggested for cofactors[42].

Methods

1-methylguanidine (2a) salt of (hydroxyimino)malononitrile (3). 1-methylguanidine (**2a**) hydrochloride salt (10.95 g, 100 mmol, 1 eq.) and malononitrile (**1**) (6.65 g, 100 mmol, 1 eq.) was dissolved in H_2O (230 mL, containing 6 mL of AcOH) in a 500 mL beaker. A solution of $NaNO_2$ (7.00 g, 101 mmol, 1.01 eq., in 20 mL of H_2O) was slowly added at room temperature. After stirring at room temperature for 2 h the reaction mixture was kept at 45 °C in an oil bath for 3–4 days open to the air until the mixture was concentrated to about 100 mL. The reaction mixture was placed in a fridge overnight at 8–10 °C. The formed yellow crystals were filtered off to give the desired product (6.70 g, 40 mmol, 40%).

Mp.: 108 °C. **^{1}H-NMR** (400 MHz, DMSO-d_6) δ = 7.23 (br m, 5H), 2.72 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 157.85, 119.50, 113.31, 107.18, 28.23. **IR** (cm^{-1}): 3405 (m), 3351 (br, m), 3197 (m), 2977 (br, m), 2229 (s), 2218 (s) 1675 (s), 1635 (s), 1465 (w) 1428 (m), 1344 (s), 1269 (s) 1226 (s), 1172 (w), 1098 (m), 915 (m), 765 (m).

Methylthioamidine (2b) salt of (hydroxyimino)malononitrile (3). *S*-Methylisothiourea (**2b**) hemisulfate salt (27.8 g, 200 mmol, 1 eq.) and malononitrile (**1**) (13.3 g, 200 mmol, 1 eq.) was dissolved in H_2O (460 mL, containing 12 mL of AcOH) in a 600 mL beaker. A solution of $NaNO_2$ (14.0 g, 202 mmol, 1.01 eq., in 40 mL of H_2O) was slowly added at room temperature. After stirring at room temperature for 2 h the reaction mixture was kept at 45 °C in an oil bath for 3–4 days open to the air until the mixture was concentrated to about 200 mL. The reaction mixture was placed in a fridge overnight at 8–10 °C. The formed yellow crystals were filtered off to give the desired product (16.7 g, 90 mmol, 45%).

Mp.: 126 °C. **^{1}H-NMR** (400 MHz, DMSO-d_6) δ 8.90 (s, 4H), 2.56 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ 171.62, 119.46, 113.29, 107.17, 13.71. **IR** (cm^{-1}): 3282(m), 3146 (br, m), 2742 (br, w), 2530 (w), 2222 (s), 2213 (s) 1698 (m), 1663 (s), 1643 (s) 1549 (m), 1450 (m), 1424 (s) 1375 (w), 1335 (s), 1269 (s), 1223 (s), 1180 (w), 1099 (m), 1076 (w), 982 (m), 970 (w), 960 (w), 897 (br, m), 801 (m), 736 (m).

Guanidine (2c) salt of (hydroxyimino)malononitrile (3). Guanidine (**2c**) hydrochloride salt (9.55 g, 100 mmol, 1 eq.) and malononitrile (**1**) (6.65 g, 100 mmol, 1 eq.) was dissolved in H_2O (230 mL, containing 6 mL of AcOH) in a 500 mL beaker. A solution of $NaNO_2$ (7.00 g, 101 mmol, 1.01 eq., in 20 mL of H_2O) was slowly added at room temperature. After stirring at room temperature for 2 h the reaction mixture was kept at 45 °C in an oil bath for 3–4 days open to the air until the mixture was concentrated to about 100 mL. The reaction mixture was placed in a fridge overnight at 8–10 °C. The formed yellow crystals were filtered off to give the desired product (8.20 g, 53 mmol, 53%).

Mp.: 159 °C. **^{1}H-NMR** (400 MHz, DMSO-d_6) δ = 6.91 (s, 6H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 158.32, 119.52, 113.32, 107.18. **IR** (cm^{-1}): 3473 (m), 3373 (m), 3172 (w), 3087 (w), 2815 (br, w), 2223 (s), 2217 (s), 1668 (m), 1641 (s), 1578 (w), 1552 (w), 1369 (w), 1343 (s), 1294 (w), 1264 (s), 1220 (s), 1140 (w), 975 (w), 792 (w), 757 (w).

Acetamidine (2d) salt of (hydroxyimino)malononitrile (3). Acetamidine (**2d**) hydrochloride salt (9.45 g, 100 mmol, 1 eq.) and malononitrile (**1**) (6.65 g, 100 mmol, 1 eq.) was dissolved in H_2O (230 mL, containing 6 mL of AcOH) in a 500 mL beaker. A solution of $NaNO_2$ (7.00 g, 101 mmol, 1.01 eq., in 20 mL of H_2O) was slowly added at room temperature. After stirring at room temperature for 2 h the reaction mixture was kept at 45 °C in an oil bath for 3–4 days open to the air until the mixture was concentrated to about 100 mL. The reaction mixture was placed in a fridge overnight at 8–10 °C. The formed yellow crystals were filtered off to give the desired product (8.80 g, 60 mmol, 60%).

Mp.: 142 °C. **^{1}H-NMR** (400 MHz, DMSO-d_6) δ = 8.60 (s, 4H), 2.11 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 168.11, 119.50, 113.31, 107.18, 18.72. **IR** (cm^{-1}): 3282 (m), 3140 (br, m), 2781 (m), 2395 (w), 2236 (s), 2217 (s) 1708 (s), 1661 (s), 1592 (br, w) 1513 (m), 1373 (s), 1351 (s) 1259 (s), 1200 (s), 1191 (s), 1160 (w), 1125 (m), 969 (w), 907 (w), 883 (w), 857 (w), 790 (s), 691 (s), 684 (s).

Synthesis of nitrosopyrimidine 4a-d from salts containing 2a-d and 3. The reaction time of the following procedures for the formation of **4a-d** depend on the crystal size of the organic salts and the heating source. Usually conversions were done in a beaker open to the air in an oil bath to simulate hot soil. Large crystals sometimes already convert into nitroso-pyrimidines by a solid-state reaction without melting. Then much longer reaction times of up to 7 days are required because of unequally distributed temperature. Alternatively, the organic salt can be converted by heating in an oven where temperature is equally distributed within the sample.

6-imino-1-methyl-5-nitroso-1,6-dihydropyrimidine-2,4-diamine (4a). 1-methylguanidine (**2a**) salt of (hydroxyimino)malononitrile (**3**) (5.00 g, 29.5 mmol, 1 eq.) was heated slowly to its melting temperature of 108 °C and kept overnight. The compound melts suddenly but becomes a dark red solid again after leaving it overnight. The quantitative reaction mixture can be directly used for the next step without purification.

For analytical reasons a small batch of 1-methylguanidine (**2a**) salt of (hydroxyimino)malononitrile (**3**) (100 mg, 0.59 mmol, 1 eq.) was reacted as described above. The reaction mixture was suspended well in water (2–3 mL) and the dark red solid was filtered off to give 6-imino-1-methyl-5-nitroso-1,6-dihydropyrimidine-2,4-diamine (59 mg, 0.35 mmol, 59%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 11.44 (s, 1H), 8.41 (s, 1H), 8.08 (br, 1H), 7.67 (br, 1H), 7.49 (s, 1H), 3.26 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 165.44, 157.64, 146.95, 137.22, 27.64. **HRMS** (ESI +): calc. for $[C_5H_9N_6O]^+$ 169.0832, found: 169.0832 $[M + H]^+$

2-(methylthio)-5-nitrosopyrimidine-4,6-diamine (4b). Methylthioamidine (**2b**) salt of (hydroxyimino)malononitrile (**3**) (5.00 g, 27 mmol, 1 eq.) was heated slowly to its melting temperature of 126 °C and kept overnight. Caution: if the product is heated too quickly above the melting temperature it decomposes with the release of MeSH! The compound melts suddenly but becomes a dark green solid again. The quantitative reaction mixture can be directly used for the next step without purification.

For analytical reasons a small batch of methylthioamidine (**2b**) salt of (hydroxyimino)malononitrile (**3**) (100 mg, 0.54 mmol, 1 eq.) was reacted as described above. The reaction mixture was suspended in water (2–3 mL) and the dark green solid was filtered off to give 2-(methylthio)-5-nitrosopyrimidine-4,6-diamine (64 mg, 0.35 mmol, 64%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ 10.18 (d, J = 4.2 Hz, 1H), 9.00 (s, 1H), 8.42 (d, J = 4.2 Hz, 1H), 8.02 (s, 1H), 2.46 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ 179.05, 164.73, 146.22, 139.43, 14.08. **HRMS** (ESI+): calc. for $[C_5H_8N_5OS]^+$ 186.0444, found: 186.0444 $[M + H]^+$

5-nitrosopyrimidine-2,4,6-triamine (4c). Guanidine (**2c**) salt of (hydroxyimino) malononitrile (**3**) (5.00 g, 32 mmol, 1 eq.) was heated slowly to its melting temperature of 159 °C and kept overnight. The compound melts suddenly but becomes a red/pinkish solid again after leaving it overnight. The quantitative reaction mixture can be directly used for the next step without purification.

For analytical reasons a small batch of guanidine (**2c**) salt of (hydroxyimino) malononitrile (**3**) (100 mg, 0.65 mmol, 1 eq.) was reacted as described above. The reaction mixture was suspended in water (2–3 mL) and the solid was filtered off to give NMR clean 5-nitrosopyrimidine-2,4,6-triamine (85 mg, 0.55 mmol, 85%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 10.26 (d, J = 5.1 Hz, 1H), 8.15 (s, 1H), 7.75 (d, J = 5.1 Hz, 1H), 7.35 (s, 1H), 7.19 (s, 2H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 166.52, 165.32, 151.43, 138.04. **HRMS** (ESI+): calc. for $[C_4H_7N_6O]^+$ 155.0676, found: 155.0676 $[M + H]^+$

2-methyl-5-nitrosopyrimidine-4,6-diamine (4d). Acetamidine (**2d**) salt of (hydroxyimino)malononitrile (**3**) (5.00 g, 32.5 mmol, 1 eq.) was heated slowly to its melting temperature of 142 °C and kept overnight. The compound melts suddenly but becomes a red/pinkish solid again after leaving it overnight. The quantitative reaction mixture can be directly used for the next step without purification.

For analytical reasons a small batch of acetamidine (**2d**) salt of (hydroxyimino) malononitrile (**3**) (100 mg, 0.65 mmol, 1 eq.) was reacted as described above. After suspension in H_2O (2–3 mL) the solid is filtered off to give NMR clean 2-methyl-5-nitrosopyrimidine-4,6-diamine (71 mg, 0.46 mmol, 71%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 10.04 (d, J = 3.2 Hz, 1H), 8.97 (s, 1H), 8.36 (d, J = 3.2 Hz, 1H), 7.96 (s, 1H), 2.20 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 175.20, 166.38, 146.76, 139.94, 26.83. **HRMS** (ESI-): calc. for $[C_5H_6N_5O]^-$ 152.0578, found: 152.0578 $[M-H]^-$

Data availability. All data generated or analyzed during this study are presented in this article and its Supplementary Information File, or are available from the corresponding author upon reasonable request. X-ray crystallographic data were also deposited at the Cambridge Crystallographic Data Centre (CCDC) under the following CCDC deposition numbers: 1574226 for 1-methylguanidine (**2a**) salt of (hydroxyimino)malononitrile (**3**); 1574223 for methylthioamidine (**2b**) salt of (hydroxyimino)malononitrile (**3**); 1574225 for guanidine (**2c**) salt of (hydroxyimino)malononitrile (**3**); 1574224 for acetamidine (**2d**) salt of (hydroxyimino) malononitrile (**3**). These can be obtained free of charge from CCDC via www.ccdc.cam.ac.uk/data_request/cif.

Received: 10 October 2017 Accepted: 14 December 2017
Published online: 11 January 2018

References

1. Benner, S. A., Kim, H.-J. & Carrigan, M. A. Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides, and RNA. *Acc. Chem. Res.* **45**, 2025–2034 (2012).
2. Budin, I. & Szostak, J. W. Expanding roles for diverse physical phenomena during the origin of life. *Annu. Rev. Biophys.* **39**, 245–263 (2010).
3. Pace, N. R. Origin of life-facing up to the physical setting. *Cell* **65**, 531–533 (1991).
4. Mulkidjanian, A. Y., Bychkov, A. Y., Dibrova, D. V., Galperin, M. Y. & Koonin, E. V. Origin of first cells at terrestrial, anoxic geothermal fields. *Proc. Natl. Acad. Sci. USA* **109**, E821–E830 (2012).
5. Djokic, T., Van Kranendonk, M. J., Campbell, K. A., Walter, M. R. & Ward, C. R. Earliest signs of life on land preserved in ca. 3.5 Ga hot spring deposits. *Nat. Commun.* **8**, 15263 (2017).
6. Damer, B. & Deamer, D. Coupled phases and combinatorial selection in fluctuating hydrothermal pools: a scenario to guide experimental approaches to the origin of cellular life. *Life* **5**, 872 (2015).
7. Deamer, D., Singaram, S., Rajamani, S., Kompanichenko, V. & Guggenheim, S. Self-assembly processes in the prebiotic environment. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **361**, 1809–1818 (2006).
8. Deamer, D., Van Kranendonk, M. & Djokic, T. The new origins of life. *Sci. Am.* **317**, 30–35 (2017).
9. Higgs, P. G. & Lehman, N. The RNA World: molecular cooperation at the origins of life. *Nat. Rev. Genet.* **16**, 7–17 (2015).
10. Wachowius, F., Attwater, J. & Holliger, P. Nucleic acids: function and potential for abiogenesis. *Q. Rev. Biophys.* **50**, e4 (2017).
11. Robertson, M. P. & Joyce, G. F. The Origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **4**, a003608 (2010).
12. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418**, 214–221 (2002).
13. Cafferty, B. J., Fialho, D. M., Khanam, J., Krishnamurthy, R. & Hud, N. V. Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. *Nat. Commun.* **7**, 11328 (2016).
14. Carell, T. et al. Structure and function of noncanonical nucleobases. *Angew. Chem. Int. Ed.* **51**, 7110–7131 (2012).
15. Rios, A. C. & Tor, Y. On the origin of the canonical nucleobases: an assessment of selection pressures across chemical and early biological evolution. *Isr. J. Chem.* **53**, 469–483 (2013).
16. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).
17. Stairs, S. et al. Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. *Nat. Commun.* **8**, 15270 (2017).
18. Becker, S. et al. A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833–836 (2016).
19. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301–307 (2015).
20. Fiore, M. & Strazewski, P. Bringing prebiotic nucleosides and nucleotides down to earth. *Angew. Chem. Int. Ed.* **55**, 13930–13933 (2016).
21. Kim, H.-J. et al. Synthesis of carbohydrates in mineral-guided prebiotic cycles. *J. Am. Chem. Soc.* **133**, 9457–9468 (2011).
22. Kofoed, J., Reymond, J.-L. & Darbre, T. Prebiotic carbohydrate synthesis: zinc-proline catalyzes direct aqueous aldol reactions of α-hydroxy aldehydes and ketones. *Org. Biomol. Chem.* **3**, 1850–1855 (2005).
23. Harsch, G., Bauer, H. & Voelter, W. Kinetik, katalyse und mechanismus der sekundärreaktion in der schlußphase der formose-reaktion. *Liebigs Ann.* **1984**, 623–635 (1984).
24. Meinert, C. et al. Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. *Science* **352**, 208–212 (2016).
25. Wächtershäuser, G. An all-purine precursor of nucleic acids. *Proc. Natl. Acad. Sci. USA* **85**, 1134–1135 (1988).
26. Kim, H.-J. & Benner, S. A. Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. *Proc. Natl. Acad. Sci. USA* **114**, 11315–11320 (2017).
27. Saladino, R. et al. Proton irradiation: a key to the challenge of N-glycosidic bond formation in a prebiotic context. *Sci. Rep.* **7**, 14709 (2017).
28. Saladino, R. et al. Meteorite-catalyzed syntheses of nucleosides and of other prebiotic compounds from formamide under proton irradiation. *Proc. Natl. Acad. Sci. USA* **112**, E2746–E2755 (2015).
29. Björk, G. R. et al. A primordial tRNA modification required for the evolution of life? *EMBO J.* **20**, 231–239 (2001).
30. Grosjean, H. *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution* (Taylor & Francis, Boca Raton, FL, 2009).
31. Engel, J. & von Hippel, P. H. Effects of methylation on the stability of nucleic acid conformations. *J. Biol. Chem.* **253**, 927–934 (1978).
32. Decatur, W. A. & Fournier, M. J. rRNA modifications and ribosome function. *Trends Biochem. Sci.* **27**, 344–351 (2002).
33. Motorin, Y. & Helm, M. tRNA stabilization by modified nucleotides. *Biochemistry* **49**, 4934–4944 (2010).
34. Helm, M. Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.* **34**, 721–733 (2006).
35. Forsythe, J. G. et al. Ester-mediated amide bond formation driven by wet–dry cycles: a possible path to polypeptides on the prebiotic earth. *Angew. Chem. Int. Ed.* **54**, 9871–9875 (2015).
36. Da Silva, L., Maurel, M.-C. & Deamer, D. Salt-promoted synthesis of RNA-like molecules in simulated hydrothermal conditions. *J. Mol. Evol.* **80**, 86–97 (2015).
37. Burcar, B. et al. Darwin's warm little pond: a one-pot reaction for prebiotic phosphorylation and the mobilization of phosphate from minerals in a urea-based solvent. *Angew. Chem. Int. Ed.* **55**, 13249–13253 (2016).
38. Kim, H.-J. et al. Evaporite borate-containing mineral ensembles make phosphate available and regiospecifically phosphorylate ribonucleosides: borate as a multifaceted problem solver in prebiotic chemistry. *Angew. Chem. Int. Ed.* **55**, 15816–15820 (2016).
39. Anantharaman, V., Koonin, E. V. & Aravind, L. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* **30**, 1427–1464 (2002).
40. Weiss, M. C. et al. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **1**, 16116 (2016).
41. Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004).
42. Jadhav, V. R. & Yarus, M. Coenzymes as coribozymes. *Biochimie* **84**, 877–888 (2002).

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for financial support through the research funding schemes SFB749, SFB1032, SPP1784, GRK2062, and the DFG Normalverfahrensprogram CA275-11/1. We thank the European Union Horizon 2020 program for funding the ERC Advanced project EPiR (741912). Further we thank the excellence cluster CiPS-M. We thank Florian Steinmann for synthetic help, Dr Markus Müller for further discussions and Dr Peter Mayer for providing X-ray structures.

Author contributions

S.B. developed the chemistry for the nitrosopyrimidine route, helped designing the study, analyzed and interpreted results and helped writing the manuscript. C.S., H.O., A.C., T.A. and M.D. performed supportive chemistry. T.C. designed the study, supervised all work, interpreted data and wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-017-02639-1.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2.5 FIFTH PUBLICATION

TITLE: Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides

AUTHORS:

S. Becker[+], J. Feldmann[+], S. Wiedemann[+], H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov, T. Carell*

*+ authors contributed equally (shared 1st authorship). * corresponding authors.*

JOURNAL, YEAR, VOLUME, AND PAGE NO.:

Science, **2019**, *366*, 76-82

SUMMARY:

RNA molecules are constructed from both purine and pyrimidine nucleotides, both of which would have likely been required for information transfer and for molecular evolution at the origin of life. Although separate pathways to both the pyrimidines and purines had until now been reported, no single, unified pathway leading to both had been demonstrated. Here, we show that the pyrimidine nucleotides can be generated together with the purines, which are themselves accessible *via* the nitrosopyrimidine chemistry already reported in our group. Synthesis of the pyrimidine nuceleosides involved an initial assembly of 3-aminoisoxazoylurea, which subsequently rearranges upon exposure to reductive conditions, thus generating the pyrimidine ring. The two pathways were consolidated through a continuous experiment involving cycling between wet and dry conditions. 5′-mono- and diphosphates were also generated upon treatment with phosphate containing minerals. Our results thus show that all four Watson-Crick bases could have plausibly been co-generated and accumulated on the early earth.

PERSONAL CONTRIBUTIONS:

- Provided supportive chemistry and assisted in the preparation of certain intermediates towards the purine assembly pathway.
- Carried out parallel investigations towards the synthesis of nucleoside phosphoramidites for the incorporation of 3-aminoisoxazoylureas into RNA (not included in the final publication).
- Contributed to the writing and preparation of the original manuscript.
- Contributed to the overall design of the study though discussions with S. Becker and T. Carell.
- Edited and revised the manuscript together with all co-authors.

PERMISION AND LICENSING

Unified prebiotically plausible synthesis of pyrimidine and purine RNA building blocks

Sidney Becker[1]*, Jonas Feldmann[1]*, Stefan Wiedemann[1]*, Hidenori Okamura[1], Christina Schneider[1], Katharina Iwan[1], Antony Crisp[1], Martin Rossa[1], Tynchtyk Amatov,[1] and Thomas Carell[1#]

[1]Center for Integrated Protein Science (CiPS[M]) at the Department of Chemistry, LMU München
Butenandtstr. 5-13, 81377 München
Fax: (+) 49 2180 77756
E-mail: Thomas.Carell@lmu.de
Homepage: www.carellgroup.de

Current affiliations:

Sidney Becker: Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, United Kingdom, CB2 1EW

Hidenori Okamura: Institute for Multidisciplinary Research for Advanced Materials, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, Miyagi 980-8577, Japan

Katharina Iwan: Centre for Translational Omics, University College London, Great Ormond Street Institute of Child Health, 30 Guilford Street, London, United Kingdom, WC1N 1EH

Tynchtyk Amatov: Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany

* These authors contributed equally

corresponding author: thomas.carell@lmu.de

One Sentence summary

Unified synthesis of pyrimidine and purine RNA building blocks under plausible prebiotic conditions.

The article was published in *Science* 04 Oct 2019: Vol. 366, Issue 6461, pp. 76-82,
DOI: 10.1126/science.aax2747

Abstract

Theories about the origin of life require chemical pathways that allow formation of life's key building blocks under prebiotically plausible conditions. Complex molecules like RNA must have originated from small molecules whose reactivity was guided by physico-chemical processes. RNA is constructed from purine and pyrimidine nucleosides, both of which are required for accurate information transfer. This is the prerequisite for Darwinian evolution. While separate pathways to purines and pyrimidines have been reported, their concurrent syntheses remain a challenge. We report the synthesis of the pyrimidine nucleosides from small molecules and ribose, driven solely by wet-dry cycles. In the presence of phosphate-containing minerals, 5'-mono- and di-phosphates also form selectively in one-pot. The pathway is compatible with purine synthesis, allowing the concurrent formation of all Watson-Crick bases.

Introduction

The discovery of catalytic RNA(*1*) and the development of replicating RNA systems(*2, 3*) have lent strong support to the concept of an RNA world(*4*). The RNA world hypothesis predicts that life started with RNAs that were able to (self)-recognize and replicate. Through a process of chemical evolution, a complex RNA and later RNA-peptide/protein world supposedly evolved, from which life ultimately emerged(*4*). A prerequisite for the RNA world is the ability to create RNA under prebiotic conditions. This requires as the first elementary step the concurrent formation of pyrimidine and purine nucleosides in the same environment. Here, they must have condensed to form information carrying polymers able to undergo Darwinian evolution. The question of how the pyrimidine and purine nucleosides could have formed together is an unsolved chemical problem, under intensive chemical investigation(*5-9*). Starting from an early atmosphere mainly composed of N_2 and CO_2,(*10*) the abiotic synthesis of life's building blocks must have occurred on the early Earth in aqueous environments, whose characteristics were determined by the minerals and chemical elements from which the early Earth's crust was made.(*11, 12*) Atmospheric chemistry, impact events and vulcanic activities must have provided the first reactive small molecules. These reacted in surface or deep-sea hydrothermal vents,(*13-15*) on mineral surfaces(*16*) or in shallow ponds.(*17*) Within these environments, volcanic activity, seasonal- or day-night cycles caused fluctuations of pH and temperature. Such fluctuating conditions provided wet-dry conditions allowing precipitation or crystallization of chemicals.(*18*) Mixing of micro-environments may have opened up new reaction pathways that led to increasing chemical complexity.

Along these geophysical boundaries, two main reaction pathways have been proposed for the formation of purine and pyrimidine nucleosides. The synthesis of the purines is possible along a continuous pathway based on the reaction of formamidopyrimidine (FaPy) precursors with ribose.(*6, 18*) For the pyrimidines a more stepwise reaction sequence involving aminooxazoles has been discovered.(*5*) These pathways provide the corresponding nucleosides under very different and partially incompatible conditions, leaving the question of how purines and pyrimidines could have formed in the same environment unanswered. Here, we report a prebiotically plausible pathway to pyrimidine nucleosides, which selectively provides the 5'-mono- and 5'-di-phosphorylated nucleosides as needed for RNA strand formation. By connecting the pathway with the reported purine route,(*6, 18*) we establish a unifying reaction network that allows for the simultaneous formation of both types of nucleosides in the same environment, driven by wet-dry cycles.

Results

Abiotic synthesis of pyrimidine nucleosides

The chemistry leading to pyrimidines starts from cyanoacetylene **1** as the key building block (Fig. 1A). Compound **1** is observed in interstellar clouds and in the atmosphere of Titan.(*19*) It has been shown to form in large quantities by electric discharge through an CH_4 / N_2 atmosphere(*20*) and is also a product of the Cu(II)-mediated reaction of HCN and acetylene in water (Fig. 1B2).(*21*) A recent report suggests that molecules such as **1** are plausible prebiotic starting materials, which could have formed in surface hydrothermal vents in significant concentrations.(*13*) We found that **1** reacts fast and cleanly with hydroxylamine **2** or hydroxylurea **3** to give 3-aminoisoxazole **4**. The reaction of **1** with **3** proceeds under slightly basic conditions (pH ~ 10) in 80 – 90% yield within 2 h. **3** is formed in almost quantitative yields from the reaction of **2** with cyanate.(*22*) Compound **4** formed robustly even if we varied the temperature (10 - 95 °C), the reactant concentrations (10 - 100 mM) or added additional compounds such as urea **5** and/or different metal ions (see below). Reaction of cyanoacetylene **1** with hydroxylamine **2** gives **4** with 17% yield after 2 h at pH 10.

While hydroxylamine **2** is an accepted building block for prebiotic amino acid syntheses,(*23*) its potential formation on the early Earth is unclear. We therefore aimed to demonstrate its prebiotic availability. **2** is ultimately produced by reduction from NO, which is formed in large quantities when lightning passes through moist atmospheres containing N_2 and CO_2 (Fig. 1B).(*10*) NO forms as the main product under these conditions and spontaneously reacts in the presence of water to nitrite (NO_2^-) and nitrate (NO_3^-), which leads to the assumption that both anions were quite abundant on the early Earth.(*24-26*). With Fe(II) as a plausible prebiotic reductant, NO_2^- is converted to NH_3 but not to NH_2OH **2**.(*26*) Formation of the latter requires a partial reduction. We now found that this can be achieved with HSO_3^-, which forms from volcanic SO_2 and water.(*27*) NO_2^- and HSO_3^- react to **2** with up to 85% yield (Fig 1B, Fig. S1).(*28*) We confirmed, that this reaction gives first the hydroxylamine disulfonate **6** (Fig. 1B), which hydrolyses to hydroxylamine **2** and HSO_4^-. We find that intermediate **6** reacts with cyanoacetylene **1** as well (88% yield, Fig. 1B, FigS2) to give the stable olefin **7**, which upon hydrolysis provides again the key intermediate **4**. The overall yield of **4** via compound **7** is 63% over these two steps. The suggested pyrimidine intermediate **4** is therefore readily available from cyanoacetylene **1** upon reaction with either **2**, **3** or **6** under prebiotic conditions (Fig 1B).

When we added urea **5** to a solution of **4**, warming (70°C - 95°C) and dry-down resulted in formation of *N*-isoxazolyl-urea **8** (Fig. 1A and 2A) in a spot-to-spot reaction that is catalysed by Zn^{2+} or Co^{2+}. These metal ions were likely present on the early Earth.(*11, 12*) In the presence of Zn^{2+}, compound **8** is formed in 88% yield after 2 d at 95 °C (at 70 °C the same yield is obtained after ~2-3 w). With Co^{2+}, 68% yield is achieved after 2 d at 95 °C. The reaction of **4** to **8** is in all cases a clean process, with the only impurity being unreacted **4** (Fig. 2A). The product **8** can be subsequently physically enriched. Addition of carbonated water to the dried down reaction mixture solubilizes **4**, **5** and **8**, leaving the metal ions as hydroxides or carbonates behind. Subsequent concentration of the supernatant leads to spontaneous crystallization of **8** (55%). This allowed us to obtain a crystal structure of **8** (Fig. S3). In order to simulate early Earth chemistry, we performed a one-pot experiment. We mixed **1** with **3**, **5**, and Zn^{2+} or Co^{2+} in a carbonate solution (pH ~10) and indeed obtained compound **4** at 95°C (80-90%). Neutralizing the solution to pH ~6-7, which may have occurred on the early Earth due to acidic rain, followed by dry-down at the same temperature provided compound **8** with yields between 56% (Zn^{2+}) and 40% (Co^{2+}). The continuous synthesis of the key building block **8** was consequently achieved in a plausible prebiotic setting that could have existed in hydrothermal vents or near volcanic activity, both of which are able to provide elevated temperatures (Fig. S3). The synthesis is also possible at lower temperatures, but with extended reaction times.

For the final step towards nucleosides, we need to assume that, due to flooding or a mixing of environments, **8** came into contact with ribose **9** (Fig. 1A and 2B) or any other sugar unit such as threose (for TNA) or glyceraldehyde (for GNA) able to form a backbone for a pairing system.(*29, 30*) When we mixed **8** with ribose **9** and warmed up the mixture to 95 °C in the presence of boric acid, we observed a fast and

high-yielding reaction to provide the ribosylated products **10a-d** in 95% yield (Fig. S4a). Other borate minerals such as synthetic lüneburgite ($Mg_3[(PO_4)_2|B_2(OH)_6]\cdot 6H_2O$)(*31*) or borax ($Na_2[B_4O_5(OH)_4]\cdot 10H_2O$) were also able to catalyze this reaction,(*32*) giving high-yields (>70%, Fig. S5). The major products are initially the α/β-pyranosides (**10c** and **10d**), which dominate over the α/β-furanosides (**10a** and **10b**, Fig. S4a). After heating the mixture under slightly basic conditions at 95°C in the presence of borates, the furanosides (54%, **10a** and **10b,** Fig. 2A) gradually became the dominant products (Fig. S4b). Under these conditions we also observed hydrolysis of **10a-d** to **8** and **9**. The accumulation of the furanosides **10a** and **10b** is best explained by complexation of their *cis*-diols with borate.(*32*)

The final step towards pyrimidine nucleosides requires reductive opening of the isoxazole N-O bond, followed by tautomerisation, intramolecular cyclization and water elimination in a cascade-like fashion (Fig. 2C and 2D). We found that this reaction occurs rapidly with Fe^{2+} in the presence of thiols (Fig. 2D).(*33*) LC-MS analysis indicated that cytidine nucleosides **11a**-**d** formed efficiently under these conditions, with the furanosidic uridine nucleosides **12a**,**b** being the corresponding deamination products formed by hydrolysis (Fig. 2C). Reductive pyrimidine formation can be performed with FeS or the mineral pyrite (FeS_2), and both have been discussed in the context of early metabolic pathways.(*15, 34*) Just 0.0001 eq. of soluble Fe^{2+} in water is sufficient for the reaction. In the absence of Fe^{2+}, pyrimidine formation was not observed. The reduction also appears to be independent of the thiol source, as the products **11a-d** and **12a**,**b** are obtained regardless of whether we used dithiothreitol (DTT), propanedithiol, mercaptoethanol or cysteine (Fig. S6).

Selective one-pot formation of 5'-nucleoside mono- and di-phosphates

The addition of naturally occurring minerals such as hydroxyapatite, colemanite or (synthetic) lüneburgite to the reductive pyrimidine-forming reaction had a strong influence on the distribution of the four cytidine isomers. Synthetic lüneburgite gave a combined high yield of 85% (Fig. 2C). The natural furanosidic β-cytidine (**11b**) and its α-anomer (**11a**) are formed under these conditions with about the same yields together with small amounts of α- and β-uridine (**12a**,**b**). Importantly, we found only small amounts of the α- and β-cytidine pyranosides (**11c** and **11d**), together with the cytosine base. Since synthetic lüneburgite is known to enable nucleotide formation in the presence of urea (Fig. 3A),(*31*) we simply added urea to the one-pot reaction mixture after pyrimidine formation and allowed the mixture to evaporate to dryness at 85°C over a period of about 20 h. LC-MS analysis of the reaction now showed formation of phosphorylated nucleosides (Fig. 3A) in remarkable 19% yield relative to cytidine (Fig. 3B and Fig. S7). We assumed that the reaction generated the α- and β-cytidine-5'-mono-phosphates **13a**/**b** and the 5'-di-phosphorylated cytidines **14a**/**b**. Due to hydrolysis we also expected some α- and β-uridine-5'-mono- and 5'-di-phosphates **15a**/**b** and **16a**/**b**. We isolated the corresponding HPLC peaks and removed the phosphate groups enzymatically (Fig. 3B). LC-MS analysis showed now the dephosphorylated furanosides **11a**/**b** and **12a**/**b** with over 94% in the nucleoside pool, which corresponds to a change of the furanoside/pyranoside ratio from initially 4:1 to now 17:1 (Fig. 3B). The formation of phosphorylated pyranosides **17** are only a minor side reaction. We found no discrimination between α- and β-anomers during the phosphorylation. The furanoside enrichment is best explained by the presence of a primary hydroxyl group in the furanosides, which is absent in the pyranosides. The enrichment of 5'-nucleoside-(mono and di)-phosphates under these one-pot conditions consequently establishes a further chemical selection step that favors the furanosides as the components of RNA. We further characterized the structures of the phosphorylated nucleosides and confirmed the formation of the 5'-α- and 5'-β-cytidine-mono- and di-phosphates (**13**/**14a** and **b**, α-/β-CMP and α-/β-CDP, Fig. S8). Additional analysis allowed identification of α,β-UDP **16a/b** (Fig. S9). Interestingly, 5'-pyrophosphates are the dominating specie within the di-phophorylated nucleoside mixture (Fig S8a).

Compatible formation of pyrimidine and purine RNA nucleosides

We next investigated if the prebiotically plausible pyrimidine and purine nucleoside pathways are compatible with each other so that they can be connected with the goal to form all Watson-Crick building blocks in the same environment solely driven by wet-dry cycles. The purine synthesis(*18*) requires as the initial step reaction of malononitrile **18** with sodium nitrite to give (hydroxyimino)malononitrile **19**. Because malononitrile **18** can be also generated from cyanoacetylene **1**, as shown by Eschenmoser,(*35*) pyrimidines and purines can be traced back to the same chemical root (Fig. 4). Compound **19** forms an organic salt with amidines **20** to give nitroso-pyrimidines **21**, and upon reduction and formylation, formamidopyrimidines (FaPys, **22**-**25**). The latter can react with ribose **9** to give ribosylated FaPys **26** and then purine nucleosides **27**-**29** (Fig. 1A).(*18*) To investigate how the chemical conditions needed for pyrimidine formation from the urea-isoxazole **8** would affect purine formation, we reacted **8** and the FaPy-compounds **22** and **23** with ribose **9** under dry-down conditions. We performed the reaction under identical conditions but in separate reaction vials (Fig. 4). Under these conditions formation of all four Watson-Crick nucleosides cytidine **11,** uridine **12**, adenosine **27** and guanosine **28**, were detected.

We next investigated if pyrimidines and purines can form simultaneously in the same environment (Fig. 5A). For this experiment, we mixed the starting materials cyanoacetylene **1**, hydroxylurea **3**, (hydroxyimino) malononitrile **19** and amidine **20** under slightly basic conditions (pH ~10). Analysis of the mixture showed indeed formation of **4** in 86% yield, despite the presence of **19** and **20**. It is surprising that the N-OH functionality of compound **19** does not interfere with the formation of **4**. Compound **4** is a liquid that can enrich from a water solution by dry-down due to its high boiling point (228°C). Interestingly, **4** can act as a solvent to facilitate the formation of **21** from the reaction of **19** with **20** under milder conditions (50 °C – 100 °C instead of 126 °C) compared to a previous procedure.(*18*) The next step requires reduction and formylation of **21** to the FaPy intermediate, which, however, cannot be performed in the presence of the isoxazole. Addition of water, eventually containing urea **5**, led to spontaneous precipitation of **21**. The supernatant containing **4** and **5** can flow away. The water insoluble **21** if brought into contact with dilute formic acid and Zn (found in Earth's crust) reacts immediately to the compounds **22** and **24** and Zn^{2+} as a side product (Fig. 5A, Fig. S10a). The reaction products are now water soluble and can potentially recombine with **4** and **5**. The side product Zn^{2+} is now catalyzing the reaction of **4** in the presence of **5** to give the *N*-isoxazolyl-urea **8** in the presence of **22** and **24** (Fig. 5A, Fig S10b). This leads to the formation of the pyrimidine and purine precursors **8**, **22** and **24**, which can be transformed into the purine and pyrimidine nucleosides. In this scenario, the intermediate **4** of the pyrimidine pathway helps formation of the purine precursor **21** while Zn^{2+} as a side product of the purine pathway mediates formation of the pyrimidine precursor **8** in a mutually synergistic way, driven by wet-dry cycles.

We combined **8** with different FaPy-intermediates and investigated if they could react in a one-pot scenario with ribose **9** to finally give the purine and pyrimidine nucleosides. To examine this, we dissolved a mixture of **8**, **22**, **25**, ribose **9** and boric acid and warmed the mixture up to 95 °C for 14 h allowing for slow evaporation of water. The solid material was then taken up with a slightly basic solution containing Fe^{2+} (0.0005 eq.) and DTT (1.5 eq.), and we allowed the mixture to warm up to 95 °C. HPLC-MS analysis proved that these conditions simultaneously provided the purine and pyrimidine nucleosides with cytidine (**11a-d**) and adenosine (**27**) as the main products. Interestingly, diaminopurine nucleosides (DA, **29**), which hydrolyse to guanosine **28**, form in this one-pot reaction as well (Fig. 5A, chromatogram). We noted additional formation of double ribosylated adenine (rib_2-A). Furthermore, the nucleoside **28** can be created in this scenario if we use **23** (R^1 = OH, R^2 = NH_2) as the starting material but the yields were lower.

Discussion

Ribose based RNA and the four canonical nucleosides A, G, C and U are central to modern life and to prebiotic hypotheses such as the "RNA world", in which RNA strands replicated and evolved to give

increasingly complex chemical systems.(*4*) Whether such RNAs were directly assembled from the canonical nucleotides (A, C, G and U bases) or if it evolved from a simpler proto-RNA system is unclear.(*36*)

Here we show that a reaction network towards the purine and pyrimdine RNA building blocks can be established, starting from simple atmospheric or volcanic molecules. Molecular complexity is generated by wet-dry cycles that can drive the chemical transformations. Therefore any environment that was able to provide wet-dry phases might have been a suitable place for the origin of RNA building blocks. Our geochemical model assumes that chemistry took place within in several basins that were needed to locally separate intermediates. We also need one or two streams of water to allow exchange of soluble molecules (Fig. 5B). Intermediates might precipitate upon fluctuations of physico-chemical parameters allowing for the separation of soluble and insoluble materials (e.g. **4** and **21**). After further reactions, which re-establish solubility, the compounds can be recombined (Fig 5B). For our scenario we need to assume that the early Earth provided environmental conditions that fluctuated between slightly acidic (pH 3), potentially caused by acidic rain (SO_2, NO_x), or basic (pH = 10) caused by carbonates. Even though most of the chemistry described here is performed at elavated temperatures, the reactions also occur at lower temperatures, but with substantially longer reaction times. We can assume that temperatures fluctuated on the early earth just like today due to day-night or seasonal cycles. Such fluctuations would certainly have brought about wet-dry cycles, akin to our modern climate of drought and rain. All the geophysical requirements needed for the reported chemistry including elevated temperatures could have existed in geothermal fields or at surface hydrothermal vents, which are plausible geological environments on early Earth.

Our proposed chemical pathways towards pyrimidines and purines begin with cyanoacetylene **1**, which could have formed in surface hydrothermal vents.(*13*) Reaction of **2**, **3** and **6** with **1** is the starting point for the pyrimidines, but if **1** reacts instead with ammonia, a pathway to malononitrile **18** as the precursor for purine synthesis is possible (Fig. 4).(*35*) Another key molecule for the synthesis of purines and pyrimdines is NO_2^-, which is needed to nitrosate malonitrile **18** to **19**.(*18*) NO_2^- is also crucial for the formation of hydroxylamine in the presence of HSO_3^-, which is formed from volcanic SO_2.(*27*) The concentration of NO_2^- that is reachable in a prebiotic setting is under debate, but it is speculated that the most likely place for its accumulation are shallow ponds, as needed for our scenario.(*17*) In general, the limited stability of NO_2^- would not be an issue, provided that it is rapidly captured by HSO_3^- upon its formation. Our model assumes a surface environment, where molecules, such as NO_2^-, HSO_3^- or urea **5**, could have been delivered by rain after their formation in the atmosphere (Fig. 5B).(*25, 37*) Importantly, our chemistry shows that robust reaction networks can be established that allow all key intermediates to be generated efficiently from relatively complex mixtures, followed by their physical enrichment or separation on the basis of their solubility in water. Wet-dry cycles govern the formation of purine and pyrimidine RNA building blocks in a scenario depicted in Fig. 5B. Of course, we will be unable to definitively prove that the described scenario took indeed place on early Earth, but the reported chemistry shows that, under plausible prebiotic conditions, mutually synergistic reaction pathways can be established in which the intermediates along one pathway help the chemistry of the other. In such a scenario, we show that the key building blocks of life can be created without the need for sophisticated isolation and purification procedures of reaction intermediates, which is common in traditional organic chemistry.

Importantly, the concurrent formation of pyrimidine and purine nucleosides in the network can be traced back to just a handful of key starting molecules such as cyanoacetylene **1**, NH_3, NH_2OH **2** (or the disulfonate **6**), HCN, urea **5**, formic acid and isocyanate plus salts such as nitrites, carbonates and borates. Metals such as Zn or Fe and their ions play an important role in our chemistry, consistent with their proposed involvement in early metabolic cycles.(*23, 38*) In particular, iron-sulfur surfaces needed for pyrimidine formation are discussed as platforms for early prebiotic chemistry.(*15, 34, 39*) The 5'-(di)phosphorylation is integrated into our pathway if phosphate minerals such as lüneburgite or struvite (Fig. S11) are present. It remains unclear, however, how ribose or any other carbohydrate, such as glycerol or threose, needed to form the backbone of RNA or pre-RNA could have formed selectively.(*29, 40*) Sugars such as ribose can be

produced non-selectively in a formose-like reaction, which is possible in a variety of different physico-chemical environments.(*32, 41-43*)

REFERENCES AND NOTES

1. J. A. Doudna, T. R. Cech, The chemical repertoire of natural ribozymes. *Nature* **418**, 222-228 (2002).
2. D. P. Horning, G. F. Joyce, Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci.* **113**, 9786-9791 (2016).
3. J. Attwater, A. Raguram, A. S. Morgunov, E. Gianni, P. Holliger, Ribozyme-catalysed RNA synthesis using triplet building blocks. *eLife* **7**, e35255 (2018).
4. W. Gilbert, Origin of life: The RNA world. *Nature* **319**, 618 (1986).
5. M. W. Powner, B. Gerland, J. D. Sutherland, Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239 (2009).
6. S. Becker *et al.*, A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833-836 (2016).
7. H.-J. Kim, S. A. Benner, Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. *Proc. Natl. Acad. Sci.* **114**, 11315-11320 (2017).
8. S. Stairs *et al.*, Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. *Nat. Commun.* **8**, 15270 (2017).
9. R. Saladino *et al.*, Meteorite-catalyzed syntheses of nucleosides and of other prebiotic compounds from formamide under proton irradiation. *Proc. Natl. Acad. Sci.* **112**, E2746-E2755 (2015).
10. J. Kasting, Earth's early atmosphere. *Science* **259**, 920-926 (1993).
11. R. M. Hazen, Paleomineralogy of the Hadean Eon: A preliminary species list. *Am. J. Sci.* **313**, 807-843 (2013).
12. E. D. Swanner *et al.*, Cobalt and marine redox evolution. *Earth. Planet. Sci. Lett.* **390**, 253-263 (2014).
13. P. B. Rimmer, O. Shorttle, Origin of Life's Building Blocks in Carbon- and Nitrogen-Rich Surface Hydrothermal Vents. *Life* **9**, 12 (2019).
14. W. Martin, J. Baross, D. Kelley, M. J. Russell, Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **6**, 805 (2008).
15. E. Camprubi, S. F. Jordan, R. Vasiliadou, N. Lane, Iron catalysis at the origin of life. *IUBMB Life* **69**, 373-381 (2017).
16. S. A. Benner, H.-J. Kim, E. Biondi, in *Prebiotic Chemistry and Chemical Evolution of Nucleic Acids,* C. Menor-Salván, Ed. (Springer International Publishing, Cham, 2018), pp. 31-83.
17. S. Ranjan, Z. R. Todd, P. B. Rimmer, D. D. Sasselov, A. R. Babbin, Nitrogen Oxide Concentrations in Natural Waters on Early Earth. *Geochem. Geophys. Geosyst.* **20**, 2021-2039 (2019).
18. S. Becker *et al.*, Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. *Nat. Commun.* **9**, 163 (2018).
19. P. Thaddeus, The prebiotic molecules observed in the interstellar gas. *Philos. Trans. Royal Soc. B* **361**, 1681-1687 (2006).
20. R. A. Sanchez, J. P. Ferris, L. E. Orgel, Cyanoacetylene in Prebiotic Synthesis. *Science* **154**, 784-785 (1966).
21. B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, J. D. Sutherland, Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301 (2015).
22. H. Kofod, On the Isomerism of Hydroxyurea. I. Kinetics of the Reaction between Hydroxylammonium Ion and Cyanate Ion. *Acta Chem. Scand.* **7**, 274-279 (1953).
23. K. B. Muchowska, S. J. Varma, J. Moran, Synthesis and breakdown of universal metabolic precursors promoted by iron. *Nature* **569**, 104-107 (2019).

24. V. S. Airapetian, A. Glocer, G. Gronoff, E. Hébrard, W. Danchi, Prebiotic chemistry and atmospheric warming of early Earth by an active young Sun. *Nat. Geosci.* **9**, 452 (2016).
25. H. J. Cleaves, J. H. Chalmers, A. Lazcano, S. L. Miller, J. L. Bada, A Reassessment of Prebiotic Organic Synthesis in Neutral Planetary Atmospheres. *Orig. Life Evol. Biosphere* **38**, 105-115 (2008).
26. D. P. Summers, S. Chang, Prebiotic ammonia from reduction of nitrite by iron (II) on the early Earth. *Nature* **365**, 630-633 (1993).
27. R. Sukrit, T. Z. R., S. J. D., S. D. D., Sulfidic Anion Concentrations on Early Earth for Surficial Origins-of-Life Chemistry. *Astrobiology* **18**, 1023-1040 (2018).
28. G. K. Rollefson, C. F. Oldershaw, The reduction of nitrites to hydroxylamine by sulfites. *J. Am. Chem. Soc.* **54**, 977-979 (1932).
29. G. F. Joyce, The antiquity of RNA-based evolution. *Nature* **418**, 214 (2002).
30. D. M. Fialho *et al.*, Glycosylation of a model proto-RNA nucleobase with non-ribose sugars: implications for the prebiotic synthesis of nucleosides. *Org. Biomol. Chem.* **16**, 1263-1271 (2018).
31. K. Hyo-Joong *et al.*, Evaporite Borate-Containing Mineral Ensembles Make Phosphate Available and Regiospecifically Phosphorylate Ribonucleosides: Borate as a Multifaceted Problem Solver in Prebiotic Chemistry. *Angew. Chem. Int. Ed.* **55**, 15816-15820 (2016).
32. A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, Borate Minerals Stabilize Ribose. *Science* **303**, 196-196 (2004).
33. M. Kijima, Y. Nambu, T. Endo, Reduction of cyclic compounds having nitrogen-oxygen linkage by dihydrolipoamide-iron(II). *J. Org. Chem.* **50**, 1140-1142 (1985).
34. G. Wächtershäuser, Before enzymes and templates: theory of surface metabolism. *Microbiol. Rev.* **52**, 452-484 (1988).
35. U. Trinks, dir. A. Eschenmoser, ETH Zurich, Zurich (1987) (ETH e-collection 10.3929/ethz-a-000413538).
36. R. Krishnamurthy, On the Emergence of RNA. *Isr. J. Chem.* **55**, 837-850 (2015).
37. J. Liebig, F. Wöhler, Untersuchungen über die Cyansäure. *Ann. Phys.* **96**, 369-400 (1830).
38. M. Preiner *et al.*, A hydrogen dependent geochemical analogue of primordial carbon and energy metabolism. *bioRxiv*, 682955 (2019).
39. C. Bonfio *et al.*, Prebiotic iron–sulfur peptide catalysts generate a pH gradient across model membranes of late protocells. *Nat. Catal.* **1**, 616-623 (2018).
40. O. Trapp, J. Teichert, F. Kruse, Direct Prebiotic Pathway to DNA Nucleosides. *Angew. Chem. Int. Ed.* **0**.
41. J. Kofoed, J.-L. Reymond, T. Darbre, Prebiotic carbohydrate synthesis: zinc–proline catalyzes direct aqueous aldol reactions of α-hydroxy aldehydes and ketones. *Org. Biomol. Chem.* **3**, 1850-1855 (2005).
42. C. Meinert *et al.*, Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. *Science* **352**, 208-212 (2016).
43. K. Usami, A. Okamoto, Hydroxyapatite: catalyst for a one-pot pentose formation. *Org. Biomol. Chem.* **15**, 8888-8893 (2017).

ACKNOWLEDGEMENTS

We thank Jonathan Kampmann for XRD measurements and Shankar Balasubramnian for supporting S.B. during the revision of the manuscript. **Funding:** We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support via the programs SFB1309 (TP-A4), SFB749 (TP-A4), SPP-1784, GRK2062/1, and CA275/11-1, the Excellence Cluster EXC114, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° EPiR 741912), and the Volkswagen foundation (Initiative “Life”: EcoRib). H. O. thanks the European Commission for a Marie Sklodowska-Curie postdoctoral fellowship (PRENUCRNA). **Author contribution:** T.C. designed and supervised research; S.B. helped to design the study, S.B., J.F., S.W. and H.O. performed the experiments. C.S., M.R., A.C. supported

the synthesis and MS quantification, K.I. performed biochemical studies, T.A. helped to design the synthesis. T.C., S.B., J.F., S.W. analysed data. T.C. and S.B. wrote the manuscript and designed the figures. **Competing interest:** The authors declare no competing interest. **Data and material availability:** The X-ray crystallographic data for isoxazoleurea **8** are deposited in the CCDC under accession number 1889652. All other data needed to support the conclusions of this manuscript are included in the main text and supporting material.

Figures

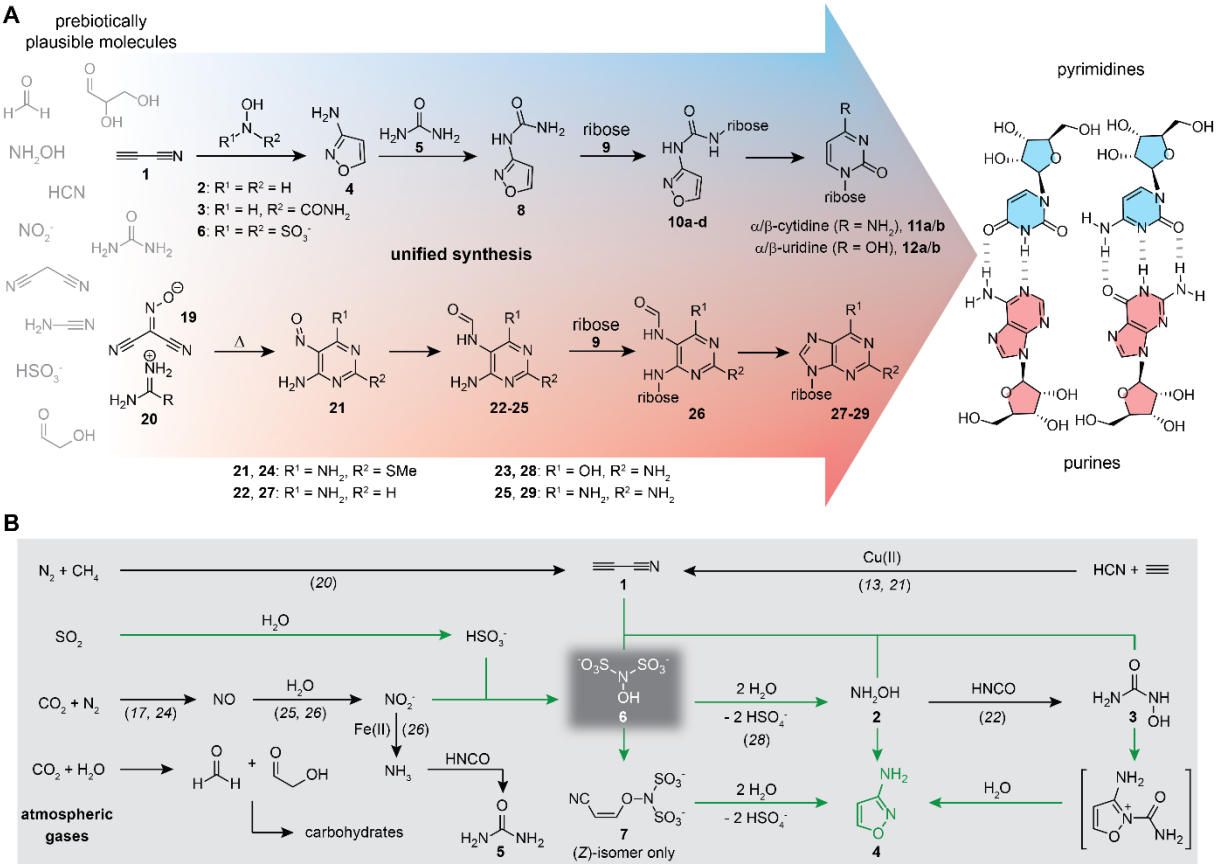


Fig. 1. Unified synthesis of pyrimidine and purine RNA building blocks. (**A**) Starting from plausible prebiotic molecules the reaction scheme depicts the route towards the pyrimdines *via* isoxazolylurea **8** (blue background) and the purines *via* formamidopyrimdines **22-25** (red background).(*18*) (**B**) Fundamental chemistry that produces the molecules needed for the pyrimidine pathway. Reactions performed in this work are shown with green arrows, while black arrows represent well known literature reactions. Formation of **4** requires reaction of **1** with hydroxylamine **2**, hydroxylurea **3** or the disulfonate **6** (dark grey box). **6** is formed from NO_2^- and SO_2/HSO_3^-.

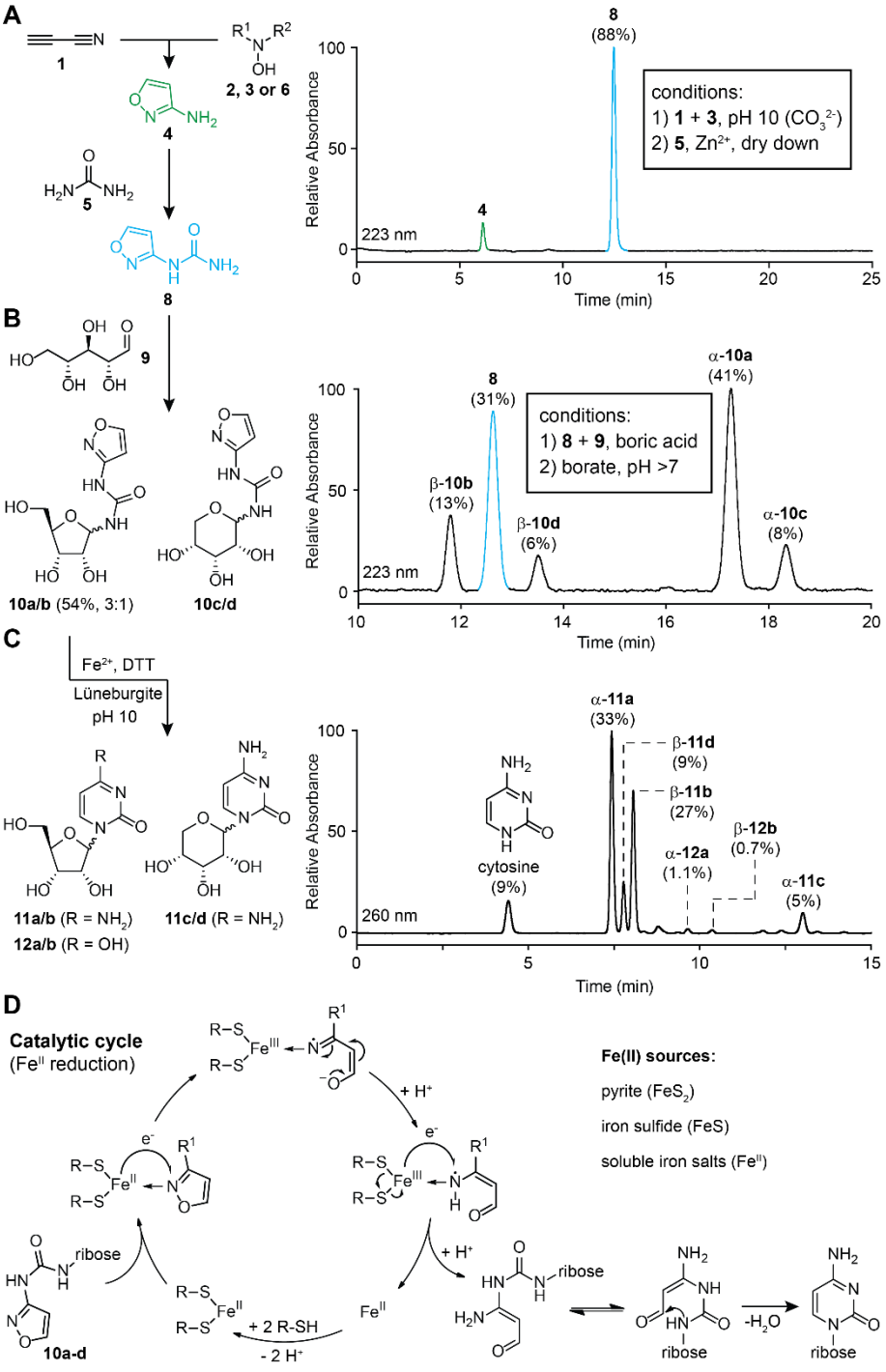


Fig. 2. Formation of pyrimidine nucleosides (**11** and **12**) from *N*-isoxazolylurea-ribosides **10a**/**b**. The different isomers are labelled according to: **a** = α-furanosyl, **b** = β-furanosyl, **c** = α-pyranosyl, and **d** = β-pyranosyl. (**A**) Formation of **4** and its conversion with urea **5** to *N*-isoxazolylurea **8**. (**B**) Ribosylation of **8** with ribose **9** and equilibration of the reaction mixture in the presence of borates gives the furanosidic isomers **10a** and **10b** (54%). (**C**) Pyrimidine nucleoside formation by reductive N-O cleavage from the compound mixture **10a**/**b** in the presence of ammonium iron(II) sulfate hexahydrate (0.0005 eq.). The HPL-chromatogram with detection at 260 nm shows formation of cytidine (C, **11a-d**) and uridine (U, **12a**/**b**). (**D**) Proposed catalytic cycle for the Fe^{2+} catalysed reduction of the N-O bond of the isoxazole moiety.

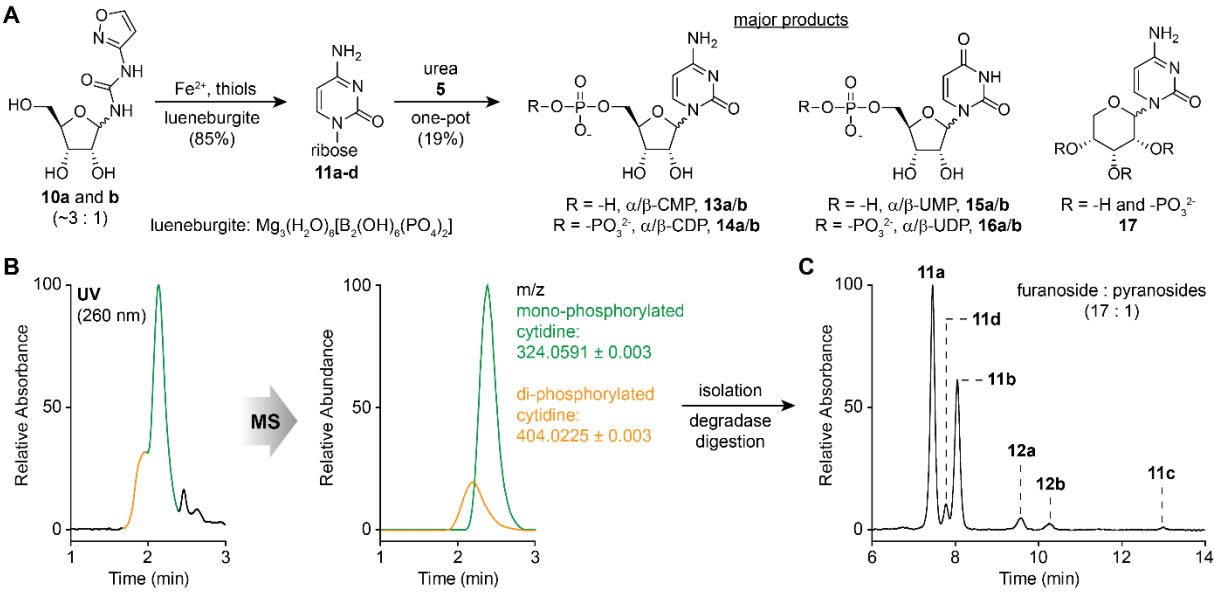


Fig. 3. One-pot nucleotide formation reaction. (**A**) One-pot synthesis of cytidine and uridine 5'-mono- and 5'-di-phosphates (**13a/b**-**16a/b**) after urea addition to the reaction mixture and allowing the mixture to dry-down at 85°C for 20 h. **a/b** represents the α- and β- anomers, respectively. (**B**) LC-MS analysis of the corresponding nucleotide peaks with UV- and MS-detection and isolation of the formed nucleotides from the prebiotic reaction, followed by an enzymatic removal of the phosphate groups. (**C**) HPLC analysis of the dephosphorylated product mixture showing predominant formation of α- and β-cytidine **11a** and **11b**.

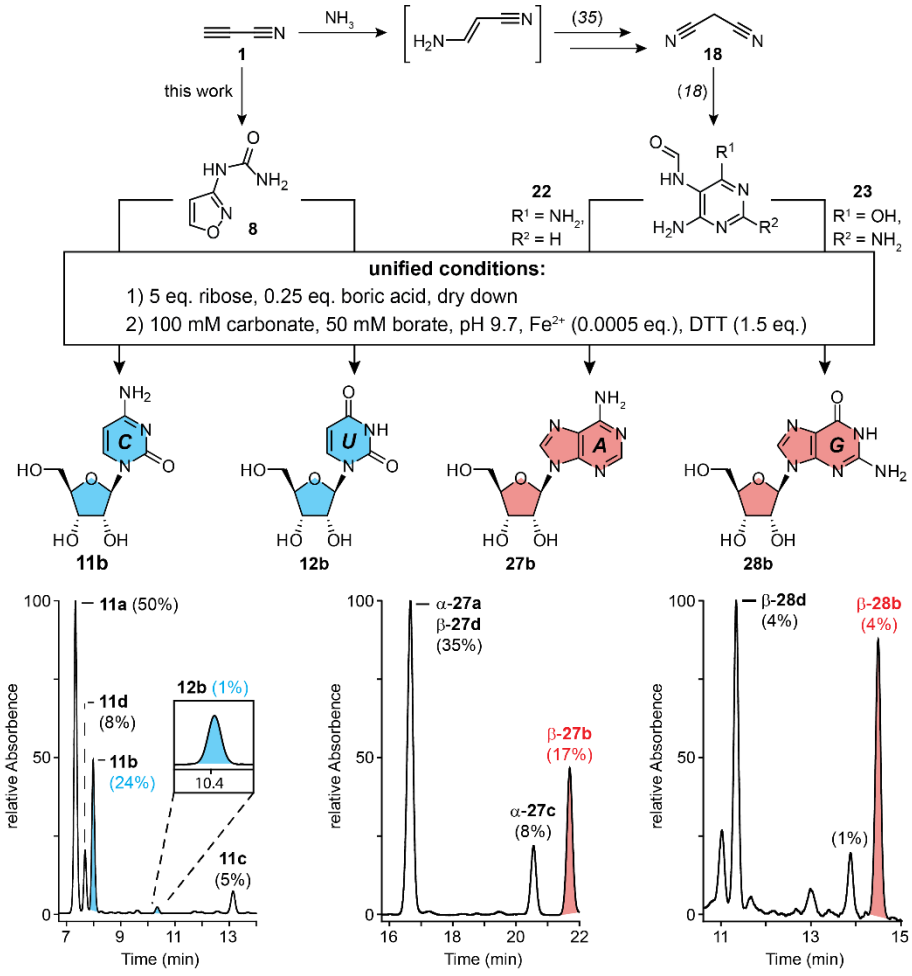


Fig. 4. Formation of all four Watson-Crick RNA building blocks in identical but parallel reactions. C (**11b**), U (**12b**), A (**27b**), and G (**28b**) are formed under the same conditions separately from **8**, **22** and **23**. HPL-chromatograms are shown with a detection at 260 nm. The nucleosides are labelled according to: **a** = α-furanosyl, **b** = β-furanosyl, **c** = α-pyranosyl, and **d** = β-pyranosyl. Canonical pyrimidine and purine RNA building blocks are labeled in blue or red, respectively.

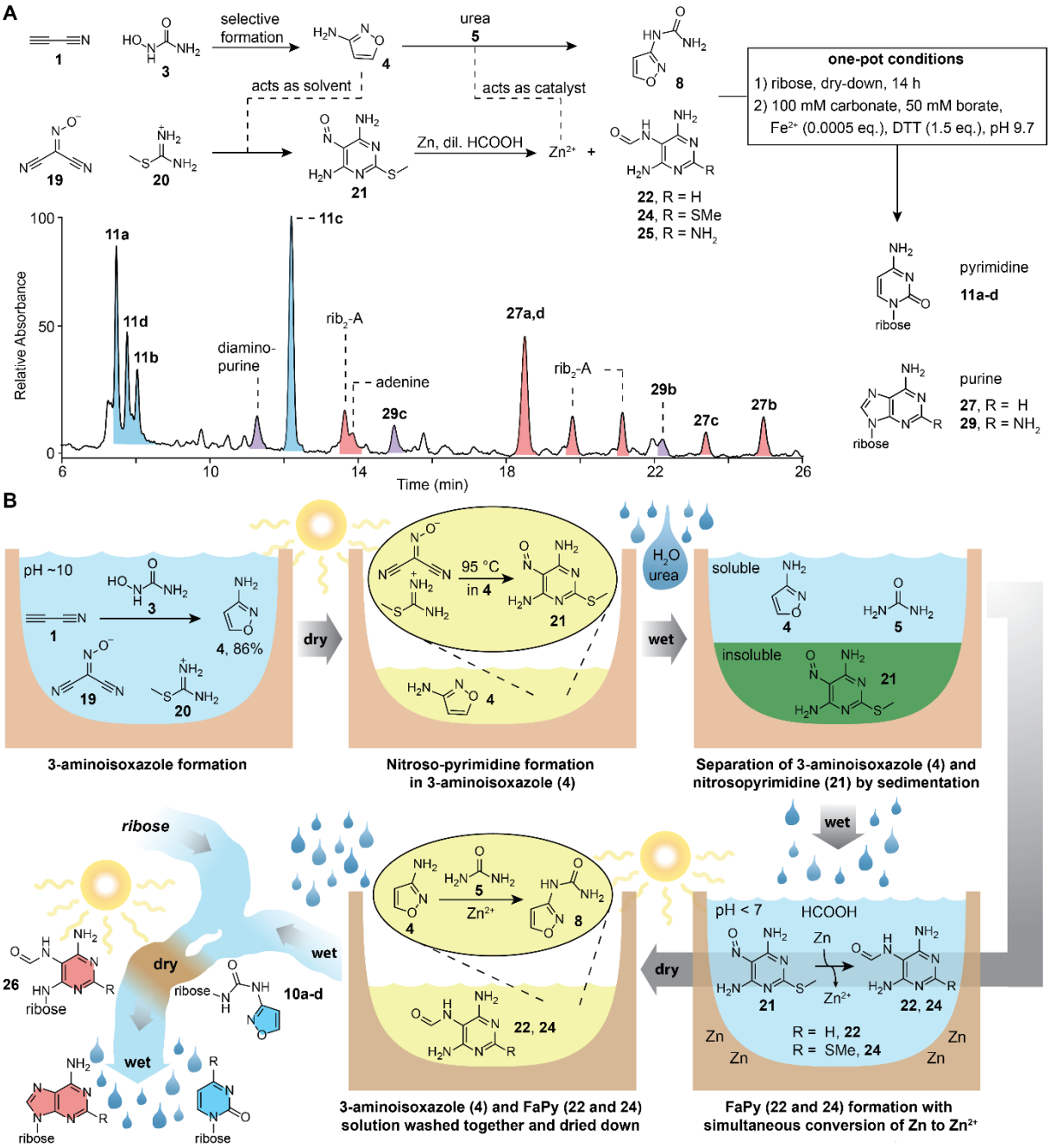


Fig. 5. Unified chemical scenario for the formation of purine and pyrimidine nucleosides. (**A**) Depiction of the connected reaction pathways to pyrimidine and purine nucleosides together with the HPLC analysis (260 nm) of the final reaction mixtures. (**B**) Proposed geochemical scenario for the simultaneous synthesis of purine and pyrimidine nucleosides, driven by wet-dry cycles. In yellow, the solvent is 3-aminoisoxazole (**4**), which can be enriched from an aqueous solution due to its high boiling point (228 °C). 2-(Methylthio)-5-nitrosopyrimidine-4,6-diamine (**21**) is a general precursor for adenosine and guanosine.(*18*) Compounds **8**, **22** and **24** are accessible in the same pot and they can react with ribose to the RNA nucleosides in a one-pot reaction. Nucleosides are labeled according to: **a** = α-furanosyl, **b** = β-furanosyl, **c** = α-pyranosyl, and **d** = β-pyranosyl.

3 UNPUBLISHED WORKS

3.1 AN ALTERNATIVE PREBIOTICALLY PLAUSIBLE PATHWAY TO CERTAIN *N*-6 CARBAMOYLATED RNA NUCLEOSIDES

3.1.1 Prologue

Deoxyribonucleic acid (DNA) and Ribonucleic acid (RNA) constitute the fundamental molecules of life that encode genetic information and act as the blueprint of all organisms. It has become increasingly clear, however, that the chemical diversity established by the four canonical Watson/Crick bases (adenine; guanine; cytosine; thymine/uracil) is not alone sufficient to account for the variety of complex functions that RNA performs beyond translation of genes. To date, over 100 modified RNA nucleosides have been discovered, with alterations to both the base-pairing and ribose moieties.[100] More recently still, evidence emerged suggesting that many of these modifications may have in fact been present in the genome of the last universal common ancestor (LUCA),[76] which prompted us to investigate whether particular modified nucleosides could have emerged in parallel to the canonical nucleosides on the early Earth. To such an end, we demonstrated in 2018 that certain methylated nucleosides can be generated in parallel to amino acid-modified nucleosides with the general formula **55** upon treatment of adenosine **33** with aminoacylated-nitrosoureas **56** in the presence of mild aqueous base (**Figure 9a**).[101]

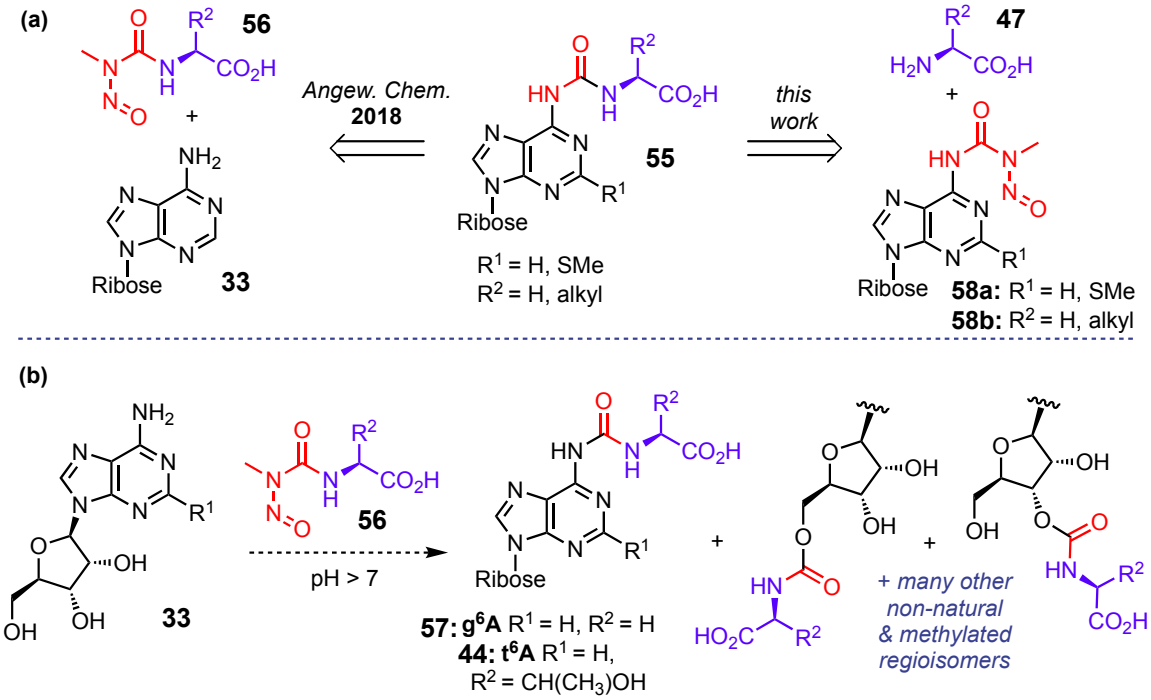
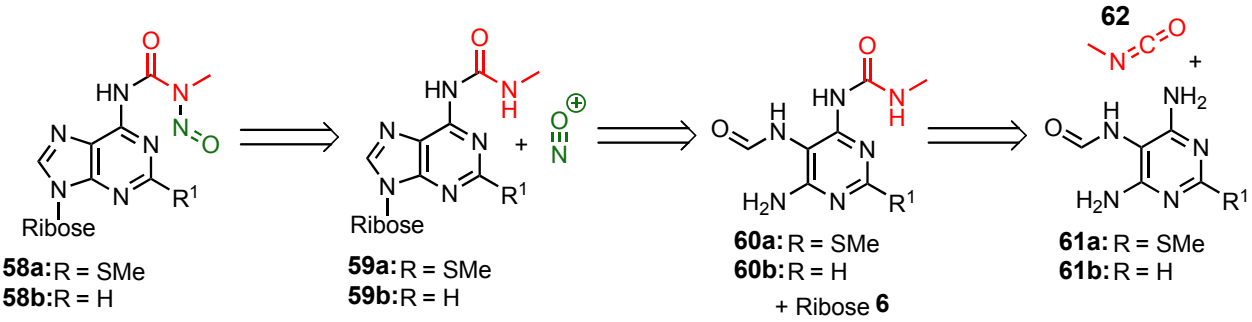


Figure 9 (a) *Comparison between our 2018 synthesis of carbamoylated adenosine nucleosides and the primary goal of this work.*[101] **(b)** *Treatment of nucleosides with aminoacylated-nitrosoureas generates a complex mixture of carbamoylated and methylated regioisomeric products.*

Although this type of model reaction was able to successfully generate the carbamoylated nucleosides g^6A **57** as well as t^6A **44**, yields were low due to the plethora of methylated and (likely) regioisomeric side products that were also generated (**Figure 9b**). This result supported our hypothesis that the canonical-nucleosides were selected from a more diverse set of bases. We nevertheless wondered whether it would be possible to generate nucleosides such as g^6A **57** and t^6A **44** more selectively and in greater yields under prebiotically plausible conditions. Accordingly, we envisioned an alternative strategy that would involve, as the critical step, *N*-nitrosation of a methylurea substituent at the *N*-6-position of adenosine (**58**), and subsequent addition of the relevant amino acid (**47**) (**Figure 9a**). We anticipated that such a strategy would also be effective for synthesising 2-thiomethyl-modified derivatives ms^2g^6A and ms^2t^6A, inspired by our earlier synthesis of ms^2A.[102]

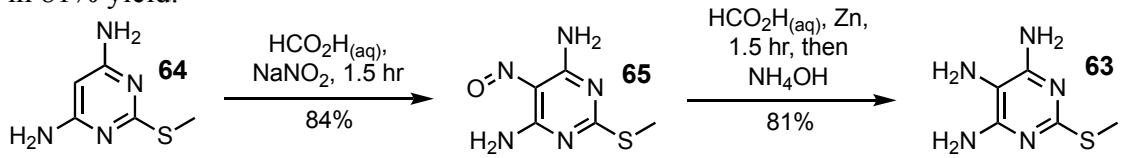
The proposed synthesis of compounds of the type **55** would firstly involve *N*-nitrosation in the presence of a methylurea-modified nucleosides **59a/b** to obtain **58a/b**, which would collapse upon exposure to base, thus forming a reactive isocyanate (**Scheme 10**). Compounds **59a/b** could themselves be obtained from the relevant formamidopyrimidine (FaPy) compounds **60a/b** *via* the dry-state ribosylation chemistry already published by our group.[70] Finally, compound **60a/b** could be obtained upon treatment of **61a/b** with the prebiotically plausible molecule, methylisocyanate **62**.[103] Our revised strategy would thus take advantage of the rotational C_{2v} symmetry present in **61a/b**, resulting as we would expect, in only a single anticipated carbamoylation product **60a/b** when one molar equivalent of methylisocyanate **62** was used.



Scheme 10 *Retrosynthetic analysis of the nitrosourea **58**.*

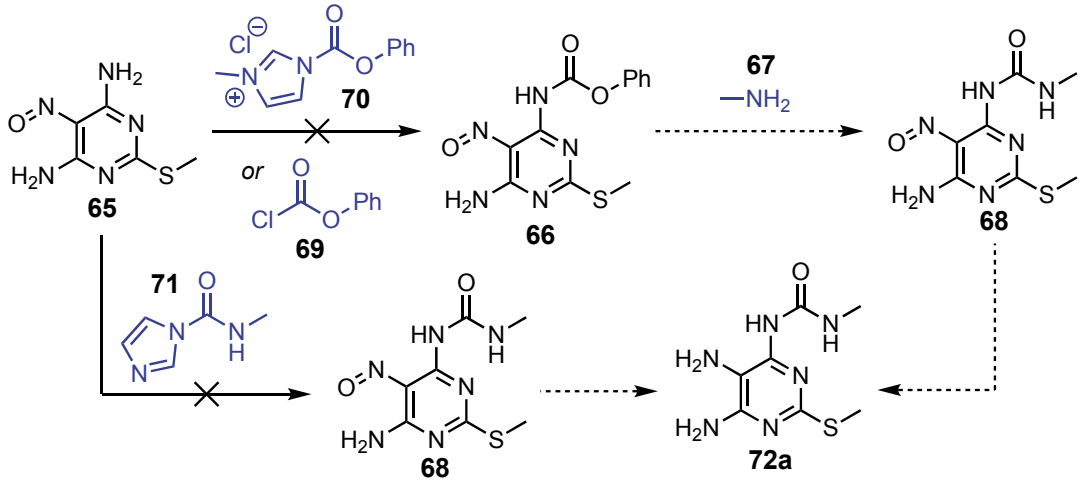
3.1.2 Results and Discussion

The first step towards establishing the prebiotically plausible synthesis of compounds with of the type **55**, would be to prepare the relevant references molecules **60a**/**b**, and **59a/b**, with which to develop model reactions. Based upon our prior syntheses of formamidopyrimidines, we anticipated that **60a** should be accessible *via* the direct, nucleophilic carbamoylation of pyrimidine **63**, followed by elaboration to give the *N*-5 formyl group. To this end, we began by synthesising compound **63** from the commercially available diamine **64** (**Scheme 11**). Upon treatment of **64** with sodium nitrite in the presence of acetic acid, a rapid reaction ensued to give compound **65** as a bright blue powder in 84% yield, whose ^{1}H NMR spectra were identical to those reported in the literature. Compound **65** was then able to be converted into **63** upon Zn-mediated reduction of the nitroso moiety to a primary amine in 81% yield.



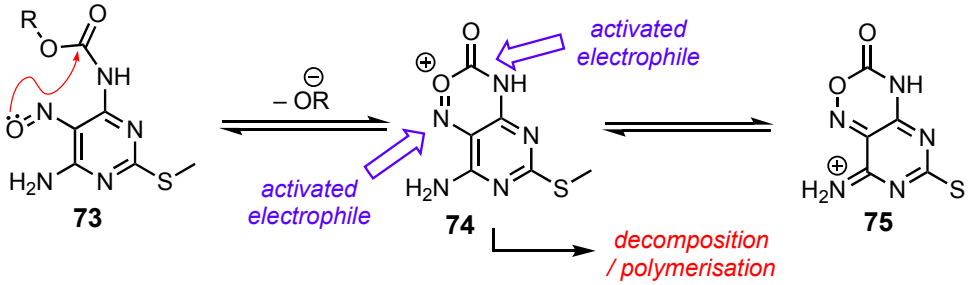
Scheme 11 *The synthesis of triaminopyrimidine **63** from diaminoprymidine **64**.*

In another preliminary strategy, we aimed to introduce the methylurea moiety *via* the direct carbamoylation of **65** to give **66**, followed by nucleophilic acyl substitution with methylamine **67**, providing **68** (**Scheme 12**). In the event, however, we observed that the treatment of **65** with either phenylchloroformate **69**, or alternatively, the less reactive 3-methyl-imidazolium carbamate **70** resulted in rapid decomposition, as observed by thin layer chromatography. Adjustments to temperature or solvent (DMSO; DMF) further did not result in isolable product, and solubility issues prohibited the use of other typical solvents such as dichloromethane. In an alternative approach, compound **65** was treated directly with *N*-methyl-1*H*-imidazole-1-carboxamide **71**, itself prepared from carbonyldiimidazole (CDI) and methylamine **67**, which gave similar results.



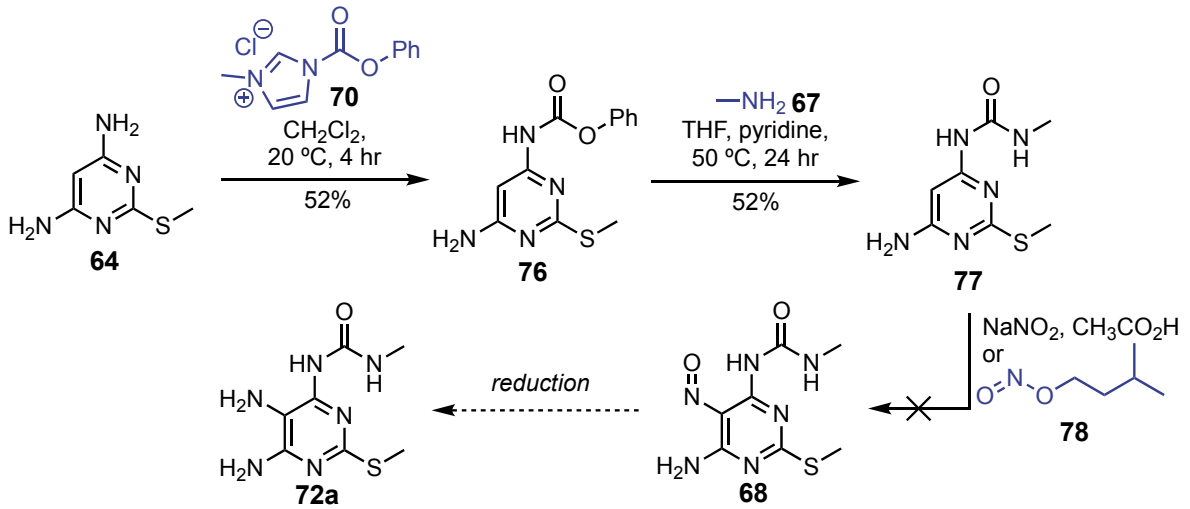
Scheme 12 *Our initial synthetic strategy for preparing the methylurea **72a** from **65**.*

A plausible mechanism accounting for the observed instability of carbamoylation products of **65** is depicted in (**Scheme 13**). Firstly, the direct product from carbamoylation **73** could undergo an intramolecular cycloannulation, as has been well documented among aryl-nitroso species such as β-nitroso-o-quinone methides.[104] The resulting doubly-activated carbamate **74**, which is itself in equilibrium with **75**, could then undergo a number of polymerisation or decomposition pathways, perhaps involving nucleophilic attack at either carbamate-carbonyl, or nitroso functional groups, such as is observed in the Liebermann nitroso reaction.[105]



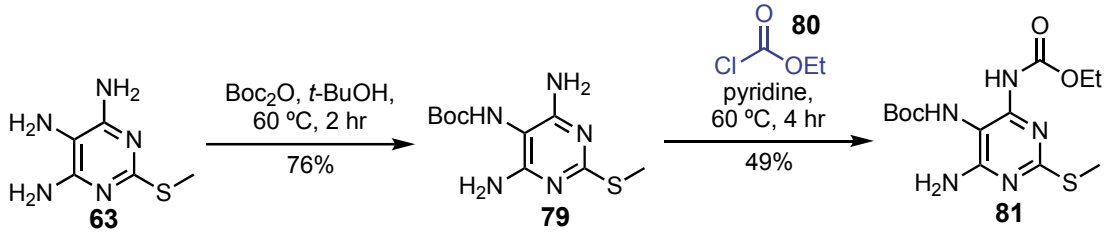
Scheme 13 *A possible mechanism responsible for the apparent instability of compounds such as 73.*

In order to circumvent the aforementioned reactivity issues, we next sought to introduce the carbamate- and urea-functionality prior to the nitroso group (**Scheme 14**). Compound **64** was accordingly treated with phenyl carbamate **70** in dichloromethane, which gave the desired product **76** in 52% yield. Notably, the decomposition that had previously been observed with compound **65** no longer appeared to be an issue. With this in mind, we next carried out the required substitution reaction of **76** with methylamine **67**, to give the methylurea **77**, also in 52% yield. Discouragingly, treatment of **77** with either of sodium nitrite or isoamylnitrite **78** did not yield formation of the desired product, but instead gave a complex mixture which we tentatively interpreted to contain homodimers, based upon the observed mass to charge ratios detected in the mixture. This outcome is particularly consistent with the observation that aryl nitroso moieties are prone to dimerization.[106]



Scheme 14 *The synthesis of methyl urea 77 from diamine 64 and its attempted nitrosation.*

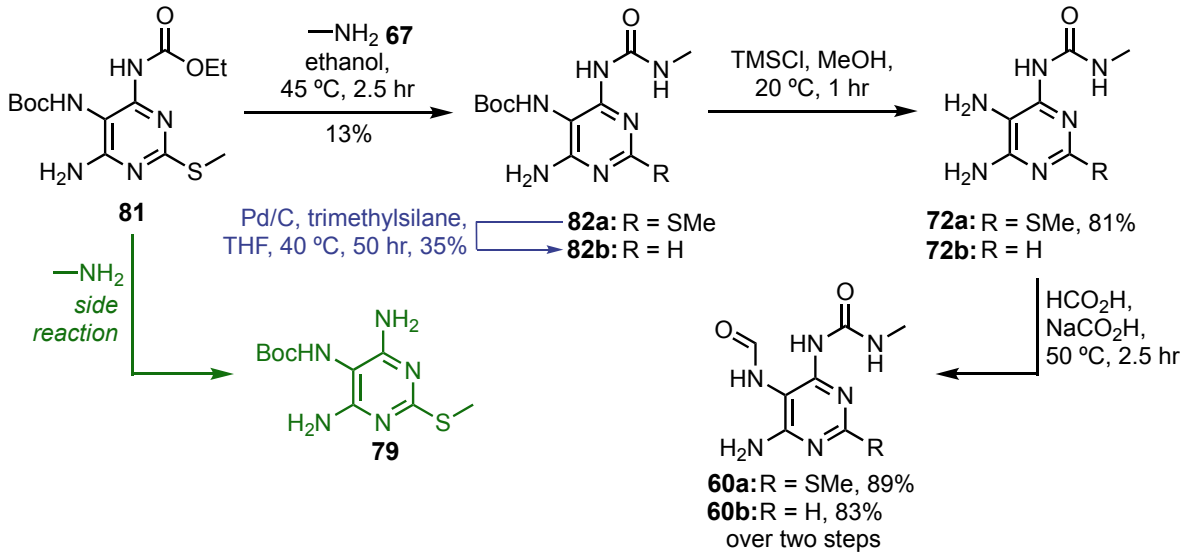
In our next approach, we aimed to avoid complications associated with the inherent reactivity and instability of the nitroso functional group by instead employing the triamine **63** (**Scheme 15**). Due to the enhanced *N*-5 nucleophilicity of **63** relative to the *N*-4 and -6 positions (resulting from protonation of the in-ring nitrogen atoms),[70] we anticipated that a protecting group for the exocyclic amine would be required. In the event, treatment of **63** with di-tert-butyl dicarbonate (Boc_2O) gave a rapid reaction at 60 ºC, resulting in the formation of **79** which precipitated from solution upon cooling to room temperature, as noted by a sudden colour change. To our gratification, subsequent treatment of **79** with ethylchloroformate **80** did not result in decomposition, as with **73**, but instead gave the desired *N*-4,5-di-carbamate **81**. The moderate yield of this reaction can perhaps be attributed to hydrolysis during purification, as reversal to give **79** was repeatedly observed during column chromatography.



Scheme 15 *Synthesis of the carbamate **81** from the triamine **63**.*

With compound **81** in hand, we would next need to convert the *N*-4 carbamate into a methylurea *via* nucleophilic acyl substitution of the ethanoate moiety **(Scheme 16**). Compound **81** was *thus* treated with a solution of ethanolic methylamine at 45 ºC, which gave the desired urea **82a** after 2.5 hours. In our hands, it was observed that the aforementioned reaction was highly sensitive to subtle changes in changes in temperature, reaction time, and concentration of methylamine. When, for example, we carried out the same reaction at 60 ºC instead of 45 ºC, only the hydrolysis product **79** was obtained. We noted similar outcomes when either concentration of methylamine was increased, or when the reaction was allowed to proceed for longer extents of time. We therefore chose to proceed, given that the current yields should be sufficient to facilitate our prebiotic experiments. Unreacted starting material was also recovered from each reaction, allowing us to resubject material in order to achieve higher-throughput. A portion of compound **82a** was then converted into **82b** *via* reductive desulfurisation using hydride, which was itself supplied by trimethylsilane in the presence of palladium on charcoal. In order to affect the deprotection of the Boc moiety, we first tried treating **82a** and **82b** with solutions of hydrogen chloride in H_2O and dioxane, although yields were limited. We then instead elected to use TMSCl, which upon exposure MeOH allowed us to generate specific concentrations of HCl *in situ*. In the event, treatment of a solution of **82a** in methanol with TMSCl gave the desired compound **72a** in 81% yield after 1 hour. The same reaction was also effective at generating **72b** from **82b**, where we opted to use the crude product directly for subsequent reactions.

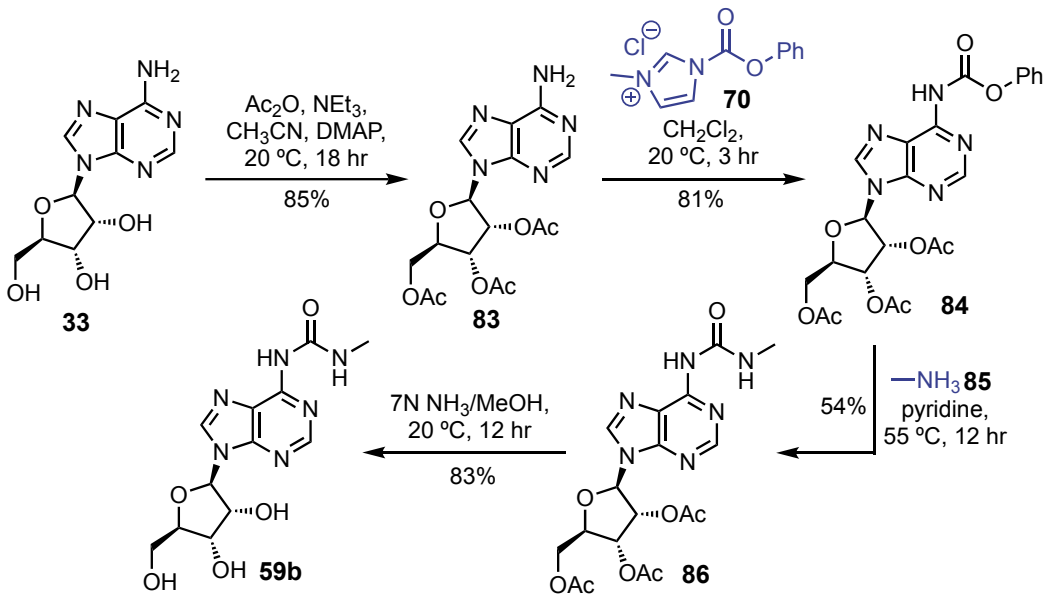
Compounds **72a** and **72b** were then successfully converted into their corresponding formamidopyrimidines (FaPys) using mixtures of sodium formate in formic acid in excellent yields (89% for **60a** and 83% over two steps in the case of **60b**).



Scheme 16 *Synthesis of compound **60a** and **60b** from the ethyl carbamate **81**.*[*]

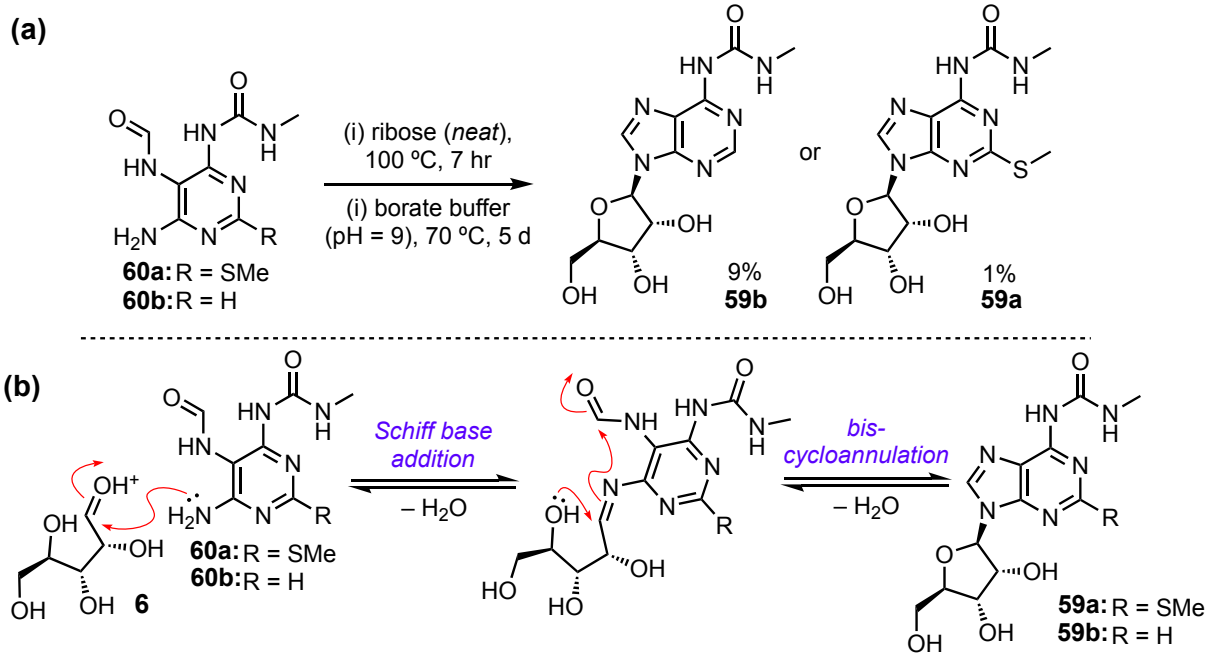
The synthetic reference material for the adenosine derivative **59b** was also prepared according to a procedure found in the chemical literature (**Scheme 17**).[107] First, the commercially available adenosine **33** was treated with acetic anhydride in the presence of 4-dimethylaminopyridine (DMAP) and triethylamine to give the acetate **83** in 85% yield, which was subsequently carbamoylated using **70** to give **84**, also in high yield. Substitution of the phenoxy group using methylamine **85** provided **86**, which was then deprotected using a 7-molar solution of ammonia in methanol to give the desired compound **59b** in 83% yield. With compounds **60a** and **60b** to hand, efforts became focused towards the ribosylation and cyclisation of the exocyclic formyl group under plausible prebiotic conditions (**Scheme 18a**). Thus, the reactions of **60a** and **60b** were set up under conditions equivalent to those previous implemented in our syntheses of purine nucleosides from FaPys.[70,102] Remarkably, the reaction of **60b** gave as the major product, the natural beta-furanoside **59b** in 9% yield, with overall 26% yield for all nucleoside products including alpha- and beta- furanosides and pyranosides. Although the yield for the ms^2-modified compound **59a** was significantly lower (1%) it should be noted that in both cases we were able to detect significant quantities of unreacted FaPy (limiting reagent), which could itself in principle be re-subjected to the same set of reaction conditions in a prebiotic context.

[*] Experimental details for the prebiotically plausible access of compounds **60a** and **60b** are outlined in the doctoral thesis of C. Schneider.[132]



Scheme 17 *Synthesis of the synthetic reference compound* ***59b****.*

The mechanism of these reactions presumably proceeds *via* the same type of process that occurs for the canonical purine nucleosides, involving firstly, Schiff base condensation of the available FaPy amino moiety with ribose **6**, and subsequent dehydrative cyclisation (either 5-exo- or 6-exo-trig) to afford the alpha- and beta- furanosyl and pyranosyl isomers (**Scheme 18b**).[70]

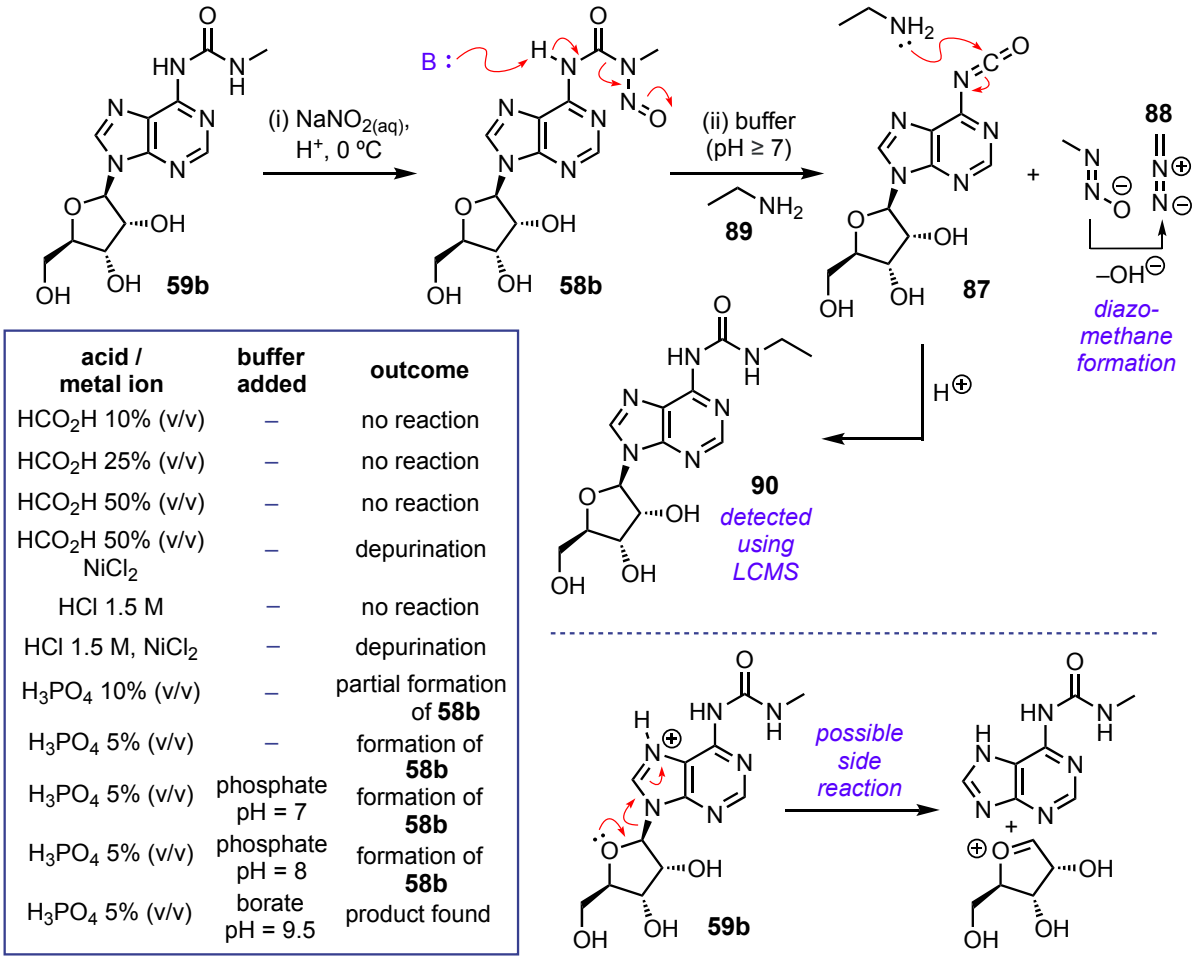


Scheme 18 (a) *Synthesis of the methylurea nucleosides under plausible prebiotic conditions.*[†] **(b)** *A probable mechanism accounting for the cycloannulation of* ***60a/b*** *in the presence of ribose* ***6****.*

[†] These two particular reactions were set up and monitored by C. Schneider with HPLC-MS analysis using material whose preparation is described here. Quantification and supporting data including preparation of **59a** can be found in her doctoral thesis.[132]

In lieu of the successful transformation of **60a/b** to give **59a/b**, we next set our interests towards testing whether **59a/b** could be converted into aminoacylated derivatives such as g^6A **57** or t^6A **44** under prebiotic conditions analogous to those presented in our other 2018 study (**Scheme 19**).[101] This would involve, as the initial step, nitrosation of the 3-methylureido nitrogen with sodium nitrite, which is known to generate the electrophilic nitrosonium cation upon exposure to dilute acid.[108] The resultant *N*-nitroso compound **58b** could then collapse to give a reactive isocyanate **87**, as well as the elements of diazomethane **88**. Subsequent nucleophilic attack from an amine (here ethylamine **89**) should then afford the desired species **90**.

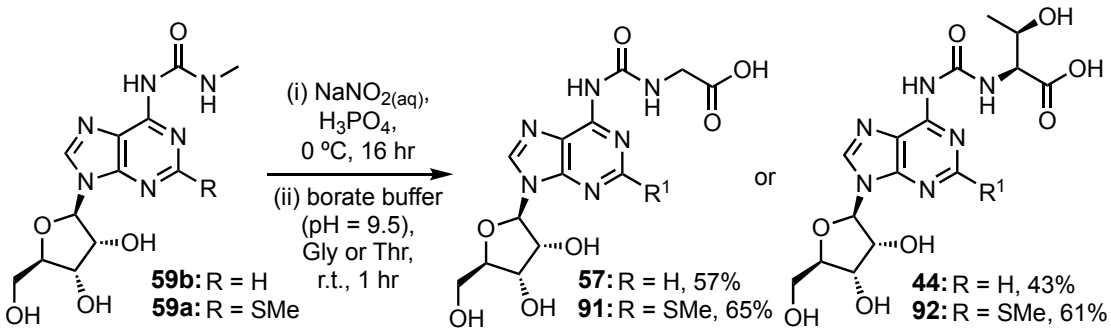
The outcomes of efforts to implement the reaction sequence just described are detailed below. In our initial trials, aqueous solutions of formic acid were chosen given the prebiotic plausibility of formate[109] and due to our former success in generating *N*-nitroso compound from amino acids under these conditions.[101] In our hands, however, formic acid was not successful in generating the desired nitroso compound **58b**, even in very high aqueous concentrations (50% v/v). Furthermore, addition of a Lewis acid ($NiCl_2$) only resulted in decomposition (likely depurination), as noted by LCMS, and TLC analyses. Similar experiments employing HCl solutions rather than formic acid gave comparable results. Gratifyingly, upon treatment of **59b** with a concentrated solution (10% v/v) of phosphoric acid, the molecular ion associated with the protonated form of **58b** (m/z = 354) became strongly visible *via* ESI mass spectrometry. We also observed disappearance of the starting material, as monitored by TLC. Upon the addition of ethylamine (delivered *via* a 70% w/w solution in H_2O), however, the desired urea **90** was not observed. We suspected that the protonated form of **58b** might be stabilised, thus preventing decay and release of 1-hydroxy-2-methyldiazene. Indeed, upon the addition of a large excess of borate buffer (pH = 9.5), the desired product **90** became observable *via* ESI MS, whereas it did not when phosphate buffer of lower pH was instead added. We also detected probable formation of adenosine, suggesting that the intermediate isocyanate **87** also undergoes hydrolysis due to H_2O.



acid / metal ion	buffer added	outcome
HCO_2H 10% (v/v)	–	no reaction
HCO_2H 25% (v/v)	–	no reaction
HCO_2H 50% (v/v)	–	no reaction
HCO_2H 50% (v/v) $NiCl_2$	–	depurination
HCl 1.5 M	–	no reaction
HCl 1.5 M, $NiCl_2$	–	depurination
H_3PO_4 10% (v/v)	–	partial formation of **58b**
H_3PO_4 5% (v/v)	–	formation of **58b**
H_3PO_4 5% (v/v)	phosphate pH = 7	formation of **58b**
H_3PO_4 5% (v/v)	phosphate pH = 8	formation of **58b**
H_3PO_4 5% (v/v)	borate pH = 9.5	product found

Scheme 19 *Preliminary investigations into the reactivity of **59b** in the presence of sodium nitrite and an amine nucleophile. Reactions were screened in order to identify conditions able to affect the transformation of **59b** to **90**.*

Finally, having identified conditions allowing for the transformation of **59b** to **90**, we set about affecting the same towards the synthesis of g^6A **57**, t^6A **44**, and their 2-thiomethylated derivatives **91** and **92**. To our satisfaction, treatment of either **59a** or **59b** with sodium nitrite and 5% phosphoric acid for 16hr, followed by the addition of either glycine or threonine in a large excess of 30 mM borate buffer (pH = 9.5) resulted in the successful formation of all four expected nucleoside products in high yields (43-65%). The main discernible by-products were once again, the de-carbamoylated nucleosides resulting from nucleophilic attack of the *in situ*-generated isocyanate by H_2O. Naturally, such products would have themselves been valuable in early Earth scenarios given that they could be recycled and have contributed towards the synthesis of canonical RNA.



Scheme 20 *Synthesis of the aminoacylated nucleosides g^{6}A **57**, t^{6}A **44**, and their 2-thiomethylated derivatives **91** and **92**, under prebiotically plausible conditions.*[‡]

In conclusion, we present here the first high-yielding prebiotically plausible synthesis of the non-canonical RNA nucleosides, g^{6}A **57**, t^{6}A **44**, and their 2-thiomethylated derivatives **91** and **92**. Compared with our previous convergent approach, which involved the use of methyl(nitroso)carbamoyl amino acid building blocks,[101] our updated method significantly improves upon existing regioselectivity issues. Future studies will be directed towards identifying strategies that allow for a continuous, one pot synthesis involving more plausible prebiotic conditions, as was already demonstrated with other non-canonical purine nucleosides.[102]

[‡] These particular reactions were set up and monitored by C. Schneider with HPLC-MS analysis, according to the reaction conditions identified in Figure 10. Quantification and supporting data can be found in her doctoral thesis.[132]

3.2 INVESTIGATIONS INTO THE aa^{6}A NUCLEOSIDE-MEDIATED ELONGATION OF POLYPEPTIDES

3.2.1 Prologue

The RNA world hypothesis posits that life first emerged from self-replicating nucleic acids.[1] Although the properties of RNA allow for the conceptual plausibility of this theory, opinions differ on whether RNA alone constituted the first self-replicating genetic system.[110] As such, various suggestions for simpler "pre-RNAs" have been made, including peptide nucleic acid (PNA), threose nucleic acid (TNA), and glycol nucleic acid (GNA).[111] Regardless of the nature of the first self-replicators, it remains self-evident that chemical evolution must have eventually led to the formation of peptides with reproducible sequence specificity. In today's biology, such a process is carried out *via* the charging of codon-specific tRNAs with amino acyl esters at the 3'-position of terminally-positioned nucleotides. Although stable within modern cells, tRNA-esters would be easily degraded under even mild prebiotic conditions.[80] We therefore set about exploring the abstract possibility that early, or pre-ribozymes facilitated peptide elongation *via* an entirely different mechanism.

Central to our investigations was the finding that the nucleoside *N*^{6}-threonylcarbamoyladenosine (t^{6}A) **44**, and its derivatives are universally conserved and found at position 37, which is 3'-adjacent to the anticodon loop of tRNAs responsible for ANN codons (**Figure 10a**).[112] Unlike tRNA-esters, which bare amino acids at the hydrolytically-fragile 3'-hydroxyl group, the amino acids present in nucleosides such as **44** are instead bound to stable *N*^{6}-functionalised ureas, thus increasing their plausibility within a prebiotic context. Also intriguingly, these modifications are believed to have been present within the tRNA of LUCA.[113] In 2013, an additional cyclic isoform (**93**) of t^{6}A was ostensibly discovered by Suzuki and co-workers in bacteria, fungi, protists, and plants, that was presumably formed *via* dehydration of t^{6}A itself **44**.[114] 5(4H)-oxazolones, which are formed through the activation of acylated α-amino acids or from the C termini of peptides, have themselves been implicated with early translation apparatus, further adding to our interest in the discovery of **93**.[77] It was, however, curious that **93** appeared to be stable enough for isolation from tRNA. This was highly surprising owing to the reactive isoimide functionality embedded within the 2-iminooxazolidin-5-one ring of **93**.[115] This confusion appeared to be resolved when in 2016 Sochacka reported crystallographic evidence for the correct structure of cyclic t^{6}A, which instead contained a hydantoin ring **94**.[116] Of interest to us, was the observation that **93** (in fact **94**) would readily hydrolyse to give t^{6}A **44** during handling of tRNA, and in certain instances, would even react with nucleophilic amines to give amides, making it a prime candidate for prebiotically plausible peptide synthesis.

A hypothetical mechanism involving modifications such as **44**, from which peptides could plausibly be generated was thus proposed (**Figure 10b**). This mechanism involves, as the initial step, cyclisation of **95** to give a hydantoin intermediate **96**, which would then be coupled *via* nucleophilic ring-opening to an appropriate peptide or amino acid **97**, giving structure **98**. Appropriate cleavage of the peptide residue **99**, perhaps tethered to another templated-RNA strand, would reset the system for further peptide synthesis upon charging of a fresh amino acid. Although purely theoretical, we believed that such a system warranted investigation, particularly since the only prebiotic translation mechanisms to have so far been explored are (to our knowledge) those involving ester or phosphate ester linkages to RNA.[77]

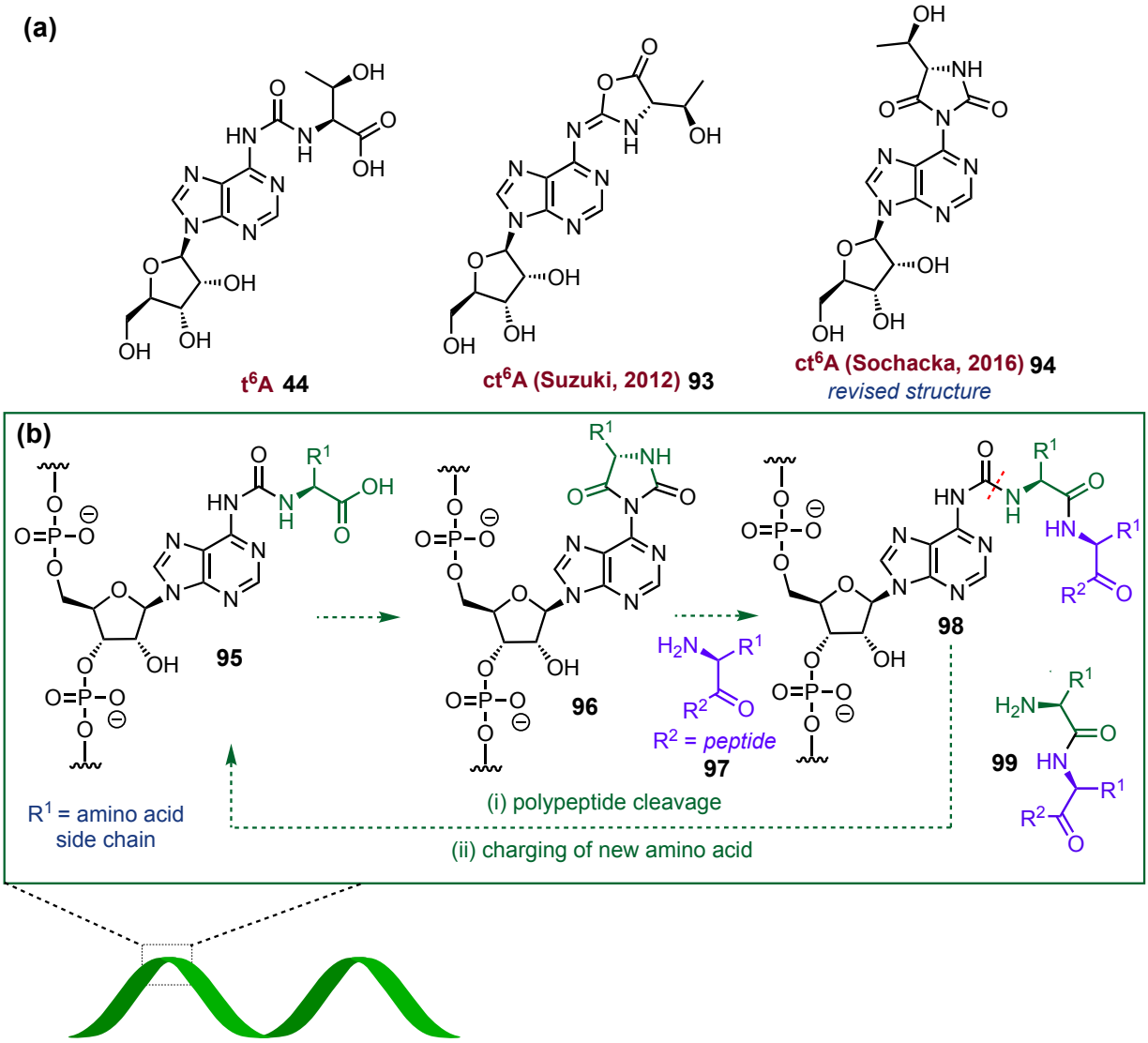
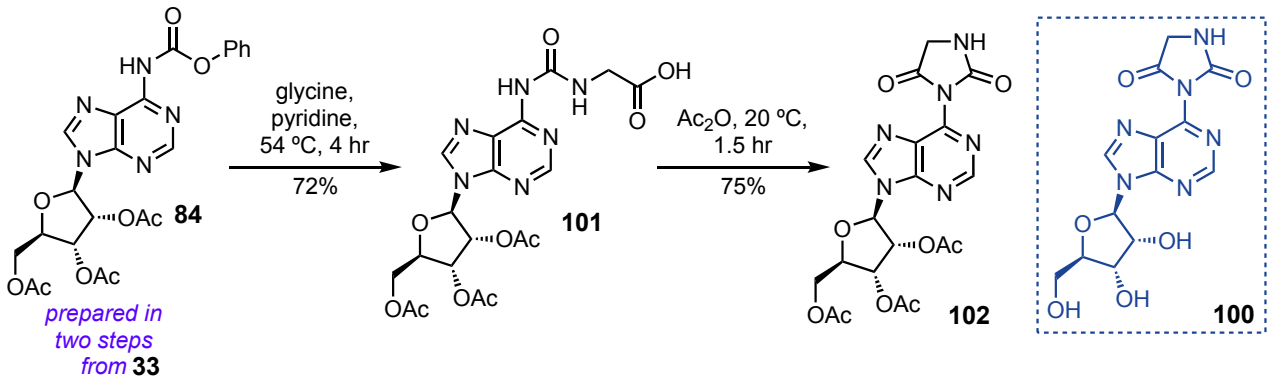


Figure 10 (a) *The structures of t⁶A **44**, ct⁶A **93** as assigned by Suzuki and co-workers, and the revised structure of ct⁶A **94** from Sochacka and co-workers.*[112,114,116] **(b)** *Speculative mechanism by which cyclic intermediates such as **95** could have plausibly facilitated the elongation of polypeptides on the early Earth.*

We thus set out collecting preliminary data to indicate whether the chemical properties of RNA modifications such as **95** could in principal facilitate a rudimentary translation system. To achieve this goal, we first aimed to synthesise the nucleosides g^{6}A **57** and its corresponding cyclic derivative in order investigate their ability to react with amino acids and generate polypeptides. This would also require demonstrating that the hydantoin-containing nucleosides are accessible under prebiotically plausible conditions. Finally, we aimed to incorporate these modifications into RNA strands using solid phase synthesis techniques and set about testing the capacity of such reactions to occur when embedded within an RNA strand.

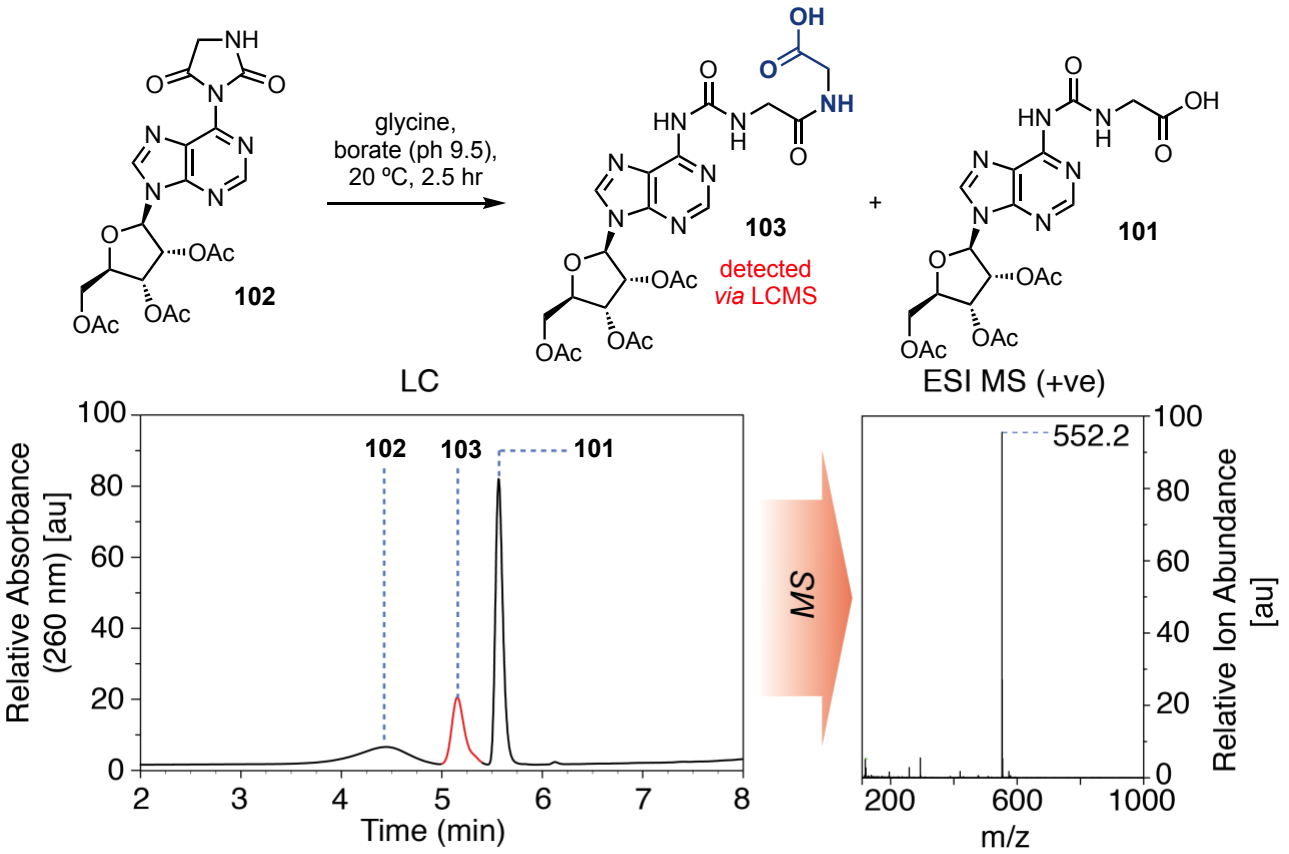
3.2.2 Results and Discussion

Given our principal goals of exploring nucleoside-mediate peptide synthesis, preparation of the cyclic derivative of g^{6}A **100** was initially pursued. An examination of the literature relating to cyclic t^{6}A **94** suggested that one way of achieving this end would be to do so by activating the terminal carboxyl group of a relevant amino acid-conjugated nucleoside using acetic anhydride.[117] Thus, the phenyl carbamate **84**, which was itself prepared from commercially available adenosine **33**, was readily converted into the corresponding amino acyl nucleoside **101** using glycine in pyridine (**Scheme 21**). The ^{1}H and ^{13}C spectroscopic data for this compound were consistent with those derived from the literature.[101] In agreement with earlier reports, treatment of the nucleoside **101** with *neat* acetic anhydride at 0 ºC gave a rapid reaction that afforded the novel nucleoside **102** in 75% yield. The hydantoin structure present in **102** was then confirmed by close comparison of the ^{1}H and ^{13}NMR spectra for the previously-reported threonine derivative.[116,117] Our results here also confirm earlier reports that hydantoin modified-nucleosides of this nature are stable in acidic medium. Compound **102** presumably forms *via* the acylation and activation of the acid **101** to give a mixed anhydride intermediate with the N^6-amide serving as a nucleophile for cyclisation, à la the Dakin–West reaction.[118]



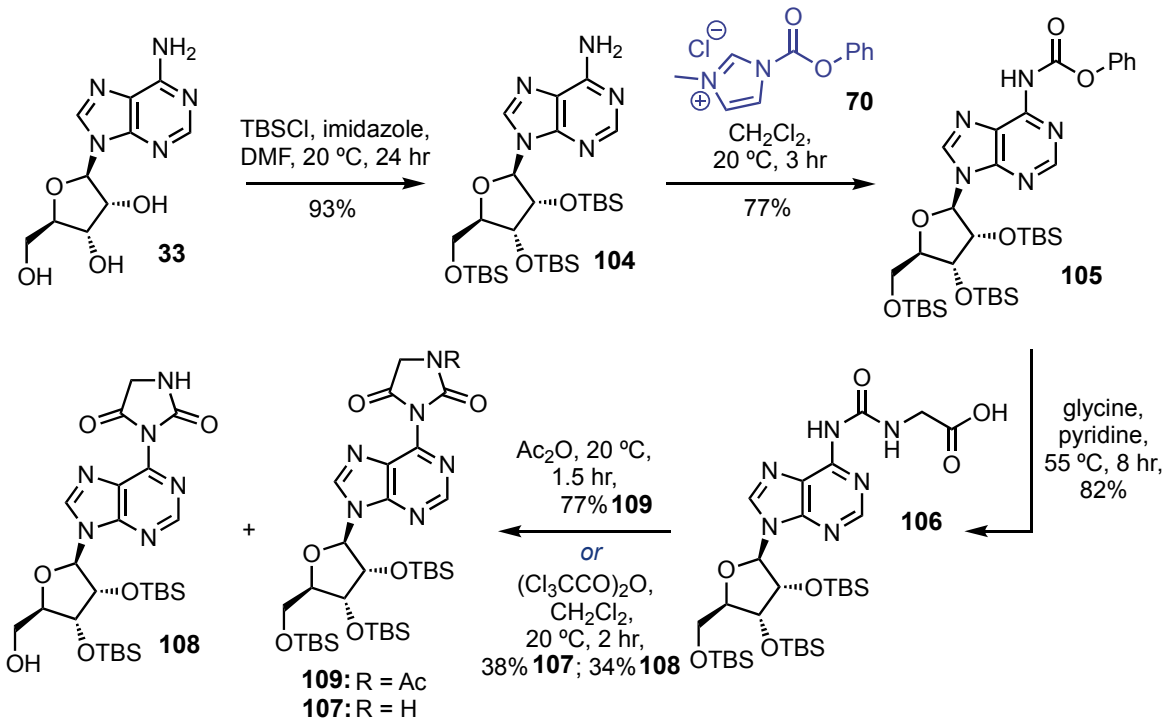
Scheme 21 *Synthesis of the cyclic-g^{6}A nucleoside **102** from the carbamate **84**, whose synthesis is described in the previous section.*

With compound **102** to hand, the nucleoside mediated formation of polypeptides under plausible prebiotic conditions was next pursued (**Scheme 22**). Thus, in a preliminary effort to investigate the capacity of the nucleoside to engage in ring-opening peptide-forming reactions, a solution of **102** in borate buffer (30 mM, pH = 9.5) was treated with an excess of glycine, and the reaction was monitored with ESI MS-coupled liquid chromatography. To our gratification, a new major species was formed in large quantities, whose molecular ion corresponded exactly to the dipeptide **103**.



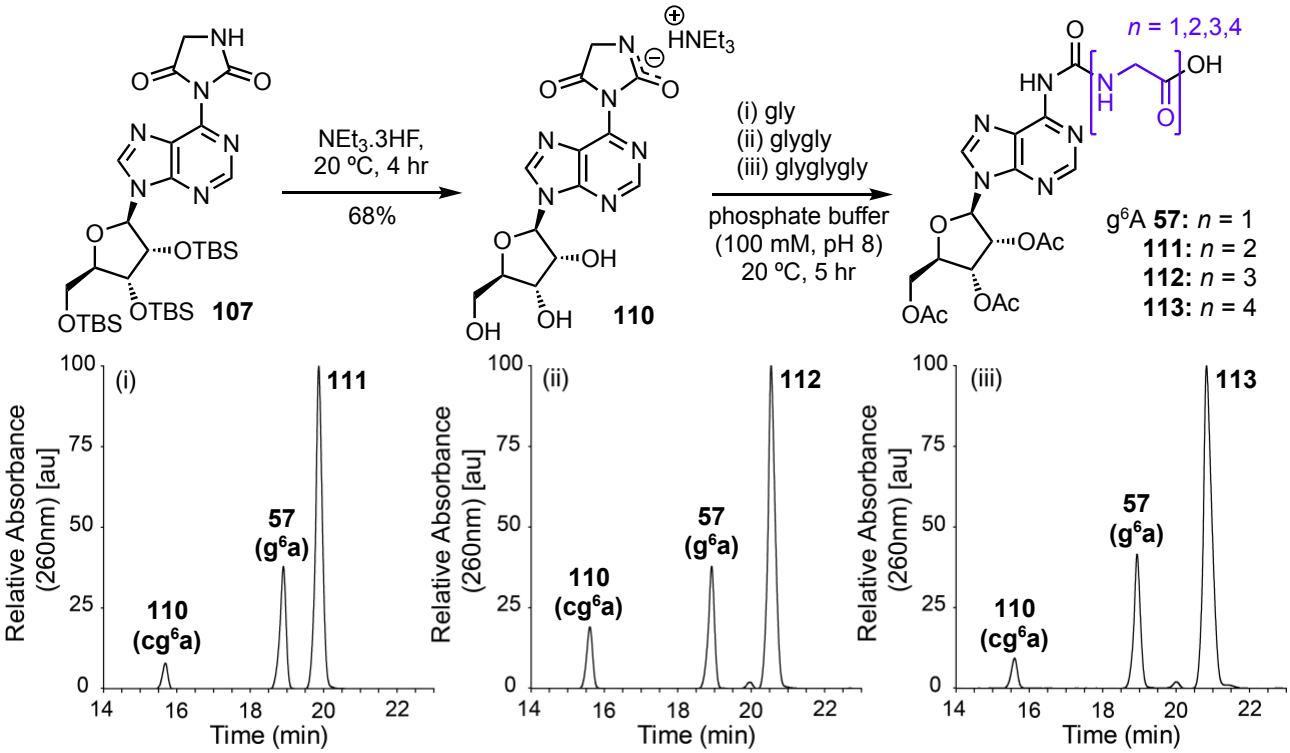
Scheme 22 *Treatment of **102** with glycine under buffered prebiotically plausible conditions likely generates the dipeptide **103** as detected by LCMS analysis, thus warranting further investigation.*

In pursuit of more realistic prebiotic conditions, we next sought to prepare the fully-deprotected derivative of the nucleoside **102** (**Scheme 23**). To this end, the silyl ether **104** was first generated from **33**, which was then subsequently converted into the carbamate **105** using similar conditions to those previously mentioned. The nucleoside **106** was then obtained in high yields upon treatment of **105** with glycine for 8 hours in *neat* pyridine. Finally, **106** was converted into the cyclic derivative **107** using trichloroacetic anhydride. The reaction also afforded **108**, which is presumably formed due to acid-promoted desilylation. When acetic anhydride was used in place of trichloroacetic anhydride, the acetate-protected compound **109** was instead obtained. With useful quantities of **107** to hand, we would next need to convert the former compound into its fully-deprotected form.



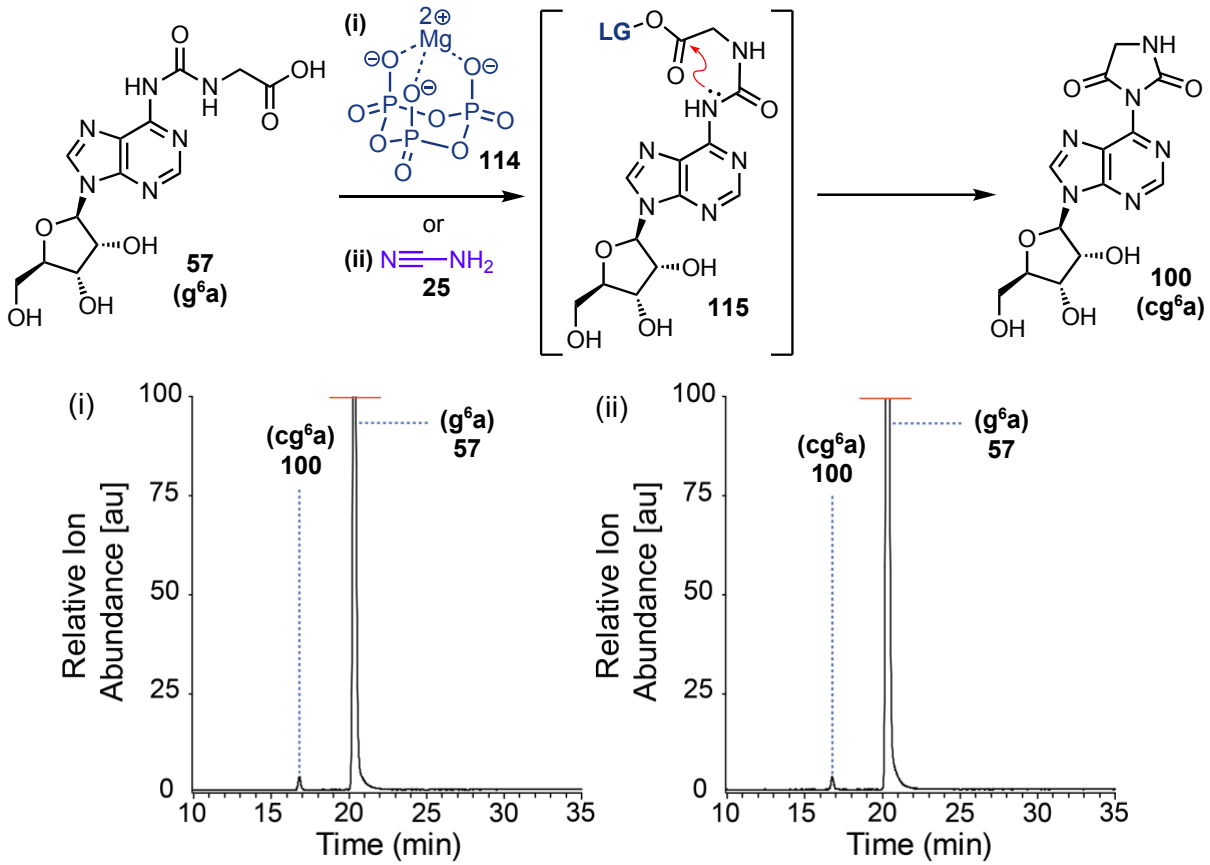
Scheme 23 *Synthesis of the cyclic-g^{6}A nucleosides **107**, **108**, and **109**, from adenosine **33**.*

To such an end, compound **107** was treated with triethylamine trihydrofluoride, which yielded the desired the free nucleoside **110** as a triethylammonium salt (**Scheme 24**). The illustrated ionic structure was proposed due to the disappearance of the hydantoin proton in the ^{1}H NMR spectrum, as well as the observed reactivity of that particular nitrogen in the presence of acetic anhydride (see compound **109**), suggesting a suitably low pK_a value. Following dissolution of **110** in phosphate buffer (pH = 8), either glycine, diglycine, or triglycine were added, and the mixture was shaken for 5 hours. After this time, samples were analysed using HRMS-coupled liquid chromatography. Satisfyingly, in all three cases, the starting material was almost entirely consumed, with the major product of the reaction being the desired peptide adduct. The identity of each nucleoside was assigned using the $[M+H]^+$ ion in the corresponding peak-picked HRMS spectrum. The free-acid **57** was also detected, which could, in a prebiotic context, naturally be recycled to give additional **107**.



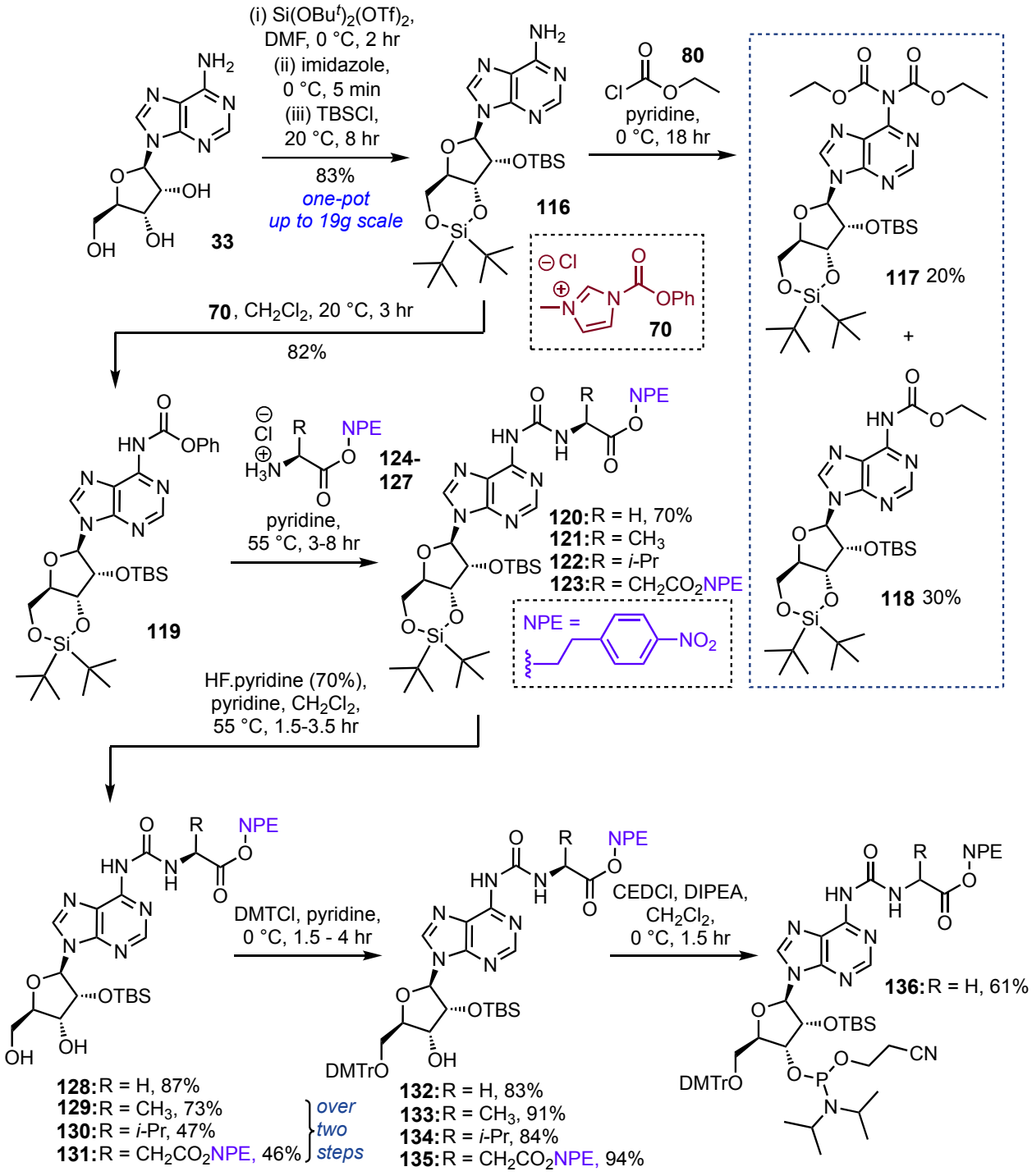
Scheme 24 *Silyl-deprotection of the cyclic-g^{6}A nucleoside **107**, and reactivity of **110** in the presence of glycine and short peptides. Reaction progress was monitored using HPLC-MS, and UV-traces (260 nm) are shown for the reactions (i), (ii), and (iii) after 5 hr.*

Another important consideration in our studies was whether or not compounds such as **100** could themselves plausibly have arisen on the prebiotic earth. The reactions of compound **57** with various prebiotically plausible acid-activators was thus investigated in buffered aqueous solutions. In our hands, both cyanamide **25** and cyclic trimetaphosphate **114**/Mg^{2+} succeeded in affecting the transformation of **57** to give **100**, although only in very minimal yields (<1%). Formation of **100** was established by co-injection of the synthetically produced sample of **110** (the triethylamine salt of **100**), and by m/z confirmation (Supporting Information). The reaction of **57** with cyanamide likely proceeds *via* the methanediimine tautomer, as has been previously employed in synthesis of the hydantoin cyclic form of t^{6}A **(95)** using polymer bound carbodiimide (EDC-P).[82] In both cases, the reaction may involve activation of the carboxylic acid moiety through an intermediate such as **115**, and subsequent substitution of the activated group. The low yields in this case could perhaps be attributed to the competing hydrolysis of **100** which gives **57**.



Scheme 25 *Reactions of g^{6}A **57** with cyclic trimetaphosphate **114** and cyanamide **25** under plausible prebiotic conditions, and analysis of the reaction mixture by HPLC-MS, showing partial formation of the cyclic-g^{6}A nucleoside **100**.*

Having established preliminary prebiotic conditions for some degree of the formation of **100**, as well as conditions that allow its reactions with amino acids and polypetides, we next set out to test similar transformations were capable of taking place directly within RNA strands. Accordingly, we began parallel syntheses of appropriate nucleoside phosphoramidite building blocks (**Scheme 26**). In addition to the nucleoside g^{6}A **57**, phosphoramidites corresponding to the modifications containing the amino acids alanine, proline, and aspartic acid were also pursued. These particular nucleosides were chosen on the basis of the prebiotic plausibility, and since alanine, proline and aspartic acid are all thought to have been among the earliest amino acids to emerge on the early earth.[119] The silyl ether **116** was thus initially prepared *via* a two-step, one-pot treatment of adenosine **33** with di-tert-butylsilyl ditriflate, followed by TBSCl, which gave the former compound **116** in 83% yield.

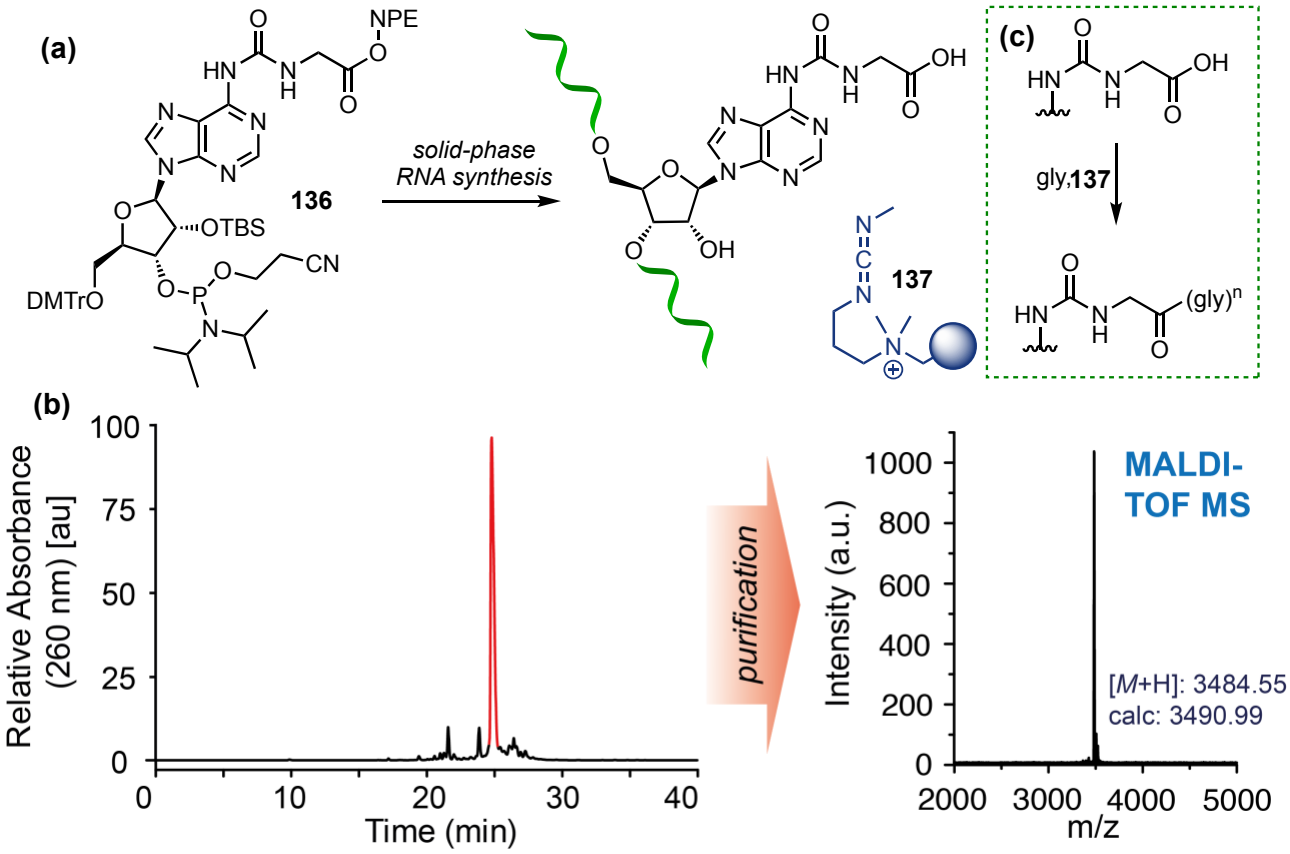


Scheme 26 *Synthesis of the phosphoramidite building block **136** and DMTr-protected compounds **133**, **134**, and **135**.*[§]

Attempts to introduce the urea functionality were initially hampered, as when ethyl chloroformate **80** was used, the reaction proved not be selective, and instead gave the unwanted bis-adduct **117** in 20% yield as well as the desired compound **118** (30% yield). Gratifyingly, treatment of **116** with the

[§] *Phosphoramidite **136** and DMTr-compounds **133** and **134** were prepared according to a slightly modified procedure compared with those that appear in the final published work.*[130]

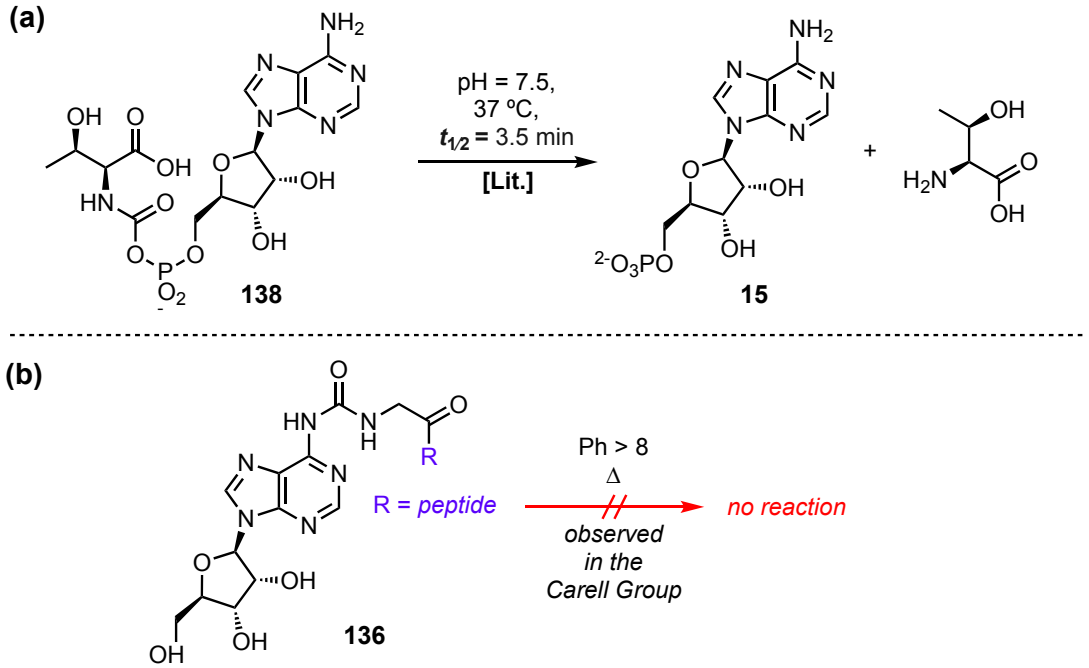
imidazolium carbamate **68** gave a rapid reaction, which provided compound **119** in 82% yield. To convert **119** in the each of the relevant urea-containing compounds **120-123**, **119** was treated with the NPE-protected amino acids **124-127**. To aid with purification, the DTBS-protecting group was then also removed using hydrogen fluoride-pyridine, which gave the very pure compounds **128-131** in up to 73% yield over two steps, and avoided the need for tedious, repeated chromatographic purification of the mixture obtained from compound **119**. DMTr protection finally gave compounds **132-135** in up to 94% yield. Having now established synthesis of the phosphoramidite precursors, compound **132** was finally converted into phosphoramidite **136** using CEDCl under standard conditions. With compound **136** in hand, incorporation of g^6A **57** into RNA was next pursued (**Scheme 27**). In keeping with expectations, **136** was successfully introduced to several RNA sequences in excellent purities, using a single 20-minute coupling, and combined with TBS-protected canonical amidites.



Scheme 27 (a) *Successful incorporation of the novel phosphoramidite building block **136** into an example RNA oligonucleotide (sequence: 5'-g⁶AAUCGCUUUUU-3'), including* **(b)** *raw-HPLC chromatogram and MALDI-TOF mass spectrum of the purified oligo, and* **(c)** *preliminary results indicated successful g⁶A-mediated glycine-polymerisation in oligos.* [**] *The preparation and characterisation of oligonucleotides is described in the Experimental section.*

[**] Investigations into EDC-mediated polymerisation of amino acids were performed by colleagues in the Carell Lab.

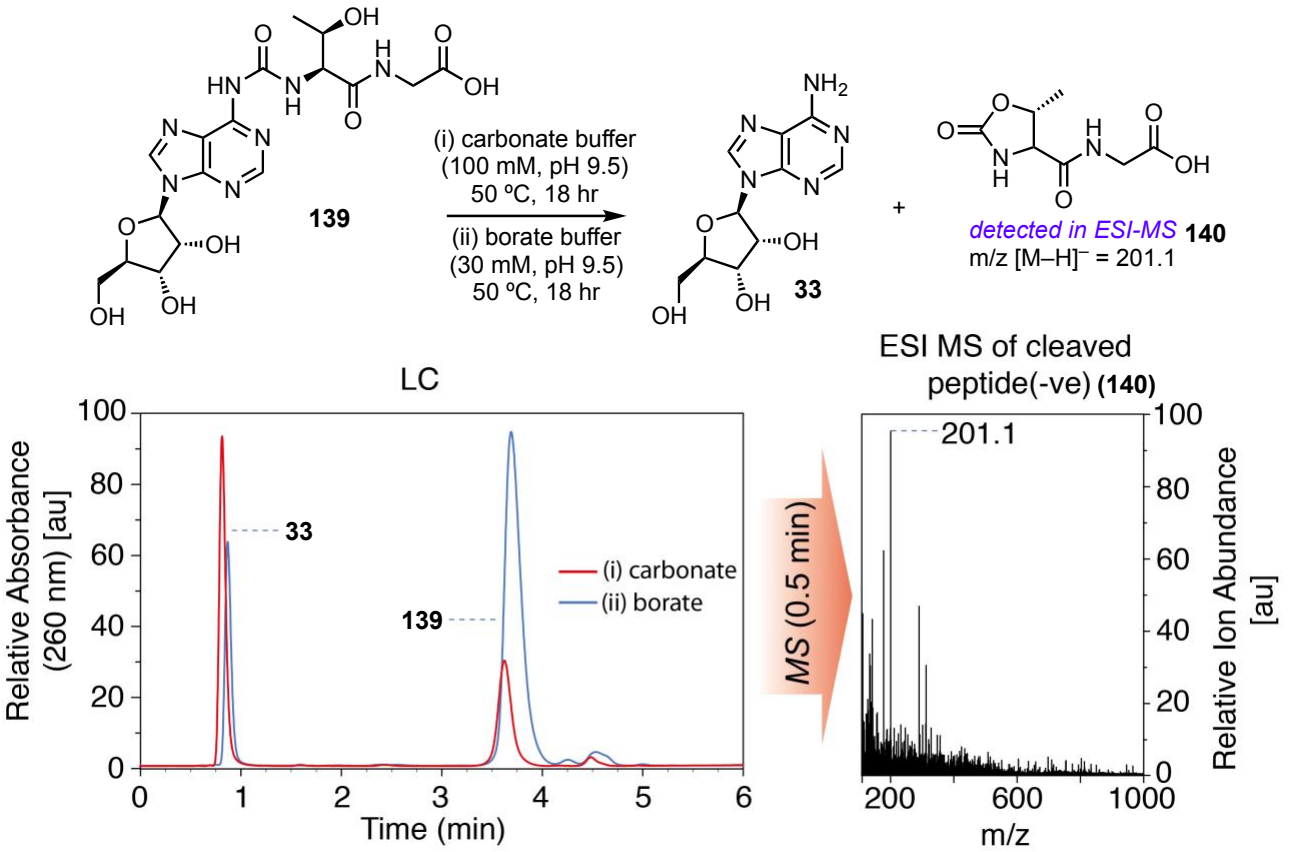
The NPE-protecting group was also able to be cleanly removed, upon simple treatment with solutions of DBU in THF. Using the method developed and presented here, parallel investigations were subsequently carried out in the Carell group, demonstrating the successful 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (**137**)-mediated formation of RNA-templated polypeptides. Further results in this area will be presented in due course. Finally, we were interested in pursuing reactions which resulted in cleavage of the synthesised peptides from an RNA-backbone. Promisingly, the literature precedent from Lauhon and co-workers indicated that threonylcarbamoyl-AMP **138** would readily detach its amino acid under mildly basic aqueous conditions.[120] In contrast however, trial reactions conducted by the Carell group seemed to indicate that related nucleosides (**136**) containing N^6-carbamoyl peptides whose first amino acid was glycine were very stable under alkaline or thermally-promoted conditions.



Scheme 28 (a) *Non-enzymatic degradation and amino acid cleavage of threonylcarbamoyl-AMP **138**, as demonstrated by Lauhon in 2012.*[120] **(b)** *Investigations carried out by colleagues in the Carell group suggested that g^6A-substituted polypeptides are resistant to alkaline and thermal hydrolysis.*

Although this result was unsurprising, given the dramatic difference in stability between a carbamoyl phosphate and a urea moiety, we came to suspect that hydroxyl group-containing side chain of threonine might partly contribute to the observed reactivity of threonylcarbamoyl-AMP **138**. To properly explore this notion, compound **139** was thus prepared from its methyl-ester, according to conditions similar to those already presented in this section (please see the **Experimental Section**). Indeed, when compound **139** was heated in aqueous solution containing a mild base, we noted

disappearance the starting material, accompanied by new formation of adenosine **33**, indicating cleavage of the peptide (**Scheme 29**). Although we did not detect formation of the free-dipeptide, we did observe formation of an ion whose m/z ratio would match the depicted compound **140**. In addition to being supported by the literature precedent,[120] this result would strongly suggest that the threonine side-chain does indeed facilitate the mild base-mediate cleavage of amino acids and peptides. This warrants further investigation, and so derivatives containing serine-, or other hydroxyl group-baring side chains will be studied in the future. Studies towards this end are already underway, and a complete synthesis of the serine-dipeptide analogue of **139** (compound **141**) can be found in the Experimental section of this thesis.



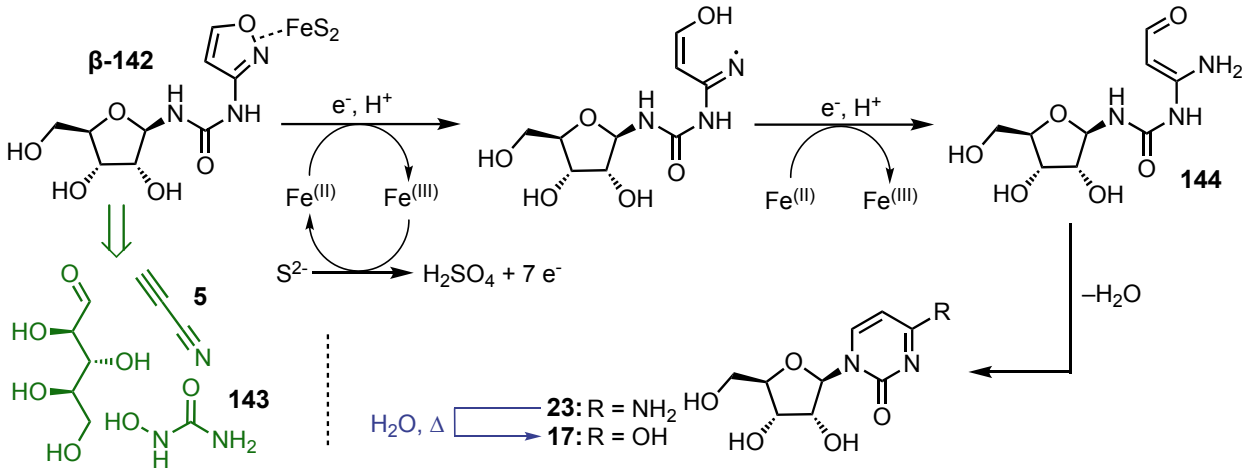
Scheme 29 *In contrast to previous observations, hydrolytic cleavage of **139** is observed under mild conditions, as indicated by LCMS analysis.*

In conclusion, we present here the first ever synthesis of the g^{6}A phosphoramidite **136** and its successful incorporation into RNA. In addition, we have developed syntheses for DMTr precursors to three other, related phosphoramidites, which we fully anticipate can also be introduced to RNA *via* the solid phase method. These resources will no doubt be of importance in future studies involving amino-acid modified nucleosides. We have also presented preliminary data showing the capacity for g^{6}A **136** to cyclise, and thus generate polypeptides under conditions compatible with the early Earth.

3.3 SYNTHESIS TOWARDS A 3-AMINOISOXAZOLE-BASED PHOSPHORAMIDITE

3.3.1 Prologue

RNA has long been thought to have heralded the first life on earth due to its ubiquity and to its unique capacity to both store information and catalyse its own replication.[24] In pursuit of evidence for the RNA world hypothesis, chemists have reported a number of plausible routes that provide both the canonical and non-canonical RNA nucleosides, each with their own distinct advantages.[53] In 2019, we rose to the challenge of generating all four Watson-Crick RNA nucleosides under a unified prebiotically plausible model - a chemical feat that had not previously been accomplished.[121] Through this endeavour, we discovered a novel pathway that provided the pyrimidine nucleosides in high yields (**Scheme 30**). The key feature associated with our new approach involved Fe-mediated reductive cleavage of the isoxazole ring present in the nucleoside **β-142**, followed by cyclisation to afford **23** and **17**. The heterocyclic portion of **β-142** was generated through a highly regioselective (3+2) cycloannulation involving 1-hydroxyurea **143** and cyanoacetylene **5**.



Scheme 30 *Prebiotically plausible synthesis of pyrimidine nucleosides, proceeding via the reductive ring-opening of 3-aminoisoxazole nucleoside* ***β-142***.[121]

Although this novel pathway appeared to involve some degree of regio- and stereochemical control (especially when reactions were carried out in the presence of borate-containing minerals), we pondered whether this control could be enhanced *via* the involvement of catalytic RNA. Specifically, we anticipated that if the nucleoside precursor **β-142** were inserted into an RNA strand, then stereochemical information could in principal be transferred from an adjacently-paired guanine nucleoside present in a complimentary RNA strand (**Figure 11**). This would involve pairing of guanine to the intermediate **144** formed upon exposure to reductive conditions. Since the α-anomer **145** should not be able to pair with its complimentary β-nucleotide, we anticipate that an enhanced stereoinductive effect would push the equilibrium towards **144**, thus favouring cyclisation to afford

the β-pyrimidine nucleotide **146**. Furthermore, due to the presence of the 5'-phosphate moieties found in RNA, unwanted pyranosyl isomers would effectively be removed as possible reaction products.

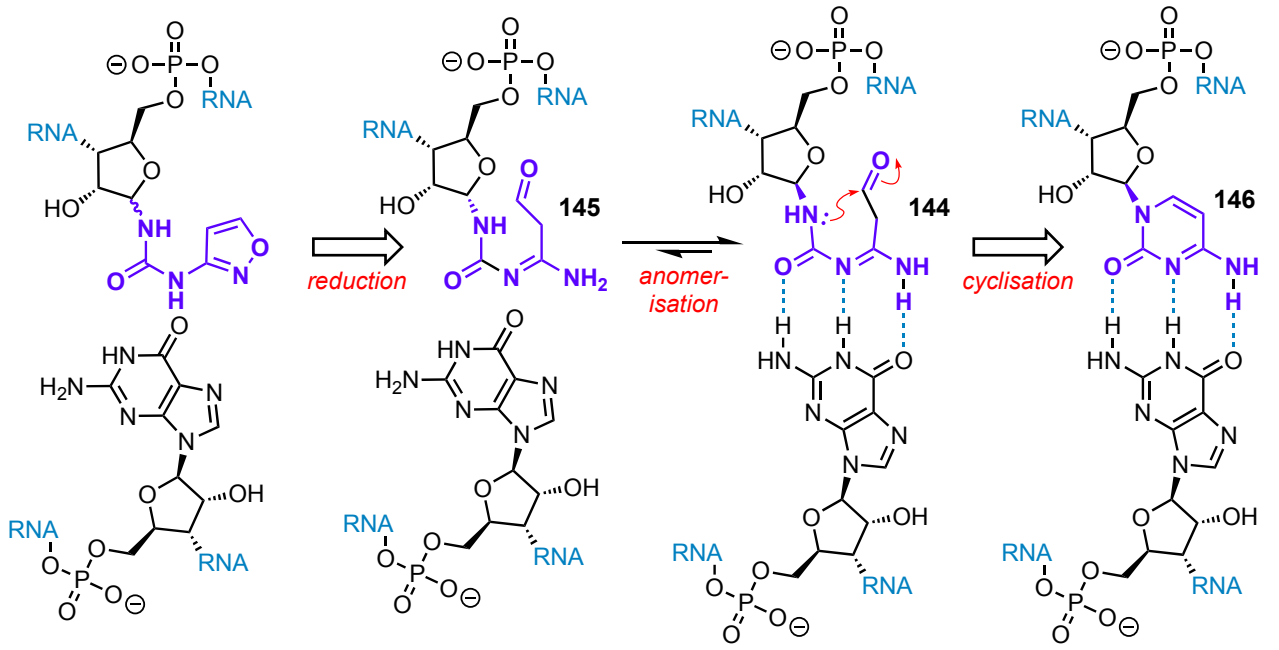
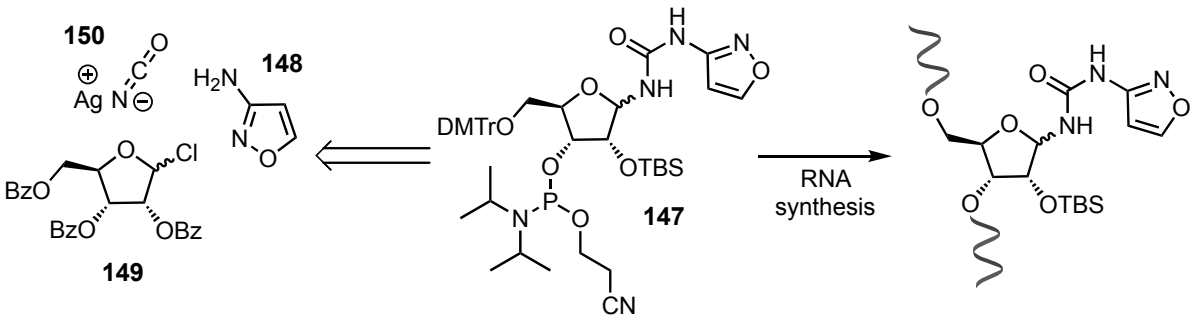


Figure 11 *The proposed stereoinductive effect of a complimentary RNA strand, in the formation of the nucleotide **146** from **144**.*

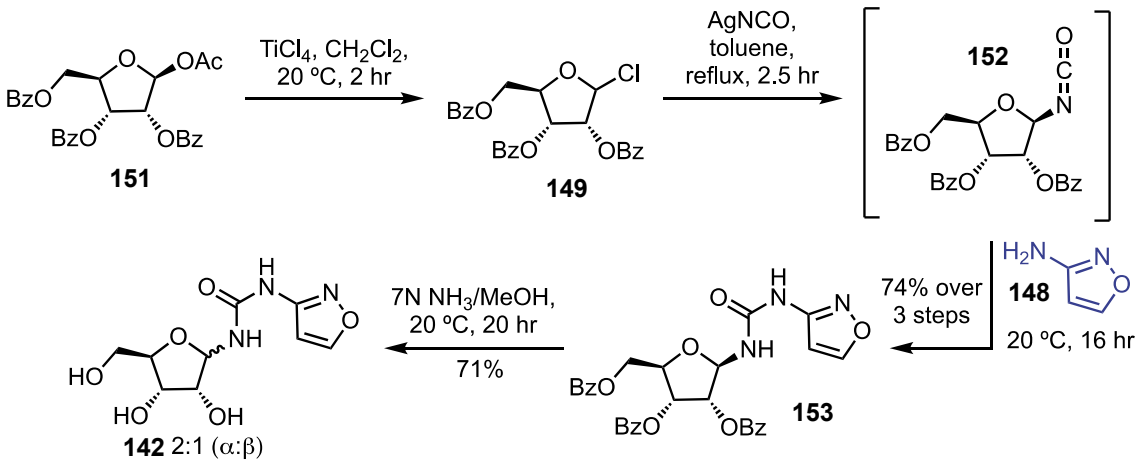
In order to explore this proposed chemical model, the nucleoside precursor **β-142** would first need to be incorporated into RNA. To this end, we aimed to develop the synthesis of a nucleoside phosphoramidite **147** with which to generate custom RNA strands containing **β-142**, and thus to study their chemical properties (**Scheme 31**). This could in principal be accomplished with either the β- *or* α- diastereomers of **β-142**, although we aimed to eventually prepare both. Our initial strategy would involve synthesis of the nucleoside **β-142** from the elements of 3-aminoisoxazole **148**, riboside **149**, and silver isocyanate **150**, and subsequent 2'/5'-protection and phosphitylation to afford **147**. Upon preparation of precursor-RNA, we would subsequently investigate whether **147** could be converted selectively into β-cytidine within an RNA strand.



Scheme 31 *Proposed synthesis of the phosphoramidite building block **147** from the elements of **148**, **149** and **150**, and its incorporation into RNA.*

3.3.2 Results and Discussion

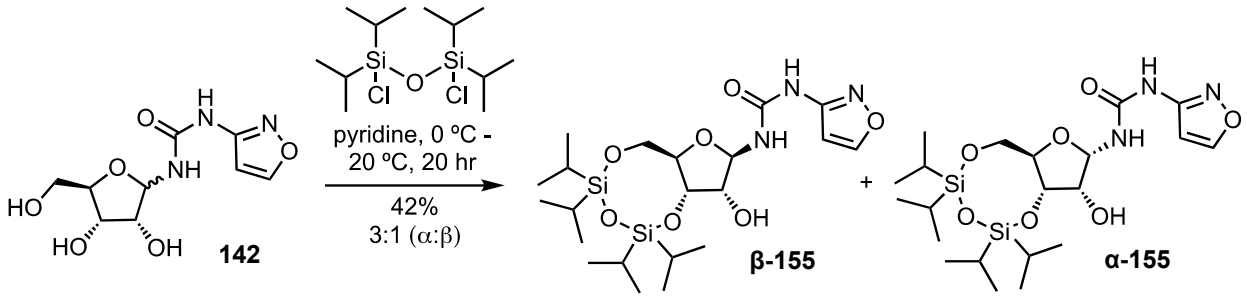
In our initial strategy towards synthesising the nucleoside phosphoramidite **147**, we aimed to take advantage of the chemistry already established in our prebiotic synthesis of pyrimidine nucleosides.[121] This would firstly involve preparation of the nucleoside precursor **142** from the commercially available starting material 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose **151**. Accordingly, synthesis of **142** was pursued (**Scheme 32**). In the event, **151** was treated with $TiCl_4$ at room temperature, which gave the crude chloride **149** as a mixture containing the β- and α-isomers, in accordance with the literature.[122] The progress of the reaction was judged by LCMS and TLC analyses. The crude mixture was subsequently treated with silver isocyanate in refluxing toluene, followed by the addition of 3-aminoisoxazole **148** to quench the *in-situ* generated intermediate **152**, thus giving **153** in 74% yield over three steps. The depicted β-stereochemistry of **153** was assigned based upon mechanistic considerations, namely, formation of β-intermediate **152** possessing the *R* configuration at the anomeric carbon as a result of neighbouring group participation from the 2'-acyl moiety.[123] Compound **153** was finally treated with a 7-molar solution of ammonia in methanol to provide the fully deprotected nucleoside **142** in 71% yield. The ^1H and ^{13}C NMR spectroscopic data derived from this anomeric mixture matched those that we had already reported.[121]



Scheme 32 *Synthesis of the 3-aminoisoxazole nucleoside precursor **142** from 1-O-acetyl-2,3,5-tri-O-benzoylribose **151**.*

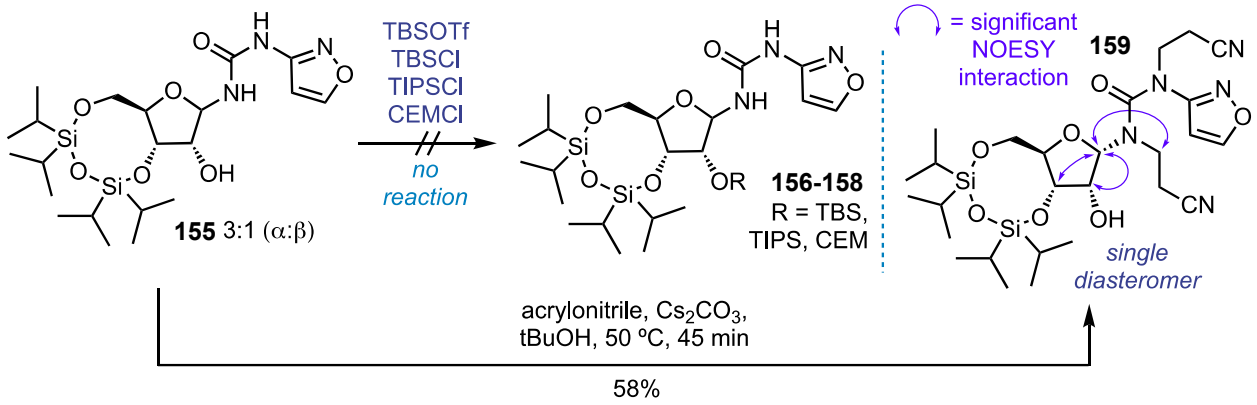
In order to introduce the necessary phosphoramidite-functionality to the 3'-hydroxyl group of the ribose moiety, we began with 3′,5′-silyl protection of **142** (**Scheme 33**). In contrast to previous efforts involving similarly substituted sugars,[124] attempts to protect **142** employing *t*-$Bu_2Si(OTf)_2$ in DMF proved unsuccessful, and gave only starting material, including when elevated temperatures or when

a large excess of reagent were involved. Gratifyingly, treatment of **142** with the Markiewicz reagent[125] **154** instead afforded a 3:1 mixture of the silyl ethers **α-155** and **β-155**. Separation of the two isomers was possible, but required repeated chromatography (silica gel, CH_2Cl_2:EtOAc = 9:1 → 4:1), resulting in substantially lower yields. Subsequent synthetic efforts were thus pursued using the 3:1 mixture with the separated standards **α-155** and **β-155** used as NMR and TLC references.



Scheme 33 *Silyl-protection of the nucleoside **142** to give a chromatographically-separable mixture of the nucleosides **α-155** and **β-155**.*

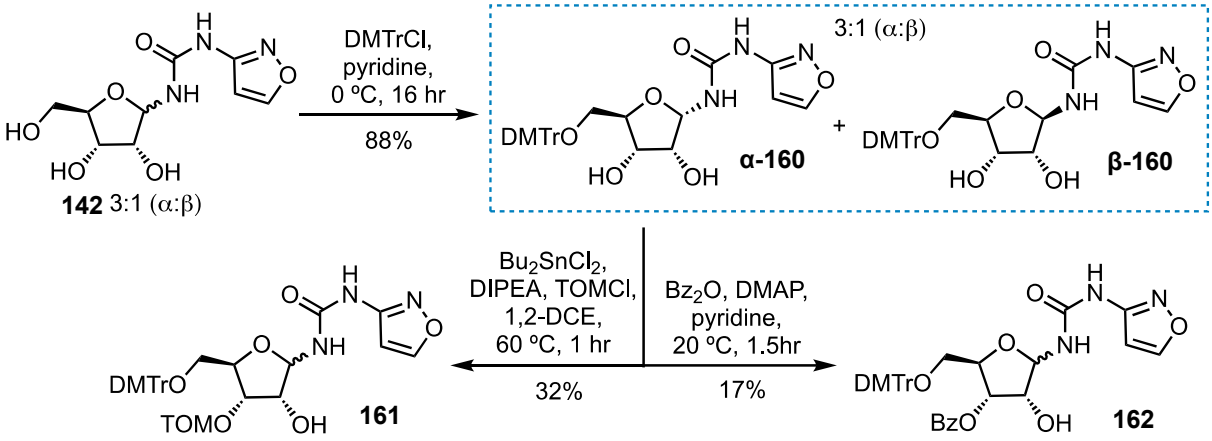
With useful quantities of **155** to hand, the focus of the project next turned towards appending a synthetically compatible 2'- protecting group for solid phase RNA synthesis (**Scheme 34**). To our regret, treatment of the isomeric mixture **155** with various silyl chlorides (TBSOTf; TBSCl; TIPSCl) under standard conditions, failed to generate any of the anticipated silyl ethers **156-158** at all. In an effort to further promote 2'-silylation, nucleophilic catalysts (DMAP; pyridine; imidazole) or Lewis acids ($AgNO_3$) were added, but still no product formation was observed. It is possible here that unfavourable steric factors contributed to higher energetic barriers for the former process, which might in turn be due to the presence the bulky Markiewicz silyl enol ether residue embedded within **155**. Accordingly, we next sought to investigate reactions using a less-bulky 2'- protecting group. Although 3-(chloromethoxy)propanenitrile (CEMCl) (itself prepared from 3-((methylthio)methoxy)propanenitrile) gave no reaction, treatment of **155** with acrylonitrile in the presence of Cs_2CO_3 at 50 ºC resulted in a rapid reaction yielding several different products (as determined by TLC analysis). After 45 minutes reaction time, these spots converged into a single isolable compound. To our surprise, structural characterisation of the isolated spot revealed the presence of a fully-deprotected 2'-hydroxy moiety. Furthermore, the reaction conditions that were employed appeared to have resulted in 1,3-bis-cyanoethyl-protection of the urea moiety, resulting in formation of compound **159** in 58% yield.



Scheme 34 *Synthesis of compound **159** from **155** and stereochemical interpretation involving 2D NOESY NMR spectral analysis.*

The illustrated structure of compound **159** and the stereochemistry at the anomeric centre, were thus assigned by 2D NOESY NMR and with comparison to the starting-material spectra. The unusual regiospecificity of the former process could perhaps due to the lower pK_a value of the isoxazole-urea-proton (~9) relative to that of the 2'-hydroxyl group (~12, as calculated using the ChemAxon pK_a plugin), although this does not explain the addition of the second equivalent of acrylonitrile. It also remains to be understood why the α-anomer is exclusively obtained from the reaction involving both **α-155** and **β-155**. To gain further insight into this matter, crystallographic evidence supporting this structure may be pursued in future studies.

Given the forgoing difficulties, we set about a more direct approach to synthesising nucleoside phosphoramidites, that would firstly involve DMTr-protection of the 5'-hydroxyl group of ribose (**Scheme 35**). Although not entirely selective, direct 5'-protection with DMTrCl has been successfully applied in a number of cases with moderate yields.[126] To our gratification, upon treatment of **142** with DMTrCl in pyridine, a rapid reaction ensued that gave compounds **α-160** and **β-160** as a 3:1 mixture of anomers with no other isolated products. The high regiospecificity of this reaction, combined with prior our difficulties in affecting the 2'-protection of **155**, indicates a potential 2'-deactivating effect, perhaps owing to hydrogen-bonding interactions with the 3-aminoxazoylurea moiety. The alpha-compound **α-160** was also successfully separated from the mixture, allowing us to achieve full structural characterisation for that isomer.



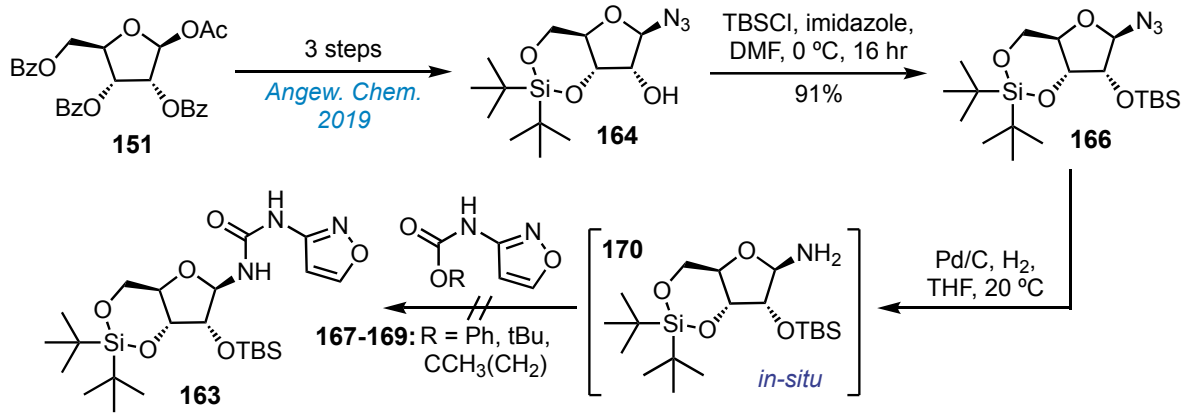
Scheme 35 *The 5'-tritylation of nucleoside **142** and attempted 2'-protection of the resulting DMTr-adducts **α-160** and **β-160**.*[††]

Once again, introduction of a suitable 2'-protecting group proved troublesome, whereby treatment of the mixture containing **α-160** and **β-160** with TOMCl in the presence of Bu_2SnCl_2 – conditions that have previously been employed in the reliable 2'-introduction of TOM functional groups[127] – instead gave exclusively the 3'-ether **161**, as was confirmed with 2D COSY and HMBC NMR spectra. In order to exclude that this result was coincidental, **160** was also subsequently treated with Bz_2O and DMAP, which similarly gave rise to the undesired 3'-benzoate **162**, thus mirroring the reactivity of Bu_2SnCl_2. Overall, these results strongly suggested a significant kinetic barrier, steric or otherwise, preventing protection of the 2'-hydroxyl moiety.

The seemingly limited reactivity of the 2'-hydroxyl group present in **160** and other related compounds prompted an investigation into whether the 1-(isoxazol-3-yl)-urea functionality could be introduced to a ribose scaffold already containing the necessary 2'-protecting group. Accordingly, a synthesis of compound **163** was pursued (**Scheme 36**). Firstly, the silyl ether **164** was prepared according to the procedure that we had already developed.[124] This involved substitution of the acetyl group present in **151** with sodium azide, followed by benzoate removal, and subsequent 3'/5'-protection using *t*-Bu_2SiCl_2 to give **165.** A TBS-protecting group was chosen in contrast to our earlier study, where a TOM-protecting group as used, due to the significantly reduced cost and availability of the necessary reagent. In the event, compound **165** was treated with TBSCl in the presence of imidazole, which gave the desired silyl ether **166** in 91% yield. In parallel, the 3-aminoxazole-substituted carbamates **167**, **168** and **169** were synthesised and characterised. In keeping with expectations, when compound **166** was treated with H_2 gas in the presence of palladium on charcoal, the primary amine **170** was

[††] Compounds **161 and 162** were originally synthesised by student S. Hübner, under supervision from myself.

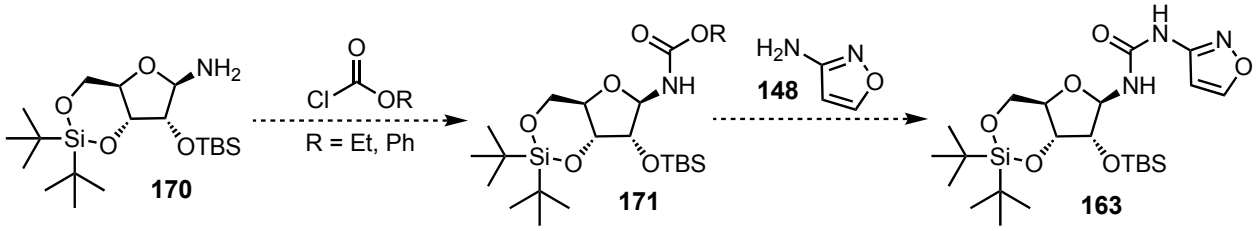
generated, which we confirmed by LCMS and TLC analyses. Subsequent treatment of the *in-situ*-generated amine **170** with carbamates **167-169** however, failed to generate the desired urea **163** in any more than trace quantities, although more optimisation is required.



151 **164** **166** **170** **163**

Scheme 36 *The synthesis of azide **166** and the attempted carbamoylation of the amine generated upon hydrogen-mediated reduction of that compound.*

In summary, some preliminary, but nevertheless encouraging results have been obtained towards the synthesis of the nucleoside phosphoramidite **147**. We envision that after a certain degree of optimisation, the azide **166** could conceivably be converted into **163** using the formerly mentioned strategy, or perhaps *via* a reactive isocyanate. Upon the successful synthesis of **163**, only three additional steps would be required – namely 3'/5'-deprotection, 5'-tritylation, and finally phosphitylation, all of which are routinely effective in phosphoramidite synthesis.[126] In the event that electrophilic additions to **170** continue to prove ineffective, we might instead consider a strategy such as that which is shown in **Scheme 37**. This would firstly involve synthesis of a reactive carbamate **171** from the amine **170**, which would upon isolation, be treated with 3-aminoisoxazole **148** to give the desired compound **163**. Studies towards this goal are now underway in the Carell laboratories.



170 **171** **148** **163**

Scheme 37 *Proposed future steps towards incorporating the 3-aminoisoxazole moiety into the fully protected riboside **170**.*

3.4 EXPERIMENTAL SECTION FOR UNPUBLISHED RESULTS

3.4.1 General Experimental

Chemicals were purchased from Sigma-Aldrich, TCI, Fluka, ABCR, Carbosynth or Acros organics and used without further purification. The solvents were of reagent grade or otherwise purified by distillation. Reactions and chromatography fractions were monitored by qualitative thin-layer chromatography (TLC) on silica gel F_{254} TLC plates from Merck KGaA. Flash column chromatography was performed on Geduran® Si60 (40-63 μm) silica gel from Merck KGaA. Reactions were conducted under a positive pressure of anhydrous nitrogen in oven-dried glassware, and at ambient room temperature, unless otherwise specified. NMR spectra were either recorded on a 400 MHz Bruker AVIIIHD spectrometer or on a 599 MHz or 800 MHz Bruker spectrometer. ^{1}H NMR shifts were calibrated to the following residual solvent signals: Chloroform-*d* (7.26 ppm), DMSO-d_6 (2.50 ppm), Methanol-d_4 (4.87 ppm), Benzene-d_6 (7.16 ppm) or Deuterium Oxide (4.79 ppm). ^{13}C NMR shifts were calibrated to the following residual solvent signals: Chloroform-*d* (77.2 ppm), DMSO-*d*6 (39.52 ppm), Methanol-d_4 (49.00 ppm), Benzene-d_6 (128.06 ppm). All NMR spectra were analysed using the program MestReNOVA 12.0.0 from Mestrelab Research S. L. Normally resolved mass spectra were measured on a LTQ FT-ICR from Thermo Finnigan GmbH. High resolution mass spectra were measured by the analytical department of the Department of Chemistry of the Ludwig-Maximilians-Universität München on the following spectrometers (ionisation mode in brackets): MAT 95 (EI) and FT-ICR MS (ESI) from Thermo Finnigan GmbH, unless otherwise specified. IR spectra were recorded on a PerkinElmer Spectrum BX II FT-IR system. All substances were directly applied as solids or oils on the ATR unit. Prebiotic reactions were carried out and analysed by LC-ESI-MS on a Thermo Finnigan LTQ Orbitrap XL and were chromatographed using a Dionex Ultimate 3000 HPLC system with a flow of 0.15 mL/min over an Interchim Uptisphere120A-3µm-HDO C18 column, unless otherwise specified. The column temperature was maintained at 30 °C. The eluting buffers were: buffer A (2 mM $HCOONH_4$ in H_2O, pH 5.5) and buffer B (2 mM $HCOONH_4$ in H_2O/MeCN 20/80). The pH values of individual buffers were adjusted using a MP 220 pH-meter (Metter Toledo). Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectra were recorded on a Bruker Autoflex II mass spectrometer.

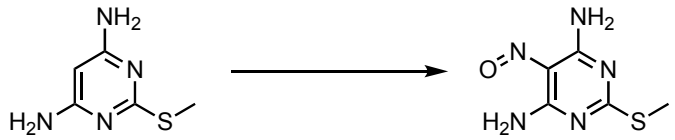
3.4.2 Synthesis of Oligonucleotides

Oligonucleotides used in this study were synthesized on a 1 μmol scale using a DNA automated synthesizer (Applied Biosystems 394 DNA/RNA Synthesizer) with standard phosphoramidite chemistry. The phosphoramidites of canonical ribonucleosides (Bz-A, Ac-C, DMF-G, U) were purchased from Glen Research and Sigma-Aldrich. NPE-protecting groups were removed by treating the CPG-bound oligonucleotides with DBU (10% w/w solution in THF) for 2 hr at room temperature. The cleavage and deprotection of the CPG bound oligonucleotides was carried out by treating the solid support with ammonia in MeOH (7 N, 1 mL) at room temperature for 4 hr. The resin was removed by centrifugation and the solution was evaporated at room temperature under reduced pressure. The residue obtained was subsequently heated with a solution of trimethylamine trihydrofluoride (98 μL) and triethylamine (60 μL) in DMSO (120 μL) at 65°C for 2 h. Upon cooling on ice, NaOAc (3.0 M, 25 μL) and *n*-BuOH (1 mL) were added. The resulting suspension was vortexed and cooled in a freezer (-20°C) for 30 min. After centrifugation, the supernatant was discarded, and the resulting oligonucleotide pellet was dried under reduced pressure. Oligonucleotides were further purified by reverse-phase HPLC using a Waters Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump) equipped with the column VP 250/32 C18 from Macherey Nagel. Preparative RP-HPLC was carried out using the following buffer system: buffer A: 100 mM NEt_3/HOAc, pH 7.0 in H_2O and buffer B: 100 mM NEt_3/HOAc in 80% (v/v) acetonitrile. A flow rate of 5 mL/min with a gradient of 0-25% of buffer B over 30 min was applied for the purifications. Analytical RP-HPLC was performed on a Waters Alliance Machine (2695 Separation Module, 2996 Photodiode Array Detector) equipped with the column Nucleosil 120-2 C18 from Macherey Nagel using a flow rate of 0.5 mL/min. A gradient of 0-25% of buffer B over 30 min was applied, except where otherwise mentioned. Molarity calculations were assisted using the software OligoAnalyzer 3.0 (Integrated DNA Technologies: https://eu.idtdna.com/calc/analyzer). Oligonucleotides were detected at the following wavelength: 260 nm. For strands containing non-canonical bases, extinction coefficients of those bases were derived from their unmodified-counterparts and employed without corrections. The structural integrity of the synthesized oligonucleotides was determined by MALDI-TOF mass measurements. Prior to MALDI-TOF MS, oligonucleotide samples were desalted on a 0.025 μm VSWP filter (Millipore) against ddH_2O and co-crystallized in a 3-hydroxypicolinic acid matrix (HPA).

3.4.3 Specific Transformations

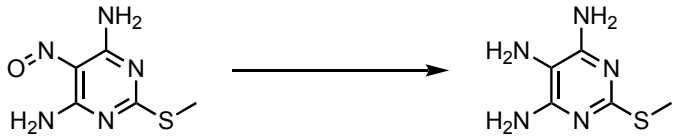
3.4.3.1 Synthesis of Formamidopyrimidines (FaPys) and Related Compounds

2-(Methylthio)-5-nitrosopyrimidine-4,6-diamine **(65)**



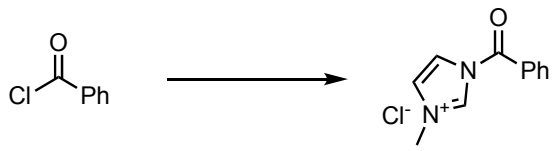
Compound **65** was synthesised according to a modified literature procedure.[128] A magnetically stirred solution containing compound **64** (11.0 g, 70.4 mmol) in a 1:11 mixture of acetic acid in water (500 mL) was treated in several portions with a solution of sodium nitrite (10.6 g, 154.4 mmol) in water (50 mL) at 0 ºC. The resulting blue suspension was stirred for 45 min at 0 ºC, and then slowly warmed to room temperature. Upon consumption of the starting material, as judged by LCMS analysis after stirring for a further 45 min, the reaction was filtered, and the resulting blue precipitate washed with water. The solid was dried for 24 h under reduced pressure to give compound **65** as a blue powder (10.9 g, 58.9 mmol, 84%). The crude product was used directly for the next reaction without further purification. The ^{1}H NMR spectral data for **65** were consistent with those previously reported in the literature.[128] $\mathbf{R_f}$ = 0.2 in 99:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.18 (s, 1H), 9.01 (s, 1H), 8.42 (s, 1H), 8.02 (s, 1H), 2.46 (s, 3H).

2-(Methylthio)pyrimidine-4,5,6-triamine **(63)**



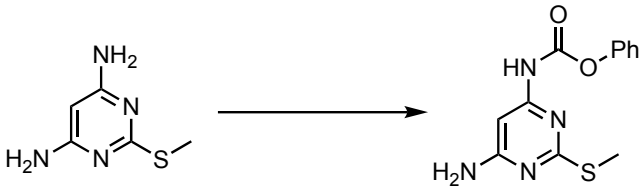
A magnetically stirred solution containing compound **65 (**10.9 g, 58.9 mmol) in a 1:2 mixture of acetic acid in Water (100 mL) was treated with zinc powder (15.0 g, 230 mmol) at 0 ºC, and maintained at that temperature for 30 min. After this time, the reaction mixture was allowed to warm to room temperature and stirred for a further 1 hr. Upon consumption of the starting material, as judged by LCMS analysis, the resulting mixture was filtered through a 2 cm plug of Celite, and the filtrate adjusted to pH 9 using small portions of 30 % ammonium hydroxide solution. The filtrate was cooled to 0 ºC and the resulting precipitate collected by filtration to give compound **63** (8.20 g, 47.9 mmol, 81%, *or* 68% over two steps) as a yellow, crystalline solid. $\mathbf{R_f}$ = 0.2 in 92:8 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 5.69 (s, 4H), 3.74 – 3.47 (m, 2H), 2.30 (s, 3H); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 156.9, 152.8, 102.8, 13.3; **IR** (ν_{max}) 3407, 3144, 2932, 1600, 1547, 1307; **HRMS** (ESI): calculated for $C_5H_{10}N_5S^+$ [M + H]$^+$: 172.0651; found 172.0651.

3-Methyl-1-(phenoxycarbonyl)-1*H*-imidazol-3-ium chloride (**70**)



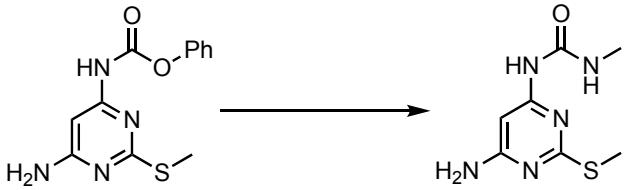
Compound **70** was synthesised according to a modified literature procedure.[129] A magnetically stirred solution of phenyl chloroformate (6.00 g, 4.81 mL, 38.3 mmol) in anhydrous CH_2Cl_2 (40 mL) was treated dropwise with 1-methylimidazole (10.6 g, 3.05 mL, 154.4 mmol) at 0 ºC. The resulting white suspension was stirred for 5 min at 0 ºC, and then slowly warmed to room temperature. Upon consumption of the starting material, as judged by LCMS analysis after stirring for a further 2 hr, the reaction was filtered, and the resulting white precipitate washed twice with anhydrous CH_2Cl_2 (2 x 20 mL). The solid material was dried for 24 h under reduced pressure to give compound **70** as a white powder (8.69 g, 36.4 mmol, 95%). The ^{1}H NMR spectral data for **70** were consistent with those previously reported in the literature.[129] Compound **70** is highly hygroscopic and should be stored under an inert atmosphere to avoid degradation. **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.17 (s, 1H), 8.38 (dd, J = 2.0 Hz, 1H), 7.96 (dd, J = 2.0 Hz, 1H), 7.62 – 7.52 (m, 2H), 7.52 – 7.41 (m, 3H), 3.99 (s, 3H).

Phenyl [6-amino-2-(methylthio)pyrimidin-4-yl]carbamate **(76)**



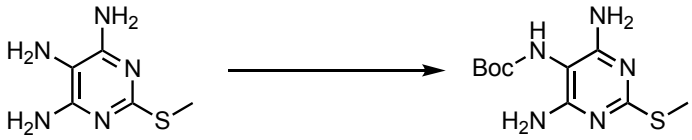
A magnetically stirred solution containing compound **64 (**1.00 g, 6.40 mmol) in anhydrous CH_2Cl_2 (100 mL) was treated with 3-methyl-1-(phenoxycarbonyl)-1*H*-imidazol-3-ium chloride **70** (1.53 g, 6.40 mmol) at 20 ºC, and maintained at that temperature for 4 hr. Upon consumption of the starting material, as judged by TLC analysis, the resulting mixture was concentrated under reduced pressure, and the residue was purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2) to afford compound **76** (921 mg, 3.33 mmol, 52%) as a colourless oil. $\mathbf{R_f}$ = 0.5 in 96:4 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.52 (s, 1H), 7.46 – 7.38 (m, 2H), 7.30 – 7.23 (m, 1H), 7.23 – 7.17 (m, 2H), 6.89 (s, 2H), 6.60 (s, 1H), 2.41 (s, 3H); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 169.9, 165.1, 157.2, 152.0, 150.7, 129.9, 126.2, 122.4, 86.1, 13.5; **IR** (ν_{max}) 3150, 2801, 1760, 1533, 1593; **HRMS** (ESI): calculated for $C_{12}H_{13}N_4O_2S^+$ [M + H]$^+$: 277.0754; found 277.0748.

1-[6-Amino-2-(methylthio)pyrimidin-4-yl]-3-methylurea **(77)**



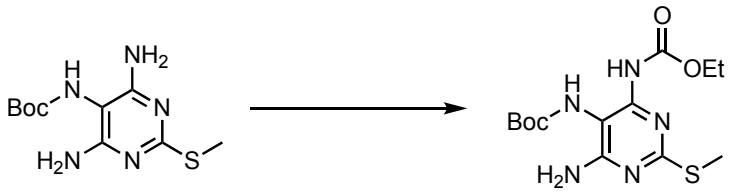
A magnetically stirred solution of compound **76** (500 mg, 1.81 mmol) in pyridine (20 mL) was treated in several portions with a 2M solution of methylamine in THF (5.43 mL, 10.86 mmol) at 20 ºC. The solution was then maintained for 24 hr at 50 ºC. Upon consumption of the starting material, as judged by LCMS analysis, the reaction mixture was subsequently filtered and the resulting precipitate washed three times with CH_2Cl_2 (5 mL). The solid was dried for 24 h under reduced pressure to give compound **77** as white powder (201 mg, 0.943 mmol, 52%) that did not require further purification. $\mathbf{R}_f$ = 0.1 in 96:4 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 9.01 (s, 1H), 7.45 (s, 1H), 6.70 (s, 2H), 6.22 (s, 1H), 2.67 (d, J = 4.5 Hz, 3H), 2.37 (s, 3H); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 168.8, 164.1, 157.8, 155.0, 83.4, 25.9, 13.1; **IR** (ν_{max}) 3446, 3160, 2927, 1689, 1589, 1491; **HRMS** (ESI): calculated for $C_7H_{12}N_5OS^+$ [M + H]$^+$: 214.0757; found 214.0757.

tert-Butyl [4,6-diamino-2-(methylthio)pyrimidin-5-yl]carbamate **(79)**



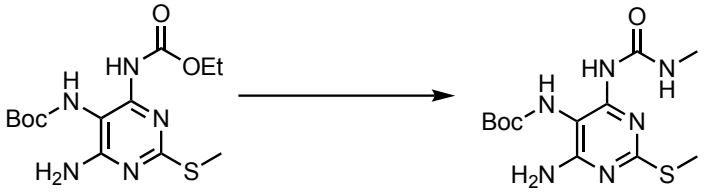
A magnetically stirred solution of compound **63** (8.20 g, 47.9 mmol) in anhydrous *t*-BuOH (200 mL) was treated dropwise with Di-*tert*-butyl dicarbonate (Boc_2O) (11.5 g, 12.1 mL, 52.7 mmol) at room temperature and the reaction mixture was heated to 60 ºC. Upon consumption of the starting material, as judged by LCMS analysis after 2 hr, the mixture was allowed to cool to room temperature, and separated into smaller portions. The individual portions were centrifuged, and the supernatant discarded. The precipitate was washed twice with anhydrous *t*-BuOH (50 mL) and then dried under reduced pressure to afford compound **79** (9.91 g, 36.5 mmol, 76%) as a white powder. $\mathbf{R}_f$ = 0.3 in 93:7 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 7.57 (s, 1H), 5.86 (s, 4H), 2.34 (s, 3H), 1.42 (s, 9H); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 166.0, 159.9, 154.1, 92.5, 78.6, 28.2, 13.4; **IR** (ν_{max}) 3475, 3312, 2935, 1680, 1610, 1478; **HRMS** (ESI): calculated for $C_{10}H_{18}N_5O_2S^+$ [M + H]$^+$: 272.1176; found 272.1176.

tert-Butyl ethyl [6-amino-2-(methylthio)pyrimidine-4,5-diyl]dicarbamate **(81)**



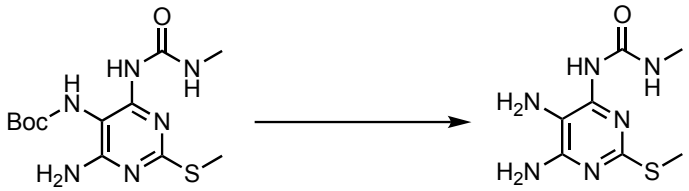
A magnetically stirred solution of compound **79** (9.90 g, 36.5 mmol) in anhydrous pyridine (60 mL) was treated dropwise with ethylchloroformate (6.74 g, 5.9 mL, 62.1 mmol) at 0 ºC. The reaction mixture was subsequently heated to 60 ºC, and maintained at that temperature for 4 h. Upon consumption of the starting material, as judged by TLC analysis, the mixture was cooled, concentrated under reduced pressure, and the residue thus obtained purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2) to afford compound **81** (6.12 g, 17.8 mmol, 49%) as a white solid. $\mathbf{R}_f$ = 0.4 in 96:4 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 7.34 (s, 1H), 7.07 (s, 1H), 5.26 (s, 2H), 4.22 (q, *J* = 7.1 Hz, 2H), 2.43 (s, 3H), 1.44 (s, 9H), 1.30 (t, *J* = 7.1 Hz, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 168.2, 161.5, 153.8, 151.1, 149.8, 128.3, 123.7, 102.5, 81.0, 62.5, 28.2, 14.3, 14.2; **IR** (ν_{max}) 3329, 2979, 1716, 1618, 1496; **HRMS** (ESI): calculated for $C_{13}H_{22}N_5O_4S^+$ [*M* + H]$^+$: 344.1387; found 344.1386.

tert-Butyl [4-amino-2-(methylthio)-6-(3-methylureido)pyrimidin-5-yl]carbamate **(82a)**



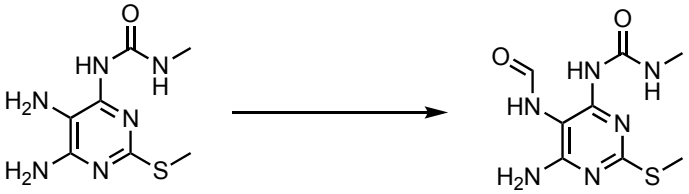
A sealed glass tube containing compound **81** (500 mg, 1.46 mmol) suspended in a solution of ethanolic methylamine (33 % w/w, 6.25 mL) was heated to 45 ºC, and magnetically stirred at that temperature for 2.5 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was concentrated under reduced pressure. The residue thus obtained was subjected to flash column chromatography (silica gel, EtOAc:Hexane = 4:1) to afford compound **82a** (64 mg, 0.19 mmol, 13%) as a white solid. $\mathbf{R}_f$ = 0.3 in 4:1 v/v EtOAc:Hexane elution; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 8.78 (d, *J* = 4.8 Hz, 1H), 7.79 (s, 1H), 7.70 (s, 1H), 6.68 (s, 2H), 2.75 (d, *J* = 4.8 Hz, 3H), 2.42 (s, 3H), 1.42 (s, 9H); **^{13}C NMR** (101 MHz, DMSO) δ 166.7, 161.2, 154.7, 154.0, 95.1, 79.7, 28.5, 26.5, 14.1; **IR** (ν_{max}) 3428, 3250, 1720, 1671, 1532, 1488; **HRMS** (ESI): calculated for $C_{12}H_{21}N_6O_3S^+$ [*M* + H]$^+$: 329.1390; found 329.1391.

1-[5,6-Diamino-2-(methylthio)pyrimidin-4-yl]-3-methylurea **(72a)**



A 15 mL-Falcon tube containing a solution of compound **82a** (15 mg, 45.7 μmol) in methanol (1.8 mL) was treated with a large excess of trimethylsilyl chloride (770 mg, 900 μL, 7.09 mmol) at 20 ºC, and the resulting solution was shaken at that temperature, during which a white precipitate appeared. Upon consumption of the starting material, as judged by LCMS analysis after 1 hr, the reaction mixture was concentrated under reduced pressure, and the solid material was re-suspended in CH_2Cl_2 (1 mL). The mixture was then centrifuged, and the supernatant discarded. The solid residue was washed once more with CH_2Cl_2 (1 mL), and then dried overnight under high vacuum to afford compound **72a** (8.5 mg, 37.2 mmol, 81%) as a white powder. $\mathbf{R}_f$ = 0.5 in 9:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 9.23 (s, 2H), 8.12 – 7.84 (m, 2H), 7.73 – 6.69 (m, 0H), 2.78 – 2.65 (m, 3H), 2.43 (s, 3H); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 156.6, 155.8, 127.5, 26.6, 13.8; **IR** (ν_{max}) 3290, 3174, 2882, 1646, 1540, 1188; **HRMS** (ESI): calculated for $C_7H_{13}N_6OS^+$ [M + H]$^+$: 229.0866; found 229.0866.

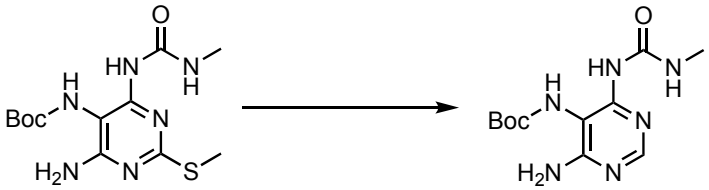
N-[4-Amino-2-(methylthio)-6-(3-methylureido)pyrimidin-5-yl]formamide **(60a)**



A 1.5 mL-safe lock tube containing a solution of compound **72a** (41 mg, 0.18 mmol) in formic acid (300 μL) was treated with sodium formate (14 mg, 0.21 mmol) and the reaction mixture was heated to 50 ºC and shaken at that temperature. Upon consumption of the starting material, as judged by TLC analysis after 2.5 hr, the mixture was concentrated under reduced pressure, and the solid material was suspended in H_2O (1 mL). The mixture was then centrifuged, and the supernatant discarded. The solid residue was washed once more with H_2O (1 mL), and then dried overnight under high vacuum to afford compound **60a** (41 mg, 0.16 mmol, 89%) as a white powder. $\mathbf{R}_f$ = 0.4 in 94:6 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 8.94 (d, J = 4.8 Hz, 1H), 8.77 (d, J = 1.3 Hz, 1H), 8.23 (s, 1H), 8.06 (d, J = 1.3 Hz, 1H), 6.76 (s, 2H), 2.75 (d, J = 4.8 Hz, 3H), 2.42 (s, 3H); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 166.6, 161.7, 160.3, 154.4, 153.5, 92.9, 26.0, 13.6; **IR** (ν_{max}) 3280,

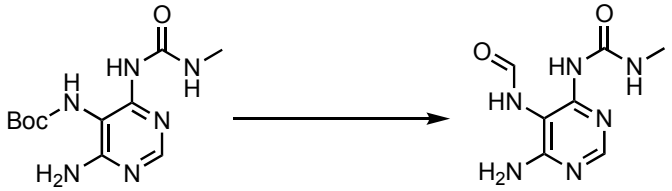
3163, 2882, 1706, 1673, 1537, 1342; **HRMS** (ESI): calculated for $C_8H_{13}N_6O_2S^+$ [*M* + H]$^+$: 257.0815; found 257.0814.

tert-Butyl [4-amino-6-(3-methylureido)pyrimidin-5-yl]carbamate **(82b)**



A sealed glass tube containing a solution of compound **82a** (30 mg, 91 μmol) in THF (8 mL) was treated with palladium on charcoal (3.2 mg, 28 μmol) and triethylsilane (109 mg, 150 μL, 0.94 mmol). The reaction mixture was heated to 40 ºC and magnetically stirred at that temperature for 50 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was concentrated under reduced pressure (*caution, smelly)*, and the residue purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2 → 96:4) to afford compound **82b** (9 mg, 31.9 μmol, 35%) as a colourless oil. $\mathbf{R_f}$ = 0.4 in 96:4 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 9.02 (d, *J* = 4.8 Hz, 1H), 8.09 – 7.85 (m, 2H), 7.66 (s, 1H), 6.60 (s, 2H), 2.75 (d, *J* = 4.8 Hz, 3H), 1.43 (s, 9H); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 161.0, 154.4, 154.0, 153.6, 153.4, 97.9, 79.4, 28.0, 26.1; **IR** (ν_{max}) 3470, 1384, 3119, 2977, 2361, 1680, 1549, 1159; **HRMS** (ESI): calculated for $C_{11}H_{19}N_6O_3^+$ [*M* + H]$^+$: 283.1513; found 283.1512.

N-[4-Amino-6-(3-methylureido)pyrimidin-5-yl]formamide **(60b)**



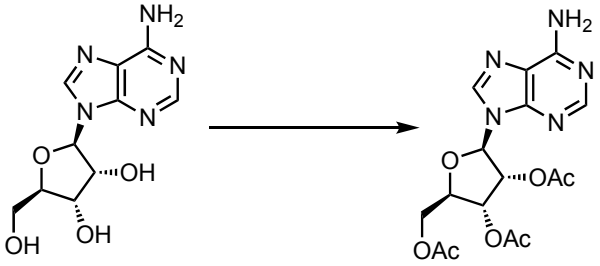
A 15 mL-Falcon tube containing a solution of compound **82b** (33 mg, 0.12 mmol) in methanol (4 mL) was treated with a large excess of trimethylsilyl chloride (1.71g, 2 mL, 15.8 mmol) at room temperature, and the reaction mixture was shaken at 25 ºC for 75 min. After this time, the reaction mixture was concentrated under reduced pressure, and the solid material was re-suspended in CH_2Cl_2 (1 mL). The mixture was then centrifuged, and the supernatant discarded. The solid residue was washed once more with CH_2Cl_2 (1 mL), and then dried overnight under high vacuum to afford a white powder that was used directly for the subsequent reaction without further purification. A 1.5 mL-safe lock tube containing a solution of the crude compound from the previous step in formic acid (300 μL) was treated with sodium formate (9.2 mg, 0.13 mmol) and the reaction mixture was heated to 50 ºC.

Upon consumption of the starting material, as judged by LCMS analysis after 2.5 hr, the mixture was cooled and concentrated under reduced pressure, and the solid material was suspended in H_2O (1 mL). The mixture was then centrifuged, and the supernatant discarded. The solid residue was washed once more with H_2O (1 mL), and then dried overnight under high vacuum to afford compound **60b** (21 mg, 0.10 mmol, 83% over two steps) as a white powder. $\mathbf{R_f}$ = 0.4 in 94:6 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (599 MHz, DMSO-d_6) δ 9.16 (d, J = 4.8 Hz, 1H), 8.98 (s, 1H), 8.14 (s, 1H), 8.10 (s, 1H), 8.05 (s, 1H), 6.69 (s, 2H), 2.75 (d, J = 4.8 Hz, 3H); **^{13}C NMR** (151 MHz, DMSO-d_6) δ 161.5, 160.6, 154.6, 154.4, 153.6, 96.3, 26.1; **IR** (ν_{max}) 3330, 3265, 3154, 1680, 1557, 1467, 1267; **HRMS** (ESI): calculated for $C_7H_{11}N_6O_2^+$ [M + H]$^+$: 211.0938; found 211.0939.

3.4.3.2 Synthesis of Modified Adenosine Nucleosides and Phosphoramidites

*Note: Phosphoramidite **136** and DMTr-compounds **134** and **135** were prepared according to a slightly modified synthesis route compared with that appearing in the final published work.*[130] *The alternative formulation is therefore also presented in this section.*

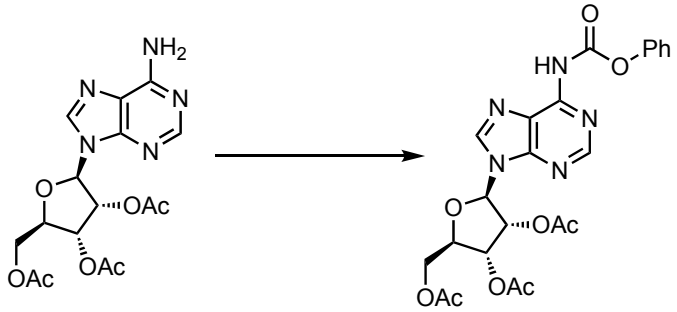
2',3',5'-Tri-*O*-acetyladenosine **(83)**



Compound **83** was synthesised according to a modified literature procedure.[131] A magnetically stirred solution containing adenosine **33** (10.0 g, 37.4 mmol) and 4-dimethylaminopyridine (1.83 g, 14.97 mmol) in anhydrous CH_3CN (250 mL) was treated with triethylamine (15.2 g, 20.9 mL, 149.7 mmol), and then acetic anhydride (15.3 g, 14.1 mL, 149.7 mmol) at 20 ºC. The resulting solution was maintained for 18 hr at room temperature. Upon completion of the reaction, as judged by TLC analysis, the mixture was cooled to 0 ºC and treated with distilled water (20 mL). The aqueous phase was then extracted with CH_2Cl_2 (3 x 100 mL), and the combined organic phases dried ($MgSO_4$), filtered, and concentrated under reduced pressure. Recrystallisation of the material thus obtained, *via* the slow cooling of a saturated solution of boiling ethanol (ca. 200 mL), gave compound **83** as amber crystals (12.5 g, 31.8 mmol, 85%). The ^{1}H NMR spectral data for **83** were consistent with those previously reported in the literature.[131] $\mathbf{R_f}$ = 0.4 in 9:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 8.36 (s, 1H), 8.17 (s, 1H), 7.40 (s, 2H), 6.20 (d, J = 5.4 Hz, 1H), 6.03 (dd, J = 5.7

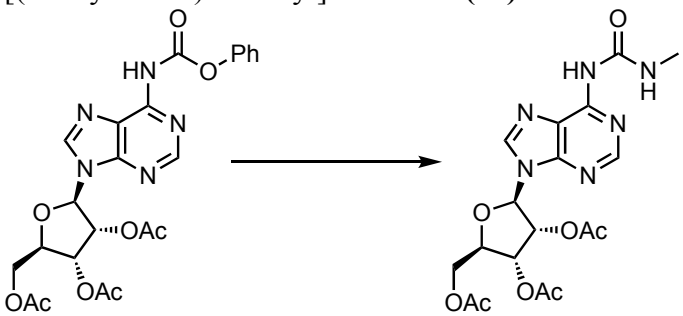
Hz, 1H), 5.63 (dd, *J* = 5.7, 4.8 Hz, 1H), 4.45 – 4.33 (m, 2H), 4.24 (dd, J = 11.7, 5.4 Hz, 1H), 2.12 (s, 3H), 2.04 (s, 3H), 2.01 (s, 3H).

2',3',5'-Tri-*O*-acetyl-N^6-(phenoxycarbonyl)adenosine **(84)**



Compound **84** was synthesised according to a modified literature procedure.[107] A magnetically stirred solution containing compound **83** (1.00 g, 2.54 mmol) in anhydrous CH_2Cl_2 (250 mL), maintained under an atmosphere of argon, was treated with 3-methyl-1-(phenoxycarbonyl)-1*H*-imidazol-3-ium chloride (**70**) (1.82 g, 8.17 mmol) at 20 ºC, and the resulting milky-white suspension was maintained for 3 hr at that temperature. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was concentrated under reduced pressure. The residue thus obtained was then purified by flash chromatography (silica gel, EtOAc:Hexane = 1:1 → 9:1) to afford compound **84** (1.06 g, 2.06 mmol, 81%) as a colourless foam. The ^{1}H NMR spectral data for **84** were consistent with those previously reported in the literature.[107] $\mathbf{R}_f$ = 0.2 in 99:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 8.81 (s, 1H), 8.51 – 8.45 (m, 1H), 8.16 (s, 1H), 7.45 – 7.37 (m, 2H), 7.30 – 7.21 (m, 5H), 6.24 (d, *J* = 5.2 Hz, 1H), 5.94 (dd, *J* = 5.2 Hz, 1H), 5.67 (dd, *J* = 5.6, 4.4 Hz, 1H), 4.55 – 4.33 (m, 4H), 2.16 (s, 3H), 2.13 (s, 4H), 2.09 (s, 3H).

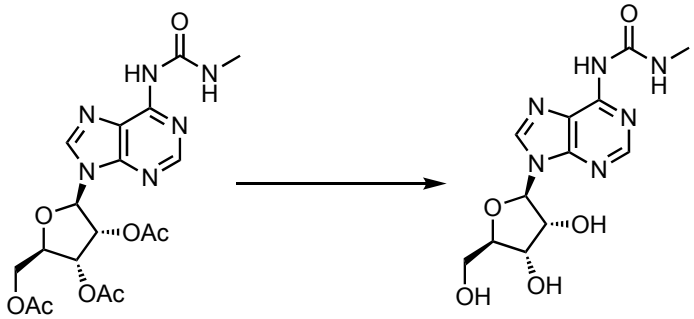
2',3',5'-Tri-*O*-acetyl-N^6-[(methylamino)carbonyl]adenosine **(86)**



Compound **86** was synthesised according to a modified literature procedure.[132] A sealed glass tube containing a magnetically stirred solution of compound **84** (420 mg, 818 μmol) in anhydrous pyridine (8 mL) was treated with a 2.0 M solution of methylamine in THF (531 μL, 1.06 mmol) at 20 ºC, and the resulting solution was heated to 55 ºC and maintained at that temperature for 12 hr. Upon

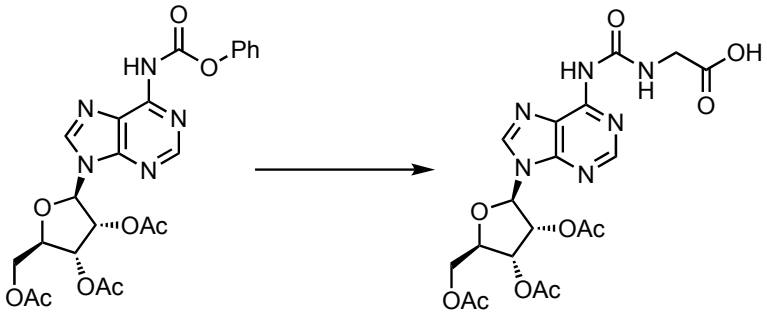
consumption of the starting material, as judged by TLC analysis, the reaction mixture was cooled to room temperature. The cooled reaction mixture was then concentrated under reduced pressure, and the residue purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 97:3) to afford compound **86** (198 mg, 442 μmol, 54%) as a colourless foam. The ^{1}H NMR spectral data for **86** were consistent with those previously reported.[132] **R_f** = 0.2 in 19:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 9.52 – 9.40 (m, 1H), 9.26 (s, 1H), 8.56 (s, 1H), 8.49 (s, 1H), 6.20 (d, *J* = 5.4 Hz, 1H), 6.00 (dd, *J* = 5.4 Hz, 1H), 5.67 (dd, *J* = 4.9 Hz, 1H), 4.52 – 4.23 (m, 3H), 2.97 (d, *J* = 4.6 Hz, 3H), 2.11 (s, 3H), 2.07 (s, 3H), 2.04 (s, 4H).

N^6-[(Methylamino)carbonyl]adenosine **(59b)**



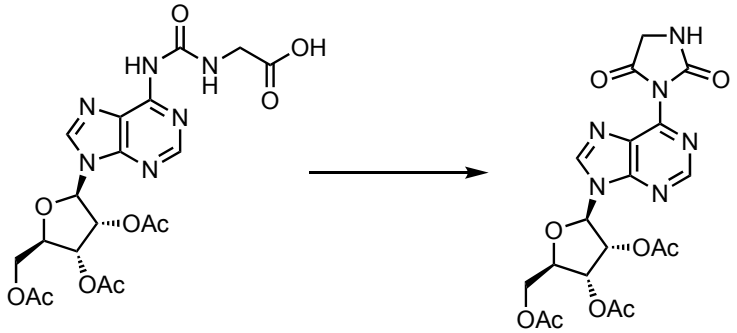
Compound **59b** was synthesised according to a modified literature procedure.[132] Compound **86** (198 mg, 0.439 mmol) was added in one-portion to a magnetically stirred 7N solution of ammonia in methanol (3.77 mL, 26.38 mmol) at 20 ºC, and the resulting mixture was maintained at that temperature for 12 hr. Upon completion of the reaction, as judged by LCMS analysis, the reaction mixture was then centrifuged, and the supernatant discarded. The solid residue was washed once with a portion of ice-cold methanol (1 mL), and then dried overnight under high vacuum to afford compound **59b** (119 mg, 0.365 mmol, 83%) as a white powder. The ^{1}H NMR spectral data for **59b** were consistent with those previously reported.[132] **R_f** = 0.1 in 9:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 9.24 (d, *J* = 4.7 Hz, 1H), 8.65 (s, 1H), 8.54 (s, 1H), 5.97 (d, *J* = 5.7 Hz, 1H), 5.59 (s, 1H), 5.31 (s, 1H), 5.15 (s, 1H), 4.59 (dd, *J* = 5.4 Hz, 1H), 4.17 (dd, *J* = 4.3 Hz, 1H), 4.02 – 3.87 (m, 1H), 3.79 – 3.65 (m, 1H), 3.65 – 3.52 (m, 1H), 2.83 (d, *J* = 4.7 Hz, 3H).

2',3',5'-Tri-*O*-acetyl-N^6-glycinylcarbamoyladenosine **(101)**



A magnetically stirred solution containing compound **84** (7.16 g, 13.96 mmol) in anhydrous pyridine (60 mL) was treated with glycine (2.62 g, 34.91 mmol) at 20 ºC, and the resulting mixture was heated to 55 ºC and maintained at that temperature for 4 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was cooled to room temperature. The cooled reaction mixture was then concentrated under reduced pressure, and the residue purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 9:1) to afford compound **101** (4.99 g, 10.10 mmol, 72%) as a colourless foam. $\mathbf{R_f}$ = 0.3 in 9:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 12.70 (s, 1H), 9.99 (s, 1H), 9.61 (dd, J = 5.7 Hz, 1H), 8.65 (s, 1H), 8.60 (s, 1H), 6.31 (d, J = 5.4 Hz, 1H), 6.04 (dd, J = 5.7 Hz, 1H), 5.64 (dd, J = 5.9, 4.7 Hz, 1H), 4.49 – 4.34 (m, 2H), 4.34 – 4.20 (m, 1H), 4.00 (d, J = 5.7 Hz, 2H), 2.13 (s, 3H), 2.05 (s, 3H), 2.02 (s, 3H); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 171.9, 170.5, 169.9, 169.8, 154.1, 151.6, 150.8, 150.6, 143.2, 120.9, 86.2, 80.0, 72.5, 70.4, 63.2, 42.1, 21.0, 20.9, 20.7; **IR** (ν_{max}) 3232, 2942, 1746, 1715, 1675, 1593, 1561; **HRMS** (ESI): calculated for $C_{19}H_{23}N_6O_{10}^+$ $[M + H]^+$: 495.1470; found 495.1472.

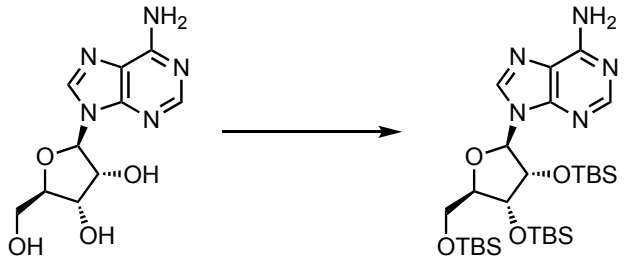
2',3',5'-Tri-*O*-acetyl-6-(3-hydantoinyl)purinosine **(102)**



To a magnetically stirred flask charged with *neat* acetic anhydride (3.24 g, 3.00 mL, 31.7 mmol), was added compound **101** (181 mg, 0.366 mmol) at 20 ºC, and the resulting solution maintained at that temperature for 1.5 hr. Upon completion of the reaction, as judged by TLC analysis, the reaction mixture was then cooled to 0 ºC and quenched with methanol (3 mL). After stirring for an additional 5 min, the reaction mixture was concentrated under reduced pressure, and the residue purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 96:4) to afford compound **102** (131 mg, 0.275

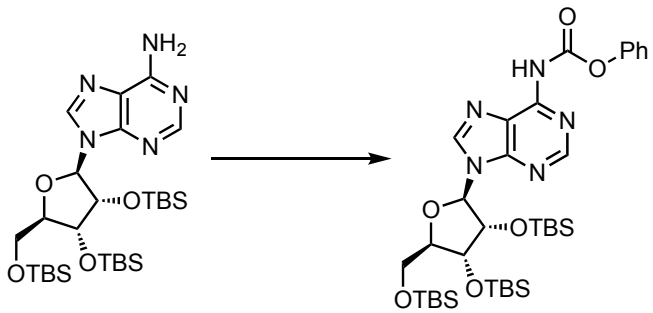
mmol, 75%) as a colourless foam. $\mathbf{R}_f$ = 0.2 in 96:4 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 9.04 (s, 1H), 8.38 (s, 1H), 7.05 (s, 1H), 6.29 (d, *J* = 5.6 Hz, 1H), 5.97 (dd, *J* = 5.6 Hz, 1H), 5.64 (dd, *J* = 5.6, 4.1 Hz, 1H), 4.49 – 4.44 (m, 1H), 4.42 – 4.37 (m, 2H), 4.25 (s, 2H), 2.15 (s, 3H), 2.11 (s, 3H), 2.07 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 170.5, 169.8, 169.5, 169.1, 155.6, 153.7, 153.0, 144.7, 144.1, 130.0, 86.5, 80.8, 73.1, 70.8, 63.2, 47.2, 20.9, 20.7, 20.5; **IR** (ν_{max}) 3311, 3122, 2936, 1726, 1602, 1579, 1380; **HRMS** (ESI): calculated for $C_{19}H_{21}N_6O_9^+$ [*M* + H]$^+$: 477.1365; found 477.1365.

2',3',5'-Tris-*O*-(*tert*-butyldimethylsilyl)adenosine **(104)**



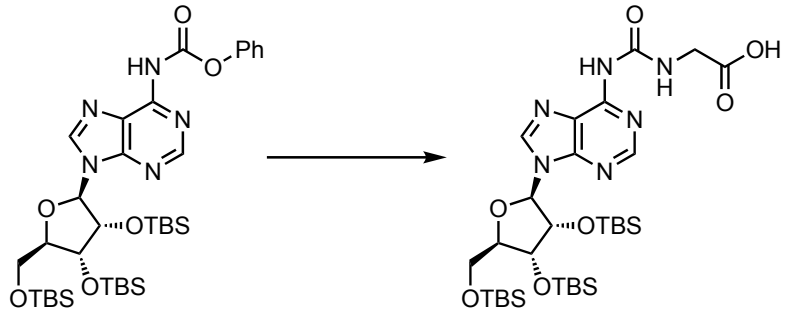
Compound **104** was synthesised according to a modified literature procedure.[133] A magnetically stirred solution containing adenosine **33** (5.00 g, 18.7 mmol) in anhydrous DMF (120 mL) was treated with *tert*-butyldimethylsilyl chloride, (11.3 g, 74.8 mmol), and then imidazole (7.64 g, 112 mmol) at 20 ºC, and the resulting mixture was maintained at that temperature for 24 hr. Upon consumption of the starting material, as judged by TLC analysis, the mixture was then treated with NH_4Cl (200 mL of a sat. *aq.* solution), the aqueous phase extracted with EtOAc (3 x 100 mL), and the combined organic phases dried ($MgSO_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:1→ 100:5) to afford compound **104** (10.6 g, 17.4 mmol, 93%) as a colourless foam. The ^{1}H and ^{13}C NMR spectral data for **104** were consistent with those previously reported in the literature.[133] $\mathbf{R}_f$ = 0.4 in 1:1 v/v EtOAc:Hexane elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 8.35 (s, 1H), 8.16 (s, 1H), 7.36 (s, 1H), 6.03 (d, *J* = 5.2 Hz, 1H), 5.52 (s, 2H), 4.69 (dd, *J* = 4.8 Hz, 1H), 4.32 (dd, *J* = 3.8 Hz, 1H), 4.13 (d, *J* = 3.8 Hz, 1H), 4.03 (dd, *J* = 11.3, 4.2 Hz, 1H), 3.79 (dd, *J* = 11.3, 2.8 Hz, 1H), 0.95 (s, 9H), 0.93 (s, 9H), 0.79 (s, 9H), 0.14 (s, 3H), 0.13 (s, 3H), 0.11 – 0.09 (m, 6H), -0.05 (s, 3H), -0.23 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 155.6, 153.1, 150.1, 139.8, 120.2, 88.5, 85.7, 76.0, 72.2, 62.7, 26.3, 26.0, 25.9, 18.7, 18.3, 18.0, -4.2, -4.5, -4.6, -4.9, -5.2; **IR** (ν_{max}) 3310, 3167, 2929, 2857, 1675, 1599, 1573, 1472; **HRMS** (ESI): calculated for $C_{28}H_{56}N_5O_4Si_3^+$ [*M* + H]$^+$: 610.3635; found 610.3638.

2',3',5'-Tris-*O*-(*tert*-butyldimethylsilyl)-N^6-(phenoxycarbonyl)adenosine **(105)**



A magnetically stirred solution containing compound **104** (1.50 g, 2.46 mmol) in anhydrous CH_2Cl_2 (55 mL) was treated with 3-methyl-1-(phenoxycarbonyl)-1*H*-imidazol-3-ium chloride (**70**) (1.47 g, 6.15 mmol) at 20 ºC, and the resulting milky-white suspension was maintained for 3 hr at that temperature. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was concentrated under reduced pressure. The residue thus obtained was then purified by flash chromatography (silica gel, EtOAc:Hexane = 1:4) to afford compound **105** (1.39 g, 1.89 mmol, 77%) as a colourless foam. $\mathbf{R_f}$ = 0.3 in 1:4 v/v EtOAc:Hexane elution; **^{1}H NMR** (400 MHz, Benzene-d_6) δ 8.51 (s, 1H), 7.98 (s, 1H), 7.14 – 7.06 (m, 2H), 6.86 – 6.78 (m, 3H), 6.10 (d, *J* = 4.3 Hz, 1H), 5.55 (s, 2H), 5.06 (dd, *J* = 4.3 Hz, 1H), 4.56 (dd, *J* = 4.3 Hz, 1H), 4.28 (ddd, *J* = 4.9, 3.1 Hz, 1H), 4.10 (dd, *J* = 11.4, 4.9 Hz, 1H), 3.75 (dd, *J* = 11.4, 3.1 Hz, 1H), 1.01 (s, 9H), 0.97 (s, 9H), 0.93 (s, 9H), 0.15 – 0.12 (m, 6H), 0.09 (s, 3H), 0.08 (s, 3H), 0.05 (s, 3H), -0.05 (s, 3H); **^{13}C NMR** (101 MHz, Benzene-d_6) δ 157.3, 156.1, 153.4, 150.1, 140.2, 129.9, 120.7, 120.3, 115.9, 89.7, 85.1, 75.4, 71.9, 62.3, 26.2, 26.1, 26.0, 18.7, 18.3, 18.2, -4.2, -4.5, -4.6, -4.6, -5.3, -5.4; **IR** (ν_{max}) 2953, 2857, 1764, 1616, 1583, 1462; **HRMS** (ESI): calculated for $C_{35}H_{60}N_5O_6Si_3^+$ [*M* + H]$^+$: 730.3846; found 730.3850.

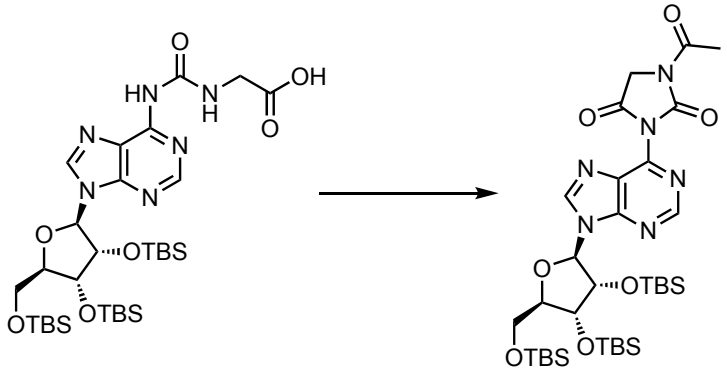
2',3',5'-Tris-*O*-(*tert*-butyldimethylsilyl)-N^6-glycinylcarbamoyladenosine **(106)**



A magnetically stirred solution containing compound **105** (1.80 g, 2.47 mmol) in anhydrous pyridine (25 mL) was treated with glycine (1.11 g, 14.8 mmol) at 20 ºC, and the resulting mixture was heated to 55 ºC and maintained at that temperature for 8 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was cooled to room temperature. The cooled reaction mixture was then concentrated under reduced pressure, and the residue purified by flash

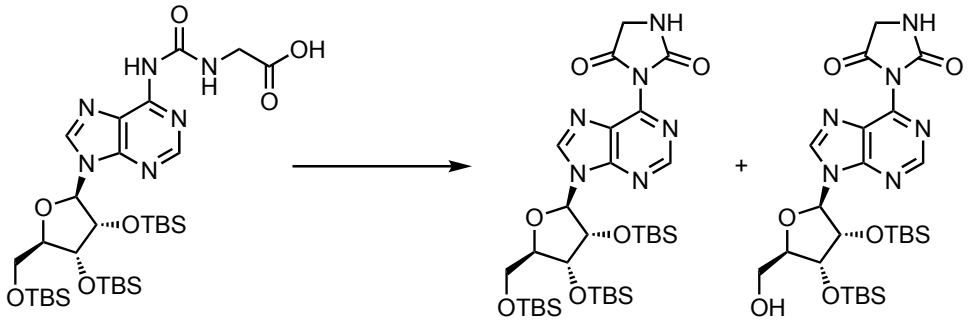
chromatography (silica gel, CH_2Cl_2:CH_3OH = 9:1) to afford compound **106** (1.44 g, 2.03 mmol, 82%) as a colourless foam. $\mathbf{R_f}$ = 0.3 in 9:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 9.94 (s, 1H), 9.64 (dd, J = 5.7 Hz, 1H), 8.63 (s, 1H), 8.55 (s, 1H), 6.02 (d, J = 6.4 Hz, 1H), 4.91 (dd, J = 6.4, 4.3 Hz, 1H), 4.34 (dd, J = 4.3, 1.8 Hz, 1H), 4.10 – 3.90 (m, 4H), 3.77 (dd, J = 10.2, 3.2 Hz, 1H), 0.93 (s, 9H), 0.89 (s, 9H), 0.70 (s, 9H), 0.14 (s, 3H), 0.12 (s, 3H), 0.08 (s, 6H), -0.10 (s, 3H), -0.39 (s, 3H); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 171.5, 153.6, 150.9, 150.4, 150.3, 142.2, 120.3, 87.0, 85.6, 74.3, 72.3, 62.5, 41.8, 25.8, 25.7, 25.4, 18.1, 17.8, 17.5, -4.7, -4.8, -4.8, -5.4, -5.5, -5.5; **IR** (ν_{max}) 2929, 2857, 1702, 1612, 1590, 1471; **HRMS** (ESI): calculated for $C_{31}H_{59}N_6O_7Si_3^+$ [M + H]$^+$: 711.3748; found 711.3750.

2',3',5'-Tris-*O*-(*tert*-butyldimethylsilyl)-6-(1-acetyl-3-hydantoinyl)purinosine **(109)**



To a magnetically stirred flask charged with *neat* acetic anhydride (2.16 g, 2.00 mL, 21.2 mmol), was added compound **106** (40 mg, 56.3 μmol) at 20 ºC, and the resulting solution maintained at that temperature for 1.5 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was then concentrated under reduced pressure, and the residue purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:1) to afford compound **109** (30 mg, 43.3 μmol, 77%) as a colourless foam. $\mathbf{R_f}$ = 0.2 in 99:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-d) δ 9.02 (s, 1H), 8.62 (s, 1H), 6.17 (d, J = 4.3 Hz, 1H), 4.57 (dd, J = 4.3 Hz, 1H), 4.33 (dd, J = 4.3 Hz, 1H), 4.21 – 4.14 (m, 1H), 4.04 (dd, J = 11.6, 3.3 Hz, 1H), 3.81 (dd, J = 11.6, 2.4 Hz, 1H), 2.65 (s, 3H), 0.95 (s, 9H), 0.93 (s, 9H), 0.81 (s, 9H), 0.14 (s, 3H), 0.14 (s, 3H), 0.11 (s, 3H), 0.10 (s, 3H), 0.00 (s, 3H), -0.15 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-d) δ 169.0, 165.3, 153.9, 152.4, 151.1, 145.3, 142.3, 130.1, 88.8, 85.3, 76.4, 71.3, 62.1, 48.3, 26.2, 25.8, 25.7, 24.7, 18.6, 18.1, 17.9, -4.3, -4.7, -4.7, -4.9, -5.3, -5.3; **IR** (ν_{max}) 2953, 2930, 2897, 1751, 1714, 1605, 1372; **HRMS** (ESI): calculated for $C_{32}H_{61}N_6O_7Si_3^+$ [M-Ac+MeOH+2H]$^+$: 725.3904; found 725.3990; calculated for $C_{32}H_{59}N_6O_7Si_3^-$ [M-Ac+MeOH]$^-$: 723.3759; found 723.3770.

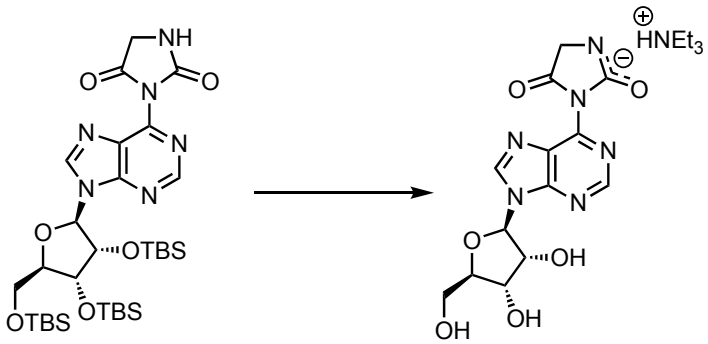
2',3',5'-Tris-*O*-(*tert*-butyldimethylsilyl)-6-(3-hydantoinyl)purinosine **(107)**
and 2',3'-Bis-*O*-(*tert*-butyldimethylsilyl)-6-(3-hydantoinyl)purinosine **(108)**



A magnetically stirred solution containing compound **106** (30 mg, 43.3 μmol) in anhydrous CH_2Cl_2 (3 mL) was treated with trifluoroacetic anhydride (222 mg, 1.05 mmol, 149 μL) at 20 ºC, and the resulting solution was maintained for 2 hr at that temperature. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was then concentrated under a positive pressure of nitrogen gas, and the residue purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 20:1). Concentration of fraction A (R_f = 0.3 in 20:1 v/v CH_2Cl_2:CH_3OH elution) afforded tris-TBS-compound **107** (11 mg, 15.9 μmol, 38%) as a clear, colourless oil. **^{1}H NMR** (599 MHz, Chloroform-*d*) δ 9.01 (s, 1H), 8.59 (s, 1H), 6.15 (d, *J* = 4.2 Hz, 1H), 6.06 (s, 1H), 4.59 (dd, *J* = 4.7, 3.7 Hz, 1H), 4.34 (dd, *J* = 4.7, 4.2 Hz, 1H), 4.27 (s, 2H), 4.19 – 4.15 (m, 1H), 4.04 (dd, *J* = 11.5, 3.3 Hz, 1H), 3.81 (dd, *J* = 11.5, 2.5 Hz, 1H), 0.95 (s, 11H), 0.93 (s, 10H), 0.83 (s, 11H), 0.14 (s, 4H), 0.14 (s, 4H), 0.11 (s, 3H), 0.10 (s, 4H), 0.01 (s, 3H), -0.11 (s, 3H); **^{13}C NMR** (151 MHz, Chloroform-*d*) δ 168.9, 155.2, 153.9, 152.6, 145.1, 143.7, 130.3, 89.1, 85.2, 76.4, 71.38, 62.2, 47.1, 26.3, 26.0, 25.9, 18.8, 18.3, 18.1, -4.1, -4.5, -4.6, -4.7, -5.1, -5.1; **IR** (ν_{max}) 2930, 2858, 1735, 1602, 1577, 1450; **HRMS** (ESI): calculated for $C_{31}H_{57}N_6O_6Si_3^+$ [*M* + H]$^+$: 693.3642; found 693.3666.

Concentration of fraction B (R_f = 0.2 in 20:1 v/v CH_2Cl_2:CH_3OH elution) afforded the alcohol **108** (10 mg, 14.4 μmol, 34%) as a clear, colourless oil. **^{1}H NMR** (599 MHz, Chloroform-*d*) δ 9.03 (s, 1H), 8.26 (s, 1H), 6.02 – 5.75 (m, 2H), 5.44 (d, *J* = 11.4 Hz, 1H), 5.00 (dd, *J* = 7.5, 4.5 Hz, 1H), 4.36 (d, *J* = 4.5 Hz, 1H), 4.28 (dd, *J* = 1.3 Hz, 2H), 4.23 – 4.19 (m, 1H), 3.98 (d, *J* = 13.0 Hz, 1H), 3.75 (dd, *J* = 12.1 Hz, 1H), 0.95 (s, 9H), 0.74 (s, 9H), 0.13 (s, 3H), 0.12 (s, 3H), -0.12 (s, 3H), -0.59 (s, 3H); **^{13}C NMR** (151 MHz, Chloroform-*d*) δ 168.2, 154.5, 152.9, 151.9, 146.2, 144.6, 131.1, 91.3, 89.2, 74.2, 73.5, 62.8, 46.8, 25.8, 25.6, 18.1, 17.8, -4.6, -4.6, -4.6, -5.8; **IR** (ν_{max}) 3323, 2930, 2857, 1790, 1732, 1602, 1578, 1449; **HRMS** (ESI): calculated for $C_{25}H_{43}N_6O_6Si_2^+$ [*M* + H]$^+$: 579.2777; found 579.2802.

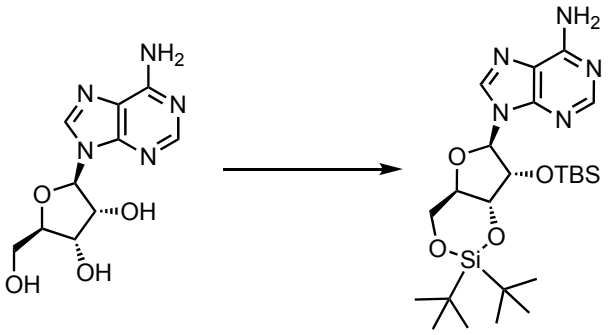
Triethylammonium 3-(purinosin-6-yl)-hydantoin-1-ide **(110)**



A 1.5 mL-safe lock tube containing a solution of compound **107** (25 mg, 36.1 μmol) was treated with *neat* trimethylamine trihydrofluoride (593 mg, 600 μL, 3.68 mmol) at 20 ºC and the reaction mixture was shaken at that temperature for 4 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction was cooled to 0 ºC and quenched by transferring the entire mixture to a separate 15 mL-Falcon tube containing methoxytrimethylsilane (4 mL), itself also maintained at 0 ºC. The mixture was then agitated using a vortex mixer, at which time a solid precipitated. The resulting suspension was then centrifuged, and the supernatant discarded. The solid residue was washed twice more with methoxytrimethylsilane (1 mL), and then dried overnight under high vacuum to afford compound **110** (11 mg, 24.4 μmol, 68%) as a pink/white powder. $\mathbf{R_f}$ = 0.05 in 19:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Deuterium Oxide) δ 8.98 (s, 1H), 8.78 (s, 1H), 6.24 (d, *J* = 5.2 Hz, 1H), 4.85 (dd, *J* = 5.2 Hz, 1H), 4.45 (dd, *J* = 4.8 Hz, 1H), 4.37 (s, 2H), 4.31 – 4.21 (m, 1H), 3.95 – 3.79 (m, 2H), 3.16 (q, *J* = 7.3 Hz, 6H), 1.24 (t, *J* = 7.3 Hz, 9H); **^{13}C NMR** (101 MHz, Deuterium Oxide)‡‡ δ 152.3, 146.6, 88.8, 73.9, 70.3, 47.0, 85.6, 61.2, 61.2, 46.8, 8.2; **IR** (ν_{max}) 3327, 2928, 1790, 1734, 1601, 1576, 1444; **HRMS** (ESI): calculated for $C_{13}H_{15}N_6O_6^+$ [*M* + H]$^+$: 351.1048; found 351.1103.

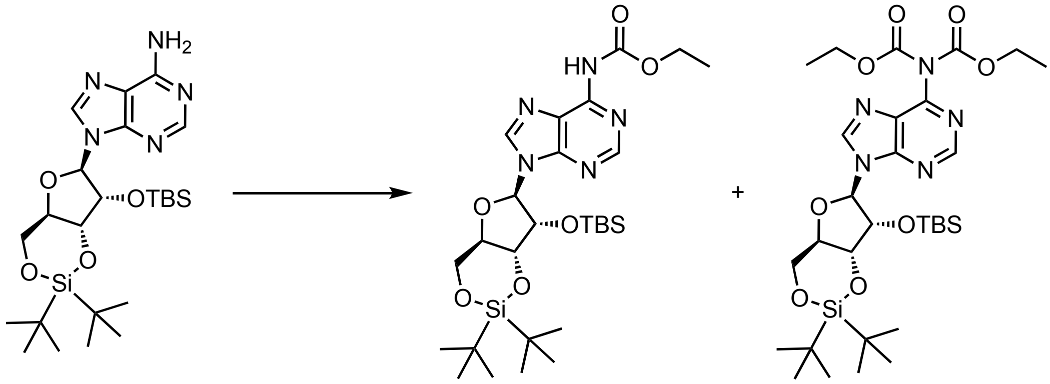
‡‡ ^{13}C NMR shifts were extracted from the 400/101 MHz HSQC NMR spectrum for this compound.

3',5'-*O*-[Bis(*tert*-butyl)silylene]-2'-*O*-(*tert*-butyldimethylsilyl)adenosine **(116)**



Compound **116** was synthesised according to a modified literature procedure.[134] A magnetically stirred solution containing adenosine (**33**) (2.00 g, 7.48 mmol) in anhydrous DMF (90 mL) was treated *dropwise* with di-*tert*-butyldichlorosilane (4.94 g, 4.90 mL, 23.2 mmol) at 0 ºC, and the resulting mixture was maintained at that temperature for 2 hr. Upon consumption of the starting material, as judged by TLC analysis, imidazole (4.08 g, 58,9 mmol) was added, and the reaction was maintained at 0 ºC for a further 5 minutes. The mixture was subsequently treated with *tert*-butyldimethylsilyl chloride (4.52 g, 30.0 mmol), and allowed to warm slowly to room temperature. After stirring for a further 8 hr, the mixture was then treated with H_2O (100 mL), the aqueous phase extracted with CH_2Cl_2 (3 x 200 mL). The combined organic phases were dried ($MgSO_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was purified by flash chromatography (silica gel, EtOAc:Hexane = 1:1) to afford compound **116** (3.25 g, 6.22 mmol, 83%) as a colourless foam. The ^1H and ^{13}C NMR spectral data for **116** were consistent with those previously reported in the literature.[134] $\mathbf{R_f}$ = 0.3 in 1:1 v/v EtOAc:Hexane elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 8.32 (s, 1H), 7.82 (s, 1H), 5.90 (s, 1H), 5.72 (s, 2H), 4.62 (d, *J* = 4.7 Hz, 1H), 4.54 (dd, *J* = 9.5, 4.7 Hz, 1H), 4.48 (dd, *J* = 9.1, 5.1 Hz, 1H), 4.21 (ddd, *J* = 10.5, 9.5, 5.1 Hz, 1H), 4.03 (dd, *J* = 10.5, 9.1 Hz, 1H), 1.08 (s, 9H), 1.04 (s, 9H), 0.93 (s, 9H), 0.16 (s, 3H), 0.14 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 155.6, 153.4, 149.5, 139.0, 120.5, 92.6, 76.0, 75.6, 74.9, 68.0, 27.7, 27.2, 26.1, 22.9, 20.5, 18.5, -4.1, -4.8; **IR** (ν_{max}) 3358, 3311, 3150, 2933, 1677, 1603, 1473; **HRMS** (ESI): calculated for $C_{24}H_{44}N_5O_4Si_2^+$ [*M* + H]$^+$: 522.2926; found 522.2927.

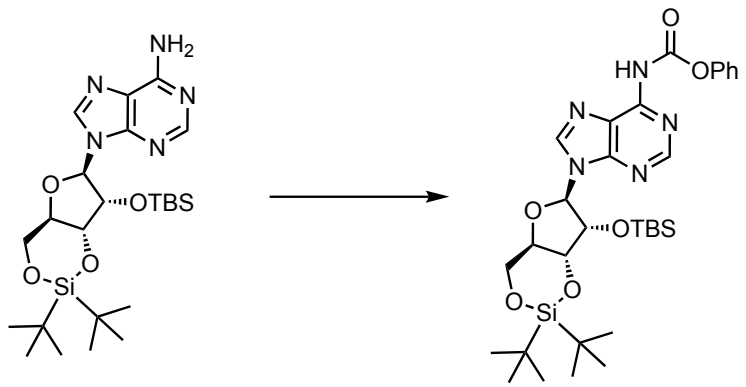
3',5'-*O*-[Bis(*tert*-butyl)silylene]-2'-*O*-(*tert*-butyldimethylsilyl)-N^6-(ethoxycarbonyl)adenosine **(118)** and 3',5'-*O*-[Bis(*tert*-butyl)silylene]-2'-*O*-(*tert*-butyldimethylsilyl)-N^6,N^6-bis(ethoxycarbonyl)adenosine **(117)**



A magnetically stirred solution containing compound (**116**) (1.80 g, 3.45 mmol) in anhydrous pyridine (50 mL) was treated *dropwise* with ethyl chloroformate (1.09 g, 960 µL, 10.1 mmol) at 0 ºC, and the resulting mixture was allowed to slowly warm to room temperature over 18 hr. Upon consumption of the starting material, as judged by TLC analysis, ethanol (3 mL) was added at 0 ºC, and the reaction was maintained at that temperature for a further 5 minutes. The reaction mixture was then concentrated under reduced pressure, and the residue purified by flash chromatography (silica gel, EtOAc:Hexane = 4:6). Concentration of fraction A (R_f = 0.5 in 4:6 v/v EtOAc:Hexane elution) afforded the bis-carbamate **117** (460 mg, 690 µmol, 20%) as a colourless foam. **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 8.86 (s, 1H), 8.11 (s, 1H), 6.00 (s, 1H), 4.64 (d, *J* = 4.6 Hz, 1H), 4.55 – 4.43 (m, 2H), 4.36 – 4.22 (m, 5H), 4.05 (dd, *J* = 10.4, 9.2 Hz, 1H), 1.24 (t, *J* = 7.1 Hz, 6H), 1.07 (s, 9H), 1.05 (s, 9H), 0.95 (s, 9H), 0.19 (s, 3H), 0.17 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 152.7, 152.4, 152.0, 149.7, 143.4, 130.0, 92.8, 76.0, 75.7, 75.0, 67.9, 64.1, 27.6, 27.2, 26.1, 22.9, 20.5, 18.5, 14.2, -4.1, -4.8; **IR** (ν_{max}) 2933, 2859, 1798, 1767, 1738, 1601, 1577; **HRMS** (ESI): calculated for $C_{30}H_{52}N_5O_8Si_2^+$ $[M + H]^+$: 666.3349; found 666.3354.

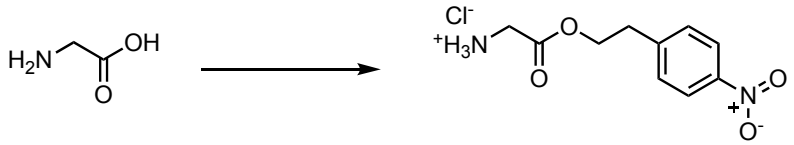
Concentration of fraction B (R_f = 0.3 in 4:6 v/v EtOAc:Hexane elution) afforded the mono-carbamate **118** (615 mg, 1.04 mmol, 30%) as a colourless foam. **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 8.72 (s, 1H), 8.00 (s, 1H), 5.97 (s, 1H), 4.61 (d, *J* = 4.6 Hz, 1H), 4.49 (dd, *J* = 9.2, 5.1 Hz, 1H), 4.46 (dd, *J* = 9.6, 4.6 Hz, 1H), 4.33 (q, *J* = 7.1 Hz, 2H), 4.23 (ddd, *J* = 10.1, 5.1 Hz, 1H), 4.03 (dd, *J* = 10.1, 9.2 Hz, 1H), 1.35 (t, *J* = 7.1 Hz, 3H), 1.07 (s, 9H), 1.04 (s, 9H), 0.93 (s, 9H), 0.16 (s, 3H), 0.14 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 153.1, 151.4, 150.8, 149.9, 141.0, 122.8, 92.6, 76.0, 75.6, 74.9, 67.9, 62.4, 27.6, 27.2, 26.1, 22.9, 20.5, 18.5, 14.6, -4.1, -4.8; **IR** (ν_{max}) 2932, 2858, 1746, 1610, 1587, 1471; **HRMS** (ESI): calculated for $C_{27}H_{48}N_5O_6Si_2^+$ $[M + H]^+$: 594.3138; found 594.3142.

3',5'-*O*-[Bis(*tert*-butyl)silylene]-2'-*O*-(*tert*-butyldimethylsilyl)-N^6-(phenoxycarbonyl)adenosine **(119)**



A magnetically stirred solution containing compound **116** (2.00 g, 3.83 mmol) in anhydrous CH_2Cl_2 (60 mL) was treated with 3-methyl-1-(phenoxycarbonyl)-1*H*-imidazol-3-ium chloride (**70**) (1.37 g, 5.75 mmol) at 20 ºC, and the resulting milky-white suspension was maintained for 3 hr at that temperature. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was concentrated under reduced pressure. The residue thus obtained was purified by flash chromatography (silica gel, EtOAc:Hexane = 1:4) to afford compound **119** (2.02 g, 3.14 mmol, 82%) as a colourless foam. $\mathbf{R}_f$ = 0.3 in 1:3 v/v EtOAc:Hexane elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 8.93 (s, 1H), 8.78 (s, 1H), 8.05 (s, 1H), 7.46 – 7.35 (m, 2H), 7.29 – 7.21 (m, 3H), 5.99 (s, 1H), 4.62 (d, *J* = 4.6 Hz, 1H), 4.54 – 4.41 (m, 2H), 4.24 (ddd, *J* = 10.0, 5.0 Hz, 1H), 4.02 (dd, *J* = 10.0 Hz, 1H), 1.07 (s, 9H), 1.05 (s, 9H), 0.93 (s, 9H), 0.17 (s, 3H), 0.15 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 153.1, 150.9, 150.4, 149.4, 149.4, 141.3, 129.6, 126.2, 123.1, 121.6, 92.7, 76.1, 75.7, 74.9, 67.9, 27.6, 27.2, 26.1, 22.9, 20.5, 18.5, -4.1, -4.8; **IR** (ν_{max}) 3333, 2932, 2858 1772, 1641, 1617, 1593, 1470; **HRMS** (ESI): calculated for $C_{31}H_{46}N_5O_6Si_2^-$ [*M* – H]$^-$: 640.2992; found 640.2991.

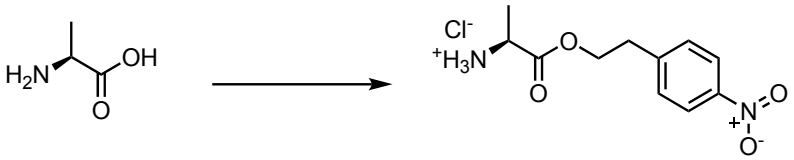
2-(4-Nitrophenethoxy)-2-oxoethan-1-aminium chloride (**124**)



Compound **124** was synthesised according to a modified literature procedure.[135] A 250 mL round-bottom flask fitted to a Dean-Stark apparatus, was charged with glycine (1.94 g, 25.8 mmol), toluene (50 mL), 2-(4-nitrophenyl)ethan-1-ol (6.5 g, 38.9 mmol), and *p*-toluenesulfonic acid (7.39 g, 42.9 mmol) at 20 ºC, in that order. The resulting vigorously-stirred mixture was then heated to reflux, and maintained at that temperature for 18 hr. Upon completion of the reaction, as noted by the appearance of immiscible water in the side-arm and distillation trap, the reaction mixture was cooled to room

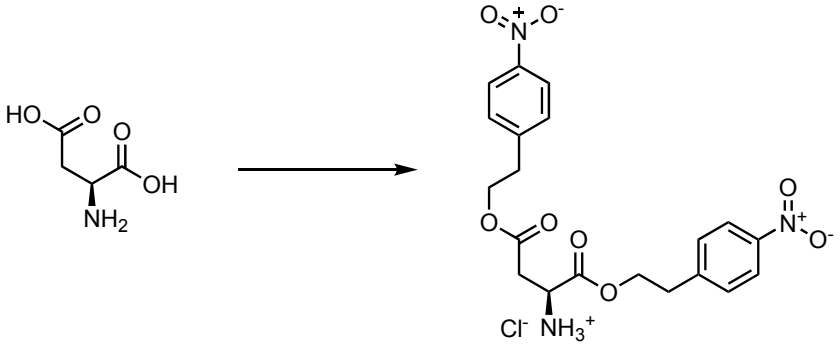
temperature. The mixture was then treated with $NaHCO_3$ (500 mL of a sat. *aq.* solution) in three equal portions, and the mixture was shaken vigorously to fully dissolve the solid material. The aqueous phase was extracted with EtOAc (4 x 100 mL), and the combined organic phases dried ($MgSO_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was then re-dissolved in the minimum amount of EtOAc (10 mL), and the product was precipitated *via* the dropwise addition of 1 mL conc. HCl at 0 ºC, to afford compound **124** (2.02 g, 3.14 mmol, 82%) as pale-yellow crystals. The ^{1}H NMR spectral data for **124** were consistent with those previously reported in the literature.[135] $\mathbf{R_f}$ = 0.2 in 96:4 v/v CH_2Cl_2:CH_3OH elution (free-base); **^{1}H NMR** (400 MHz, DMSO-d_6) δ 8.24 – 8.14 (m, 2H), 8.07 (s, 3H), 7.65 – 7.52 (m, 2H), 4.44 (t, J = 6.4 Hz, 2H), 3.76 (s, 2H), 3.09 (t, J = 6.4 Hz, 2H).

(*S*)-1-(4-Nitrophenethoxy)-1-oxopropan-2-aminium chloride **(125)**



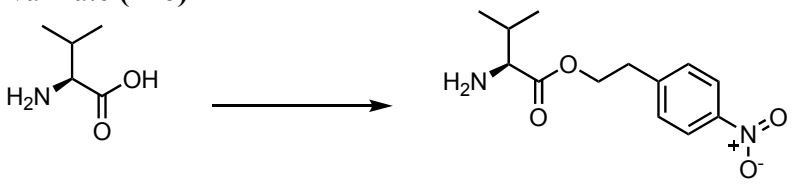
A 250 mL round-bottom flask fitted to a Dean-Stark apparatus, was charged with L-alanine (2.00 g, 22.5 mmol), toluene (50 mL), 2-(4-nitrophenyl)ethan-1-ol (5.63 g, 33.67 mmol), and *p*-toluenesulfonic acid (6.42 g, 37.3 mmol) at 20 ºC, in that order. The resulting vigorously-stirred mixture was then heated to reflux, and maintained at that temperature for 18 hr. Upon completion of the reaction, as noted by the appearance of immiscible water in the side-arm and distillation trap, the reaction mixture was cooled to room temperature. The mixture was then treated with $NaHCO_3$ (500 mL of a sat. *aq.* solution) in three equal portions, and the mixture was shaken vigorously to fully dissolve the solid material. The aqueous phase was extracted with EtOAc (4 x 100 mL), and the combined organic phases dried ($MgSO_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was then re-dissolved in the minimum amount of EtOAc (10 mL), and the product was precipitated *via* the dropwise addition of 2 mL conc. HCl at 0 ºC, to afford compound **125** (5.58 g, 20.3 mmol, 90%) as pale-yellow crystals. $\mathbf{R_f}$ = 0.2 in 96:4 v/v CH_2Cl_2:CH_3OH elution (free-base); **^{1}H NMR** (400 MHz, DMSO-d_6) δ 8.60 (s, 3H), 8.25 – 8.06 (m, 2H), 7.70 – 7.47 (m, 2H), 4.52 – 4.33 (m, 2H), 4.00 (q, J = 7.2 Hz, 1H), 3.10 (dd, J = 6.3 Hz, 2H), 1.33 (d, J = 7.2 Hz, 3H); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 170.0, 146.4, 146.3, 130.4, 123.5, 65.3, 47.8, 33.9, 15.7; **IR** (ν_{max}) 3400-2700 (br), 1731, 1598, 1510, 1344; **HRMS** (ESI): calculated for $C_{11}H_{15}N_2O_4^+$ [M + H]$^+$: 239.1026; found 239.1023.

(*S*)-1,4-Bis(4-nitrophenethoxy)-1,4-dioxobutan-2-aminium chloride **(127)**



A 250 mL round-bottom flask fitted to a Dean-Stark apparatus, was charged with L-aspartic acid (2.00 g, 15.0 mmol), toluene (50 mL), 2-(4-nitrophenyl)ethan-1-ol (7.54 g, 45.1 mmol), and *p*-toluenesulfonic acid (4.30 g, 24.9 mmol) at 20 ºC, in that order. The resulting vigorously-stirred mixture was then heated to reflux, and maintained at that temperature for 18 hr. Upon completion of the reaction, as noted by the appearance of immiscible water in the side-arm and distillation trap, the reaction mixture was cooled to room temperature. The mixture was then treated with $NaHCO_3$ (500 mL of a sat. *aq.* solution) in three equal portions, and the mixture was shaken vigorously to fully dissolve the solid material. The aqueous phase was extracted with EtOAc (2 x 100 mL), and then CH_2Cl_2 (2 x 100 mL), and the combined organic phases were dried ($MgSO_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was then re-dissolved in the minimum amount of EtOAc (10 mL), and the product was precipitated *via* the dropwise addition of 2 mL conc. HCl at 0 ºC, to afford compound **127** (6.74 g, 14.4 mmol, 96%) as pale-yellow crystals. $\mathbf{R}_f$ = 0.3 in 96:4 v/v CH_2Cl_2:CH_3OH elution (free-base); **^{1}H NMR** (400 MHz, DMSO-d_6) δ 8.61 (s, 3H), 8.26 – 8.05 (m, 4H), 7.66 – 7.42 (m, 4H), 4.37 (dd, J = 6.3 Hz, 2H), 4.31 – 4.19 (m, 3H), 3.10 – 2.97 (m, 4H), 2.94 – 2.86 (m, 2H); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 169.0, 168.2, 146.3, 146.3, 146.2, 130.3, 130.3, 130.2, 123.5, 123.4, 65.6, 64.6, 48.3, 34.1, 33.8, 33.7; **IR** (ν_{max}) 3400-2700 (br), 1742, 1720, 1601, 1533, 1516, 1349; **HRMS** (ESI): calculated for $C_{20}H_{22}N_3O_8^+$ [M + H]$^+$: 432.1401; found 432.1392.

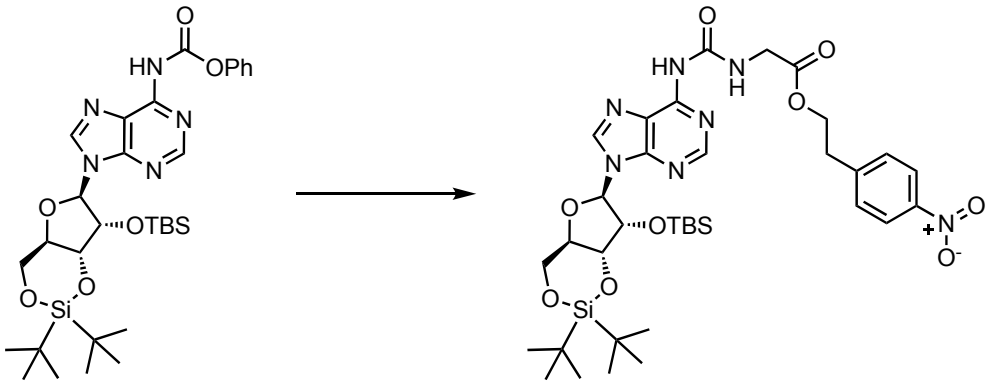
4-Nitrophenethyl L-valinate **(126)**



A 250 mL round-bottom flask fitted to a Dean-Stark apparatus, was charged with L-valine (2.00 g, 17.1 mmol), toluene (50 mL), 2-(4-nitrophenyl)ethan-1-ol (4.28 g, 25.6 mmol), and *p*-toluenesulfonic acid (4.88 g, 28.3 mmol) at 20 ºC, in that order. The resulting vigorously-stirred mixture was then

heated to reflux, and maintained at that temperature for 18 hr. Upon completion of the reaction, as noted by the appearance of immiscible water in the side-arm and distillation trap, the reaction mixture was cooled to room temperature. The mixture was then treated with $NaHCO_3$ (500 mL of a sat. *aq.* solution) in three equal portions, and the mixture was shaken vigorously to fully dissolve the solid material. The aqueous phase was extracted with EtOAc (2 x 100 mL), and then CH_2Cl_2 (2 x 100 mL), and the combined organic phases were dried ($MgSO_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2 → 96:4) to afford compound **126** (3.94 g, 14.8 mmol, 87%) as a pale-yellow solid. $\mathbf{R}_f$ = 0.3 in 96:4 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 8.34 – 8.03 (m, 2H), 7.49 – 7.36 (m, 2H), 4.50 – 4.31 (m, 2H), 3.25 (d, *J* = 5.0 Hz, 1H), 3.08 (dd, *J* = 6.7 Hz, 2H), 1.99 – 1.87 (m, 1H), 0.91 (d, *J* = 6.7 Hz, 3H), 0.81 (d, *J* = 6.7 Hz, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 147.1, 145.7, 129.9, 124.0, 64.2, 60.1, 35.1, 32.3, 19.4, 17.2; **IR** (ν_{max}) 3388, 2961, 2873, 1729, 1600, 1516, 1466; **HRMS** (ESI): calculated for $C_{13}H_{19}N_2O_4^+$ $[M + H]^+$: 267.1339; found 267.1338.

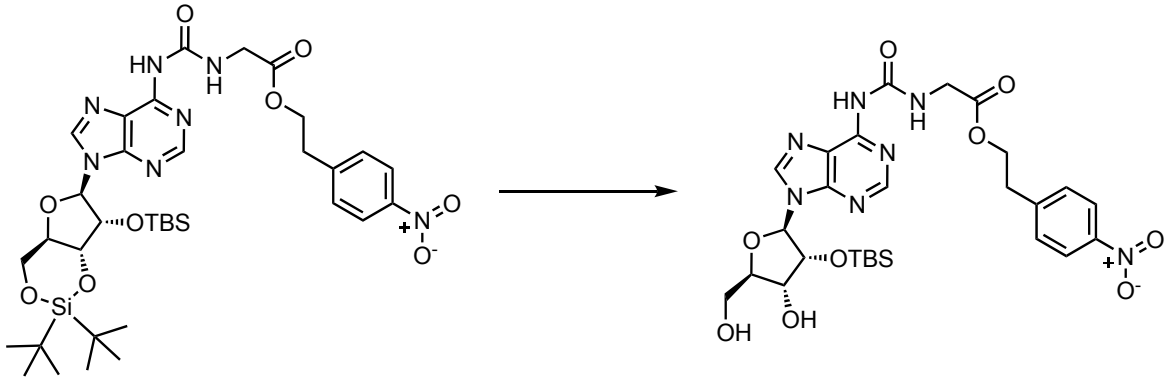
4-Nitrophenethyl ({3',5'-*O*-[bis(*tert*-butyl)silylene]-2'-*O*-[*tert*-butyldimethylsilyl]adenosin-N^6-yl}carbonyl)glycinate **(120)**



A magnetically stirred solution containing compound **119** (4.00 g, 6.23 mmol) in anhydrous pyridine (50 mL) was treated with compound **124** (1.95 g, 7.48 mmol) at 20 ºC, and the resulting mixture was heated to 55 ºC and maintained at that temperature for 16 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was cooled to room temperature. The cooled reaction mixture was then concentrated under reduced pressure, and the residue purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2) to afford compound **120** (3.35 g, 4.34 mmol, 70%) as a colourless foam. $\mathbf{R}_f$ = 0.2 in 98:2 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 10.00 (dd, *J* = 5.6 Hz, 1H), 8.78 (s, 1H), 8.50 (s, 1H), 8.23 (s, 1H), 8.13 – 8.05 (m, 2H), 7.42 – 7.34 (m, 2H), 6.00 (s, 1H), 4.59 (d, *J* = 4.6 Hz, 1H), 4.53 – 4.41 (m, 4H), 4.22 (dd, *J* = 21.3, 5.3 Hz, 3H), 4.07 (dd, *J* = 10.5, 9.2 Hz, 1H), 3.09 (dd, *J* = 6.6 Hz, 2H), 1.08 (s,

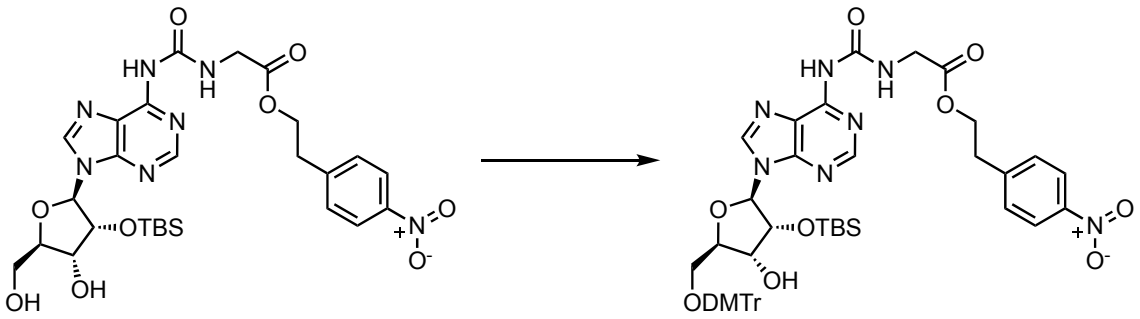
9H), 1.04 (s, 9H), 0.94 (s, 9H), 0.17 (s, 3H), 0.15 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 170.0, 154.4, 151.2, 150.3, 149.9, 147.0, 145.5, 141.7, 129.9, 123.9, 121.1, 92.5, 76.0, 75.8, 75.0, 67.9, 64.8, 42.3, 35.0, 27.7, 27.2, 26.1, 22.9, 20.5, 18.5, -4.1, -4.8; **IR** (ν_{max}) 3233, 2932, 2857, 1749, 1702, 1611, 1587, 1519, 1468; **HRMS** (ESI): calculated for $C_{35}H_{54}N_7O_9Si_2^+$ [*M* + H]$^+$: 772.3516; found 772.3521.

4-Nitrophenethyl {[2'-*O*-(*tert*-butyldimethylsilyl)adenosin-*N*^{6}-yl]carbonyl}glycinate **(128)**



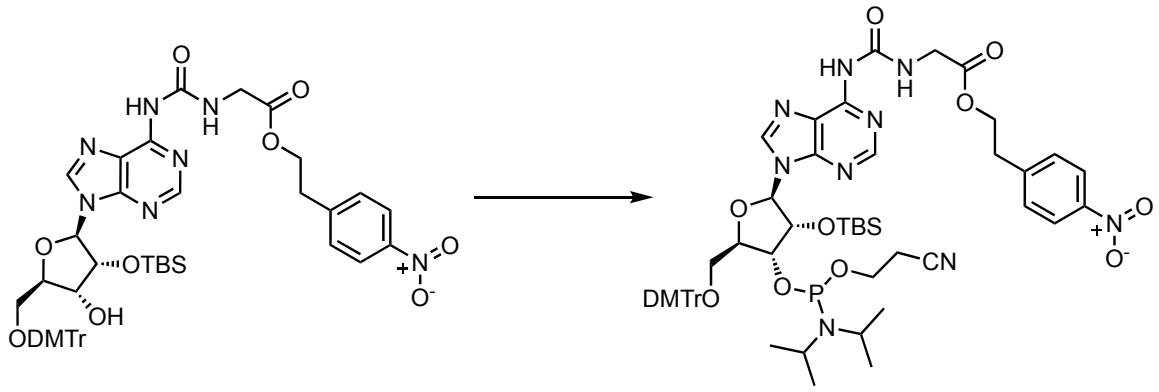
A 50 mL-falcon tube containing a solution of compound **120** (3.35 g, 4.34 mmol) in CH_2Cl_2 (20 mL) was treated at 0 °C with pyridine (0.91 mL) followed by HF-pyridine (70% w/w, 563 μL, 619 mg, 21.6 mmol) and the resulting mixture was maintained at that temperature with stirring for 3.5 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction was then quenched by dropwise addition of methoxytrimethylsilane (8 mL), itself also maintained at 0 ºC. After stirring for an additional 20 min, the reaction mixture was concentrated under reduced pressure, and the residue purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2→ 97:3) to afford compound **128** (2.38 g, 3.77 mmol, 87%) as a colourless foam. $\mathbf{R_f}$ = 0.3 in 97:3 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 9.91 (dd, *J* = 5.6 Hz, 1H), 8.86 (s, 1H), 8.54 (s, 1H), 8.21 (s, 1H), 8.17 – 8.10 (m, 2H), 7.43 – 7.36 (m, 2H), 5.92 – 5.76 (m, 2H), 5.07 (dd, *J* = 7.2, 4.7 Hz, 1H), 4.44 (dd, *J* = 6.7 Hz, 2H), 4.39 – 4.32 (m, 2H), 4.27 – 4.12 (m, 2H), 4.01 – 3.92 (m, 1H), 3.09 (dd, *J* = 6.7 Hz, 2H), 2.92 (s, 1H), 0.79 (s, 9H), -0.19 (s, 3H), -0.38 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 170.0, 154.0, 150.9, 150.9, 149.5, 147.1, 145.4, 143.6, 129.9, 124.0, 122.1, 91.3, 87.7, 74.8, 72.9, 64.9, 63.4, 42.2, 35.0, 25.7, 18.0, -5.2, -5.3; **IR** (ν_{max}) 3242, 2927, 2855, 1747, 1696, 1610, 1588, 1540, 1517, 1469; **HRMS** (ESI): calculated for $C_{27}H_{38}N_7O_9Si^+$ [*M* + H]$^+$: 632.2495; found 632.2492.

4-Nitrophenethyl {[5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-(*tert*-butyldimethylsilyl)adenosin-N^6-yl]carbonyl}glycinate **(132)**



A magnetically stirred solution containing compound **128** (1.87 g, 2.96 mmol) in anhydrous pyridine (30 mL) was treated with 4,4'-dimethoxytrityl chloride (1.50 g, 4.44 mmol) at 0 ºC, and the resulting pale-orange/yellow solution was maintained at that temperature for 4 hr. Upon consumption of the starting material, as judged by TLC analysis, the mixture was concentrated under reduced pressure. Excess pyridine was removed *via* azeotropic coevaporation using toluene. The residue thus obtained was purified by flash chromatography (silica gel, 0.1% pyridine, CH_2Cl_2:CH_3OH = 100:0 → 99:1) to afford compound **132** (2.31 g, 2.47 mmol, 83%) as a colourless foam. $\mathbf{R}_f$ = 0.2 in 99:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 9.89 (dd, *J* = 5.6 Hz, 1H), 8.46 (s, 1H), 8.17 (s, 1H), 8.16 – 8.11 (m, 2H), 7.88 (s, 1H), 7.48 – 7.21 (m, 11H), 6.84 – 6.79 (m, 4H), 6.07 (d, *J* = 5.2 Hz, 1H), 4.96 (dd, *J* = 5.1 Hz, 1H), 4.45 (dd, *J* = 6.7 Hz, 2H), 4.41 – 4.33 (m, 1H), 4.30 – 4.24 (m, 1H), 4.19 (d, *J* = 5.6 Hz, 2H), 3.80 – 3.78 (m, 6H), 3.53 (dd, *J* = 10.7, 2.9 Hz, 1H), 3.39 (dd, *J* = 10.7, 3.7 Hz, 1H), 3.10 (dd, *J* = 6.7 Hz, 2H), 2.68 (d, *J* = 4.2 Hz, 1H), 0.84 (s, 9H), 0.00 (s, 3H), -0.14 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 170.0, 158.7, 154.2, 151.2, 150.5, 150.2, 147.0, 145.5, 144.7, 142.0, 136.1, 135.7, 130.2, 129.9, 128.3, 128.0, 127.1, 120.9, 113.3, 88.6, 86.8, 84.3, 75.8, 71.5, 64.8, 63.4, 55.3, 42.2, 34.9, 25.7, 18.0, -4.8, -5.1; **IR** (ν_{max}) 3232, 3059, 2930, 2856, 1748, 1701, 1608, 1509, 1468; **HRMS** (ESI): calculated for $C_{48}H_{56}N_7O_{11}Si^+$ [*M* + H]$^+$: 934.3802; found 934.3797.

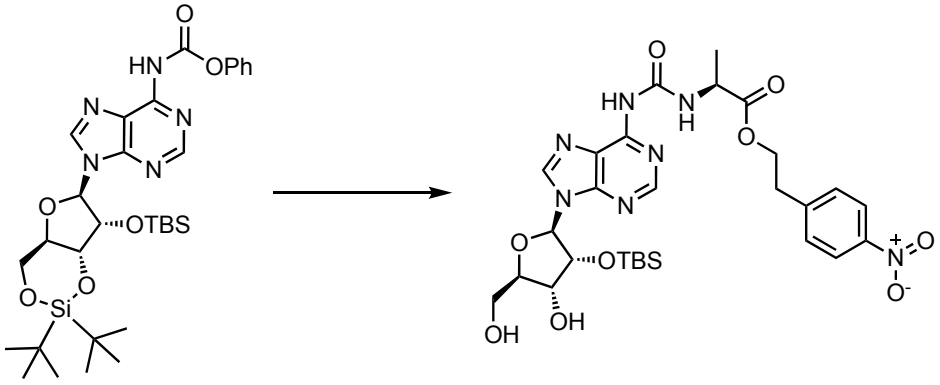
4-Nitrophenethyl {[5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-(*tert*-butyldimethylsilyl)adenosin-N^6-yl]carbonyl}glycinate, 3'-[2-cyanoethyl *N,N*-bis(1-methylethyl)phosphoramidite] (**136**)



A magnetically stirred solution containing compound **132** (200 mg, 214 μmol) in anhydrous CH_2Cl_2 (7 mL) was treated with *N,N*-diisopropylethylamine (42 mg, 56 μL, 321 μmol), followed by 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (86 mg, 81 μL, 364 μmol) at 0 ºC, and the resulting pale-yellow solution was stirred at room temperature for 1.5 hr. Upon consumption of the starting material, as judged by TLC analysis, the mixture was carefully concentrated under a blanket of nitrogen gas. The residue thus obtained was purified by flash chromatography (HPLC-grade solvent,[§§] silica gel, 0.1% pyridine, EtOAc:Hexane = 7:3) to afford compound **136** (148 mg, 131 μmol, 61%) as a colourless oil. A dry, white-powder was obtained *via* the sublimation of a solution of **136** in frozen benzene (6 mL) under high-vacuum. $\mathbf{R_f}$ = 0.3 in 1:1 v/v EtOAc:Hexane elution; **^{1}H NMR** (400 MHz, Benzene-d_6) diastereomeric mixture δ 9.84 – 9.74 (m, 2H), 8.42 – 8.35 (m, 2H), 8.24 (s, 2H), 8.21 – 8.14 (m, 2H), 7.78 – 7.68 (m, 7H), 7.61 – 7.18 (m, 6H), 7.12 – 7.05 (m, 1H), 6.83 – 6.51 (m, 9H), 6.27 – 6.15 (m, 2H), 5.35 – 5.24 (m, 2H), 4.79 – 4.56 (m, 4H), 3.97 – 3.76 (m, 10H), 3.66 – 3.04 (m, 18H), 1.21 – 1.00 (m, 24H), 0.97 – 0.87 (m, 18H), 0.20 – 0.09 (m, 6H), 0.06 – -0.02 (m, 6H); **^{31}P NMR** (162 MHz, Benzene-d_6) diastereomeric mixture δ 150.5, 149.2; **IR** (ν_{max}) 3233, 2930, 1747, 1704, 1608, 1587, 1509, 1467, 1345; **HRMS** (ESI): calculated for $C_{57}H_{73}N_9O_{12}PSi^+$ [*M* + H]$^+$: 1134.4880; found 1134.4909.

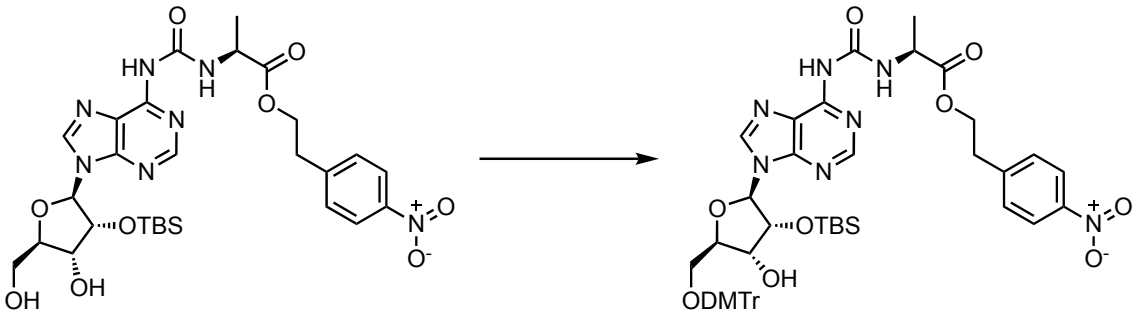
[§§] Decomposition to the corresponding H-phosphonate was observed, except when HPLC-grade solvents were used. Exclusion of oxygen or H_2O during purification meanwhile gave no improvement to the reaction outcome. Purification was also possible, for example, when an aqueous extraction was performed.

4-Nitrophenethyl {[2'-*O*-(*tert*-butyldimethylsilyl)adenosin-N^6-yl]carbonyl}-L-alaninate **(129)**



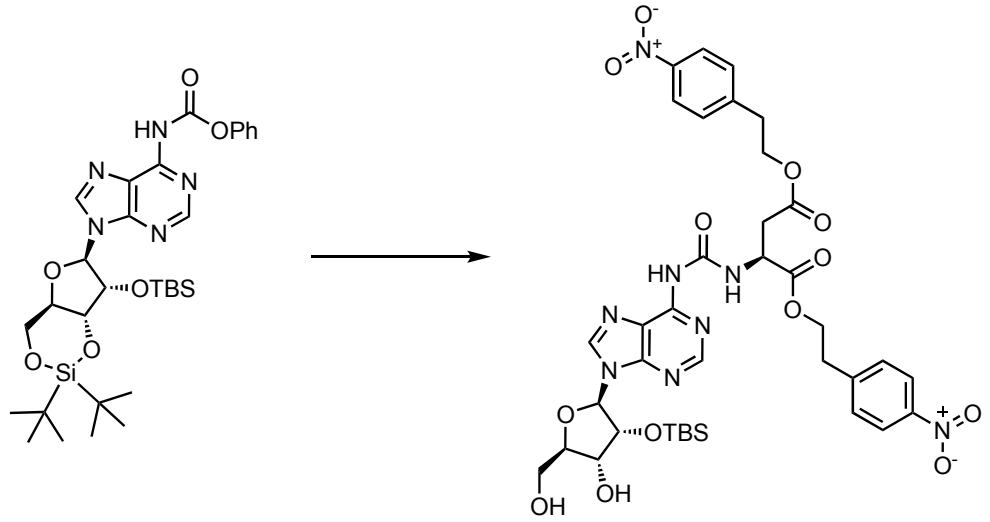
A magnetically stirred solution containing compound **119** (689 mg, 1.07 mmol) in anhydrous pyridine (10 mL) was treated with compound **125** (590 mg, 2.15 mmol) at 20 ºC, and the resulting mixture was heated to 55 ºC and maintained at that temperature for 16 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was cooled to room temperature. The cooled reaction mixture was then concentrated under reduced pressure, and the residue subjected to flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2). The mixture obtained was used directly in the next reaction without further purification. A 15 mL-Falcon tube containing the residue obtained in the previous reaction, suspended in CH_2Cl_2 (5 mL) was treated at 0 °C with pyridine (230 µL) followed by HF-pyridine (70% w/w, 111 µL, 122 mg, 4.26 mmol) and the resulting mixture was maintained at that temperature with stirring for 3.5 hr. The reaction was then quenched by dropwise addition of methoxytrimethylsilane (2 mL), itself also maintained at 0 ºC. After stirring for an additional 20 min, the reaction mixture was concentrated under reduced pressure, and the residue purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2→ 97:3) to afford compound **129** (408 mg, 785 µmol, 73% over two steps) as a colourless foam. $\mathbf{R_f}$ = 0.2 in 97:3 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 9.82 (d, *J* = 7.1 Hz, 1H), 8.56 (s, 1H), 8.29 (s, 1H), 8.19 – 8.12 (m, 2H), 8.06 (s, 1H), 7.44 – 7.39 (m, 2H), 5.88 (dd, *J* = 12.0, 2.1 Hz, 1H), 5.83 (d, *J* = 7.3 Hz, 1H), 5.09 (dd, *J* = 7.3, 4.8 Hz, 1H), 4.68 – 4.57 (m, 1H), 4.49 – 4.42 (m, 2H), 4.40 – 4.34 (m, 2H), 4.01 – 3.91 (m, 1H), 3.85 – 3.71 (m, 1H), 3.11 (dd, *J* = 6.6 Hz, 2H), 2.85 (s, 1H), 1.51 (d, *J* = 7.2 Hz, 3H), 0.80 (s, 9H), -0.18 (s, 3H), -0.39 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 173.1, 153.0, 151.1, 150.9, 149.4, 147.1, 145.5, 143.2, 130.0, 124.0, 122.2, 91.4, 87.8, 74.7, 72.9, 64.9, 63.5, 49.4, 35.0, 25.7, 18.6, 18.0, -5.2, -5.2; **IR** (ν_{max}) 3238, 2930, 2857, 1742, 1649, 1612, 1588, 1518, 1469, 1344; **HRMS** (ESI): calculated for $C_{28}H_{40}N_7O_9Si^+$ [*M* + H]$^+$: 646.2651; found 646.2648.

4-Nitrophenethyl {[5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-(*tert*-butyldimethylsilyl)adenosin-N^6-yl]carbonyl}-L-alaninate (**133**)



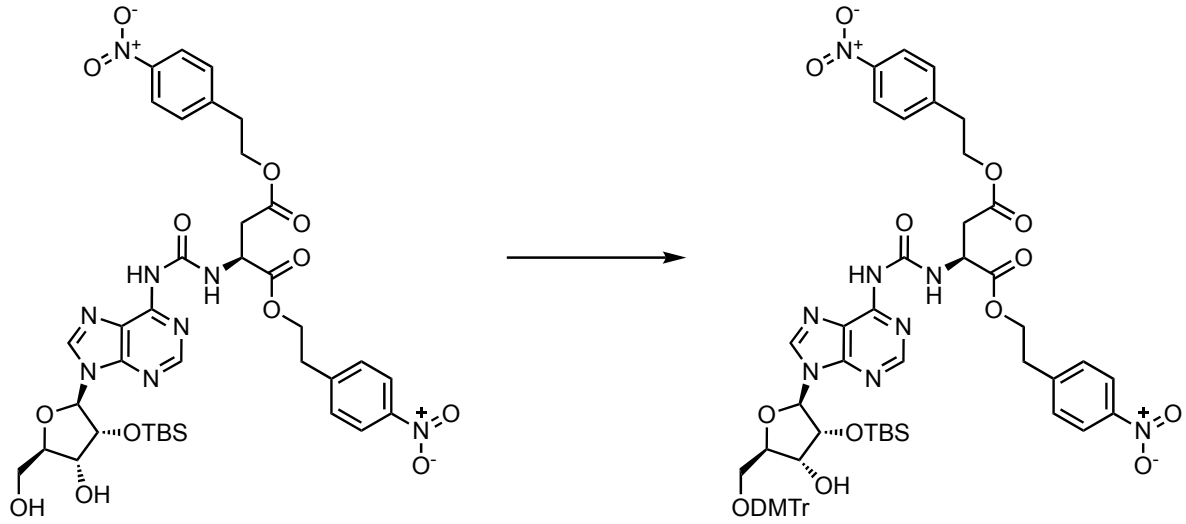
A magnetically stirred solution containing compound **129** (123 mg, 190 μmol) in anhydrous pyridine (8 mL) was treated with 4,4'-dimethoxytrityl chloride (129 mg, 381 μmol) at 0 ºC, and the resulting pale-orange/yellow solution was maintained at that temperature for 1.5 hr. Upon consumption of the starting material, as judged by TLC analysis, the mixture was concentrated under reduced pressure. Excess pyridine was removed *via* azeotropic coevaporation using toluene. The residue thus obtained was purified by flash chromatography (silica gel, 0.1% pyridine, CH_2Cl_2:CH_3OH = 100:0 → 99:1) to afford compound **133** (165 mg, 174 μmol, 91%) as a colourless foam. $\mathbf{R}_f$ = 0.2 in 99:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (599 MHz, Chloroform-*d*) δ 9.86 (d, *J* = 7.2 Hz, 1H), 8.46 (s, 1H), 8.16 (s, 1H), 8.13 – 8.10 (m, 2H), 7.91 (s, 1H), 7.46 – 7.37 (m, 4H), 7.36 – 7.31 (m, 4H), 7.30 – 7.25 (m, 2H), 7.24 – 7.20 (m, 1H), 6.83 – 6.79 (m, 4H), 6.06 (d, *J* = 5.2 Hz, 1H), 5.00 (dd, *J* = 5.2 Hz, 1H), 4.66 – 4.59 (m, 1H), 4.47 – 4.41 (m, 2H), 4.39 – 4.35 (m, 1H), 4.27 (ddd, *J* = 3.5 Hz, 1H), 3.53 (dd, *J* = 10.7, 3.1 Hz, 1H), 3.39 (dd, *J* = 10.8, 3.7 Hz, 1H), 3.10 (dd, *J* = 6.6 Hz, 2H), 2.68 (d, *J* = 4.2 Hz, 1H), 1.50 (d, *J* = 7.3 Hz, 3H), 0.85 (s, 9H), 0.00 (s, 3H), -0.14 (s, 3H); **^{13}C NMR** (151 MHz, Chloroform-*d*) δ 172.8, 158.6, 153.0, 151.2, 150.3, 150.0, 149.9, 146.9, 145.3, 144.6, 141.5, 135.6, 130.1, 129.8, 128.1, 127.9, 127.0, 123.7, 120.8, 113.2, 88.5, 86.7, 84.3, 75.7, 71.4, 64.6, 63.3, 55.2, 49.1, 34.8, 25.5, 18.4, 17.9, -5.0, -5.2; **IR** (ν_{max}) 3235, 2930, 2856, 1742, 1700, 1608, 1587, 1519, 1468, 1344; **HRMS** (ESI): calculated for $C_{49}H_{58}N_7O_{11}Si^+$ [*M* + H]$^+$: 948.3958; found 948.3971.

Bis(4-nitrophenethyl) {[2'-*O*-(*tert*-butyldimethylsilyl)adenosin-N^6-yl]carbonyl}-L-aspartate (**131)**



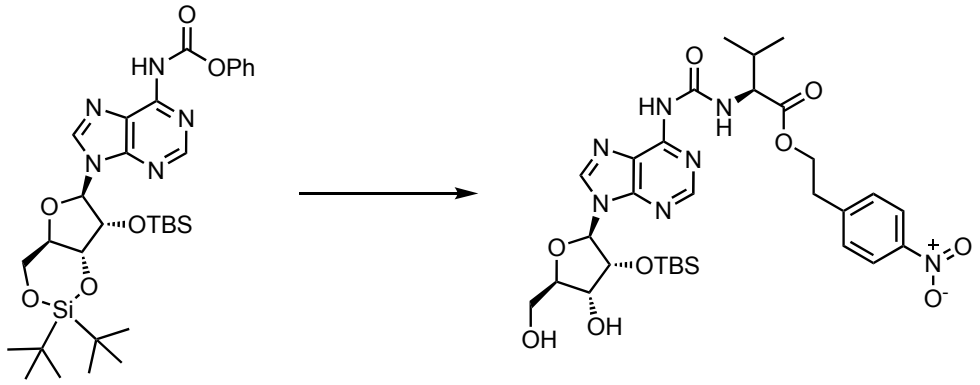
A magnetically stirred solution containing compound **119** (2.46 g, 3.83 mmol) in anhydrous pyridine (35 mL) was treated with compound **127** (2.69 g, 5.75 mmol) at 20 ºC, and the resulting mixture was heated to 55 ºC and maintained at that temperature for 16 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was cooled to room temperature. The cooled reaction mixture was then concentrated under reduced pressure, and the residue subjected to flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2). The mixture obtained was used directly in the next reaction without further purification. A 15 mL-Falcon tube containing the residue obtained in the previous reaction, suspended in CH_2Cl_2 (15 mL) was treated at 0 °C with pyridine (450 µL) followed by HF-pyridine (70% w/w, 278 µL, 306 mg, 10.7 mmol) and the resulting mixture was maintained at that temperature with stirring for 3.5 hr. The reaction was then quenched by dropwise addition of methoxytrimethylsilane (4 mL), itself also maintained at 0 ºC. After stirring for an additional 20 min, the reaction mixture was concentrated under reduced pressure, and the residue purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2→ 97:3) to afford compound **131** (1.49 g, 1.77 mmol, 46% over two steps) as a colourless foam. $\mathbf{R_f}$ = 0.2 in 97:3 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 10.25 (d, *J* = 7.9 Hz, 1H), 8.70 (s, 1H), 8.55 – 8.45 (m, 1H), 8.17 (s, 1H), 8.15 – 8.02 (m, 4H), 7.41 – 7.30 (m, 4H), 5.85 (d, *J* = 7.2 Hz, 1H), 5.82 – 5.74 (m, 1H), 5.12 – 5.03 (m, 1H), 4.96 – 4.85 (m, 1H), 4.51 – 4.23 (m, 6H), 4.02 – 3.92 (m, 1H), 3.77 (dd, *J* = 12.3 Hz, 1H), 3.13 – 2.93 (m, 6H), 2.89 (s, 1H), 0.79 (s, 9H), -0.17 (s, 3H), -0.38 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 170.7, 170.7, 153.4, 150.9, 150.8, 149.5, 147.0, 145.5, 145.4, 143.6, 129.9, 129.9, 123.9, 123.9, 122.1, 91.4, 87.7, 74.8, 72.8, 65.3, 64.6, 63.4, 49.9, 36.6, 34.9, 34.9, 26.0, 25.7, 18.0, -5.2, -5.2; **IR** (ν_{max}) 3234, 2951, 2929, 2856, 1734, 1695, 1608, 1516, 1469, 1343; **HRMS** (ESI): calculated for $C_{37}H_{47}N_8O_{13}Si^+$ [*M* + H]$^+$: 839.3026; found 839.3033.

Bis(4-nitrophenethyl) {[5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-(*tert*-butyldimethylsilyl)adenosin-N^6-yl]carbonyl}-L-aspartate **(135)**



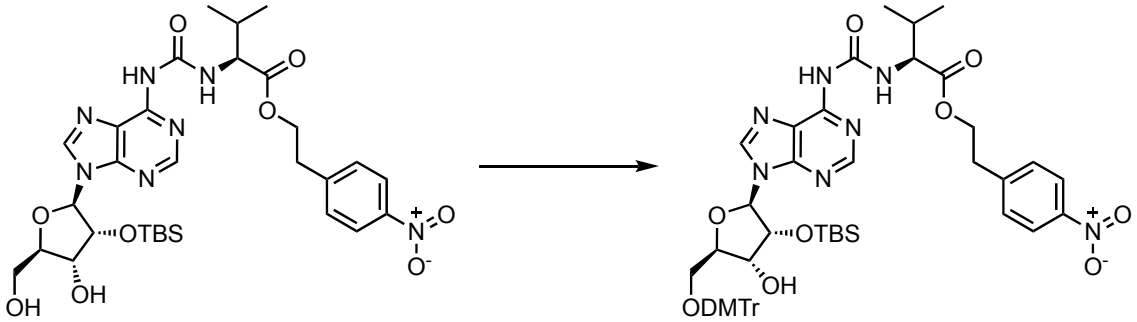
A magnetically stirred solution containing compound **131** (541 mg, 645 μmol) in anhydrous pyridine (10 mL) was treated with 4,4'-dimethoxytrityl chloride (328 mg, 967 μmol) at 0 ºC, and the resulting pale-orange/yellow solution was maintained at that temperature for 4 hr. Upon consumption of the starting material, as judged by TLC analysis, the mixture was concentrated under reduced pressure. Excess pyridine was removed *via* azeotropic coevaporation using toluene. The residue thus obtained was purified by flash chromatography (silica gel, 0.1% pyridine, CH_2Cl_2:CH_3OH = 100:0 → 99:1) to afford compound **135** (693 mg, 607 μmol, 94%) as a colourless foam. $\mathbf{R}_f$ = 0.2 in 99:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 10.21 (d, *J* = 8.0 Hz, 1H), 8.41 (s, 1H), 8.20 (s, 1H), 8.11 – 8.06 (m, 2H), 8.06 – 8.00 (m, 2H), 7.95 (s, 1H), 7.48 – 7.41 (m, 2H), 7.41 – 7.27 (m, 8H), 7.25 – 7.17 (m, 2H), 6.86 – 6.76 (m, 4H), 6.07 (d, *J* = 4.7 Hz, 1H), 4.96 (dd, *J* = 4.9 Hz, 1H), 4.93 – 4.86 (m, 1H), 4.51 – 4.28 (m, 5H), 4.25 (ddd, *J* = 3.9 Hz, 1H), 3.78 (s, 6H), 3.54 (dd, *J* = 10.8, 2.9 Hz, 1H), 3.40 (dd, *J* = 10.8, 3.9 Hz, 1H), 3.11 – 2.96 (m, 6H), 2.64 (d, *J* = 4.8 Hz, 1H), 0.85 (s, 9H), 0.03 (s, 3H), -0.09 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 170.8, 170.8, 158.8, 153.5, 151.2, 150.5, 150.1, 149.9, 147.0, 147.0, 145.5, 145.4, 144.8, 141.8, 136.1, 135.7, 130.3, 129.9, 129.9, 129.9, 128.3, 128.1, 127.2, 124.0, 123.9, 123.8, 120.9, 113.4, 88.8, 86.9, 84.3, 76.0, 71.4, 65.2, 64.6, 63.4, 55.4, 49.8, 36.7, 35.0, 34.9, 25.8, 18.1, -4.7, -5.01; **IR** (ν_{max}) 3209, 2929, 2856, 1735, 1699, 1607, 1517, 1467, 1344; **HRMS** (ESI): calculated for $C_{58}H_{65}N_8O_{15}Si^+$ [*M* + H]$^+$: 1141.4333; found 1141.4354.

4-Nitrophenethyl {[2'-*O*-(*tert*-butyldimethylsilyl)adenosin-N^6-yl]carbonyl}-L-valinate **(130)**



A magnetically stirred solution containing compound **119** (2.63 g, 4.10 mmol) in anhydrous pyridine (35 mL) was treated with compound **126** (1.31 g, 4.92 mmol) at 20 ºC, and the resulting mixture was heated to 55 ºC and maintained at that temperature for 16 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was cooled to room temperature. The cooled reaction mixture was then concentrated under reduced pressure, and the residue subjected to flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2). The mixture obtained was used directly in the next reaction without further purification. A 15 mL-Falcon tube containing the residue obtained in the previous reaction, suspended in CH_2Cl_2 (15 mL) was treated at 0 °C with pyridine (580 µL) followed by HF-pyridine (70% w/w, 358 µL, 394 mg, 13.8 mmol) and the resulting mixture was maintained at that temperature with stirring for 1.5 hr. The reaction was then quenched by dropwise addition of methoxytrimethylsilane (4.8 mL), itself also maintained at 0 ºC. After stirring for an additional 10 min, the reaction mixture was concentrated under reduced pressure, and the residue purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2→ 97:3) to afford compound **130** (1.31 g, 1.94 mmol, 47% over two steps) as a colourless foam. $\mathbf{R}_f$ = 0.3 in 97:3 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 9.91 (d, *J* = 8.2 Hz, 1H), 8.57 (s, 1H), 8.41 (s, 1H), 8.19 – 8.13 (m, 2H), 8.10 (s, 1H), 7.44 – 7.35 (m, 2H), 5.95 – 5.78 (m, 2H), 5.10 (dd, *J* = 7.3, 4.8 Hz, 1H), 4.53 (dd, *J* = 8.2, 4.7 Hz, 1H), 4.44 (dd, *J* = 6.6 Hz, 2H), 4.41 – 4.33 (m, 2H), 4.00 – 3.87 (m, 1H), 3.76 (dd, *J* = 12.4 Hz, 1H), 3.10 (dd, *J* = 6.8 Hz, 2H), 2.88 (s, 1H), 2.31 – 2.16 (m, 1H), 1.00 (d, *J* = 6.9 Hz, 3H), 0.95 (d, *J* = 6.9 Hz, 3H), 0.80 (s, 9H), -0.17 (s, 3H), -0.38 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 172.1, 153.6, 151.1, 151.0, 149.4, 147.1, 145.6, 143.3, 130.0, 129.9, 124.0, 122.2, 91.4, 87.7, 74.7, 72.9, 64.8, 63.5, 58.8, 35.1, 31.1, 25.7, 19.5, 18.0, -5.2, -5.2; **IR** (ν_{max}) 3234, 2956, 2857, 1739, 1697, 1612, 1588, 1519, 1469, 1345; **HRMS** (ESI): calculated for $C_{30}H_{44}N_7O_9Si^+$ [*M* + H]$^+$: 674.2964; found 674.2960.

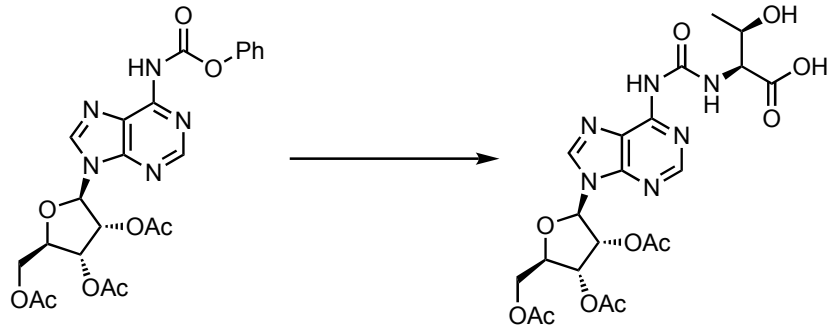
4-Nitrophenethyl {[5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-(*tert*-butyldimethylsilyl)adenosin-N^6-yl]carbonyl}-L-valinate **(134)**



A magnetically stirred solution containing compound **130** (90 mg, 134 μmol) in anhydrous pyridine (7 mL) was treated with 4,4'-dimethoxytrityl chloride (68 mg, 200 μmol) at 0 ºC, and the resulting pale-orange/yellow solution was maintained at that temperature for 2.5 hr. Upon consumption of the starting material, as judged by TLC analysis, the mixture was concentrated under reduced pressure. Excess pyridine was removed *via* azeotropic coevaporation using toluene. The residue thus obtained was purified by flash chromatography (silica gel, 0.1% pyridine, CH_2Cl_2:CH_3OH = 100:0 → 99:1) to afford compound **134** (110 mg, 113 μmol, 84%) as a colourless foam. $\mathbf{R}_f$ = 0.2 in 99:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 9.91 (d, *J* = 8.3 Hz, 1H), 8.45 (s, 1H), 8.15 (s, 1H), 8.14 – 8.08 (m, 2H), 7.82 (s, 1H), 7.48 – 7.37 (m, 4H), 7.36 – 7.31 (m, 4H), 7.25 – 7.19 (m, 2H), 6.91 – 6.74 (m, 4H), 6.06 (d, *J* = 5.1 Hz, 1H), 4.99 (dd, *J* = 5.1 Hz, 1H), 4.53 (dd, *J* = 8.3, 4.7 Hz, 1H), 4.47 – 4.41 (m, 2H), 4.37 (ddd, *J* = 4.4 Hz, 1H), 4.30 – 4.24 (m, 1H), 3.79 (s, 6H), 3.53 (dd, *J* = 10.7, 3.0 Hz, 1H), 3.39 (dd, *J* = 10.7, 3.9 Hz, 1H), 3.09 (dd, *J* = 6.6 Hz, 2H), 2.67 (d, *J* = 4.3 Hz, 1H), 2.31 – 2.20 (m, 1H), 1.01 (d, *J* = 6.9 Hz, 3H), 0.96 (d, *J* = 6.9 Hz, 3H), 0.85 (s, 9H), 0.01 (s, 3H), -0.12 (s, 3H); **^{13}C NMR** (101 MHz, Chloroform-*d*)*** δ 151.3, 141.4, 123.9, 128.3, 128.3, 130.0, 130.3, 128.1, 113.4, 88.6, 75.9, 58.8, 64.2, 71.6, 84.4, 55.4, 63.4, 63.4, 35.0, 30.9, 19.4, 18.0, 25.7, -5.1, -4.9; **IR** (ν_{max}) 3235, 2955, 2929, 2856, 1740, 1703, 1609, 1586, 1519, 1509, 1467, 1345; **HRMS** (ESI): calculated for $C_{51}H_{62}N_7O_{11}Si^+$ $[M + H]^+$: 976.4271; found 976.4282.

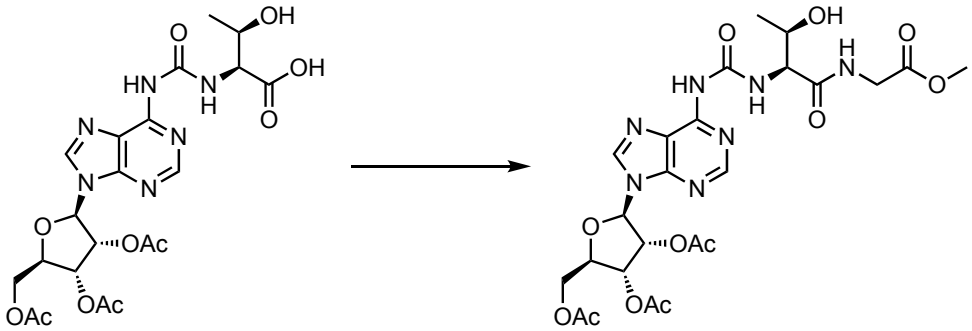
*** These ^{13}C NMR shifts were extracted from the 400/101 MHz HSQC NMR spectrum.

2',3',5'-Tri-*O*-acetyl-N^6-(L-threonyl)carbamoyladenosine (**172**)



Compound **172** was synthesised according to a modified literature procedure.[136] A magnetically stirred solution containing compound **84** (1.96 g, 3.82 mmol) in anhydrous pyridine (10 mL) was treated with L-threonine (910 mg, 7.64 mmol) at 20 ºC, and the resulting mixture was heated to 55 ºC and maintained at that temperature for 16 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was cooled to room temperature. The cooled reaction mixture was then concentrated under reduced pressure, and the residue purified by flash chromatography (silica gel, 0.1% CH_3CO_2H, CH_2Cl_2:CH_3OH = 9:1→ 8:2) to afford compound **172** (1.94 g, 3.60 mmol, 94%) as a colourless foam. The ^{1}H NMR spectral data for **172** were consistent with those previously reported in the literature.[136] $\mathbf{R}_f$ = 0.2 in 9:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Methanol-d_4) δ 8.58 (s, 1H), 8.41 (s, 1H), 7.93 (s, 1H), 7.19 (s, 1H), 7.10 (s, 1H), 6.28 (s, 1H), 6.03 (s, 1H), 5.80 – 5.63 (m, 1H), 4.53 – 4.28 (m, 4H), 3.80 – 3.77 (m, 2H), 2.14 (s, 3H), 2.10 – 2.01 (m, 6H), 1.32 – 1.24 (m, 3H).

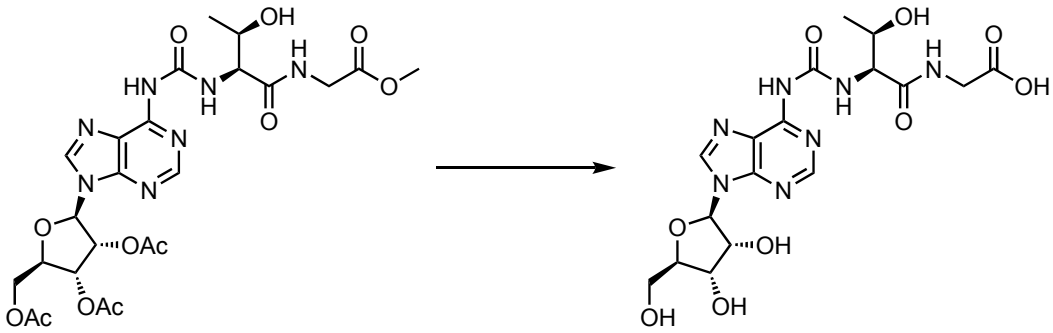
Methyl {[(2',3',5'-Tri-*O*-acetyl-adenosin-N^6-yl)carbonyl]-L-threonyl}glycinate (**173**)



A magnetically stirred solution containing compound **172** (30 mg, 56 µmol) in anhydrous DMF (2 mL) was treated with glycine methyl ester (10 mg, 111 µmol), 1-hydroxybenzotriazole (HOBt) hydrate (11 mg, 72 µmol), 1-ethyl-3-(3′-dimethylaminopropyl)carbodiimide (EDC hydrochloride, 11 mg, 72 µmol), and finally *N,N*-diisopropylethylamine (24 µL, 18 mg, 138 µmol) at 20 ºC, and the resulting solution was maintained at that temperature for 16 hr. Upon consumption of the starting material, as judged by TLC analysis, the mixture was treated with NH_4Cl (10 mL of a sat. *aq.* solution),

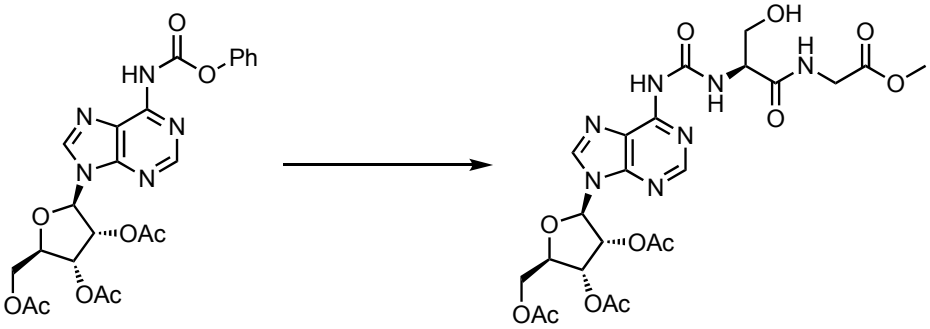
the aqueous phase extracted with CH_2Cl_2 (3 x 10 mL), and the combined organic phases dried ($MgSO_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2 → 97:3 → 96:4 → 95:5) to afford compound **173** (20 mg, 33 μmol, 59%) as a colourless oil. $\mathbf{R}_f$ = 0.3 in 92:8 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Methanol-d_4) δ 8.61 (s, 1H), 8.49 (dd, J = 5.9 Hz, 1H), 8.44 (s, 1H), 6.30 (d, J = 5.1 Hz, 1H), 6.05 (dd, J = 5.3 Hz, 1H), 5.73 (dd, J = 5.3 Hz, 1H), 4.52 – 4.32 (m, 5H), 4.06 – 3.97 (m, 2H), 3.72 (s, 3H), 2.15 (s, 3H), 2.07 (s, 3H), 2.06 (s, 3H), 1.30 (d, J = 6.5 Hz, 3H); **^{13}C NMR** (101 MHz, Methanol-d_4) δ 173.6, 172.2, 171.9, 171.4, 171.2, 156.3, 152.6, 151.7, 151.6, 143.8, 121.8, 88.2, 81.7, 74.4, 71.9, 68.5, 64.2, 61.3, 52.7, 42.0, 20.7, 20.5, 20.3, 20.2; **IR** (ν_{max}) 3236, 3123, 2957, 1742, 1664, 1611, 1589, 1518, 1468; **HRMS** (ESI): calculated for $C_{24}H_{32}N_7O_{12}^+$ [M + H]$^+$: 610.2103; found 610.2099.

N^6-{[(Glycinyl)-L-threonyl]carbamoyl}adenosine (**139**)



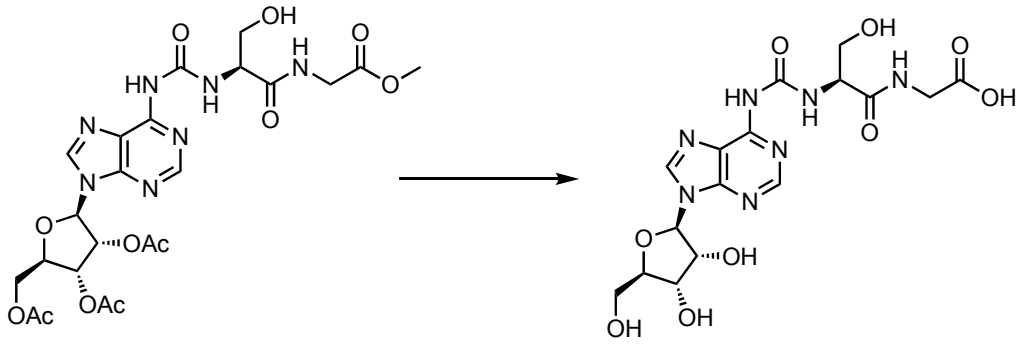
A magnetically stirred solution containing compound **173** (20 mg, 33 μmol) in 1:1 water/methanol (3 mL) was treated with lithium hydroxide (4 mg, 164 μmol) at 20 ºC, and the resulting mixture was maintained at that temperature for 1 hr. Upon consumption of the starting material, as judged by LCMS analysis, the reaction mixture concentrated under reduced pressure (lyophilised). The residue thus obtained was purified by reverse phase chromatography (Sep-Pak C18 cartridge, 4 x 1.5 mL H_2O, then 4 x 1.5 mL CH_3CN) to afford, after lyophilisation of the relevant fractions, compound **139** (12 mg, 26 μmol, 78%) as a white solid. $\mathbf{R}_f$ = 0.05 in 92:8 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Deuterium Oxide) δ 8.62 (d, J = 2.9 Hz, 1H), 8.46 (s, 1H), 6.11 (d, J = 5.7 Hz, 1H), 4.51 – 4.37 (m, 3H), 4.27 (ddd, J = 3.4 Hz, 1H), 3.96 – 3.71 (m, 4H), 1.28 (d, J = 6.3 Hz, 3H); **^{13}C NMR** (101 MHz, Deuterium Oxide) δ 207.4, 176.4, 172.4, 155.9, 151.2, 150.1, 149.8, 142.6, 88.4, 85.6, 73.7, 70.4, 67.4, 61.3, 59.9, 43.4, 18.8; **IR** (ν_{max}) 3224, 2936, 1659, 1611, 1590, 1529, 1468, 1396, 1297; **HRMS** (ESI): calculated for $C_{17}H_{24}N_7O_9^+$ [M + H]$^+$: 470.1630; found 470.1629.

Methyl {[(2',3',5'-Tri-*O*-acetyl-adenosin-N^6-yl)carbonyl]-L-seryl}glycinate (**174**)



A magnetically stirred solution containing compound **84** (378 mg, 736 μmol) in anhydrous pyridine (5 mL) was treated with L-serine (93 mg, 883 μmol) at 20 ºC, and the resulting mixture was heated to 55 ºC and maintained at that temperature for 16 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was cooled to room temperature. The cooled reaction mixture was then concentrated under reduced pressure, and the residue subjected to flash chromatography (silica gel, 0.1% CH_3CO_2H, CH_2Cl_2:CH_3OH = 9:1→ 8:2) to afford a colourless oil. A magnetically stirred solution containing the residue obtained from the previous reaction in anhydrous DMF (3 mL) was treated with glycine methyl ester hydrochloride (31 mg, 248 μmol), 1-hydroxybenzotriazole (HOBt) hydrate 38 mg, 248 μmol), 1-ethyl-3-(3′-dimethylaminopropyl)carbodiimide (EDC hydrochloride, 48 mg, 248 μmol), and finally *N,N*-diisopropylethylamine (126 μL, 94 mg, 725 μmol) at 20 ºC, and the resulting solution was maintained at that temperature for 16 hr. Upon consumption of the starting material, as judged by TLC analysis, the mixture was treated with NH_4Cl (10 mL of a sat. *aq.* solution), the aqueous phase extracted with CH_2Cl_2 (3 x 10 mL), and the combined organic phases dried ($MgSO_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2 → 97:3 → 96:4 → 95:5 → 94:6 → 93:7 → 92:8) to afford compound **174** (34 mg, 57 μmol, 8% over two steps) as a colourless oil. $\mathbf{R_f}$ = 0.3 in 92:8 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (800 MHz, Methanol-d_4) δ 8.58 (s, 1H), 8.43 (s, 1H), 6.28 (d, J = 5.0 Hz, 1H), 6.04 (dd, J = 5.7, 5.0 Hz, 1H), 5.73 (dd, J = 5.4 Hz, 1H), 4.57 (dd, J = 4.7 Hz, 1H), 4.50 – 4.43 (m, 2H), 4.41 – 4.34 (m, 1H), 4.10 – 3.98 (m, 3H), 3.94 – 3.88 (m, 1H), 3.72 (s, 3H), 2.15 (s, 3H), 2.07 (s, 3H), 2.05 (s, 3H); **^{13}C NMR** (201 MHz, Methanol-d_4) δ 173.2, 172.2, 171.8, 171.4, 171.2, 155.9, 152.6, 151.6, 151.5, 143.8, 121.8, 88.2, 81.6, 74.4, 71.9, 64.1, 63.5, 57.6, 52.7, 42.1, 20.7, 20.5, 20.3; **IR** (ν_{max}) 3240, 2951, 1741, 1661, 1611, 1589, 1518, 1467,1367; **HRMS** (ESI): calculated for $C_{19}H_{26}N_7O_{10}^+$ [M – 2Ac + 3H]$^+$: 512.1736; found 512.1735; calculated for $C_{17}H_{24}N_7O_9^+$ [M – 3Ac + 4H]$^+$: 470.1630; found 470.1629.

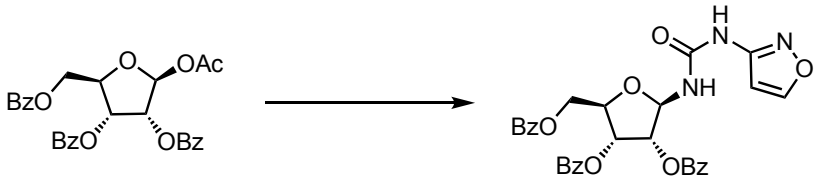
N^6-{[(Glycinyl)-L-seryl]carbamoyl}adenosine (**141**)



A magnetically stirred solution containing compound **174** (20 mg, 34 μmol) in 1:1 water/methanol (3 mL) was treated with lithium hydroxide (4 mg, 168 μmol) at 20 ºC, and the resulting mixture was maintained at that temperature for 1 hr. Upon consumption of the starting material, as judged by LCMS analysis, the reaction mixture was neutralised to pH = 7 *via* the careful addition of HCl (ca. 100 μL of a 1.0 M aqueous solution) and concentrated under reduced pressure (lyophilised). The residue thus obtained was purified by reverse phase chromatography (Sep-Pak C18 cartridge, 4 x 1.5 mL H_2O, then 4 x 1.5 mL CH_3CN) to afford, after lyophilisation of the relevant fractions, compound **141** (8 mg, 18 μmol, 52%) as a white solid. $\mathbf{R}_f$ = 0.05 in 92:8 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Deuterium Oxide) δ 8.43 (s, 1H), 6.02 (d, *J* = 5.8 Hz, 1H), 4.72 (dd, *J* = 5.5 Hz, 1H), 4.54 (dd, *J* = 4.4 Hz, 1H), 4.39 (dd, *J* = 5.2, 3.7 Hz, 1H), 4.25 (ddd, *J* = 3.4 Hz, 1H), 4.05 (dd, *J* = 11.8, 4.5 Hz, 1H), 4.00 – 3.74 (m, 6H); **^{13}C NMR** (101 MHz, Deuterium Oxide) δ 182.2, 172.7, 156.0, 151.7, 150.4, 150.2, 143.1, 121.0, 89.1, 86.3, 74.5, 71.1, 62.5, 62.0, 56.9, 44.1; **IR** (ν_{max}) 3299, 2920, 2850, 1713, 1651, 1588, 1469, 1424; **HRMS** (ESI): calculated for $C_{16}H_{22}N_7O_9^+$ [*M* + H]$^+$: 456.1474; [*M* + H]$^+$ ion for this compound was not able to be identified.

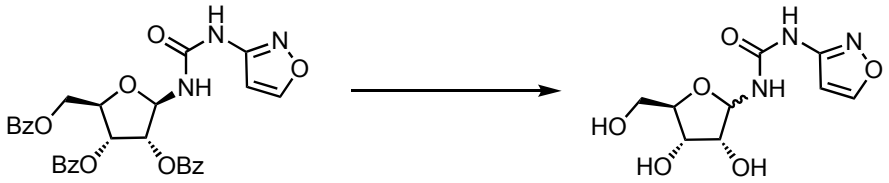
3.4.3.3 Synthesis of Pyrimidine-Precursor Nucleosides

N-(Isoxazol-3-yl)-*N'*-(2',3',5'-tri-*O*-benzoyl-β-D-ribofuranosyl)urea (**153**)



Compound **153** was synthesised according to a modified literature procedure.[121] A magnetically stirred solution containing 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose **(151)** (17.0g g, 33.8 mmol) in anhydrous CH_2Cl_2 (300 mL) was treated with $TiCl_4$ (7.70 g, 4.46 mL, 40.6 mmol) at 20 ºC and the resulting solution was maintained at that temperature for 2 hr. Upon consumption of the starting material, as judged by TLC analysis, the mixture was treated with H_2O (250 mL) at 0 ºC and the suspension was filtered through a 5 cm pad of Celite. The aqueous phase was extracted with CH_2Cl_2 (3 x 100 mL), and the combined organic phases dried ($MgSO_4$), filtered, and concentrated under reduced pressure to obtain a white solid. The solid was then dissolved in anhydrous toluene (300 mL), and the magnetically-stirred solution treated with AgNCO (6.25 g, 41.6 mmol) at 20 ºC. This mixture was then brought to reflux, and maintained at that temperature for 2.5 hr, during which, AgCl was observed to precipitate. The cooled reaction mixture was then filtered once more through a 5 cm pad of Celite and the frit was washed with ca. 50 mL anhydrous toluene. The filtrate was then treated at 20 ºC with 3-aminoiosoxazole (3.40 g, 2.99 mL, 40.4 mmol) and stirred for 16 hr at that temperature. The cloudy reaction mixture was then separated into smaller portions, which were subsequently centrifuged, and the supernatants were discarded. The precipitate was washed twice with anhydrous toluene (50 mL) and then dried under reduced pressure to afford compound **153** (14.3 g, 25.0 mmol, 74%) as a white powder containing an impurity in the aromatic region of the ^{1}H NMR. This material was used directly in further reactions without additional purification. $\mathbf{R_f}$ = 0.3 in 98:2 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Methanol-d_4) δ 8.43 (d, J = 1.8 Hz, 1H), 8.12 – 8.08 (m, 2H), 7.96 – 7.90 (m, 4H), 7.61 – 7.52 (m, 3H), 7.50 – 7.43 (m, 2H), 7.42 – 7.35 (m, 4H), 6.70 (d, J = 1.8 Hz, 1H), 5.91 (d, J = 6.0 Hz, 1H), 5.88 – 5.84 (m, 1H), 5.68 (dd, J = 6.0 Hz, 1H), 4.73 – 4.51 (m, 3H); **^{13}C NMR** (101 MHz, Methanol-d_4) δ 167.6, 166.8, 160.5, 159.6, 159.4, 155.7, 134.8, 134.7, 134.5, 130.9, 130.8, 130.8, 130.7, 130.7, 130.3, 130.1, 129.9, 129.7, 129.7, 129.6, 129.5, 129.2, 126.3, 99.4, 98.1, 84.8, 80.1, 75.3, 73.0, 65.5; **IR** (ν_{max}) 3299, 2911, 1713, 1693, 1557, 1451, 1267; **HRMS** (ESI): calculated for $C_{30}H_{26}N_3O_9^+$ [M + H]$^+$: 572.1664; found 572.1666.

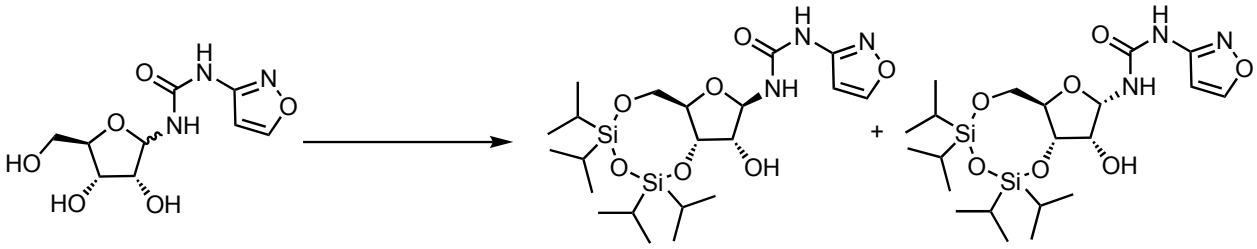
N-(Isoxazol-3-yl)-*N'*-(D-ribofuranosyl)urea (**142**)



Compound **142** was synthesised according to a modified literature procedure.[121] Compound **153** (11.7 g, 20.5 mmol) was added in one-portion to a magnetically stirred 7N solution of ammonia in methanol (300 mL, 2.10 mol) at 20 ºC, and the resulting mixture was maintained at that temperature for 20 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was concentrated under reduced pressure. The residue thus obtained was then purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 19:1 → 9:1 → 4:1) to afford compound **142** (3.76 g, 14.5 mmol, 71%, 2:1 α:β) as a white solid. The ^1H and ^{13}C NMR spectral data for **142** were consistent with those previously reported in the literature.[121] $\mathbf{R_f}$ = 0.4 in 4:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, DMSO-d_6) diastereomeric mixture δ 9.96 (s, 1H), 9.62 (s, 1H), 8.79 (d, J = 1.8 Hz, 1H), 8.77 (d, J = 1.7 Hz, 1H), 8.14 – 7.95 (m, 1H), 7.71 – 7.51 (m, 1H), 7.34 (d, J = 9.5 Hz, 1H), 7.07 (d, J = 9.6 Hz, 1H), 6.85 (d, J = 1.7 Hz, 1H), 6.84 (dd, J = 7.6, 1.7 Hz, 1H), 5.60 (dd, J = 9.5, 4.3 Hz, 1H), 5.51 (d, J = 4.8 Hz, 1H), 5.29 (dd, J = 9.6, 5.5 Hz, 1H), 5.19 (d, J = 6.2 Hz, 1H), 5.13 (d, J = 6.0 Hz, 1H), 5.03 (d, J = 5.2 Hz, 1H), 4.89 (dd, J = 5.6 Hz, 1H), 4.79 (dd, J = 5.7 Hz, 1H), 4.22 (ddd, J = 5.2 Hz, 1H), 4.06 – 3.90 (m, 4H), 3.86 – 3.73 (m, 4H), 3.65 – 3.49 (m, 2H); **^{13}C NMR** (101 MHz, DMSO-d_6) diastereomeric mixture δ 159.8, 159.6, 158.4, 158.3, 153.5, 153.3, 98.4, 98.4, 84.4, 83.6, 81.9, 80.6, 74.3, 71.2, 70.4, 70.2, 61.8, 61.6; **IR** (ν_{max}) 3299, 2937, 1673, 1600, 1532, 1479, 1405; **HRMS** (ESI): calculated for $C_9H_{14}N_3O_6^+$ [M + H]$^+$: 260.0877; found 260.0881.

N-(Isoxazol-3-yl)-*N'*-{3',5'-*O*-[1,1,3,3-tetrakis(1-methylethyl)-1,3-disiloxanediyl]-β-D-ribofuranosyl}urea (**β-155**)

and *N*-(Isoxazol-3-yl)-*N'*-{3',5'-*O*-[1,1,3,3-tetrakis(1-methylethyl)-1,3-disiloxanediyl]-α-D-ribofuranosyl}urea (**α-155)**

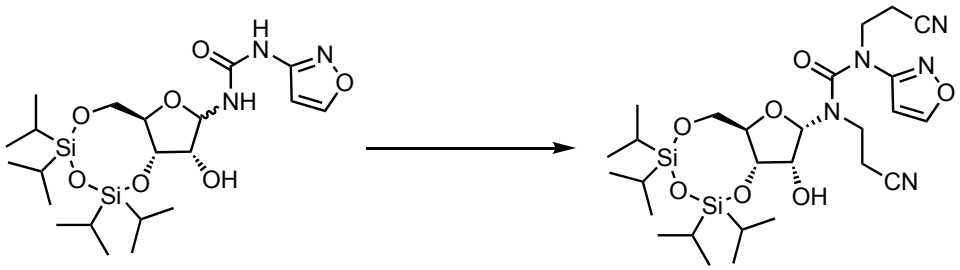


A magnetically stirred solution containing compound **142** (51 mg, 197 μmol, 2:1 α:β) in pyridine (5 mL) was treated at 0 ºC with 1,3-dichloro-1,1,3,3-tetraisopropyldisiloxane (TIPDSiCl_2) (74 mg, 76 μL, 236 μmol) and the resulting mixture was maintained at room temperature for 20 hr. Upon consumption of the starting material, as judged by TLC analysis, the mixture was treated with NH_4Cl (20 mL of a sat. *aq.* solution), the aqueous phase extracted with CH_2Cl_2 (3 x 14 mL), and the combined organic phases dried ($MgSO_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was purified by flash chromatography (silica gel, CH_2Cl_2:EtOAc = 9:1 → 4:1) to afford compound **155** (41 mg, 81.7 μmol, 42%, 3:1 α:β) as a colourless foam. Analytical samples of α-**155** and **β-155** were obtained for spectroscopic characterisation by applying repeated chromatography to the 3:1 α:β mixture.

Concentration of fraction A (R_f = 0.2 in 9:1 v/v CH_2Cl_2:EtOAc elution) afforded compound **α-155**. **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 8.92 (s, 1H), 8.19 (d, *J* = 1.7 Hz, 1H), 6.45 (s, 1H), 5.82 (dd, *J* = 8.5, 4.4 Hz, 1H), 4.35 (dd, *J* = 7.1, 5.5 Hz, 1H), 4.23 (ddd, *J* = 5.0, 1.9 Hz, 1H), 4.06 – 3.84 (m, 3H), 3.22 (d, *J* = 2.2 Hz, 1H), 1.13 – 1.01 (m, 28H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 158.4, 158.3, 154.5, 98.5, 81.5, 80.0, 72.9, 70.5, 62.6, 17.6, 17.5, 17.4, 17.2, 17.1, 17.1, 13.5, 13.3, 13.0, 12.8; **IR** (ν_{max}) 3332, 2944, 2867, 1673, 1602, 1535, 1465; **HRMS** (ESI): calculated for $C_{21}H_{40}N_3O_7Si_2^+$ [*M* + H]$^+$: 502.2399; found 502.2399.

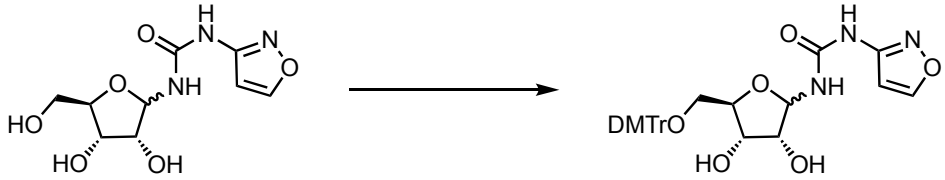
Concentration of fraction B (R_f = 0.2 in 9:1 v/v CH_2Cl_2:EtOAc elution) afforded compound **β-155**). **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 9.28 (s, 1H), 8.17 (d, *J* = 1.8 Hz, 1H), 7.71 (s, 1H), 6.39 (s, 1H), 5.45 (dd, *J* = 8.3, 3.7 Hz, 1H), 4.40 (dd, *J* = 6.2 Hz, 1H), 4.08 – 3.85 (m, 4H), 3.13 (d, *J* = 4.2 Hz, 1H), 1.17 – 0.94 (m, 28H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 158.4, 158.2, 154.5, 98.4, 86.0, 82.1, 74.8, 71.8, 63.5, 17.6, 17.6, 17.5, 17.5, 17.4, 17.3, 17.2, 17.2, 13.5, 13.4, 13.0, 12.8; **IR** (ν_{max}) 3333, 2944, 2867, 1674, 1596, 1536, 1465; **HRMS** (ESI): calculated for $C_{21}H_{40}N_3O_7Si_2^+$ [*M* + H]$^+$: 502.2399; found 502.2399.

N,N'-Bis(2-cyanoethyl)-*N*-(isoxazol-3-yl)-*N'*-{3',5'-*O*-[1,1,3,3-tetrakis(1-methylethyl)-1,3-disiloxanediyl]-α-D-ribofuranosyl}urea (**159**)



A 1.5 mL-safe lock tube containing a solution of solution of compound **155** (10 mg, 19.9 μmol, 3:1 α:β) in *t*-BuOH (800 μL) was treated at 0 ºC with caesium carbonate (16 mg, 49.1 μmol), followed by acrylonitrile (42 mg, 52 μL, 797 μmol), and the resulting mixture was heated to 50 ºC shaken at that temperature for 45 min. Upon consumption of the starting material, as judged by LCMS analysis, the mixture was carefully decanted using a micropipette, and the supernatant concentrated under a steady stream of nitrogen. The residue thus obtained was purified by flash chromatography (silica gel, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2) to afford compound **159** (7 mg, 11.5 μmol, 58%) as a colourless oil. $\mathbf{R_f}$ = 0.2 in 99:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Methanol-d_4) δ 8.21 (d, J = 1.8 Hz, 1H), 5.96 (d, J = 1.8 Hz, 1H), 5.71 (d, J = 5.2 Hz, 1H), 4.96 (dd, J = 5.2 Hz, 1H), 4.33 (dd, J = 9.3, 5.2 Hz, 1H), 4.14 – 4.03 (m, 2H), 3.77 – 3.69 (m, 2H), 3.55 (ddd, J = 14.4, 6.1 Hz, 1H), 3.49 – 3.43 (m, 2H), 2.86 – 2.80 (m, 2H), 2.80 – 2.73 (m, 2H), 1.17 – 1.04 (m, 28H); **^{13}C NMR** (101 MHz, Methanol-d_4) δ 159.7, 159.4, 119.2, 97.3, 89.6, 78.9, 78.0, 72.4, 60.7, 40.7, 39.8, 18.1, 18.0 17.8, 17.8, 17.7, 17.7, 17.7, 17.6, 17.4, 17.4, 14.6, 14.2, 14.0, 13.8; **IR** (ν_{max}) 3380, 2945, 2856, 1759, 1685, 1595, 1540, 1464; **HRMS** (ESI): calculated for $C_{27}H_{44}N_5O_7Si_2^-$ [M – H]$^-$: 606.2785; found 606.2793.

N-(Isoxazol-3-yl)-*N'*-[5'-*O*-(4,4'-dimethoxytrityl)-D-ribofuranosyl]urea (**160**)

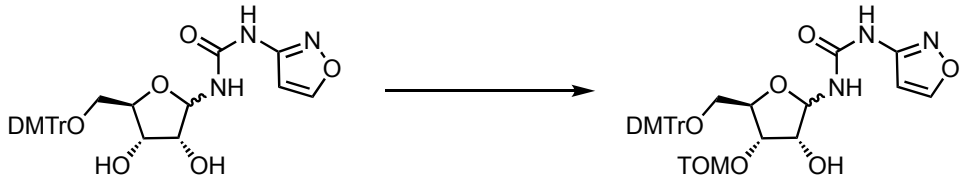


A magnetically stirred solution containing compound **142** (1.00 g, 3.86 mmol) in anhydrous pyridine (39 mL) was treated with 4,4'-dimethoxytrityl chloride (1.57 g, 4.63 mol) at 0 ºC, and the resulting pale-orange/yellow solution was maintained at room temperature for 16 hr. Upon consumption of the starting material, as judged by TLC analysis, the mixture was concentrated under reduced pressure. The residue thus obtained was purified by flash chromatography (silica gel, 0.2% pyridine, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2 → 97:3 → 96:4 → 95:5) to afford compound **160** (1.91 g,

3.40 mmol, 88%, 3:1 α:β) as a colourless foam. An analytical sample of α-**160** was obtained for spectroscopic characterisation by applying repeated chromatography to the 3:1 α:β mixture.

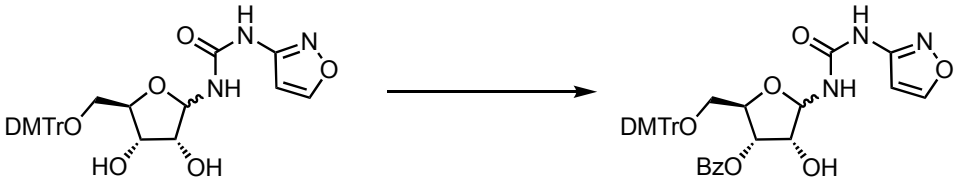
Concentration of fraction A (R_f = 0.3 in 1:4 v/v CH_2Cl_2:EtOAc elution) afforded compound **α-160**. **^{1}H NMR** (400 MHz, DMSO-*d*$_6$) δ 9.94 (s, 1H), 8.69 (d, *J* = 1.8 Hz, 1H), 7.49 – 7.17 (m, 10H), 6.95 – 6.84 (m, 4H), 6.78 (s, 1H), 5.63 (dd, *J* = 9.6, 3.3 Hz, 1H), 5.54 (d, *J* = 4.0 Hz, 1H), 5.09 (d, *J* = 4.8 Hz, 1H), 3.95 4.00 – 3.89 (m, 3H), 3.73 (s, 6H), 3.19 – 3.08 (m, 1H), 3.01 – 2.90 (m, 1H); **^{13}C NMR** (101 MHz, DMSO-*d*$_6$) δ 159.7, 158.4, 158.1, 153.3, 145.1, 135.9, 135.8, 129.8, 128.1, 127.9, 127.8, 127.6, 126.7, 113.2, 98.4, 85.3, 80.9, 80.0, 71.9, 70.3, 64.5, 55.1, 55.0; **IR** (ν_{max}) 3335, 2930, 1678, 1606, 1508, 1249; **HRMS** (ESI): calculated for $C_{30}H_{30}N_3O_8^-$ [*M* – H]$^-$: 560.2038; found 560.2045.

N-(Isoxazol-3-yl)-*N'*-{5'-*O*-[4,4'-dimethoxytrityl]-3'-*O*-[(triisopropyl)silyloxymethyl]-D-ribofuranosyl}urea (**161**)



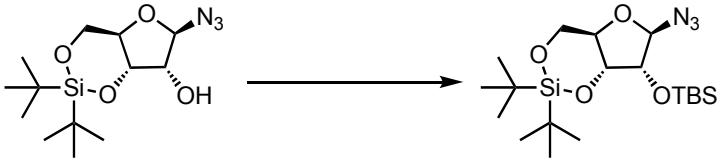
A magnetically stirred solution containing compound **160** (20 mg, 35.7 μmol) in 1,2-dichloroethane (140 μL) was treated at 20 ºC with *N,N*-diisopropylethylamine (16 mg, 21 μL, 124 μmol), followed by $SnBu_2Cl_2$ (12 mg, 39.2 μmol) and the resulting mixture was maintained at room temperature for 1 hr. After this time, (triisopropylsiloxy)methyl chloride (TOMCl) (10 mg, 10.4 μL, 45 μmol) was added and the reaction was stirred at 60 ºC for a further 1 hr. Upon consumption of the starting material, as judged by TLC analysis, the cooled mixture was purified by flash chromatography (silica gel, 0.2% pyridine, CH_2Cl_2:CH_3OH = 100:0 → 99:1) to afford compound **161** (9 mg, 11.6 μmol, 32 %, 2:1 α:β) as a colourless oil. **^{1}H NMR** (400 MHz, Chloroform-*d*) diastereomeric mixture δ 8.39 – 8.31 (m, 2H), 8.17 – 8.08 (m, 2H), 7.48 – 7.17 (m, 20H), 6.87 – 6.78 (m, 8H), 6.52 – 6.35 (m, 2H), 6.02 – 5.85 (m, 2H), 5.09 – 4.82 (m, 4H), 4.45 – 4.05 (m, 6H), 3.84 – 3.75 (m, 12H), 3.49 – 3.07 (m, 6H), 1.35 – 0.96 (m, 42H); **^{13}C NMR** (101 MHz, Chloroform-*d*) diastereomeric mixture δ 158.8, 158.7, 158.6, 158.6, 158.4, 158.3, 158.2, 154.3, 145.0, 136.2, 136.0, 135.8, 130.4, 130.2, 128.3, 128.1, 128.0, 126.9, 113.4, 113.3, 98.4, 90.7, 86.3, 80.8, 80.3, 70.3, 63.5, 55.4, 18.0, 12.0, 12.0; **IR** (ν_{max}) 3343, 2943, 2866, 1678, 1606, 1536, 1508, 1248; **HRMS** (ESI): calculated for $C_{40}H_{52}N_3O_9Si^-$ [*M* – H]$^-$: 746.3478; found 746.3500.

N-(Isoxazol-3-yl)-*N'*-[3'-*O*-benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-D-ribofuranosyl]urea (**162**)



A magnetically stirred solution containing compound **160** (100 mg, 178 μmol) in pyridine (1.8 mL) was treated with benzoic anhydride (40 mg, 0.178 mmol) followed by 4-dimethylaminopyridine (DMAP) (2 mg, 18 μmol) at 20 ºC, and the resulting solution was maintained at that temperature for 1.5 hr. Upon consumption of the starting material, as judged by TLC analysis, the reaction was quenched with MeOH (100 μL) and then diluted with CH_2Cl_2 (20 mL) and brine (20 mL of a sat. *aq.* solution). The aqueous phase extracted with CH_2Cl_2 (3 x 10 mL), and the combined organic phases dried ($MgSO_4$), filtered, and concentrated under reduced pressure. The residue thus obtained was purified by flash chromatography (silica gel, 0.2% pyridine, CH_2Cl_2:CH_3OH = 100:0 → 99:1 → 98:2 → 97:3 → 96:4 → 95:5) to afford compound **162** (20 mg, 30 μmol, 17%, 3:1 α:β) as a colourless solid. $\mathbf{R_f}$ = 0.3 in 19:1 v/v CH_2Cl_2:CH_3OH elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) diastereomeric mixture δ 9.02 (d, *J* = 142.3 Hz, 2H), 8.16 – 8.00 (m, 10H), 7.58 – 7.14 (m, 16H), 6.86 – 6.71 (m, 8H), 6.53 – 6.26 (m, 2H), 6.11 – 5.71 (m, 2H), 5.49 – 5.40 (m, 2H), 4.87 – 4.50 (m, 2H), 4.48 – 4.27 (m, 2H), 3.73 (s, 12H), 3.49 – 3.20 (m, 4H); **^{13}C NMR** (101 MHz, Chloroform-*d*) diastereomeric mixture δ 169.8, 166.4, 166.2, 158.6, 158.6, 158.6, 158.4, 158.2, 154.8, 144.8, 136.0, 135.9, 133.5, 132.8, 130.3, 130.2, 130.2, 130.0, 129.5, 129.4, 128.6, 128.4, 128.3, 128.1, 128.0, 126.9, 113.3, 98.5, 86.7, 86.5, 82.2, 81.1, 80.1, 75.2, 74.6, 70.0, 63.8; **IR** (ν_{max}) 3310, 2934, 1717, 1675, 1603, 1540, 1507; **HRMS** (ESI): calculated for $C_{37}H_{34}N_3O_9^-$ [*M* – H]$^-$: 664.2301; found 664.2308.

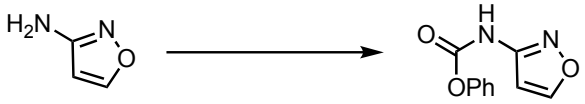
3,5-*O*-[Bis(*tert*-butyl)silylene]-2-*O*-(*tert*-butyldimethylsilyl)-β-D-ribofuranosyl azide (**166**)



A magnetically stirred solution containing compound **164** [124] (2.42 g, 7.67 mmol) in anhydrous DMF (19 mL) was treated at 0 ºC with imidazole (2.31 g, 15.3 mmol) and then *tert*-butyldimethylsilyl chloride (2.09 g, 30.7 mmol) and the resulting mixture was maintained at that temperature for 16 hr. Upon consumption of the starting material, as judged by TLC analysis, the mixture was diluted with CH_2Cl_2 (100 mL) and treated with treated brine (150 mL of a sat. *aq.* solution). The aqueous phase was extracted with CH_2Cl_2 (3 x 100 mL), and the combined organic phases dried (Na_2SO_4), filtered, and concentrated under reduced pressure. The residue thus obtained was purified by flash

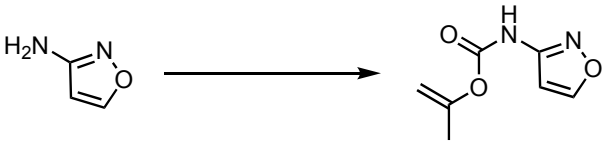
chromatography (silica gel, Hexane:EtOAc = 100:0 → 99:1) to afford compound **166** (3.02 g, 7.03 mmol, 91%) as a colourless oil. $\mathbf{R}_f$ = 0.2 in 99:1 v/v Hexane:EtOAc elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 5.07 (s, 1H), 4.45 – 4.36 (m, 1H), 4.12 (ddd, *J* = 10.8, 9.6, 5.1 Hz, 1H), 4.01 (d, *J* = 4.1 Hz, 1H), 3.95 – 3.87 (m, 2H), 1.05 (s, 9H), 1.00 (s, 9H), 0.91 (s, 9H), 0.13 (s, 3H), 0.12 (s, 3H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 96.3, 76.4, 76.0, 74.7, 68.5, 27.6, 27.2, 26.0, 22.9, 20.5, 18.5, -4.2, -5.0; **IR** (ν_{max}) 2933, 1859, 2109, 1471, 1253; **HRMS** (ESI): calculated for $C_{19}H_{40}N_3O_5Si_2^-$ $[M + OH]^-$: 446.2512; found 446.2520.

Phenyl isoxazol-3-ylcarbamate (**167**)



Compound **167** was synthesised according to a modified literature procedure.[137] A magnetically stirred solution containing phenyl chloroformate (1.96 g, 1.57 mL, 12.5 mmol) and pyridine (1.03 g, 1.06 mL, 13.1 mmol) in anhydrous CH_3CN (25 mL) was treated dropwise with 3-aminoisoxazole (1.00 g, 879 µL, 11.9 mmol) at 0 ºC and the resulting suspension was maintained at that temperature for 3.5 hr. Upon consumption of the starting material, as judged by LCMS analysis, the mixture was then treated with H_2O (20 mL), and the resulting white precipitate collected quickly by filtration. The solid material was washed once with ice-cold anhydrous CH_3CN (10 mL) and dried for 24 h under reduced pressure to give compound **167** as a white powder (1.75 g, 8.56 mmol, 72%). **^{1}H NMR** (400 MHz, Benzene-d_6) δ 7.54 (s, 1H), 7.29 (d, *J* = 1.6 Hz, 1H), 7.08 – 6.96 (m, 4H), 6.94 – 6.86 (m, 1H), 6.73 (s, 1H).

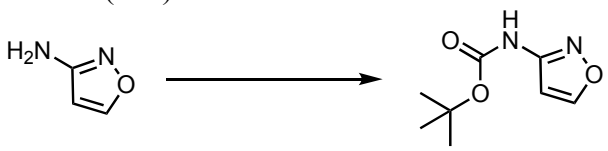
Prop-1-en-2-yl isoxazol-3-ylcarbamate (**169**)



A magnetically stirred solution containing 3-aminoisoxazole (332 mg, 3.95 mmol) and pyridine (348 mg, 354 µL, 4.40 mmol) in anhydrous CH_3CN (8 mL) was treated at 0 ºC with isopropenyl chloroformate (506 mg, 458 µL, 4.20 mmol) and the resulting mixture was maintained at that temperature for 4 hr. Upon consumption of the starting material, as judged by LCMS analysis, the reaction mixture was then concentrated under reduce pressure, and the residue thus obtained was purified *via* slow (evaporative) re-crystallisation from a 50% v/v solution of CH_3OH/H_2O (6 mL) to obtain afford compound **169** (579 mg, 3.44 mmol, 87%) as colourless needles. **^{1}H NMR** (400 MHz, Methanol-d_4) δ 8.50 (d, *J* = 1.8 Hz, 1H), 6.78 (d, *J* = 1.8 Hz, 1H), 4.86 – 4.58 (m, 2H), 1.98 (d, *J* =

1.1 Hz, 3H); **^{13}C NMR** (101 MHz, Methanol-d_4) δ 160.8, 159.4, 154.1, 102.5, 99.2, 19.5; **IR** (ν_{max}) 3245, 3186, 2955, 1743, 1674, 1614, 1486, 1404; **HRMS** (ESI): calculated for $C_7H_7N_2O_3^-$ [M + H]$^+$: 167.0462; found 167.0461.

tert-Butyl isoxazol-3-ylcarbamate (**168**)



A magnetically stirred solution containing 3-aminoisoxazole (101 mg, 88 μL, 1.20 mmol) in anhydrous pyridine (6 mL) was treated dropwise at 20 ºC with di-*tert*-butyl dicarbonate (Boc_2O) (262 mg, 276 μL, 1.20 mmol), followed by 4-dimethylaminopyridine (DMAP) (15 mg, 119 μmol) and the resulting mixture was maintained at that temperature for 75 min. Upon consumption of the starting material, as judged by TLC analysis, the reaction mixture was then concentrated under reduce pressure, and the residue thus obtained was purified by flash chromatography (silica gel, CH_2Cl_2) to afford compound **168** (120 mg, 651 μmol, 55%) as a colourless solid. $\mathbf{R_f}$ = 0.4 in CH_2Cl_2 elution; **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 8.37 (d, J = 1.7 Hz, 1H), 6.37 (d, J = 1.7 Hz, 1H), 1.46 (s, 9H); **^{13}C NMR** (101 MHz, Chloroform-*d*) δ 159.6, 158.7, 149.9, 103.6, 84.3, 27.9; **IR** (ν_{max}) 2980, 1770, 1723, 1567, 1470, 1469; **HRMS** (ESI): calculated for $C_8H_{13}N_2O_3^+$ [M + H]$^+$: 185.0921; found 185.0923.

3.4.3.4 Prebiotic Synthesis Experiments

Electrophilic cyclisation of compound **57**

A set of two experiments were carried out in order to investigate the general capacity of the compound **57** to undergo electrophilic cyclisation reactions in the presence of different activators (sodium trimetaphosphate and cyanamide) under prebiotically plausible conditions. To this end, a 1.5 mL-safe lock tube charged with compound **57** (2 mg, 5.43 μmol) suspended in phosphate buffer (100 μL, 100 mM, pH = 8.0) was treated with the either cyanamide (10 mg, 238 μmol), or sodium trimetaphosphate (10 mg, 32.7 μmol) and magnesium chloride (10 mg, 105 μmol) and shaken at 25 ºC for 18 hr. After this time, a 1 μL aliquot was taken from the reaction mixture and subjected to HPLC-MS analysis. The gradient for both samples was from 100% to 80% buffer A over 45 min; the elution was monitored at 260 nm (Dionex Ultimate 3000 Diode Array Detector); and the chromatographic eluent was directly injected into the ion source without prior splitting. Ions were scanned by use of a positive polarity mode over a full-scan range of m/z 50-1000 with a resolution of 30000. The retention time of each nucleoside was confirmed by the relevant [M+H] ion in the HRMS spectrum and by co-injection of an independently synthesised standard (compound **110**; Supporting Information). The results from this analysis are presented in **Scheme 25** (Unpublished Results section).

Nucleophilic ring-opening of compound **102**

A preliminary test reaction was carried out in order to investigate the general capacity of the hydantoin **102** to undergo nucleophilic ring-opening reactions in the presence of amino acids. To this end, a magnetically stirred solution containing compound **102** (28 mg, 59 μmol) in borate buffer (3 mL, 30 mM, pH = 9.5) was treated with glycine (30 mg, 400 μmol) at 20 ºC, and the resulting mixture was maintained at that temperature for 2.5 hr. At this time, a 1 μL aliquot was taken from the reaction mixture and subjected to LCMS analysis. The gradient for all samples was from 95% to 80% buffer A over 5 min; the elution was monitored at 260 nm; and the chromatographic eluent was directly injected into the ion source without prior splitting. Ions were scanned by use of a positive polarity mode over a full-scan range of m/z 150-1000. The retention time of each nucleoside was tentatively assigned by the relevant [M+H] ion in the ESI MS spectrum without co-injection of an independently synthesised standard, with the exception of the starting material. The results from this analysis are presented in **Scheme 22** (Unpublished Results section).

Nucleophilic ring-openings of compound **110**

A set of three preliminary test reactions were carried out in order to investigate the general capacity of the hydantoin **110** to undergo nucleophilic ring-openings in the presence of amino acids (glycine) and short peptides (glygly and glyglygly) under prebiotically plausible conditions. To this end, a 1.5 mL-safe lock tube charged with the relevant amino acid or peptide (133 μmol) suspended in phosphate buffer (100 μL, 100 mM, pH = 8.0) was treated with compound **110** (2 mg, 4.42 μmol) at 20 ºC, and the resulting mixture was shaken at that temperature for 5 hr. After this time, a 1 μL aliquot was taken from the reaction mixture and subjected to HPLC-MS analysis. The gradient for all samples was from 100% to 80% buffer A over 45 min; the elution was monitored at 260 nm (Dionex Ultimate 3000 Diode Array Detector); and the chromatographic eluent was directly injected into the ion source without prior splitting. Ions were scanned by use of a positive polarity mode over a full-scan range of m/z 50-1000 with a resolution of 30000. The retention time of each nucleoside was tentatively assigned based upon the relevant [M+H] ion in the HRMS spectrum without co-injection of an independently synthesised standard, with the exception of the starting material. The results from this analysis are presented in **Scheme 24** (Unpublished Results section).

Peptide cleavage of compound **139**

A set of two preliminary test reactions were carried out in order to investigate the general capacity of the adenosine-dipeptide **139** to undergo base-promoted acyl substitution reactions, thus providing free adenosine **33** under prebiotically plausible conditions. To this end, a 1.5 mL-safe lock tube charged with compound **139** (1 mg, 2.13 μmol) suspended in either carbonate buffer (1.0 mL, 100 mM, pH = 9.5), or borate buffer (1.0 mL, 30 mM, pH = 9.5) was heated to 50 ºC, and shaken at that temperature for 18 hr. After this time, a 1 μL aliquot was taken from the reaction mixture and subjected to LCMS analysis. The gradient for all samples was from 95% to 80% buffer A over 5 min; the elution was monitored at 260 nm; and the chromatographic eluent was directly injected into the ion source without prior splitting. Ions were scanned by use of a positive polarity mode over a full-scan range of m/z 150-1000. The retention time of each compound was tentatively assigned based upon the relevant [M+H] ion in the ESI MS spectrum without co-injection of an independently synthesised standard, with the exception of the starting material. The results from this analysis are presented in **Scheme 29** (Unpublished Results section).

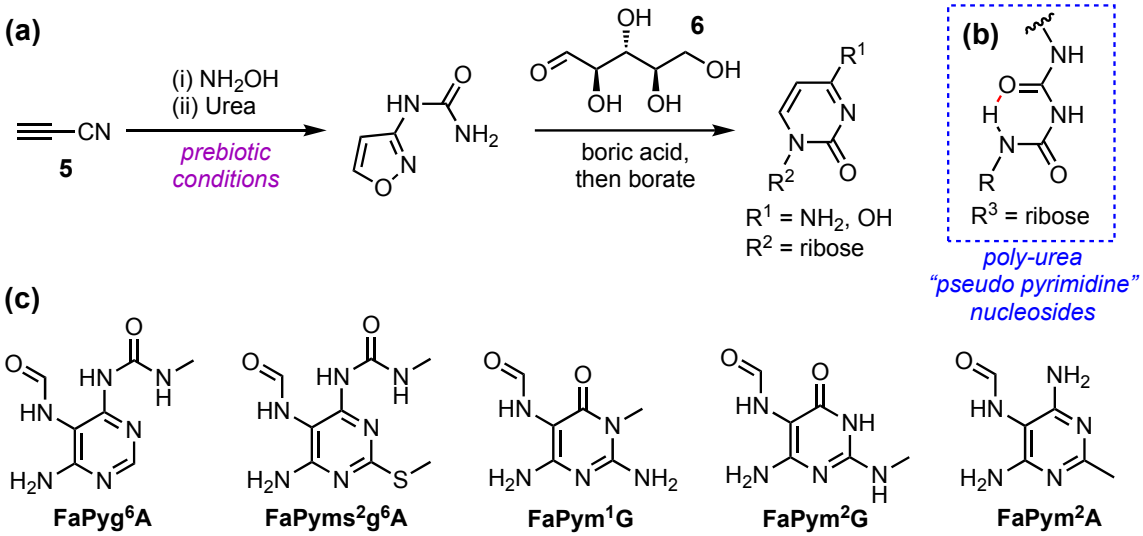
4 THESIS DISCUSSION

4.1 Geochemical Constraints upon Prebiotic Nucleoside Synthesis

This work has been aimed towards addressing several major open questions in the field of origins of life research. The first of these questions relates to the nature and potentially independent origins of the building blocks that constitute life's genetic code. Specifically, we set out to determine whether a complete set of Watson-Crick RNA nucleosides, as well as their related non-canonical variants, could plausibly have arisen under conditions compatible with the Early earth. RNA molecules, which are constructed from both purine and pyrimidine nucleotides, are thought to have been the first fundamental biopolymers and those that heralded life on Earth.[24] In the context of the earliest stages of life's evolution, it is necessary to consider chemical pathways that predated enzymatic- or ribozymatic reactions. This begs the question of what other natural occurring processes could possibly have given rise to the complex organic networks that could have facilitated the first RNA synthesis. Some have argued that the prebiotic pathways leading to building blocks such as those described here were unlikely to have emerged from discontinuous processes that depended upon rare events.[138] Ultimately, however, some chain of improbable reactions must have eventually led to the highly-complex informational molecules that make up RNA. In pursuit of prebiotically plausible routes to the different nucleosides discussed throughout this thesis, we set about exploring syntheses leading to those molecules, starting with the smallest practical building blocks that were likely to have been available on the early earth.

Although independent high-yielding pathways leading to either purines or pyrimidines have been demonstrated,[67,70] no unified pathway providing an entire set of ribonucleosides was known at time of these investigations. Through the work described in this thesis, we are now able report a novel pathway leading to the pyrimidine nucleosides that is compatible with our previously reported FaPy chemistry, and which even yielded a complete collection of Watson-Crick nucleosides (rA, rU, rC and rG) under a unified set of reaction conditions (**Scheme 38**).[121] In addition to this discovery, we were particularly interested in the role of urea at life's origin. While pursuing further chemical pathways stemming from Wöhler's serendipitous synthesis of urea in 1828,[38] we were thus able to discover an entirely new class of polyurea-containing "pseudo pyrimidine" nucleosides whose nucleobase moieties are held together entirely through hydrogen bonding interactions, and which bare a prominent resemblance to the canonical pyrimidines.[124] In certain instances, these pseudopyrimdines were even found to possess highly specific base pairing properties. One of these bases (triuret), for example, was found to display wobble base interactions[139] when paired with

guanine or inosine within complimentary RNA strands. Additionally, we expanded upon the existing FaPy route to purine nucleosides,[70] leading to the discovery of several additional pathways that provided a plethora of non-canonical (modified) bases.



Scheme 38 (a) *Prebiotic synthesis of pyrimidine nucleosides, under conditions compatible with the previously-reported FaPy chemistry.*[121] **(b)** *Structure of a new class of ribonucleoside containing poly-urea functionality.*[124] **(c)** *A selection of modified-FaPys that provide non-canonical nucleosides upon dehydrative reactions in the presence of ribose.*[102]

Given the diversity of small molecules that are described in these pathways, namely in the formation of the nucleobase portion of the nucleosides, a discussion on the inorganic origins of their building blocks is warranted. It is well-understood that any search for past life on Earth or beyond should reflect the associated environmental constraints that may have permitted those chemical pathways. One such environmental limitation upon prebiotic chemistry would have been the nature of gases emitted by volcanoes into the Hadean atmosphere.[140] Isotopic data, for example, has demonstrated that Earth likely accreted from oxygen-poor material containing CO and CH_4, as opposed to CO_2, and with NH_3 rather than N_2.[140] Also likely abundant during Earth's accretion were reducing gasses such as H_2S and H_2.[141] Remarkably, this composition of gasses is very close to that which was originally envisaged by Stanley Miller during his electrical discharge experiments that formed aldehydes, cyanide and amino acids.[43] Deviating from Miller's hypothesis, however, it is now known that Earth's mantle as well as its resultant atmosphere were only temporarily reducing. Rather, the planet experienced a rapid (less than 100 million years long) restructuring event as early as 4.51 Ga, in which the molten iron collapsed within Earth's core, where it has remained since.[142] This event is generally thought to have transformed Earth's Hadean atmosphere, generating a far more oxidizing composition of gases than were present immediately after Earth's accretion.

Emissions from Hadean volcanoes could also have fuelled the production of small molecules such as H_2O, SO_2, N_2, and CO_2. These molecules may have been directly involved in the assembly of nucleosides, through the formation of semi-activated building blocks such as nitrogen oxides and sulphites (**Figure 12**). Fixation of nitrogen in the prebiotic atmosphere could have been achieved *via* the lightning-mediated coupling of N_2 to H_2O, resulting in the production of NO gas.[143] As demonstrated experimentally, capture of NO can be facilitated by condensation with SO_2/HSO_3^- to generate a disulfonate **175**, which reacts upon exposure to cyanoacetylene to generate the key pyrimidine building block 3-aminoisoxazole **148**.[121]

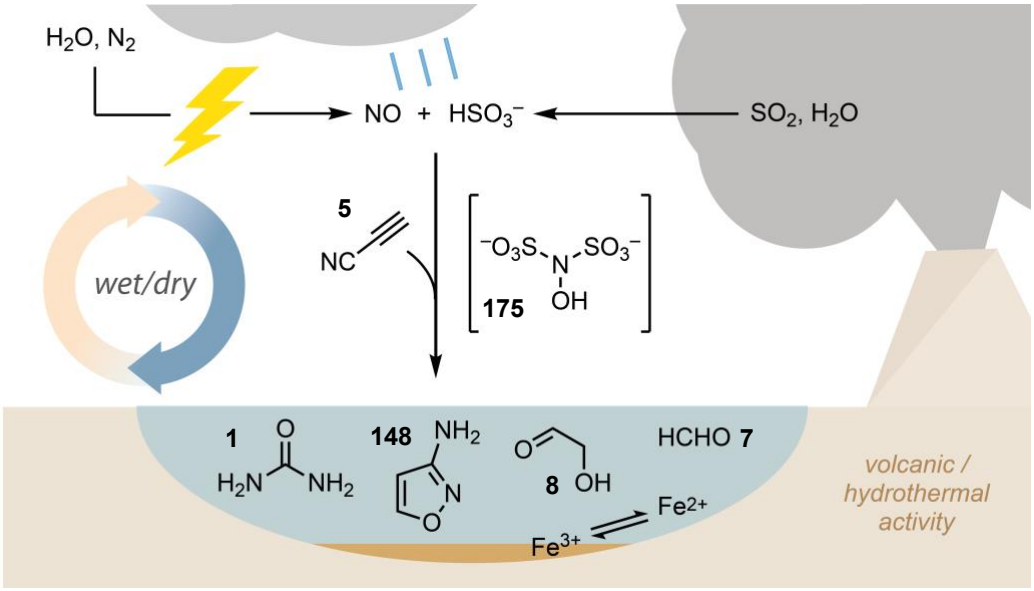


Figure 12 *The hypothetical formation of nucleoside building blocks including 3-aminoisoxazole **148** via natural events resulting from Hadean volcanic emissions.*

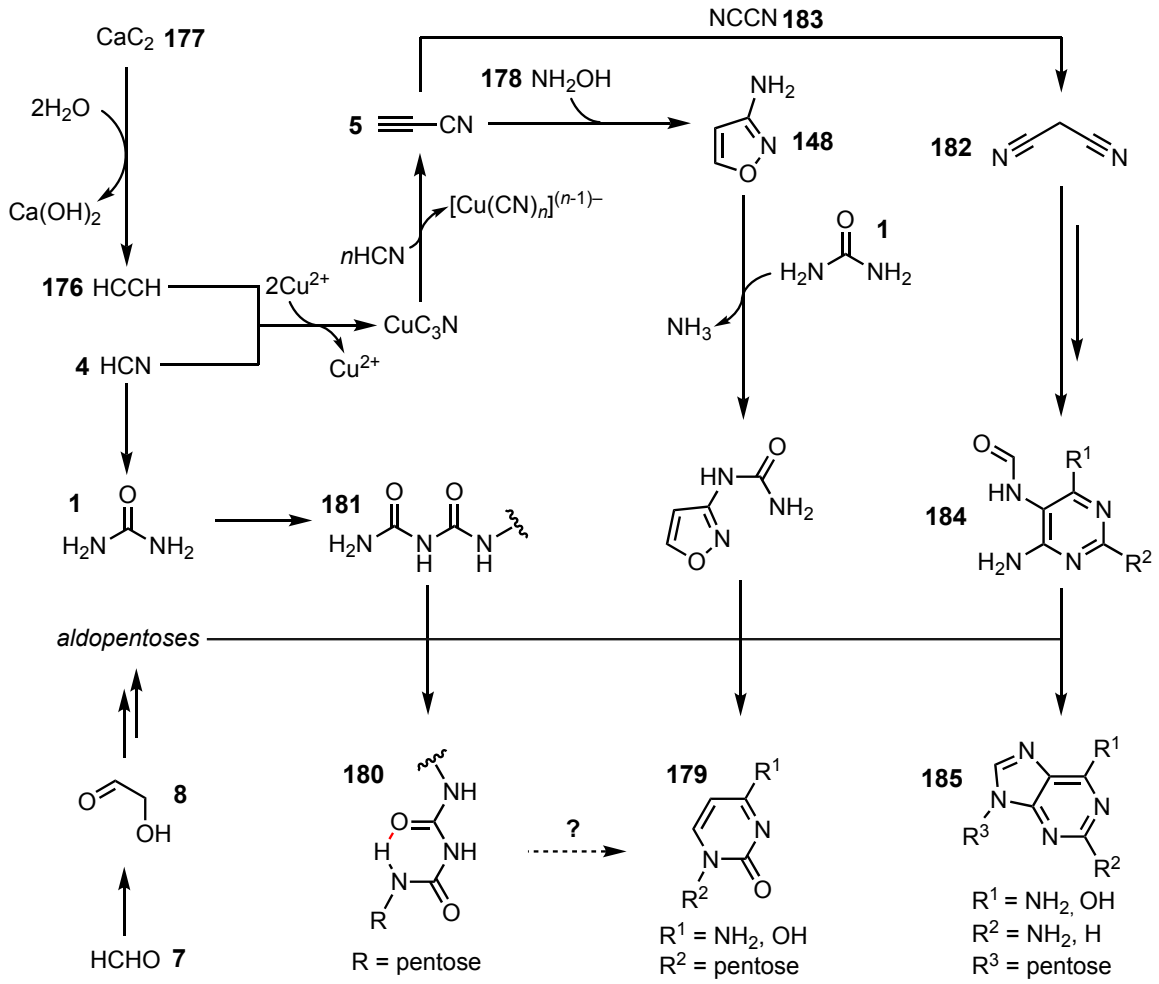
In order to invoke chemistry such as that which is described here, a substantial feedstock of HCN must also have been delivered to Earth's surface in a concentrated manner. To accomplish this, rain would not alone have been sufficient, as it would not have provided the steady concentrations of HCN that would have been necessary to accomplish known prebiotically plausible reactions.[144] A more plausible alternative, would involve the capture of HCN by Fe^{II} in groundwater to provide ferrocyanide $[Fe(CN)_6]^{4-}$.[144] This type of transformation would have been readily achievable on the early Earth, given the likely abundance of ferrous iron and various mechanisms by which HCN could have been supplied.[145] The formation of ferrocyanide is furthermore a very favourable process, and can be expected in bodies of water containing Fe^{2+} that are exposed to trace HCN.[144] Ferrocyanide further provides the advantage that it can facilitate the reductive homologation of HCN in the presence of sulphite (also critical to our pyrimidine synthesis), to provide simple sugars, hydroxy acids, and amino acid precursors.[145] The heating of dry-ferrocyanide salts that are distributed within

sedimentary terrain can also generate important sources of inorganic-carbon, such as calcium cyanamide and calcium carbide ($CaNCN + CaC_2$).[146] Finally, the presence of nitrite (from $NO^{\bullet}$) could facilitate further chemistry beyond that of hydroxylamine, which can be generated *via* the Raschig process, or by those described here.[147] Phosphate, the probable precursor to nucleotides and phospholipids, may have been available at many stages of the Hadean Earth's timeline due to its stability, and could derive from anoxic corrosion of minerals such as meteoric schreibersite or apatite.[148] It is understood that the success and sequence of reactions arising from such inorganic components would depend very much on their order of addition. This therefore raises the question of whether the majority of important prebiotic chemistry occurred predominantly in "one pot", or whether these processes were facilitated by sequential delivery of reagents to support selective synthesis.[146] If the latter were to be true, then this could have been supported by certain geochemical mechanisms such as wet and dry cycles.

4.2 Wet-Dry Cycles and Hypothetical Conditions for Prebiotic Nucleoside Synthesis

Another key finding described within this doctoral thesis, relates to how early nucleoside-forming processes were likely contingent upon wet-dry cycles to drive their activity.[102,121,124] Extant life on Earth is understood to rely heavily upon proteins, which facilitate the various chemical reactions necessary to keep our cells alive. Enzymes are able to accomplish these tasks by supplying a steady input of chemical- and photochemical energy, thereby allowing chemical processes inside of those cells to exist out of thermodynamic equilibrium with their neighbouring environments. Formation of biomolecules from monomers is often meanwhile characterised by relatively large free energies of formation, with ΔGf° values ranging between +2 to +4 kCal/mol for peptides, or as high as +5.3 kCal/mol in the case of DNA phosphodiester bonds.[149–151] While these large endergonic barriers are overcome by enzymes in *de novo* processes, such would have been impossible for the formation of the earliest prebiotic nucleosides and RNA. At least in the case of prebiotic ribosylations, it was found here that wet-dry cycles are tremendously effective in overcoming these barriers, and thus delivering a variety of canonical, and modified nucleosides in substantial yields.[102,121,124] In the "wet-dry cycle" model, phases of cool temperatures that accumulate liquid H_2O are iteratively followed by periods that are hot and result in evaporation. The "dry" phase of each cycle thus drives thermodynamically-uphill condensation reactions, whereas the "wet" phase delivers fresh solvent, permitting better diffusion of reactants than would otherwise be possible in solid mixtures.[152] A recognisable limitation of wet-dry cycles in the context of prebiotic chemistry relates to their dependence upon unpredictable rehydration events like rainstorms or flooding. The prebiotic plausibility of wet-dry cycles is however clearly feasible given the magmatic, water-rich environment that likely existed on

the earth between 2.4 to 4 billion years ago.[153] If life indeed emerged on the Hadean Earth shortly after the decline of Late Heavy Bombardment, as is frequently invoked,[148,154] then geothermal activity, impact heating, surface water, and indeed rain and flooding would have been present elements simultaneously at life's origin.[155] Based on these environmental factors, a hypothetical prebiotic scenario leading to formation of the aforementioned canonical and noncanonical nucleosides is presented in **Scheme 39**.



Scheme 39 *Plausible prebiotic routes to purine and pyrimidine nucleosides, as investigated in this thesis.*

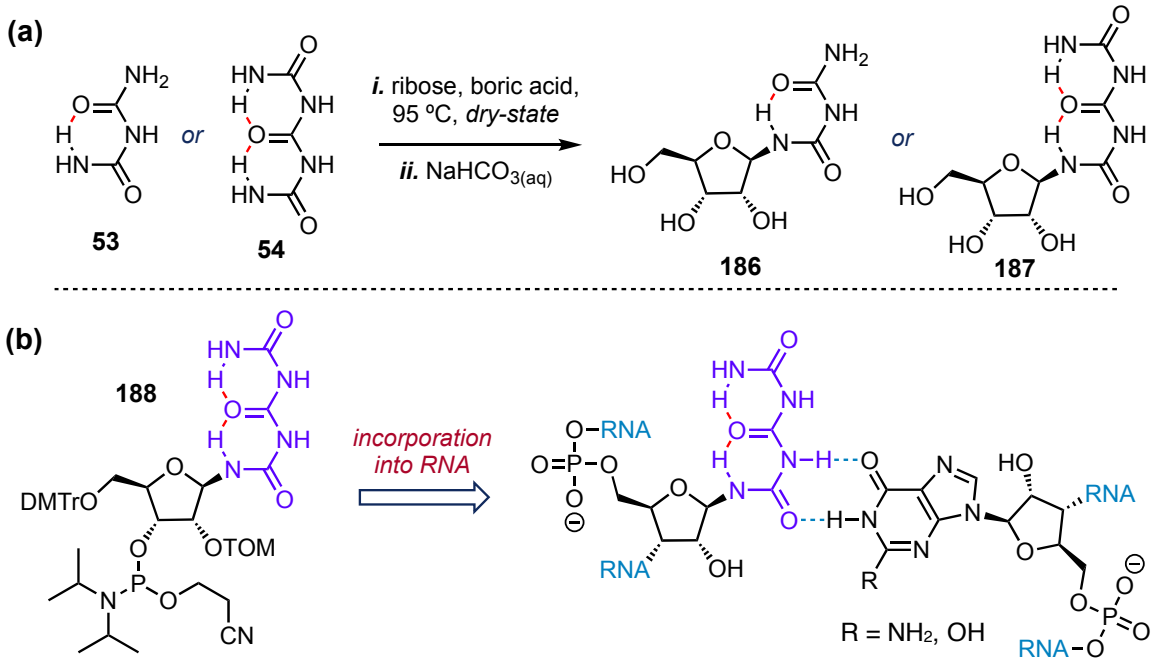
As originally proven by Wöhler,[156] acetylene **176** can be generated under prebiotic conditions *via* the hydration of CaC_2 **177**, which can itself arise from heating $CaCN_2$ to elevated temperatures (1000 ºC).[144] Acetylene **176** is also known to react with hydrogen cyanide **4** in the presence of copper (II) to generate cyanoacetylene **5**, *via* its Cu-coordinated precursor.[157] Cyanoacetylene **5** formed through this pathway is reportedly highly stable, and can be released into mildly acidic aqueous solutions (for example through the dissolution of HCN **4**) to generate highly concentrated mixtures available for nucleoside formation. Upon its reaction with the elements of hydroxylamine **178**, either as the disulfonate **175**, or as hydroxylurea **143**, 3-aminoisoxazole **148** can be obtained in substantial yields

(up to 90 % under alkaline conditions).[121] This key building block can then be further modified to obtain the pyrimidine ribonucleosides **179** *via* a sequence of condensation and redox reactions driven by wet-dry cycles, as demonstrated in this work. The discussed poly-urea nucleosides **180** can also be obtained *via* similar chemistry, when instead employing nucleophiles of the form **181**, which are the pyrolysis products of urea **1**. Malonitrile **182** can meanwhile be obtained *via* the gas phase reaction of acetonitrile and cyanogen (NCCN) **183**, the former of which is itself derived from cyanoacetylene **5** under prebiotic conditions, as described by Trinks and Eschenmoser.[71] Aminomalonitrile, a candidate precursor to amino acids as well as for guanosine, can also be readily obtained by trimerization of HCN **4**.[158] These building blocks can then be elaborated to provide formamidopyrimdines **184**, the precursors to purines **185**.[70] With these components present, it would thus be possible to form both pyrimidine and purine nucleosides, as well as such compounds as **180**, *via* the chemistry described in our published works.[102,121,124]

4.3 Poly-urea Wöhler RNA as an Informational Biopolymer

The immense complexity in identifying prebiotically plausible routes to the canonical nucleosides has led to the idea that RNA itself may have been predated by proto-RNA derived from simpler, more readily available starting materials.[83] This prospect also implies that such a polymer could spontaneously assemble from simple precursors.[86] Of particular interest to us in this regard were the prebiotic origins of urea and its possible role in the formation of proto-RNA. Inspired by revolutionary studies from Eschenmoser and others,[88] we therefore asked the question of whether an RNA-like polymer could plausibly arise exclusively from urea and simple carbohydrates. In one of the most celebrated experiments in history, Wöhler unintentionally discovered that urea can be prepared under abiotic conditions, whilst attempting to study the properties of ammonium cyanate salt.[38] In addition to Wöhler's synthesis, urea and as its related compound guanidine can be prepared efficiently from cyanide and cyanamide under plausible prebiotic conditions, as shown by Oró and Kimball in 1961.[159] Urea can also be generated in large quantities through spark-discharge experiments such as those used in the pioneering works of Miller.[43] The prebiotic plausibility of urea is thus well established, and one can reasonable assume that urea may have been abundantly available as a nucleoside precursor on the Early earth.[39] This availability leads to interesting scenarios, such as the formation of eutectic solutions with unique properties that can favour a variety of prebiotic reactions.[160] An abundance of further evidence suggests that urea may have been a critical component in chemical evolution, including its ability to affect phosphorylation reactions in the presence of inorganic phosphate.[161–163]

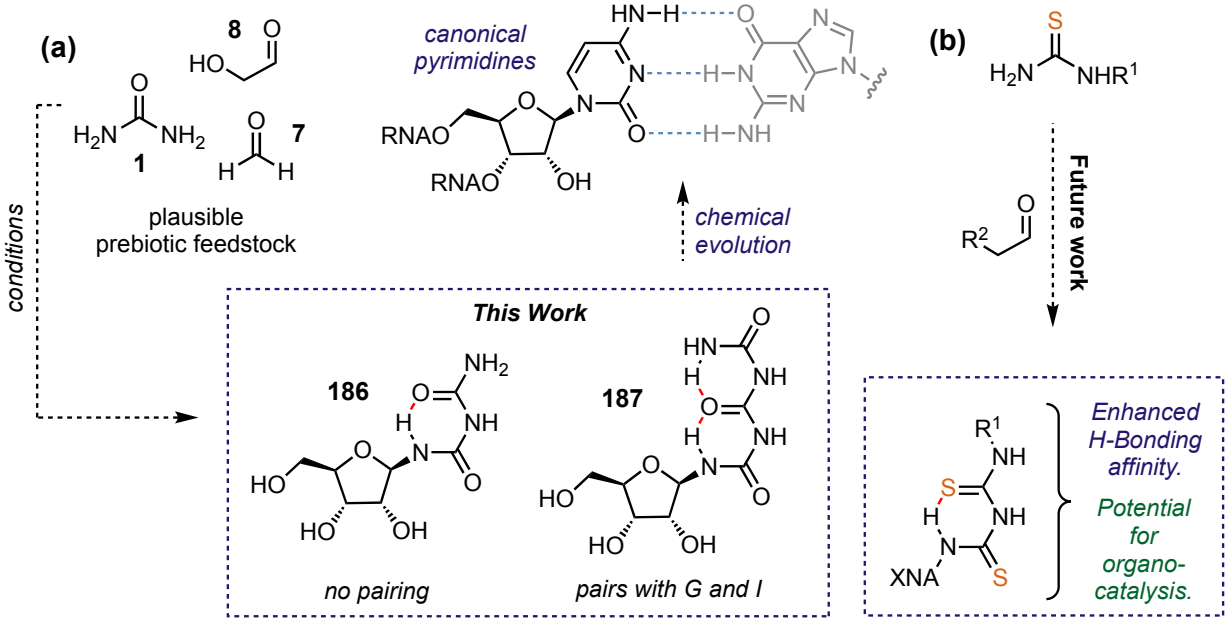
Given the unique physicochemical properties of urea **1**, together with its probable prebiotic availability, we set out to investigate whether an informational-, H-bonding biopolymer derived from **1** could have plausibly existed on the early Earth. Since **1** has itself been known to react directly with ribose **6** under mild acidic conditions,[164] it was anticipated that the same would be true for biuret **53** and triuret **54**, the pyrolysis products of urea **1**. To this end, we first showed that the H-bonded nucleosides **186** and **187** can be generated by treating either biuret or triuret with ribose under prebiotically plausible dry-down conditions (**Scheme 40**).[124] Remarkably, in addition to detecting both products in our experiments, we also observed that the resultant nucleosides were very stable, perhaps due to the nature of their intramolecular, pi-conjugated network of H-bonds.[165] This stability was subsequently verified *via* synthesis of phosphoramidite derivatives of **186** and **187**, and their incorporation into RNA strands. Ensuing ^{1}H NMR and UV melting studies not only confirmed the anticipated planar conformation of the folded pseudobase **187**, but also revealed its capacity to base-pair with the canonical nucleoside guanosine and related non-canonical nucleoside inosine. In reference to the pioneering work of Friedrich Wöhler, we termed the modified oligonucleotides *Wöhler RNA*.



Scheme 40 (a) *Synthesis of the nucleosides **186** and **187** under prebiotically plausible conditions.* **(b)** *Incorporation of the triuret phosphoramidite **188** into RNA, and its base-pairing interactions with guanosine and inosine.*

Our results here emphasise the important question of whether the first information-encoding oligomers were derived from different building blocks to the canonical nucleobases. If we consider RNA itself to be a product of Darwinian selection for properties that are best suited for contemporary

life, then the earliest proto-RNAs may have possessed very different functional characteristics. Hydrolytic stability, ease of assembly, availability of precursor molecules, and versatility in forming catalytic structures are examples of properties that would have better suited proto-RNA, and that may have since been refined by evolution.[86] In this respect, structures resembling **186** and **187** are perhaps more likely candidates for the first information-encoding monomers. Future studies will therefore focus on whether building blocks such as **53** and **54** can be generated from even simpler aldose-precursors (**Scheme 41a**). One potential shortcoming of our model is the fact that triuret-G/I base-pairs are not as stable as canonical Watson-Crick base pairs, being only comparable in strength to UG wobble pairs. It should be noted, however, that informational flexibility may have presented distinct advantages during the earliest stages of life.[83] If for example, a proto-RNA could shift conformation to achieve different pairing characteristics, then individual recognition units might be able to enhance the stability of the polymer as a whole. This type of degeneracy is also not unheard of in contemporary biology, where in translation for example, each particular amino acid may be decoded by more than just one codon.[166] Certain primordial conditions, such as the presence of high-salt concentrations may have also favoured such alternative recognition units in the absence of highly specific base-pairs.



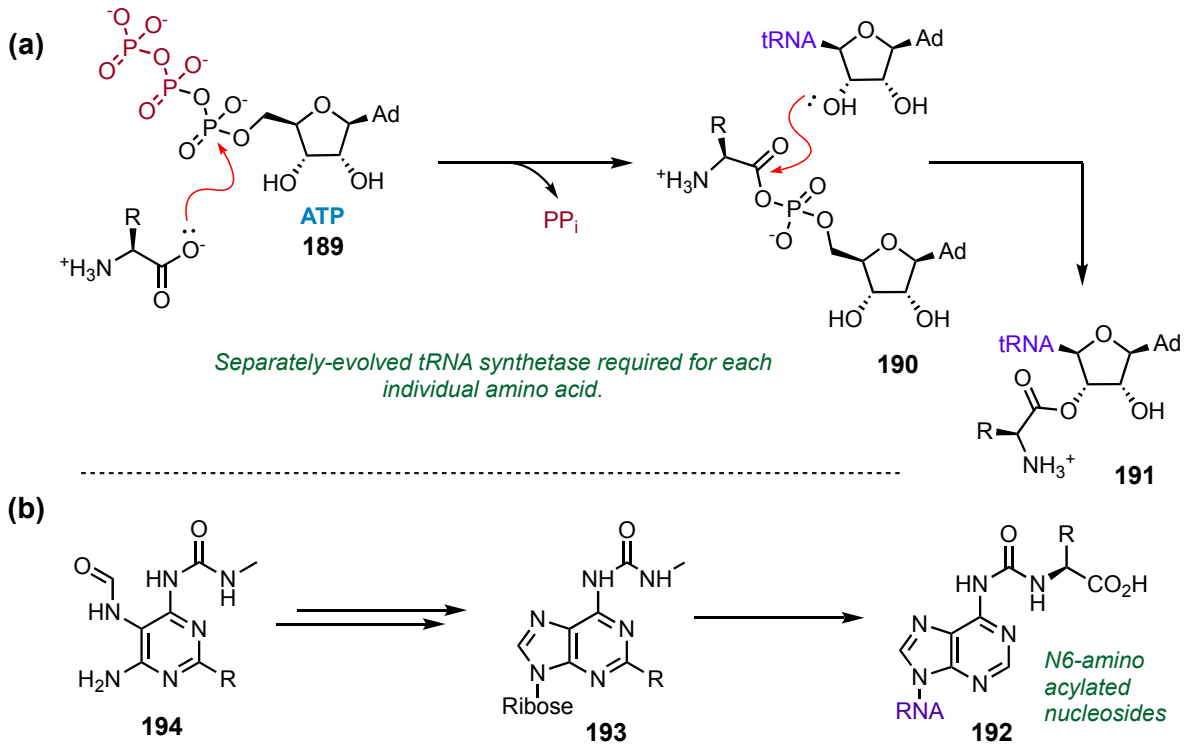
Scheme 41 (a) *A hypothetical route to Wöhler RNA nucleosides, deriving from simple precursors, and possible evolutionary transition to Watson-Crick bases.* **(b)** *Future objectives of exploring XNAs derived from thiourea and simple carbohydrates.*

In order to explore this notion further, future studies will partly focus on identifying conditions that allow nucleobases such as **186** and **187** to act as a universal recognition unit in the presence of other

nucleobases. To this ends, chemical modifications that enhance H-bonding affinity and might support catalysis, such as the incorporation of thiourea moieties, will also be explored (**Scheme 41b**). Finally, we aim to explore whether even simpler biopolymers derived from plausible backbones including GNA, TNA, and FNA are possible.[167] These findings will be reported in due course.

4.4 Noncanonical Nucleosides and Coevolution of the Translational System

As already noted, the chemical origins and evolution of the translational system is a tremendous and formidable problem. Central, yet unanswered questions include: why is the code universal; why is it derived from triplets; and why are there no more or less than 20 amino acids that are universally encoded?[168] To begin to tackle these problems, we aimed to approach the evolution of translation from a chemical standpoint. If the RNA World hypothesis were to be true, then one can also reasonably assume that informational decoding in the earliest translation systems was carried out by RNA molecules that were the evolutionary precursors of today's tRNAs.[168] It is therefore also conceivable that these proto-tRNAs possessed entirely different primary structural components to modern tRNAs. In the contemporary translational system, 20 unique and separately-evolved aminoacyl–tRNA synthetase (aaRS) each recognise a separate tRNA, each of which are catalytically charged with an associated amino acid **(Scheme 42a)**. This aminoacyl-charging occurs by ATP (**189**)-mediated transfer *via* **190** to obtain the appropriate 3'-aminoacylated derivative **191**.



Scheme 42 (a) *De-novo synthesis of aminoacyl–tRNA by aminoacyl–tRNA synthetases*. **(b)** *Synthesis of N^6-aminoacylcarbamoyl adenosine nucleosides, as shown in this work.*

In order to circumvent the prebiotic chemical constraints associated with contemporary translation and peptide synthesis, we instead sought to explore an entirely new chemical pathway. Here, we showed that N^6-aminoacylated adenosine nucleosides **192** can be obtained directly from N^6-carbamoylated building blocks (**193)**, which are themselves derived from FaPys **194 (Scheme 42b)**.[102] We also showed that that the same type of transformation can be achieved by direct carbamoylation of adenosine nucleosides, with nucleophilic substitution of N-methyl-N-nitrosoureas.[101] We next developed syntheses of phosphoramidite building blocks that allowed us to incorporate the aforementioned N^6-aminoacylated adenosine nucleosides into RNA strands.[130] Through these studies, we presented how alternating aminoacylated adenine bases positioned within an RNA-duplex might align in the major groove, thus being primed for peptide formation. Given appropriate activation conditions, for example *via* catalytically induced aminoacyl-cyclisation, it might in principal be possible to generate peptides entirely from modified RNA. With this in mind, future investigations will be aimed towards accomplishing chemical peptide synthesis from RNA. These studies are now underway in the Carell laboratories.

4.5 Concluding Remarks

The RNA World hypothesis has for decades been the most widely accepted view for the chemical origins of life on Earth.[19] Through this and other work, there is now ample evidence to show that RNA's building blocks are capable of arising in the absence of cellular life, given the right set of chemical conditions. Furthermore, this evidence undeniably supports the conclusion that many different non-canonical nucleotides were also likely present on the Early earth.[169] What remains yet unclear, is how simpler biopolymers eventually transitioned to those present in life today. When considering nucleoside syntheses involving the coupling of trivalent nucleophiles to sugars, such as those explored in this thesis, the exact chemical mechanism by which ribose emerged also becomes ever-pressing. In our view, the best hope for substantial progress in these and related efforts, lie with technically challenging but conceptually clear experimentation. That is to say, only through experimental synthesis can we truly delineate which prebiotic chemical pathways were once possible. It will therefore be exciting to see how the field develops in the coming decades.

5 REFERENCES

Citation Convention: *Angew. Chem. Int. Ed.*

[1] L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99–123.

[2] D. Prieur, G. Erauso, C. Jeanthon, *Planet. Space Sci.* **1995**, *43*, 115–122.

[3] R. N. Glud, F. Wenzhöfer, M. Middelboe, K. Oguri, R. Turnewitsch, D. E. Canfield, H. Kitazato, *Nat. Geosci.* **2013**, *6*, 284–288.

[4] D. C. Demirjian, F. Morís-Varas, C. S. Cassidy, *Curr. Opin. Chem. Biol.* **2001**, *5*, 144–151.

[5] M. S. Dodd, D. Papineau, T. Grenne, J. F. Slack, M. Rittner, F. Pirajno, J. O'Neil, C. T. S. Little, *Nature* **2017**, *543*, 60–64.

[6] N. Lane, *Life Ascending: The Ten Great Inventions of Evolution*, Profile Books, **2010**.

[7] J. Geiss, A. P. Rossi, *Astron. Astrophys. Rev.* **2013**, *21*, 21–68.

[8] N. H. Sleep, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*.

[9] A. C. Barr, *J. Geophys. Res. Planets* **2016**, *121*, 1573–1601.

[10] B. Cavalazzi, R. Barbieri, in *Cnidaria, Past, Present Futur.*, **2016**, pp. 3–13.

[11] E. G. Nisbet, N. H. Sleep, *Nature* **2001**, *409*, 1083–1091.

[12] M. F. Hohmann-Marriott, R. E. Blankenship, *Annu. Rev. Plant Biol.* **2011**, *62*, 515–548.

[13] W. F. Martin, S. Garg, V. Zimorski, *Philos. Trans. R. Soc. B Biol. Sci.* **2015**, *370*.

[14] D. E. G. Briggs, *Curr. Biol.* **2015**, *25*, R864–R868.

[15] F. Tian, O. B. Toon, A. A. Pavlov, H. De Sterck, *Science* **2005**, *308*, 1014–1017.

[16] A. Eschenmoser, M. Volkan Kisakürek, *Helv. Chim. Acta* **1996**, *79*, 1249–1259.

[17] S. A. Benner, *Astrobiology* **2010**, *10*, 1021–1030.

[18] C. E. Cleland, C. F. Chyba, *Orig. Life Evol. Biosph.* **2002**, *32*, 387–393.

[19] G. F. Joyce, in *Orig. Life Cent. Concepts*, **1994**.

[20] J. W. Szostak, *J. Biomol. Struct. Dyn.* **2012**, *29*, 599–600.

[21] F. Crick, *Nature* **1970**, *227*, 561–563.

[22] E. V. Koonin, *Biol. Direct* **2012**, *7*.

[23] D. M. J. Lilley, *Philos. Trans. R. Soc. B Biol. Sci.* **2011**, *366*, 2910–2917.

[24] W. Gilbert, *Nature* **1986**, *319*, 618.

[25] F. H. C. Crick, *J. Mol. Biol.* **1968**, *38*, 367–379.

[26] L. E. Orgel, *J. Mol. Biol.* **1968**, *38*, 381–393.

[27] T. R. Cech, *Angew. Chem. Int. Ed.* **1990**, *29*, 759–768.

[28] S. Altman, *Angew. Chem. Int. Ed.* **1990**, *29*, 749–758.

[29] R. Wheeler, "Structure of haemoglobin," can be found under

https://commons.wikimedia.org/wiki/File:1GZX_Haemoglobin.png, **2007**.

[30] R. Weiss, J. Cherry, *Cold Spring Harb. Monogr. Ser.* **1993**, *24*, 71–89.

[31] M. Di Giulio, *J. Mol. Evol.* **1997**, *45*, 571–578.

[32] L. E. Orgel, J. E. Sulston, *Kimball, AP Orò, J* **1971**, 89–94.

[33] L. E. Orgel, *J. Theor. Biol.* **1986**, *123*, 127–149.

[34] H. B. White, *J. Mol. Evol.* **1976**, *7*, 101–104.

[35] H. B. White, in *The Pyridine Nucleotide Coenzymes*, **1982**, pp. 1–17.

[36] A. Kirschning, *Angew. Chem. Accept. Author Manuscr.* **2020**.

[37] E. Domingo, in *Virus as Popul.*, **2020**, pp. 1–33.

[38] F. Wöhler, *Ann. Phys.* **1828**, *88*, 253–256.

[39] C. Menor-Salván, **2018**, pp. 85–142.

[40] W. H. Warren, *J. Chem. Educ.* **1928**, *5*, 1539–1553.

[41] A. Oparin, *Proischogdenie Zhizni. Translation: "Origin of Life,"* Moscow, **1924**.

[42] K. Kvenvolden, J. Lawless, K. Pering, E. Peterson, J. Flores, C. Ponnamperuma, I. R. Kaplan, C. Moore, *Nature* **1970**, *228*, 923–926.

[43] S. L. Miller, *Science* **1953**, *117*, 528–529.

[44] T. M. McCollom, *Annu. Rev. Earth Planet. Sci.* **2013**, *41*, 207–229.

[45] A. P. Johnson, H. J. Cleaves, J. P. Dworkin, D. P. Glavin, A. Lazcano, J. L. Bada, *Science* **2008**, *322*, 404.

[46] P. H. Abelson, *Proc. Natl. Acad. Sci.* **1966**, *55*, 1365–1372.

[47] *Science* **1966**, *154*, 784–785.

[48] J. P. Ferris, W. J. Hagan, *Tetrahedron* **1984**, *40*, 1093–1120.

[49] A. Butlerow, *Justus Liebigs Ann. Chem.* **1861**, *120*, 295–298.

[50] S. Cannizzaro, *Justus Liebigs Ann. Chem.* **1853**, *88*, 129–130.

[51] C. A. L. de Bruyn, W. A. van Ekenstein, *Recl. des Trav. Chim. des Pays-Bas* **1895**, *14*, 203–216.

[52] I. V. Delidovich, A. N. Simonov, O. P. Taran, V. N. Parmon, *ChemSusChem* **2014**, *7*, 1833–1846.

[53] M. Yadav, R. Kumar, R. Krishnamurthy, *Chem. Rev.* **2020**.

[54] M. Haas, S. Lamour, S. B. Christ, O. Trapp, *Commun. Chem.* **2020**, *3*, 140.

[55] J. S. Teichert, F. M. Kruse, O. Trapp, *Angew. Chem. Int. Ed.* **2019**, *58*, 9944–9947.

[56] E. Torrents, P. Aloy, I. Gibert, F. Rodríguez-Trelles, *J. Mol. Evol.* **2002**, *55*, 138–152.

[57] S. Miller, H. Cleaves, *Syst. Biol. (Stevenage).* **2006**, *1*, 3–56.

[58] A. W. Schwartz, A. B. Voet, M. Van Der Veen, *Orig. Life* **1984**, *14*, 91–98.

[59] J. Oró, *Biochem. Biophys. Res. Commun.* **1960**, *2*, 407–412.

[60] R. Saladino, C. Crestini, G. Costanzo, R. Negri, E. Di Mauro, *Bioorganic Med. Chem.* **2001**, *9*, 1249–1253.

[61] M. Levy, S. L. Miller, K. Brinton, J. L. Bada, *Icarus* **2000**, *145*, 609–613.

[62] A. W. Schwartz, H. Joosten, A. B. Voet, *BioSystems* **1982**, *15*, 191–193.

[63] R. Sanchez, J. Ferris, L. E. Orgel, *Science* **1966**, *153*, 72–73.

[64] H. J. Cleaves, K. E. Nelson, S. L. Miller, *Naturwissenschaften* **2006**, *93*, 228–231.

[65] J. P. Ferris, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1968**, *33*, 693–704.

[66] R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1970**, *47*, 531–543.

[67] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239–242.

[68] S. Stairs, A. Nikmal, D.-K. Bucˇar, S.-L. Zheng, J. W. Szostak, M. W. Powner, *Nat. Commun.* **2017**, *8*, 15270.

[69] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1972**, *67*, 25–33.

[70] S. Becker, I. Thoma, A. Deutsch, T. Gehrke, P. Mayer, H. Zipse, T. Carell, *Science* **2016**, *352*, 833–836.

[71] U. P. Trinks, A. Eschenmoser, Zur Chemie Der Aminopyrimidine, ETH Zurich, **1987**.

[72] S. A. Benner, H. J. Kim, M. A. Carrigan, *Acc. Chem. Res.* **2012**, *45*, 2025–2034.

[73] B. E. Prieur, *Comptes Rendus l'Academie des Sci. - Ser. IIc Chem.* **2001**, *4*, 667–670.

[74] A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, *Science* **2004**, *303*, 196.

[75] H. Grosjean, *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*, **2009**.

[76] M. C. Weiss, F. L. Sousa, N. Mrnjavac, S. Neukirchen, M. Roettger, S. Nelson-Sathi, W. F. Martin, *Nat. Microbiol.* **2016**, *1*, 16116.

[77] Z. Liu, L. Rigger, J. C. Rossi, J. D. Sutherland, R. Pascal, *Chem. - A Eur. J.* **2016**, *22*, 14940–14949.

[78] S. H. Kim, F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H. J. Wang, N. C. Seeman, A. Rich, *Science* **1974**, *185*, 435–440.

[79] Yikrazuul, "X-ray structure of the tRNAPhe from yeast," can be found under https://commons.wikimedia.org/wiki/File:TRNA-Phe_yeast_1ehz.png, **2010**.

[80] F. Schuber, M. Pinck, *Biochimie* **1974**, *56*, 383–390.

[81] S. E. Walker, K. Fredrick, *Methods* **2008**, *44*, 81–86.

[82] M. Matuszewski, K. Debiec, E. Sochacka, *Chem. Commun.* **2017**, *53*, 7945–7948.

[83] N. V. Hud, B. J. Cafferty, R. Krishnamurthy, L. D. Williams, *Chem. Biol.* **2013**, *20*, 466–474.

[84] J. P. Ferris, *Orig. Life Evol. Biosph.* **1993**, *23*, 307–315.

[85] G. F. Joyce, A. W. Schwartz, S. L. Miller, L. E. Orgel, *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 4398–4402.

[86] A. E. Engelhart, N. V. Hud, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*.

[87] L. Zhang, A. Peritz, E. Meggers, *J. Am. Chem. Soc.* **2005**, *127*, 4174–4175.

[88] K. U. Schoning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser, *Science* **2000**, *290*, 1347–1351.

[89] Y. W. Yang, S. Zhang, E. O. McCullum, J. C. Chaput, *J. Mol. Evol.* **2007**.

[90] P. E. Nielsen, *Chem. Biodivers.* **2007**, *4*, 1996–2002.

[91] B. D. Heuberger, C. Switzer, *Chembiochem* **2008**, *9*, 2779–2783.

[92] K. Groebke, J. Hunziker, W. Fraser, L. Peng, U. Diederichsen, K. Zimmermann, A. Holzner, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1998**, *81*, 375–474.

[93] B. Han, B. Jaun, R. Krishnamurthy, A. Eschenmoser, *Org. Lett.* **2004**, *6*, 3691–3694.

[94] J. A. Piccirilli, S. A. Benner, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33–37.

[95] V. M. Kolb, J. P. Dworkin, S. L. Miller, *J. Mol. Evol.* **1994**, *38*, 549–557.

[96] B. J. Cafferty, D. M. Fialho, J. Khanam, R. Krishnamurthy, N. V. Hud, *Nat. Commun.* **2016**, *7*, 11328.

[97] A. Diez-Martinez, E. K. Kim, R. Krishnamurthy, *J. Org. Chem.* **2015**, *80*, 7066–7075.

[98] Y. Ni, F. Becquart, J. Chen, M. Taha, *Macromolecules* **2013**, *46*, 1066–1074.

[99] I. M. Lagoja, P. Herdewijn, *Chem. Biodivers.* **2007**, *4*, 818–822.

[100] T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110–7131.

[101] C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier, T. Carell, *Angew. Chem. Int. Ed.* **2018**, *57*, 5943–5946.

[102] S. Becker, C. Schneider, H. Okamura, A. Crisp, T. Amatov, M. Dejmek, T. Carell, *Nat. Commun.* **2018**, *9*, 163.

[103] S. Becker, C. Schneider, A. Crisp, T. Carell, *Nat. Commun.* **2018**, *9*, 5174.

[104] D. Tzeli, P. G. Tsoungas, I. D. Petsalakis, P. Kozielewicz, M. Zloh, *Tetrahedron* **2015**, *71*, 359–369.

[105] T. A. Turney, *J. Org. Chem.* **1957**, *22*, 1692–1693.

[106] I. Halasz, E. Mestrovic, H. Cicak, Z. Mihalic, H. Vancik, *J. Org. Chem.* **2005**, *70*, 8461–8467.

[107] H. Sigmund, W. Pfleiderer, *Helv. Chim. Acta* **1994**, *77*, 1267–1280.

[108] L. Bering, A. P. Antonchick, *Tetrahedron* **2019**, *75*, 1131–1143.

[109] S. J. Varma, K. B. Muchowska, P. Chatelain, J. Moran, *Nat. Ecol. Evol.* **2018**, *2*, 1019–1024.

[110] T. R. Cech, *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, 1–5.

[111] L. Orgel, *Science* **2000**, *290*, 1306–1307.

[112] M. P. Schweizer, G. B. Chheda, L. Baczynskyj, R. H. Halit, *Biochemistry* **1969**, *8*, 3283–3289.

[113] P. Forterre, *Front. Microbiol.* **2015**, *6*.

[114] K. Miyauchi, S. Kimura, T. Suzuki, *Nat. Chem. Biol.* **2013**, *9*, 105–111.

[115] A. A. B. Y. Dic, *Side Reactions Upon Amino Acid / Peptide Carboxyl Activation*, **2016**.

[116] M. Matuszewski, J. Wojciechowski, K. Miyauchi, Z. Gdaniec, W. M. Wolf, T. Suzuki, E. Sochacka, *Nucleic Acids Res.* **2017**, *45*, 2137–2149.

[117] M. Matuszewski, E. Sochacka, *Bioorganic Med. Chem. Lett.* **2014**, *24*, 2703–2706.

[118] L. Dalla Vechia, R. O. M. A. de Souza, L. S. de Mariz e Miranda, *Tetrahedron* **2018**, *74*, 4359–4371.

[119] E. N. Trifonov, *Gene* **2000**, *261*, 139–151.

[120] C. T. Lauhon, *Biochemistry* **2012**, *51*, 8950–8963.

[121] S. Becker, J. Feldmann, S. Wiedemann, H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov, T. Carell, *Science* **2019**, *366*, 76–82.

[122] R. Zelli, W. Zeinyeh, R. Haudecoeur, J. Alliot, B. Boucherle, I. Callebaut, J. L. Décout, *Org. Lett.* **2017**, *19*, 6360–6363.

[123] H. Vorbrüggen, C. Ruh-Pohlenz, in *Org. React.*, **1999**, pp. 1–630.

[124] H. Okamura, A. Crisp, S. Hübner, S. Becker, P. Rovó, T. Carell, *Angew. Chem. Int. Ed.* **2019**, *58*, 18691–18696.

[125] W. T. Markiewicz, M. Wiewiórowski, *Nucleic Acids Res.* **1978**, *1*, s185–s190.

[126] Á. Somoza, *Chem. Soc. Rev.* **2008**, *37*, 2668–2675.

[127] S. Pitsch, P. A. Weiss, L. Jenny, A. Stutz, X. Wu, *Helv. Chim. Acta* **2001**, *84*, 3773–3795.

[128] E. Pontiki, D. Hadjipavlou-Litina, A. Patsilinakos, T. M. Tran, C. M. Marson, *Future Med. Chem.* **2015**, *7*, 1937–1951.

[129] F. Himmelsbach, B. S. Schulz, T. Trichtinger, R. Charubala, W. Pfleiderer, *Tetrahedron* **1984**, *40*, 59–72.

[130] T. Carell, M. Nainytė, F. Müller, G. Ganazzoli, C.-Y. Chan, A. Crisp, D. Globisch, *Chem. – A Eur. J.* **2020**, Accepted Author Manuscript.

[131] T. Bavaro, S. Rocchietti, D. Ubiali, M. Filice, M. Terreni, M. Pregnolato, *European J. Org. Chem.* **2009**, 1967–1975.

[132] C. Schneider, Untersuchungen Zur Präbiotischen Entstehung von RNA Und Ihrer Modifikationen, Ludwig-Maximilian-University of Munich, **2019**.

[133] Y. Hisamatsu, K. Hasada, F. Amano, Y. Tsubota, Y. Wasada-Tsutsui, N. Shirai, S. I. Ikeda, K. Odashima, *Chem. - A Eur. J.* **2006**, *12*, 7733–7741.

[134] V. Serebryany, L. Beigelman, *Nucleosides, Nucleotides and Nucleic Acids* **2003**, *22*, 1007–1009.

[135] G. Leszczynska, P. Leonczak, A. Dziergowska, A. Malkiewicz, *Nucleosides, Nucleotides and Nucleic Acids* **2013**, *32*, 599–616.

[136] M. Sundaram, P. F. Crain, D. R. Davis, *J. Org. Chem.* **2000**, *65*, 5609–5614.

[137] J. M. Keith, R. Apodaca, M. Tichenor, W. Xiao, W. Jones, J. Pierce, M. Seierstad, J. Palmer, M. Webb, M. Karbarz, et al., *ACS Med. Chem. Lett.* **2012**, *3*, 823–827.

[138] P. L. Luisi, *Isr. J. Chem.* **2015**, *55*, 906–918.

[139] F. H. C. Crick, *J. Mol. Biol.* **1966**, *19*, 548–555.

[140] S. A. Benner, H. J. Kim, E. Biondi, *Life* **2019**, *9*.

[141] D. Trail, E. B. Watson, N. D. Tailby, *Nature* **2011**, *480*, 79–82.

[142] D. J. Stevenson, *Science* **1981**, *214*, 611–619.

[143] Y. L. Yung, M. B. McElroy, *Science* **1979**, *203*, 1002–1004.

[144] D. D. Sasselov, J. P. Grotzinger, J. D. Sutherland, *Sci. Adv.* **2020**, *6*.

[145] J. Xu, D. J. Ritson, S. Ranjan, Z. R. Todd, D. D. Sasselov, J. D. Sutherland, *Chem. Commun.* **2018**, *54*, 5566–5569.

[146] D. J. Ritson, C. Battilocchio, S. V. Ley, J. D. Sutherland, *Nat. Commun.* **2018**, *9*.

[147] F. Raschig, *Berichte der Dtsch. Chem. Gesellschaft (A B Ser.* **1923**, *56*, 206–208.

[148] M. A. Pasek, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 853–858.

[149] I. Mamajanov, P. J. Macdonald, J. Ying, D. M. Duncanson, G. R. Dowdy, C. A. Walker, A. E. Engelhart, F. M. Fernández, M. A. Grover, N. V. Hud, et al., *Macromolecules* **2014**, *47*, 1334–1343.

[150] R. Bruce Martin, *Biopolymers* **1998**, *45*, 351–353.

[151] K. S. Dickson, C. M. Burns, J. P. Richardson, *J. Biol. Chem.* **2000**, *275*, 15828–15831.

[152] T. D. Campbell, R. Febrian, J. T. McCarthy, H. E. Kleinschmidt, J. G. Forsythe, P. J. Bracher, *Nat. Commun.* **2019**, *10*.

[153] K. Zahnle, L. Schaefer, B. Fegley, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*.

[154] W. F. Bottke, M. D. Norman, *Annu. Rev. Earth Planet. Sci.* **2017**, *45*, 619–647.

[155] B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, J. D. Sutherland, *Nat. Chem.* **2015**, *7*, 301–307.

[156] F. Wöhler, *Justus Liebigs Ann. Chem.* **1862**, *124*, 220–220.

[157] J. D. Sutherland, *Angew. Chem. Int. Ed.* **2016**, *55*, 104–121.

[158] C. Kuroda, R. Kobayashi, S. Kurebayashi, K. Tazawa, A. Masuda, R. Takeuchi, A. Washio, H. Onuki, *Nat. Prod. Commun.* **2020**, *15*.

[159] J. Oró, A. P. Kimball, *Arch. Biochem. Biophys.* **1961**, *94*, 217–227.

[160] C. Menor-Salván, M. R. Marín-Yaseli, *Chem. Soc. Rev.* **2012**, *41*, 5404–5415.

[161] R. Lohrmann, L. E. Orgel, *Science* **1971**, *171*, 490–494.

[162] C. Menor-Salvan, M. Bouza, D. M. Fialho, B. T. Burcar, F. M. Fernández, N. V. Hud, *ChemBioChem* **2020**.

[163] J. L. Lago, B. T. Burcar, N. V. Hud, R. Febrian, C. Mehta, P. J. Bracher, Z. D. Atlas, M. A. Pasek, *Orig. Life Evol. Biosph.* **2020**, *50*.

[164] O. Björnberg, M. Vodnala, V. Domkin, A. Hofer, A. Rasmussen, G. Andersen, J. Piškur, *Nucleosides, Nucleotides and Nucleic Acids* **2010**, *29*, 433–437.

[165] Y. V. Novakovskaya, *Struct. Chem.* **2012**, *23*, 1253–1266.

[166] A. S. Rodin, E. Szathmáry, S. N. Rodin, *Biol. Direct* **2011**, *6*.

[167] C. H. Tsai, J. Chen, J. W. Szostak, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14598–14603.

[168] E. V. Koonin, A. S. Novozhilov, *Annu. Rev. Genet.* **2017**, *51*, 45–62.

[169] D. M. Fialho, T. P. Roche, N. V. Hud, *Chem. Rev.* **2020**, *120*, 4806–4830.

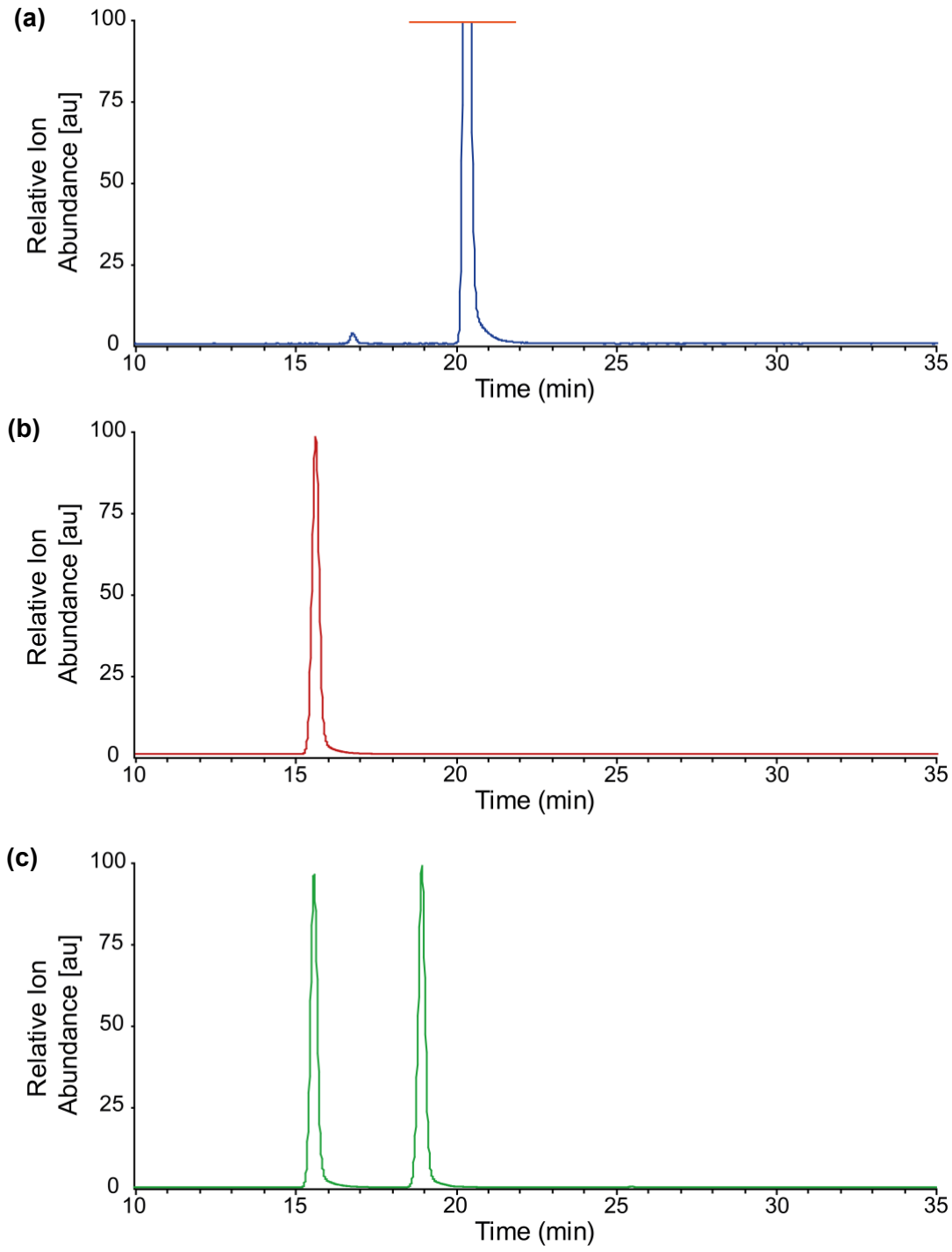


Figure S1 *Supplementary data to* ***Scheme 25*** *showing HPLC-MS chromatogram data for* **(a)** *the reaction of compound* ***57*** *with cyclic trimetaphosphate in the presence of* $MgCl_2$, **(b)** *compound* ***110*** *only, and* **(c)** *co-injection of the reaction mixture from the first chromatogram, and compound* ***110****. Compound identities were further confirmed by their identical [M+H] ion measurements.*

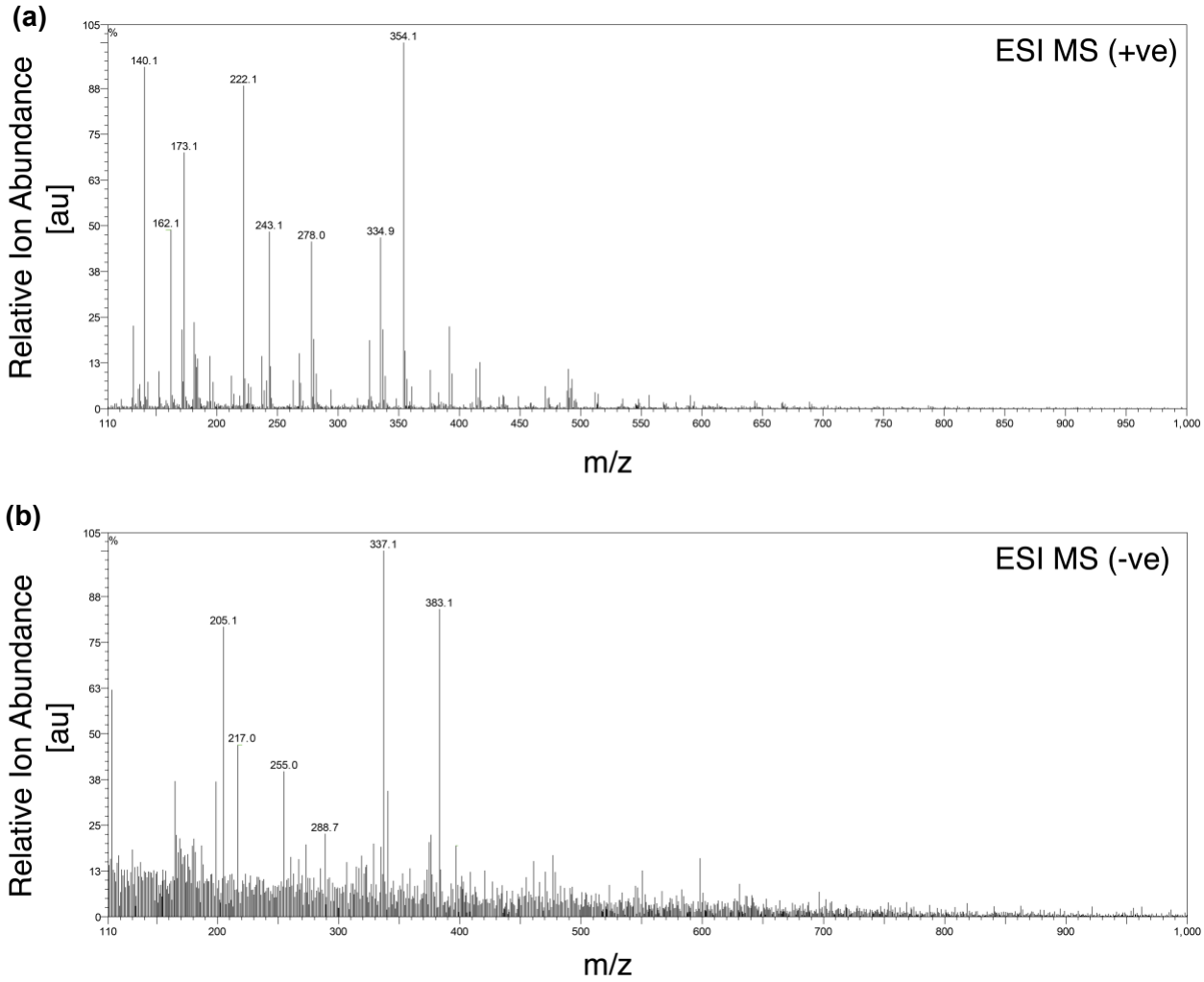
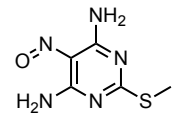
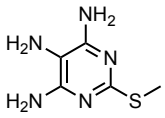


Figure S2 *Supplementary data to* ***Scheme 19*** *showing preliminary ESI data for* **(a)** *formation of compound* ***58b*** *(m/z = 354.1), as well as* **(b)** *formation of compound* ***59b*** *(m/z = 337.1).*

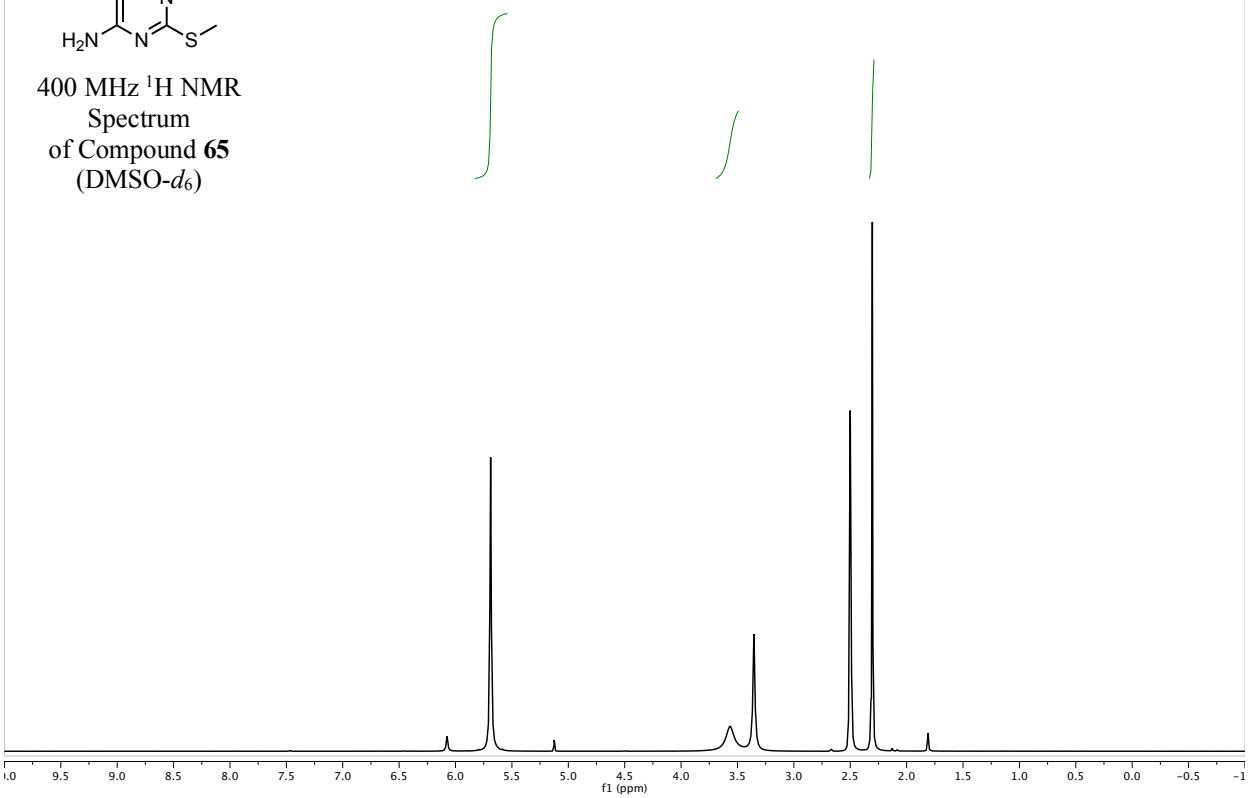
SUPPORTING NMR SPECTRA FOR UNPUBLISHED RESULTS

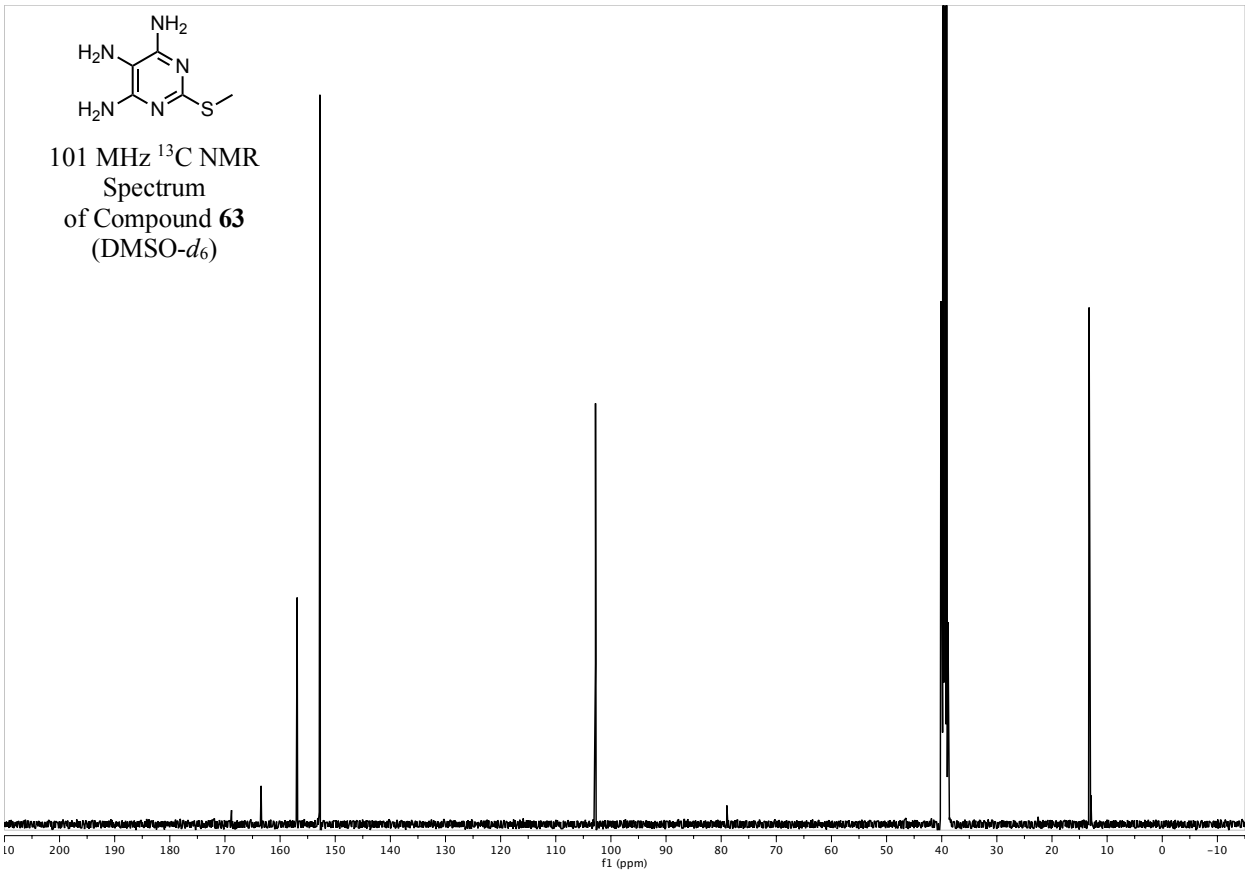


400 MHz ^{1}H NMR Spectrum of Compound **65** (DMSO-d_6)

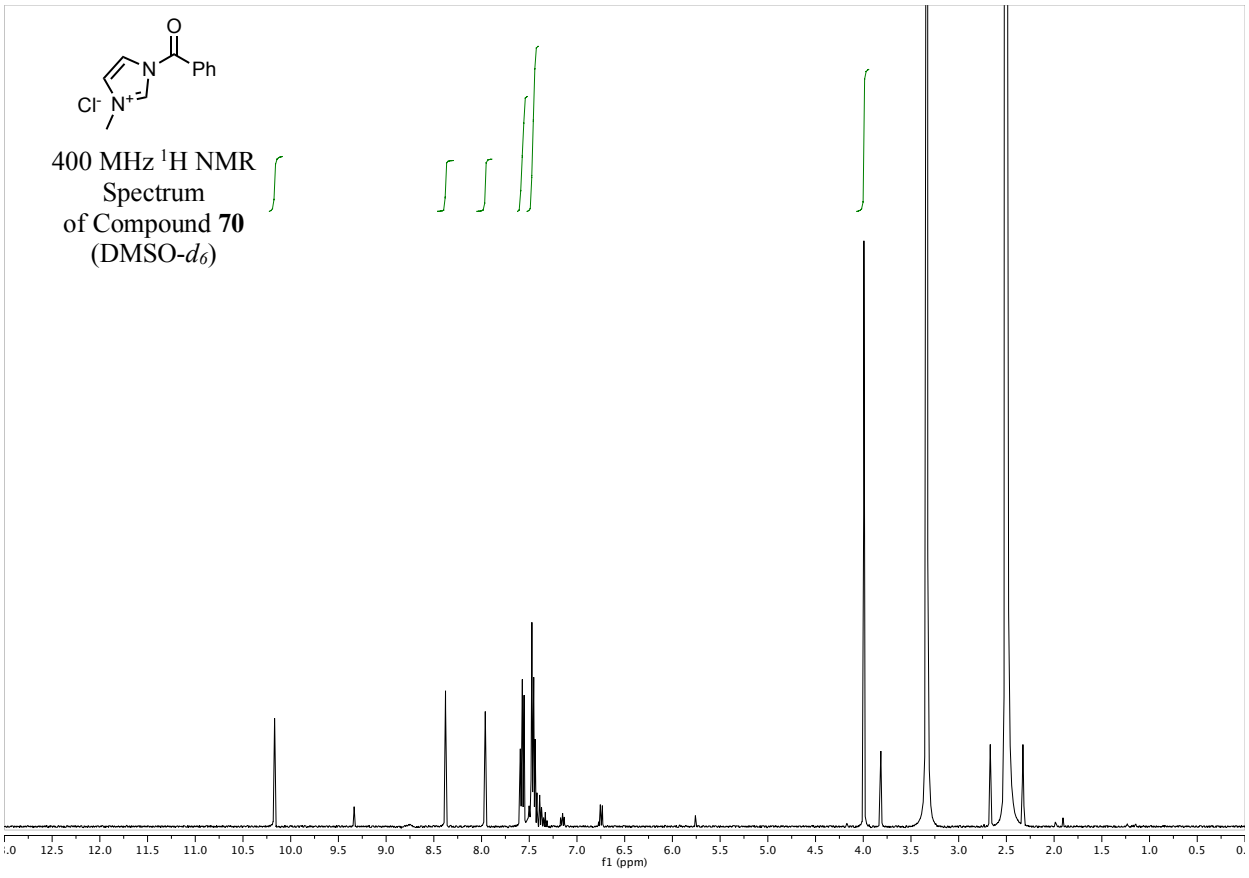


400 MHz ^{1}H NMR Spectrum of Compound **65** (DMSO-d_6)

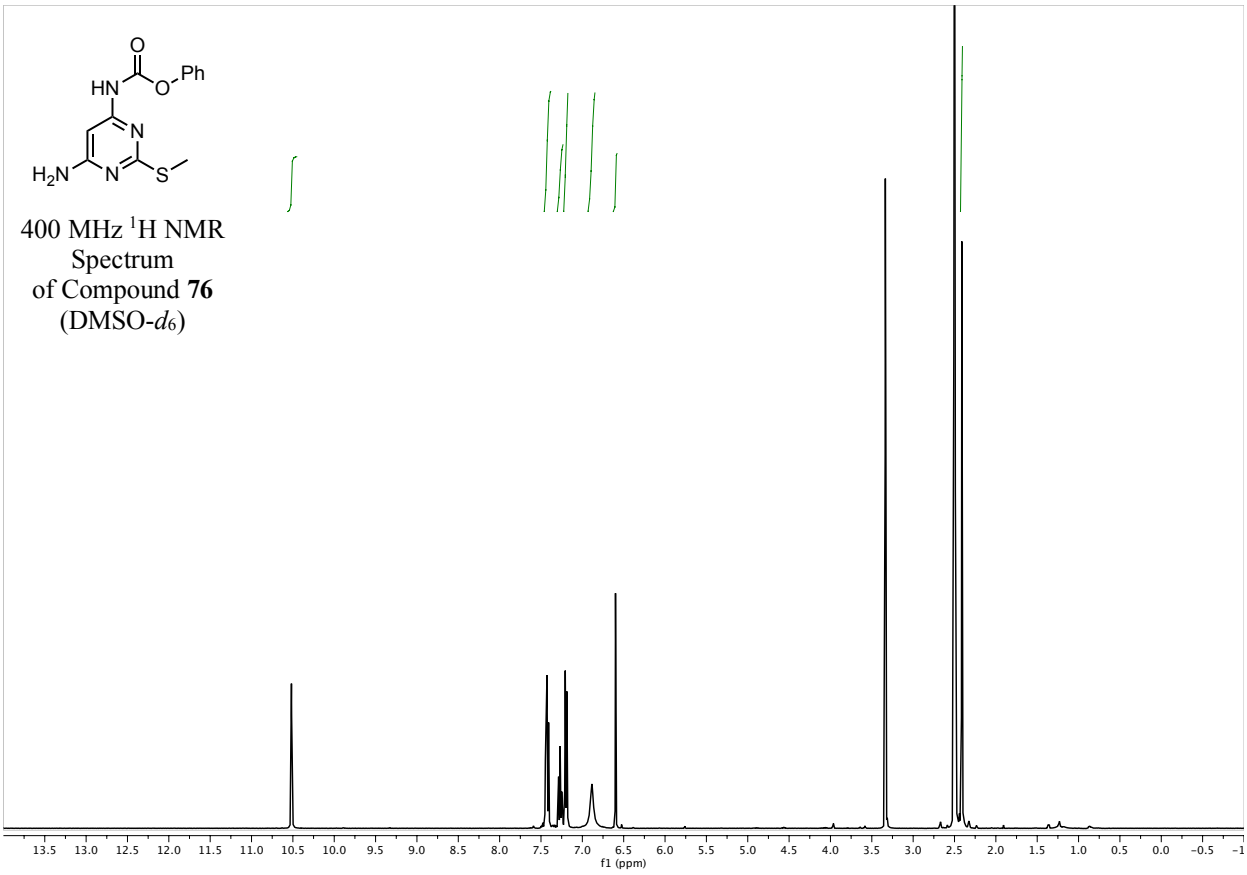




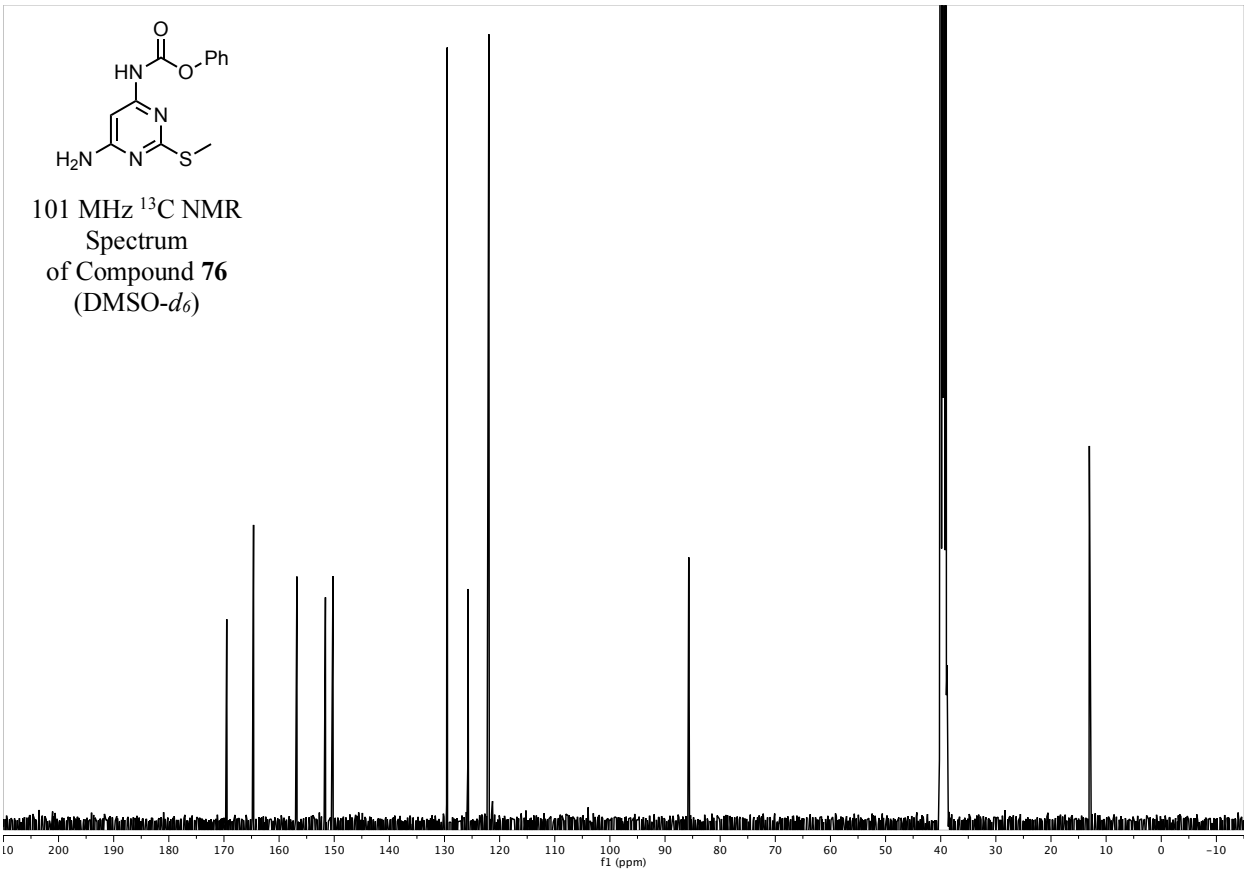
101 MHz ^{13}C NMR Spectrum of Compound **63** (DMSO-d_6)



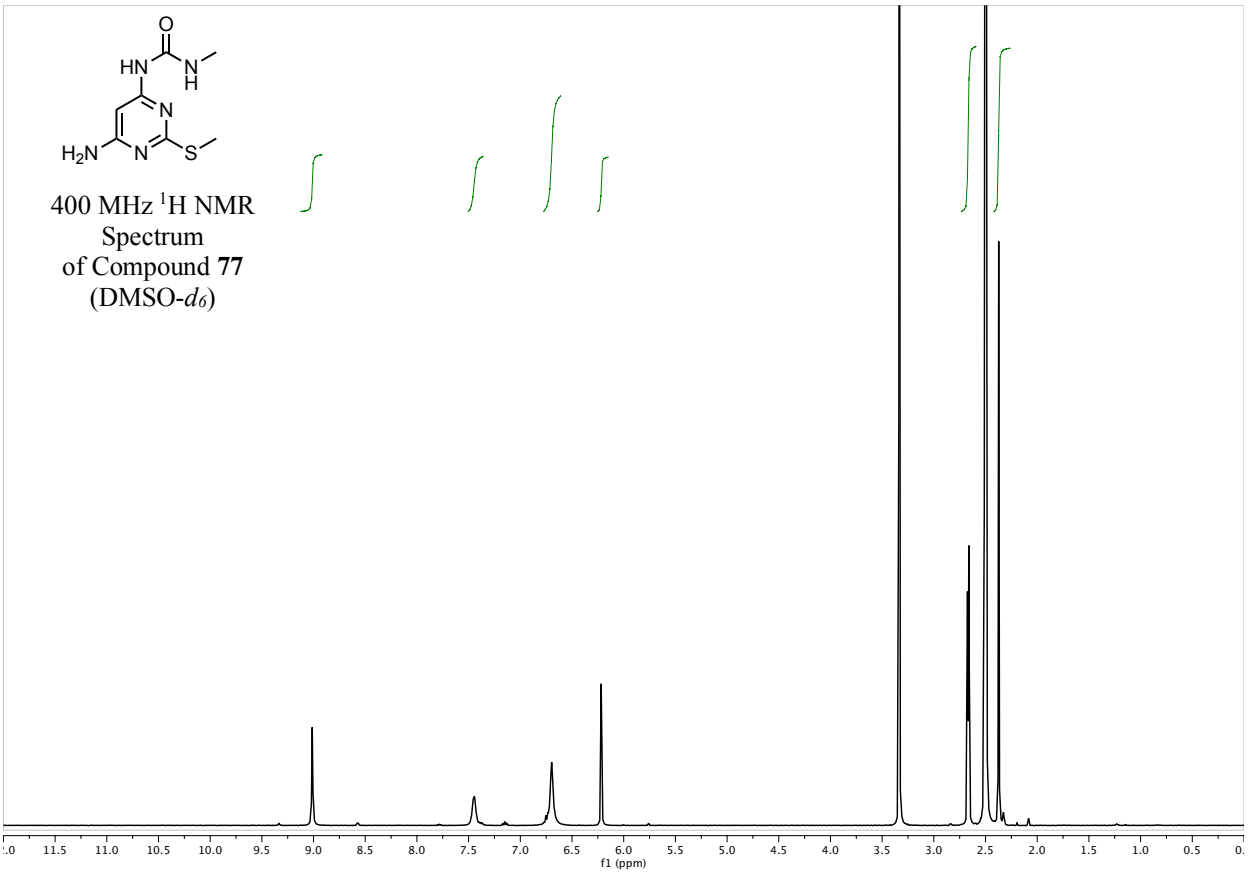
400 MHz ^{1}H NMR Spectrum of Compound **70** (DMSO-d_6)



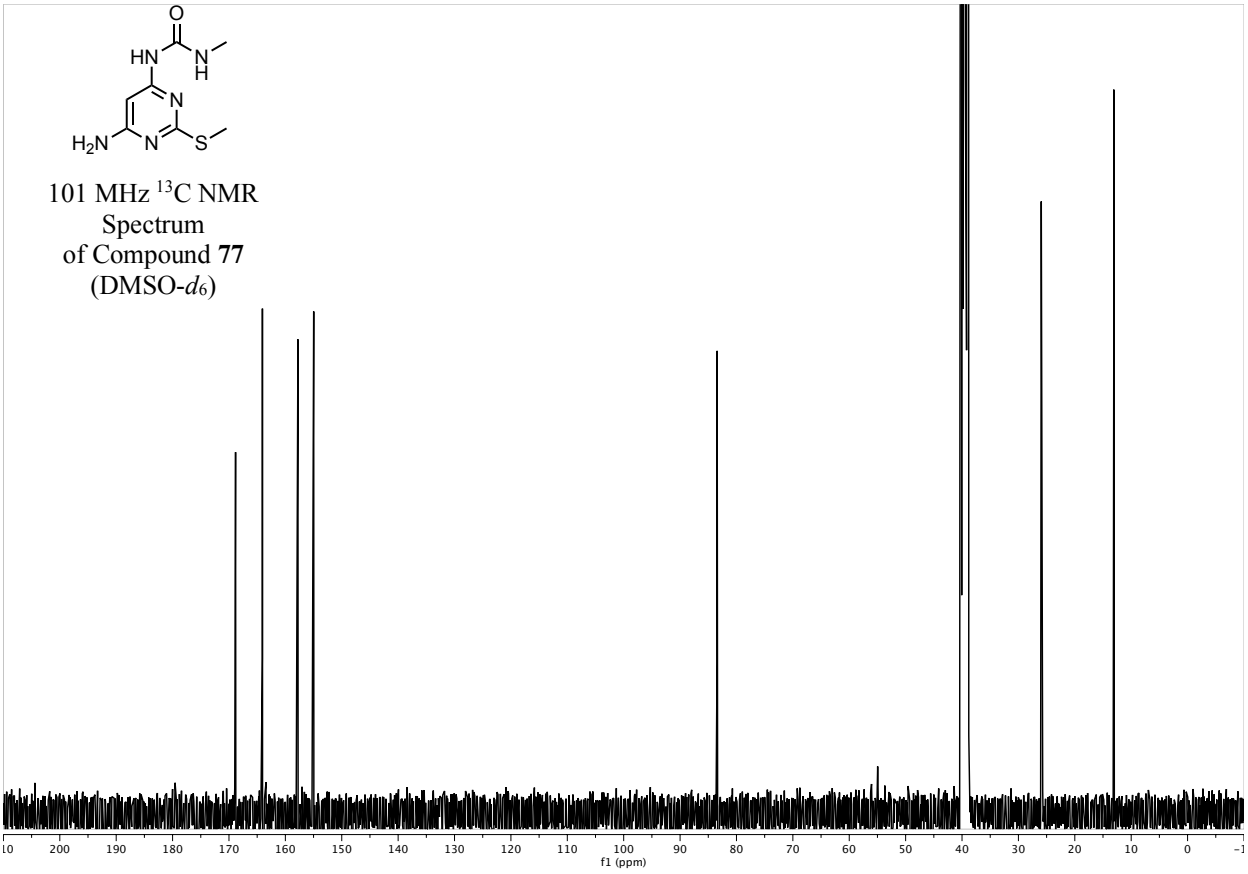
400 MHz ^{1}H NMR Spectrum of Compound **76** (DMSO-d_6)



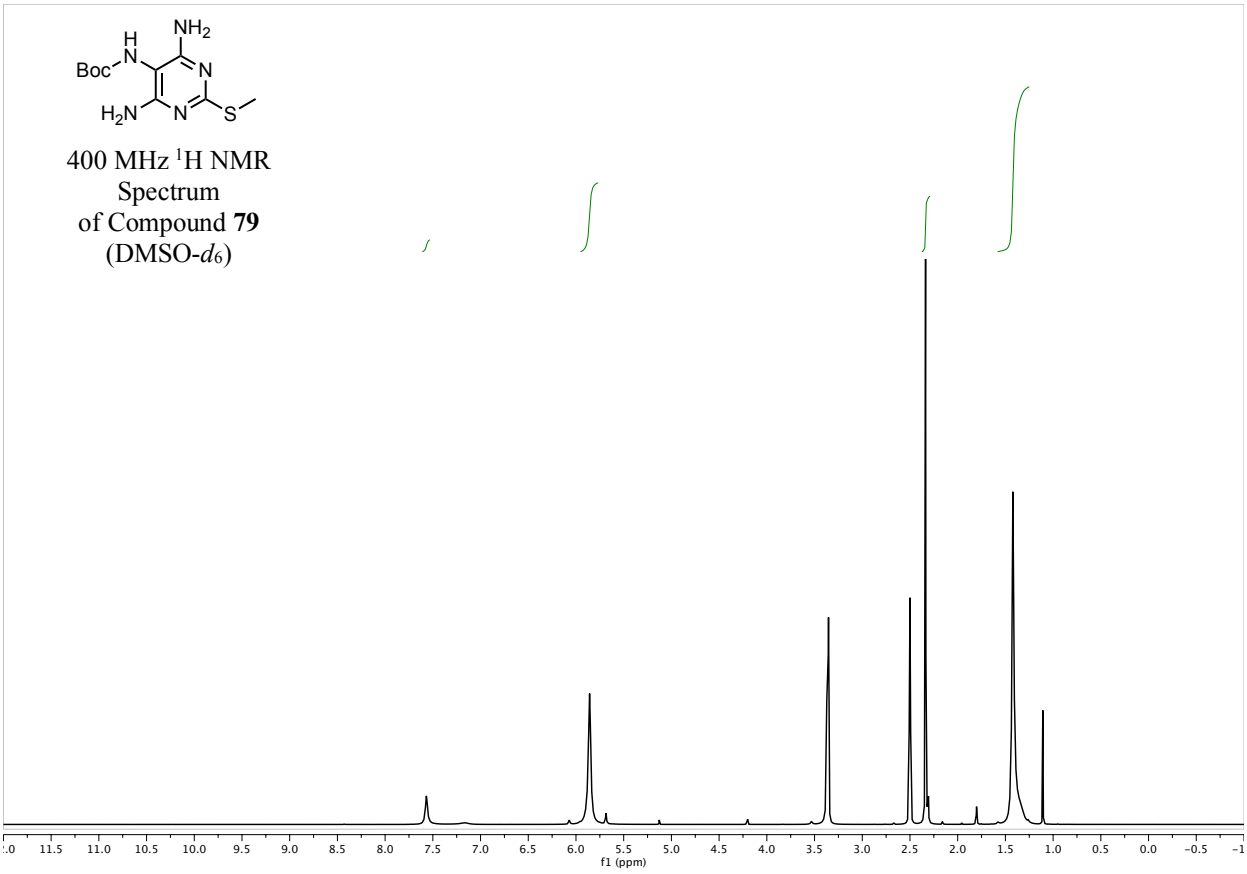
101 MHz ^{13}C NMR Spectrum of Compound **76** (DMSO-d_6)



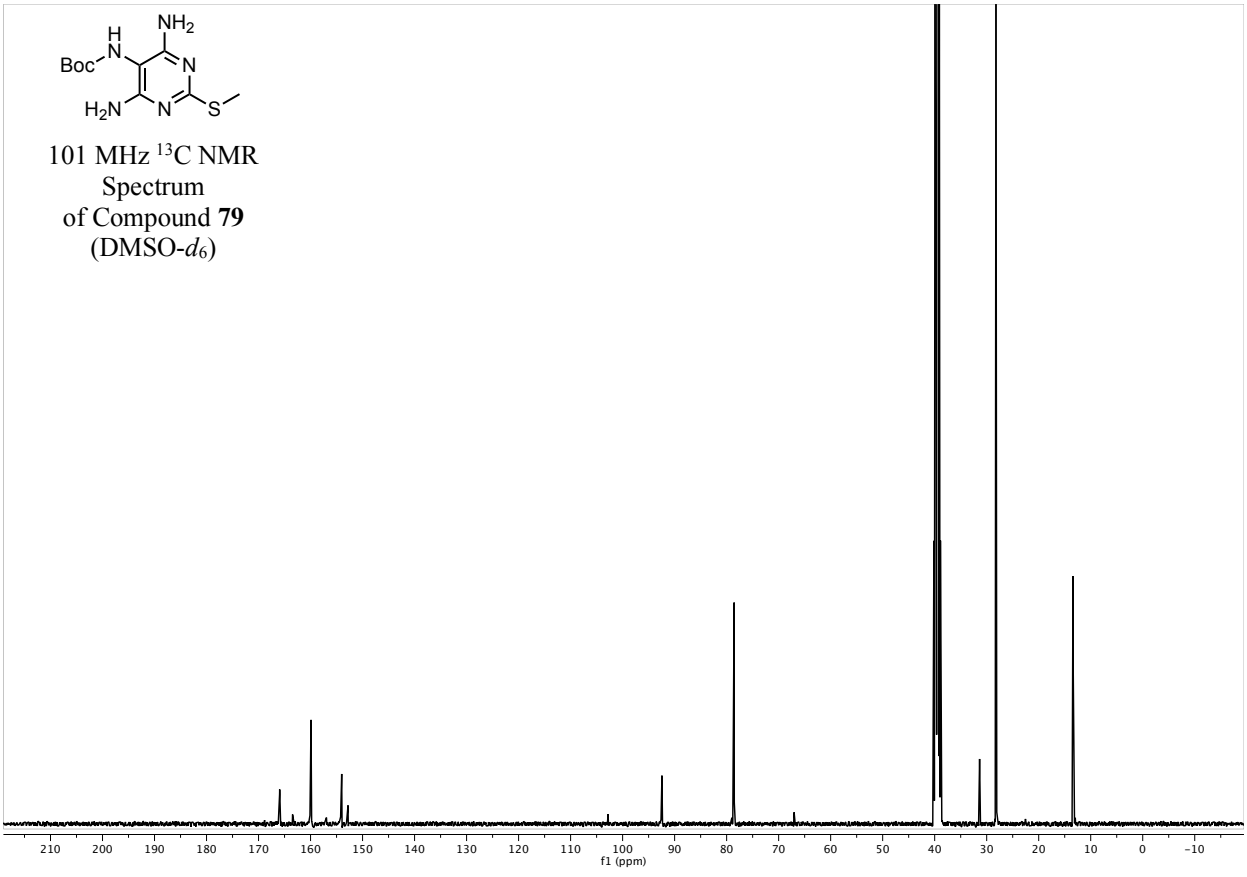
400 MHz ^{1}H NMR Spectrum of Compound **77** (DMSO-d_6)



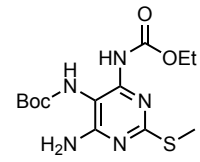
101 MHz ^{13}C NMR Spectrum of Compound **77** (DMSO-d_6)



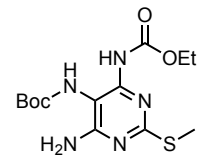
400 MHz ^{1}H NMR Spectrum of Compound **79** (DMSO-d_6)



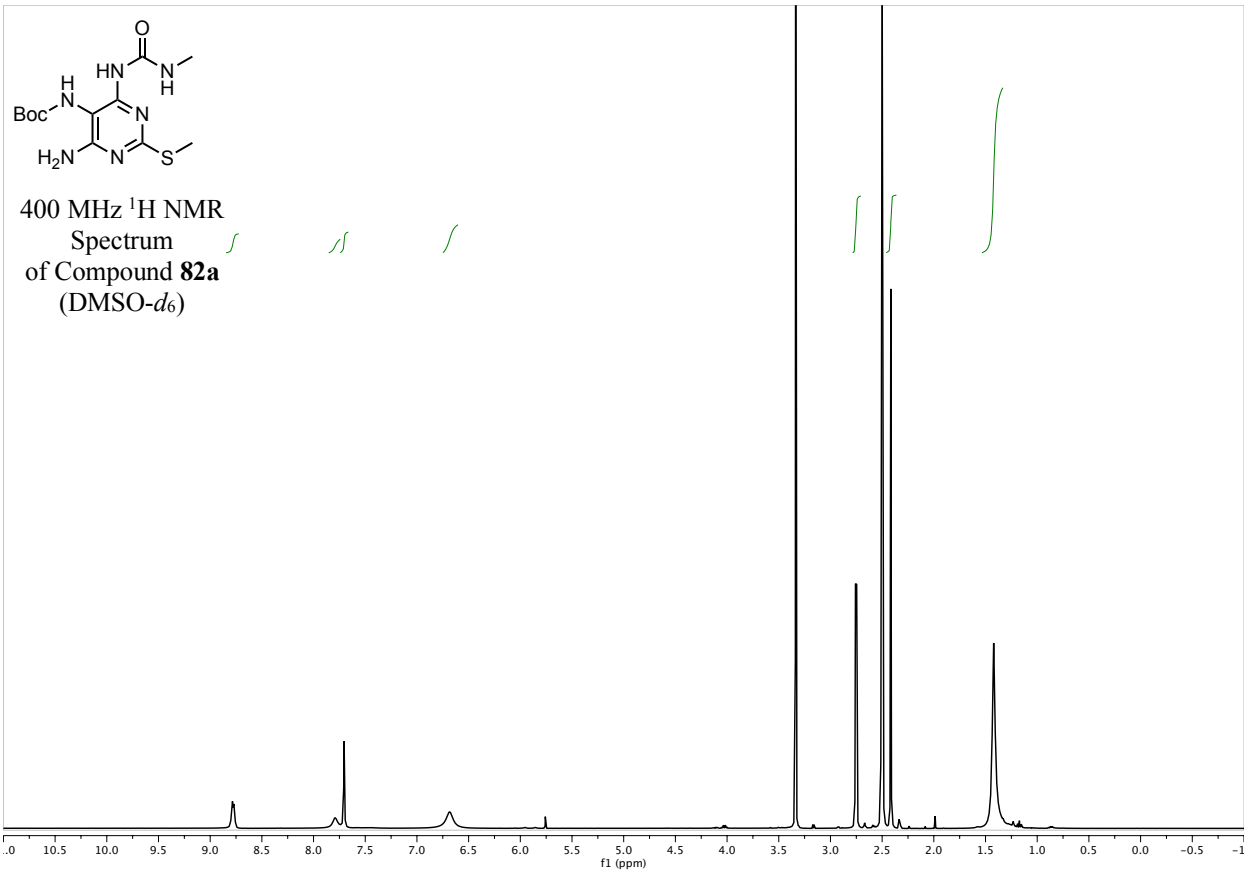
101 MHz ^{13}C NMR Spectrum of Compound **79** (DMSO-d_6)



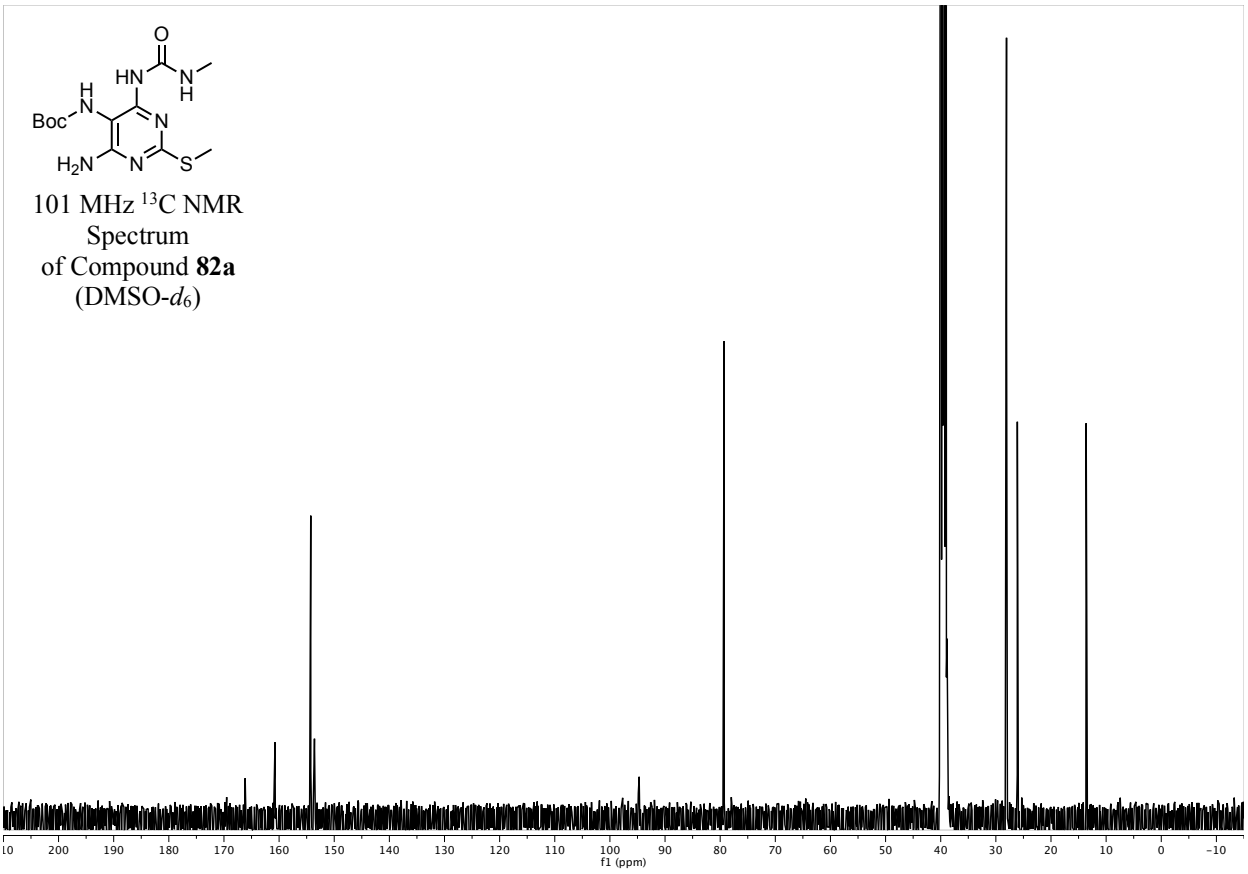
400 MHz ^{1}H NMR Spectrum of Compound **81** (Chloroform-*d*)



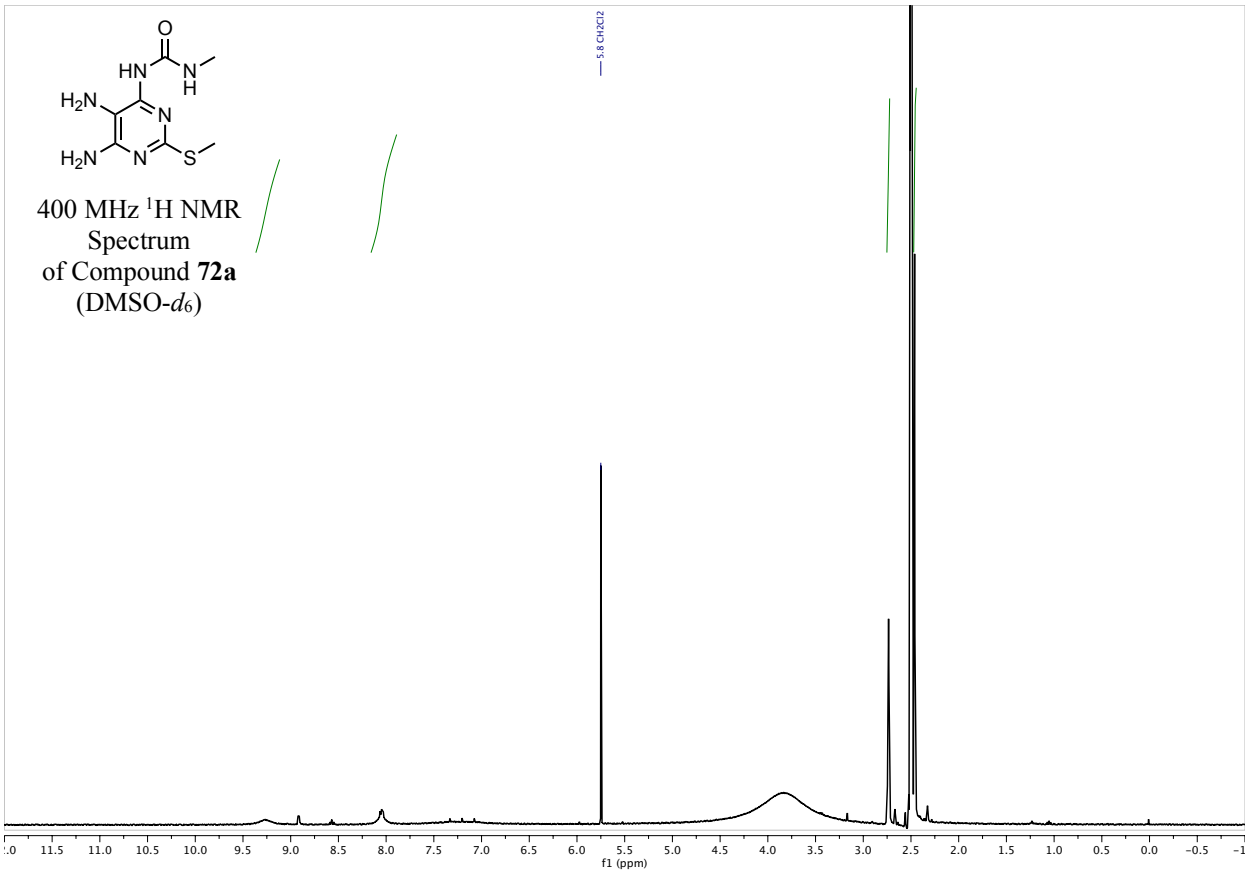
101 MHz ^{13}C NMR Spectrum of Compound **81** (Chloroform-*d*)



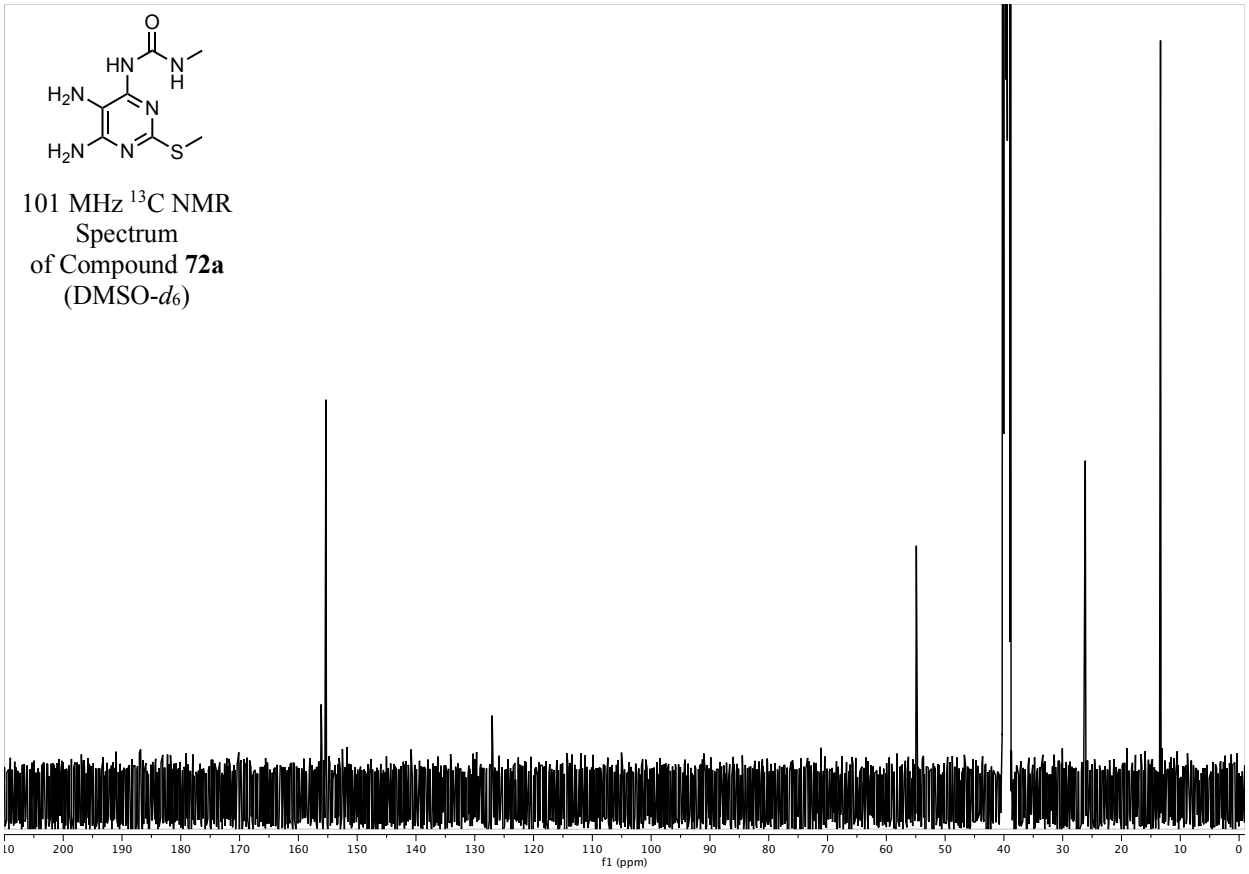
400 MHz ^{1}H NMR Spectrum of Compound **82a** (DMSO-d_6)



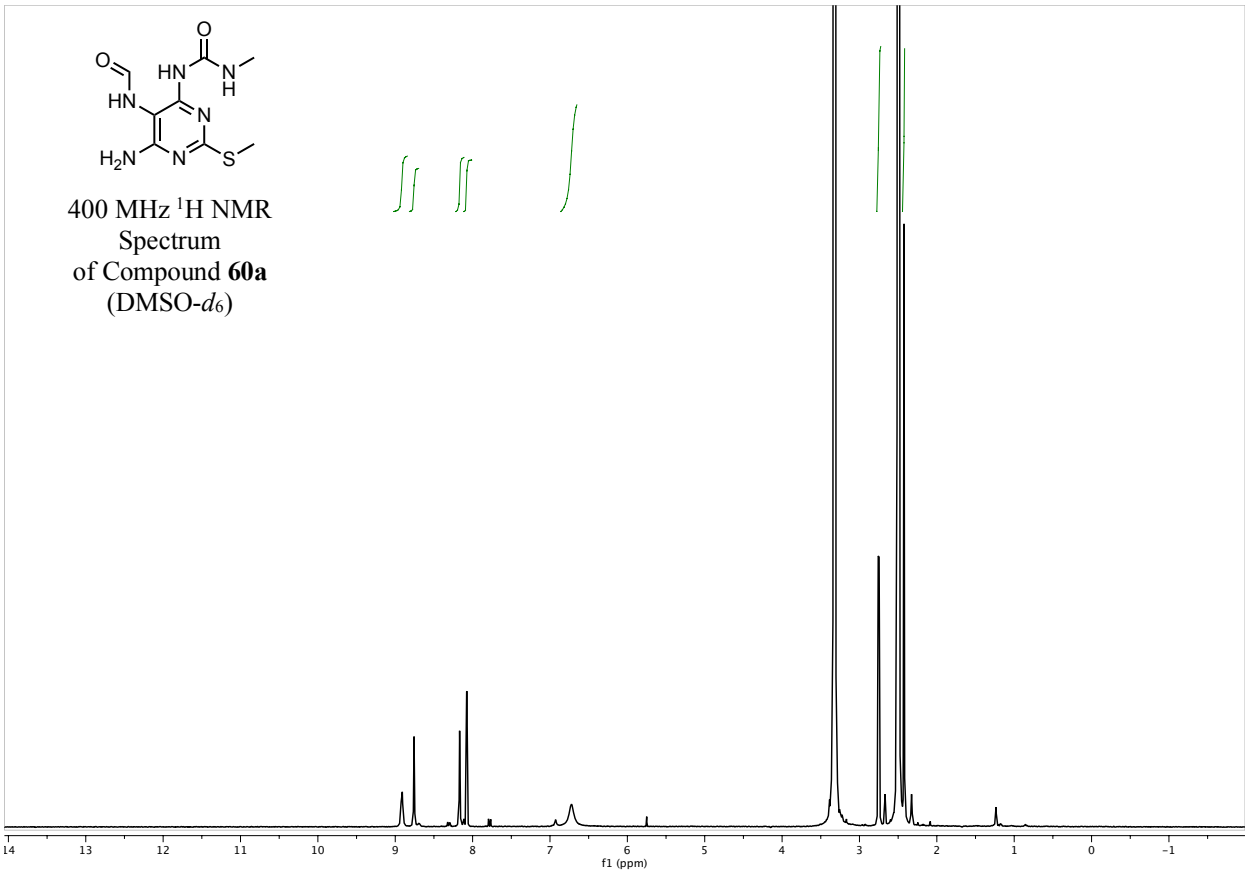
101 MHz ^{13}C NMR Spectrum of Compound **82a** (DMSO-d_6)



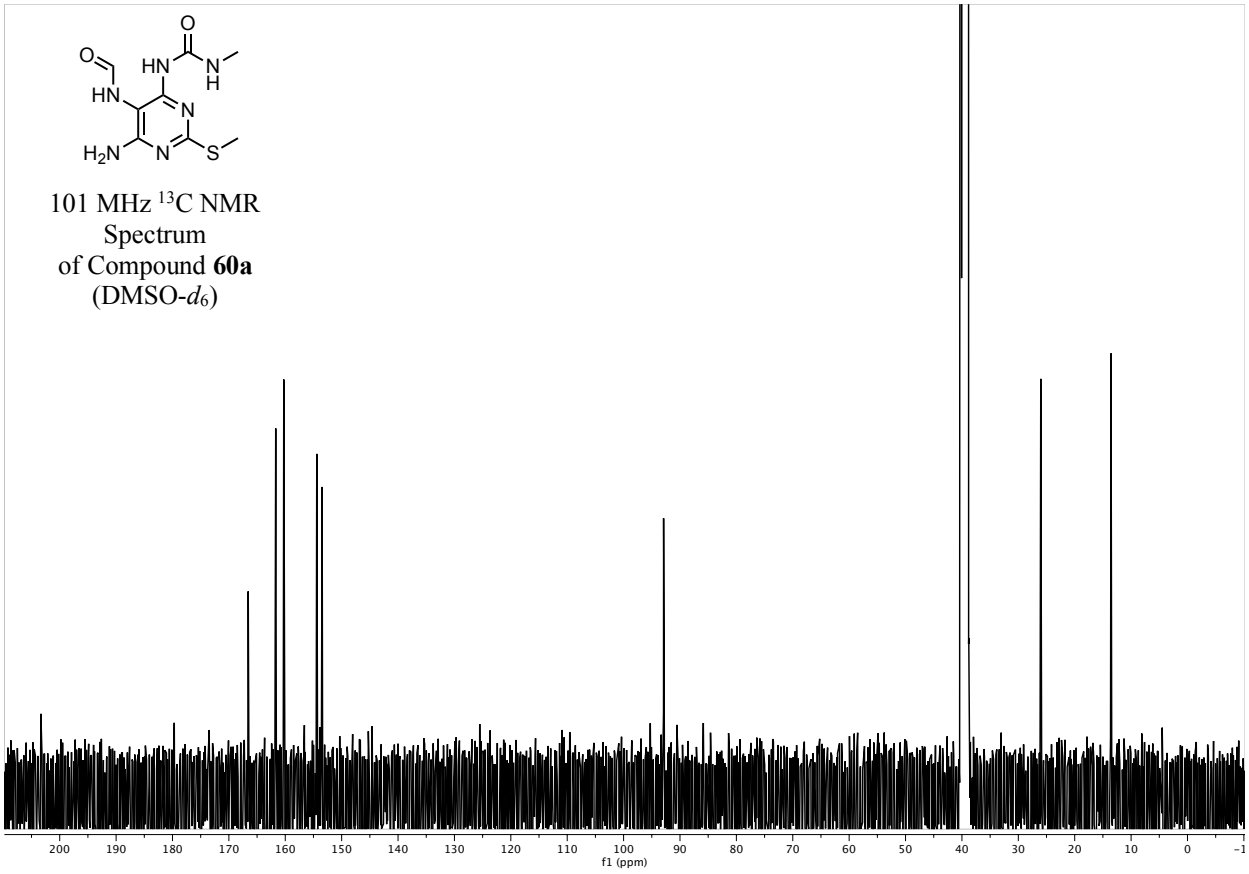
400 MHz ^{1}H NMR Spectrum of Compound **72a** (DMSO-d_6)



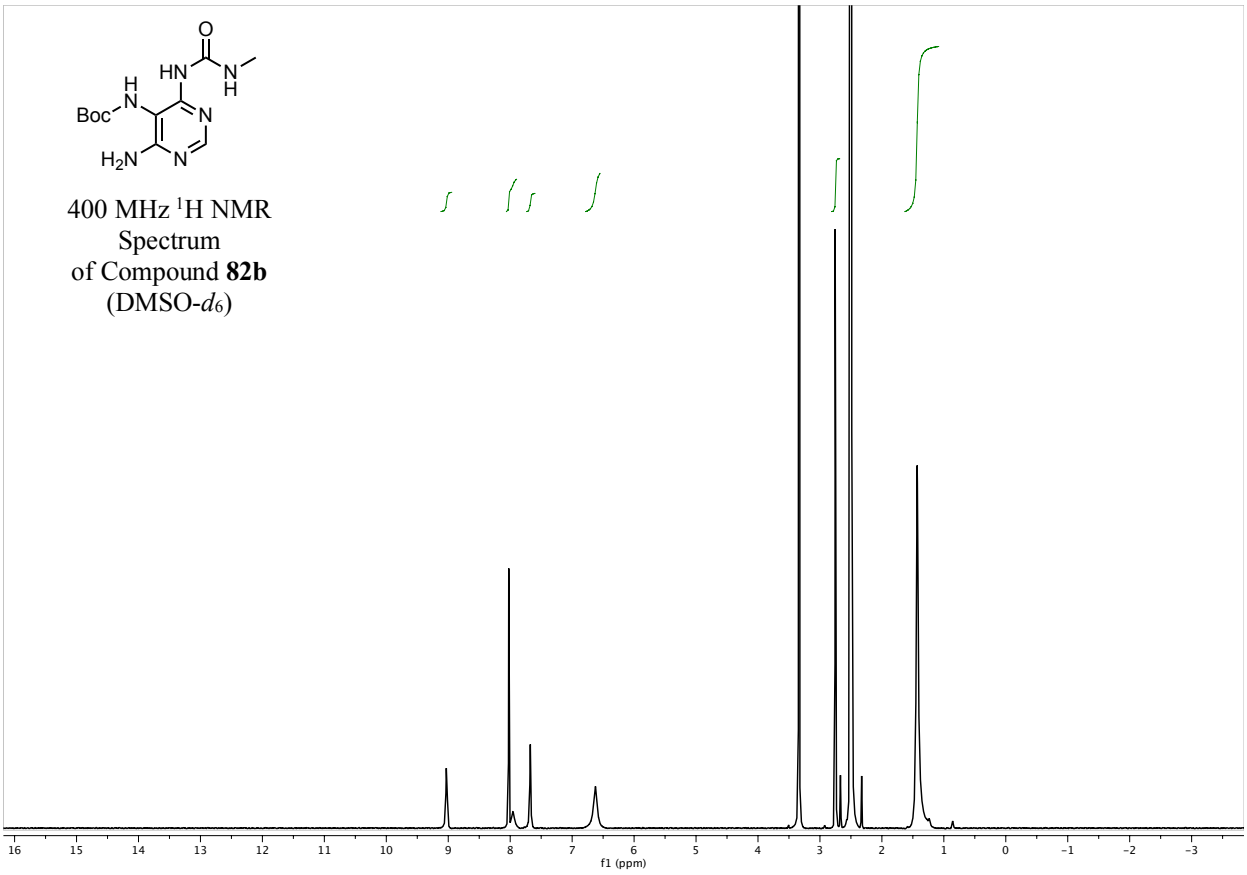
101 MHz ^{13}C NMR Spectrum of Compound **72a** (DMSO-d_6)



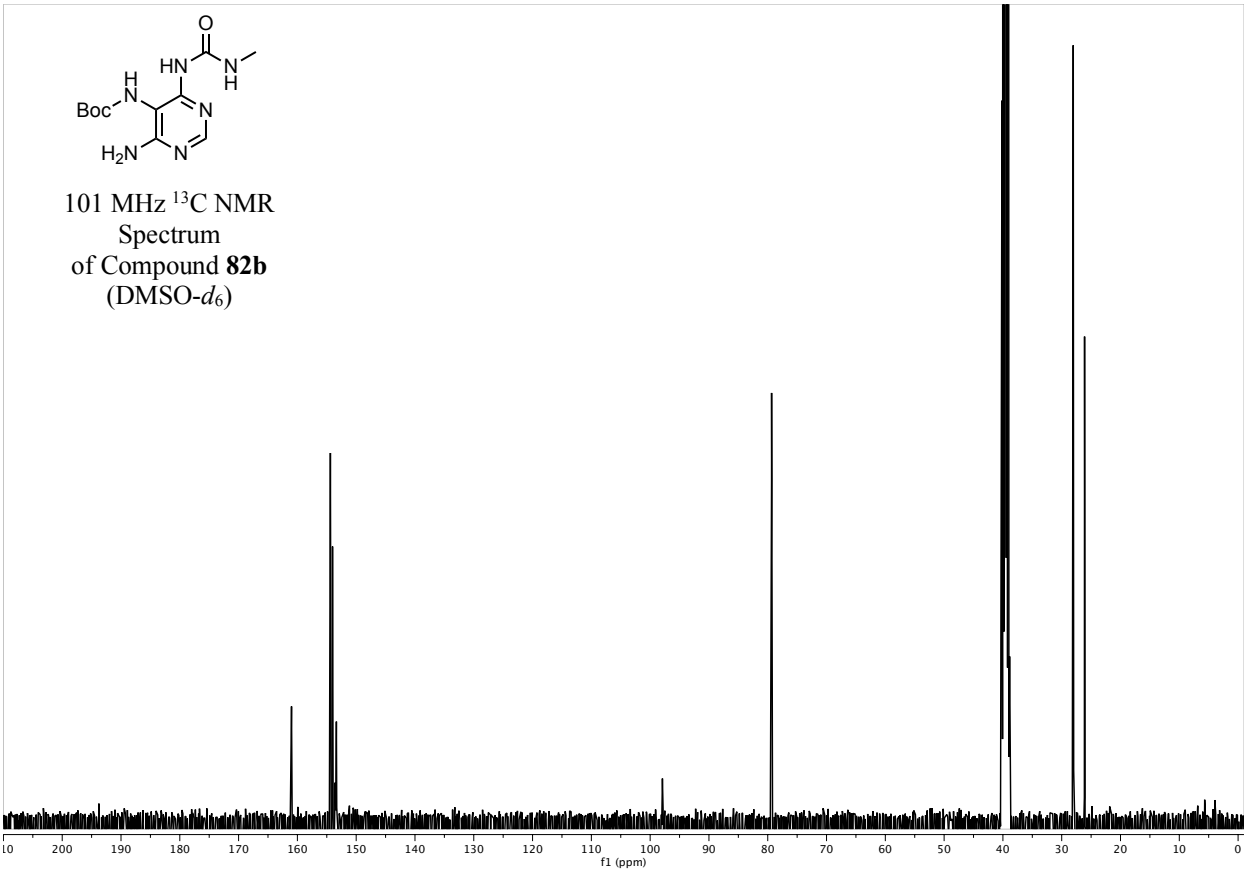
400 MHz ^{1}H NMR Spectrum of Compound **60a** (DMSO-d_6)



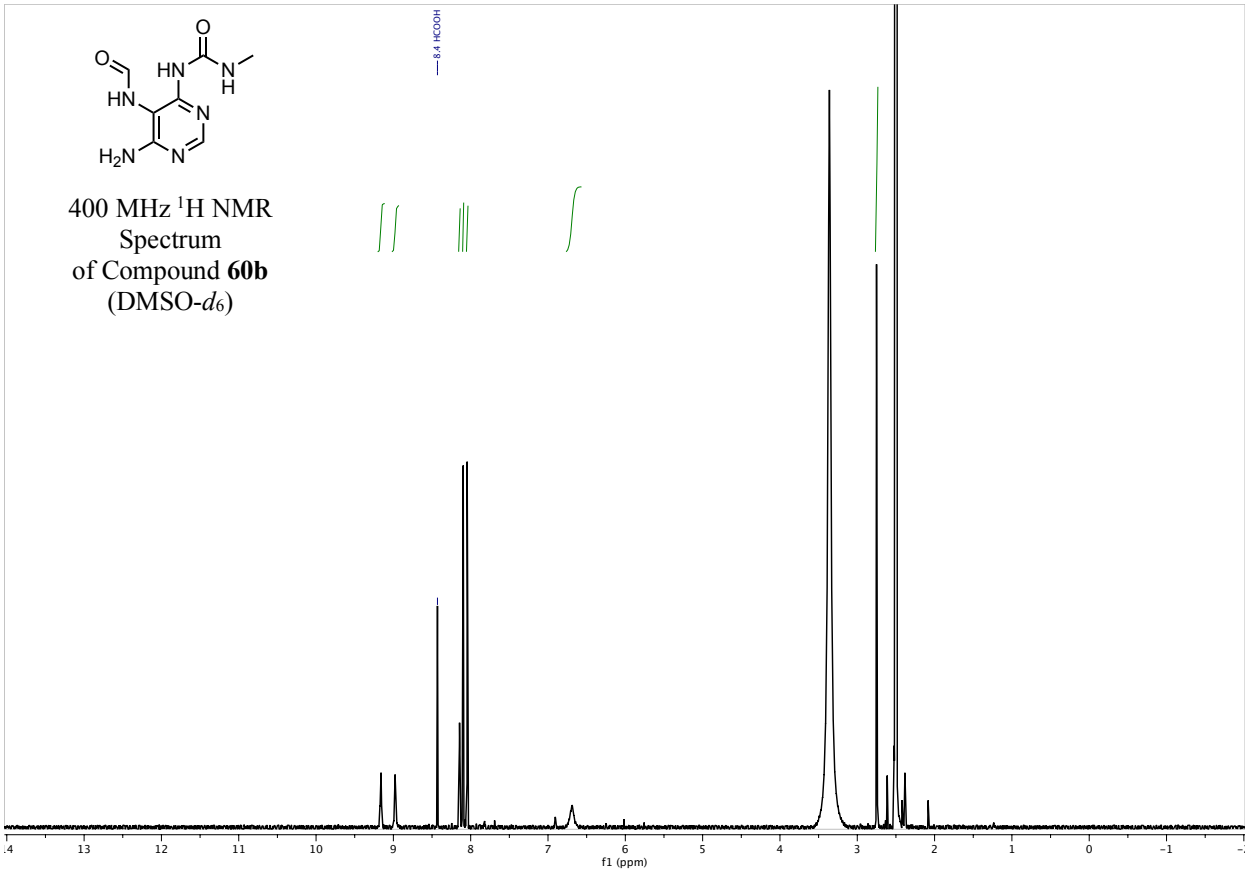
101 MHz ^{13}C NMR Spectrum of Compound **60a** (DMSO-d_6)



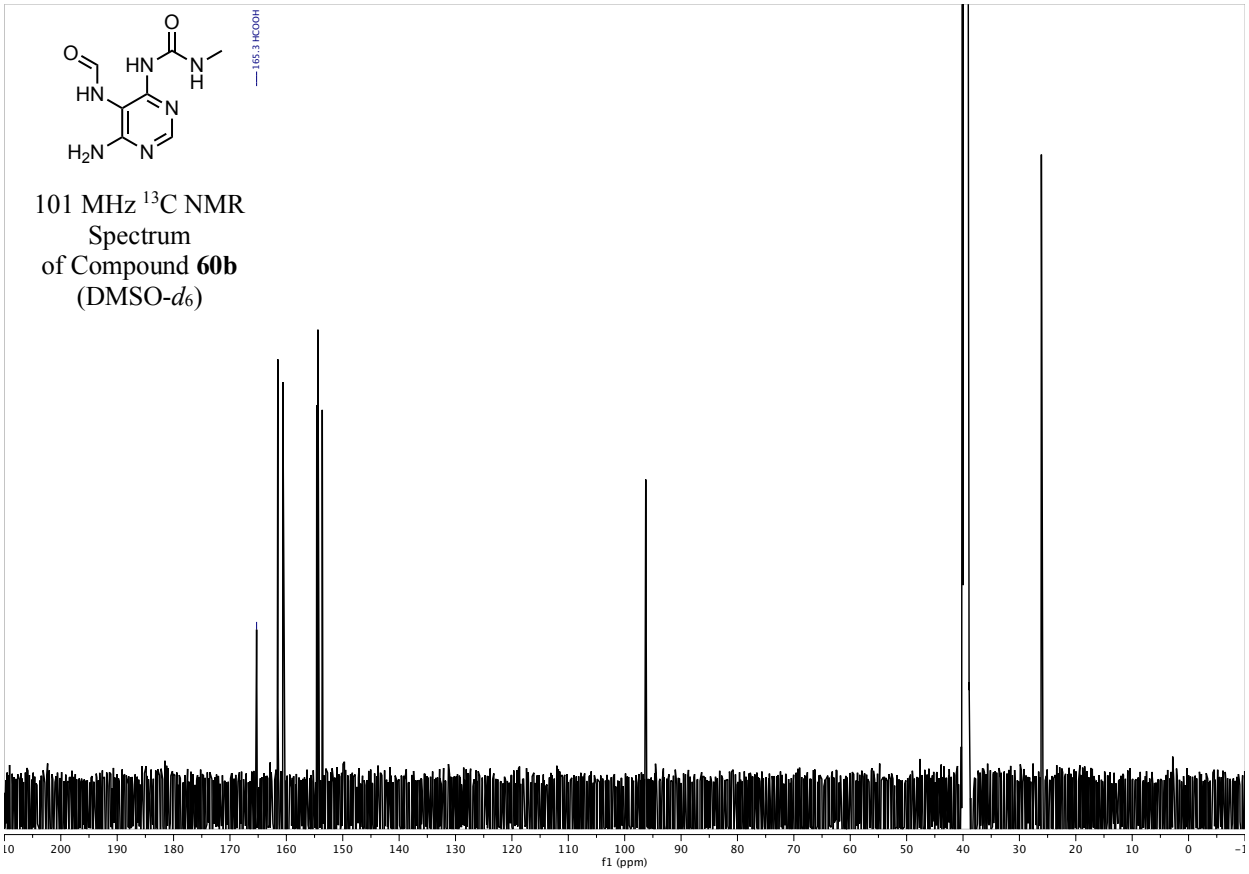
400 MHz ^{1}H NMR Spectrum of Compound **82b** (DMSO-d_6)



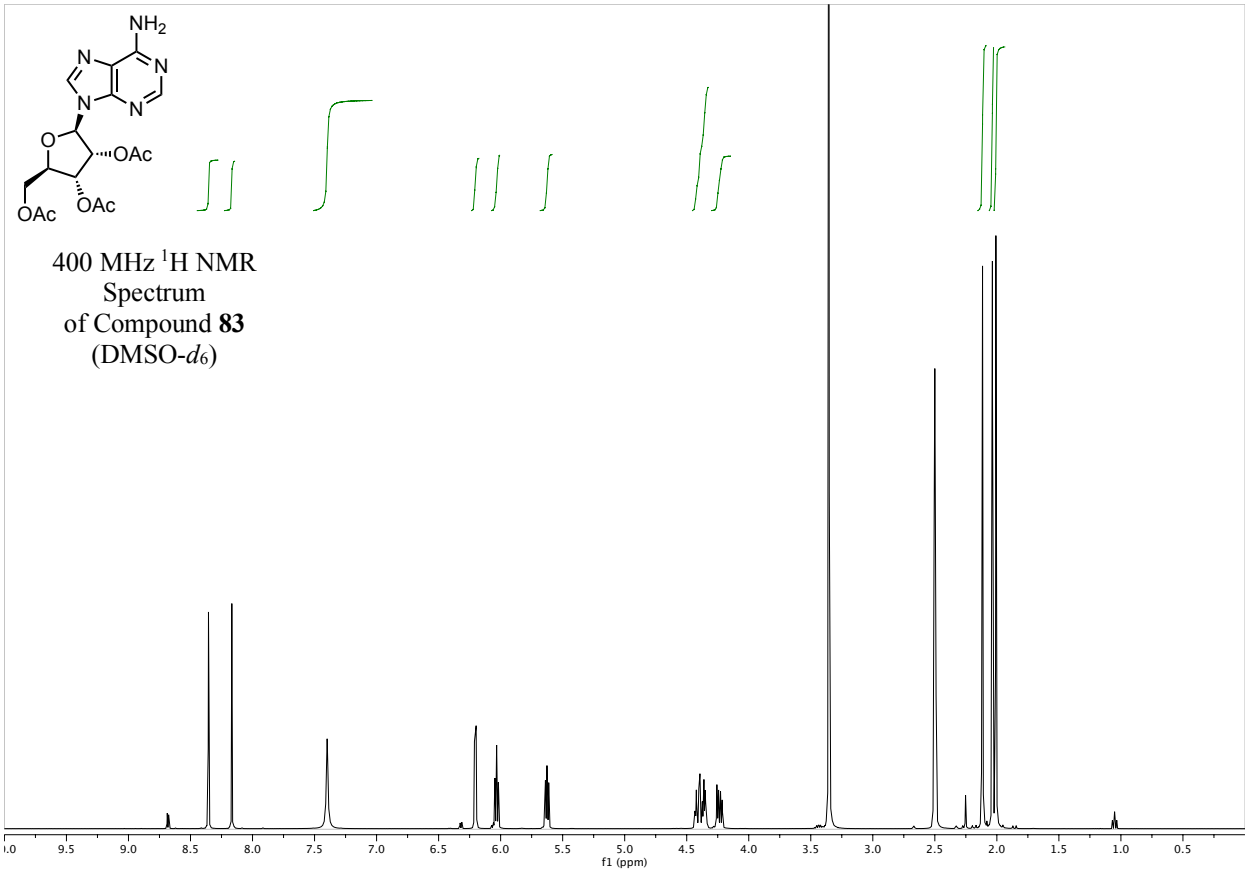
101 MHz ^{13}C NMR Spectrum of Compound **82b** (DMSO-d_6)



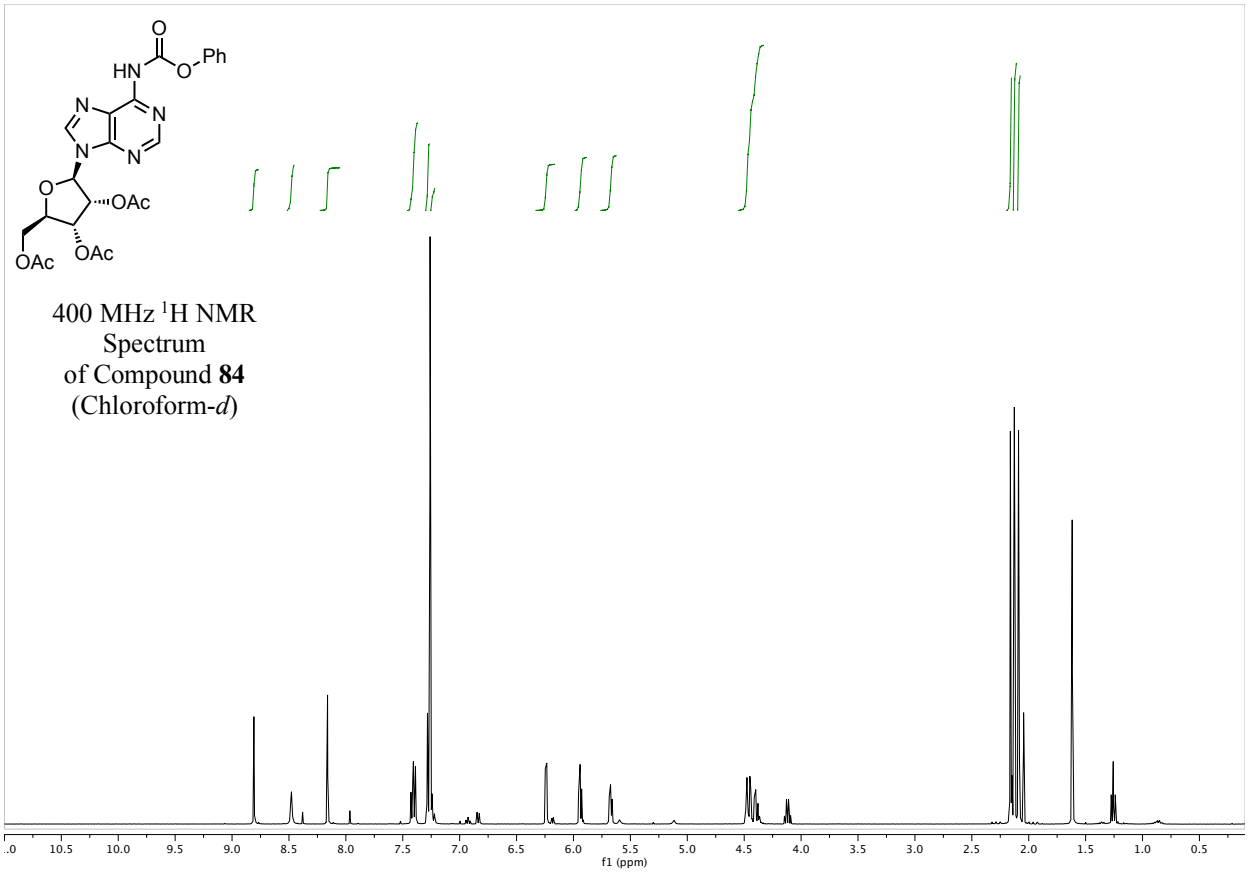
400 MHz ^{1}H NMR Spectrum of Compound **60b** (DMSO-d_6)



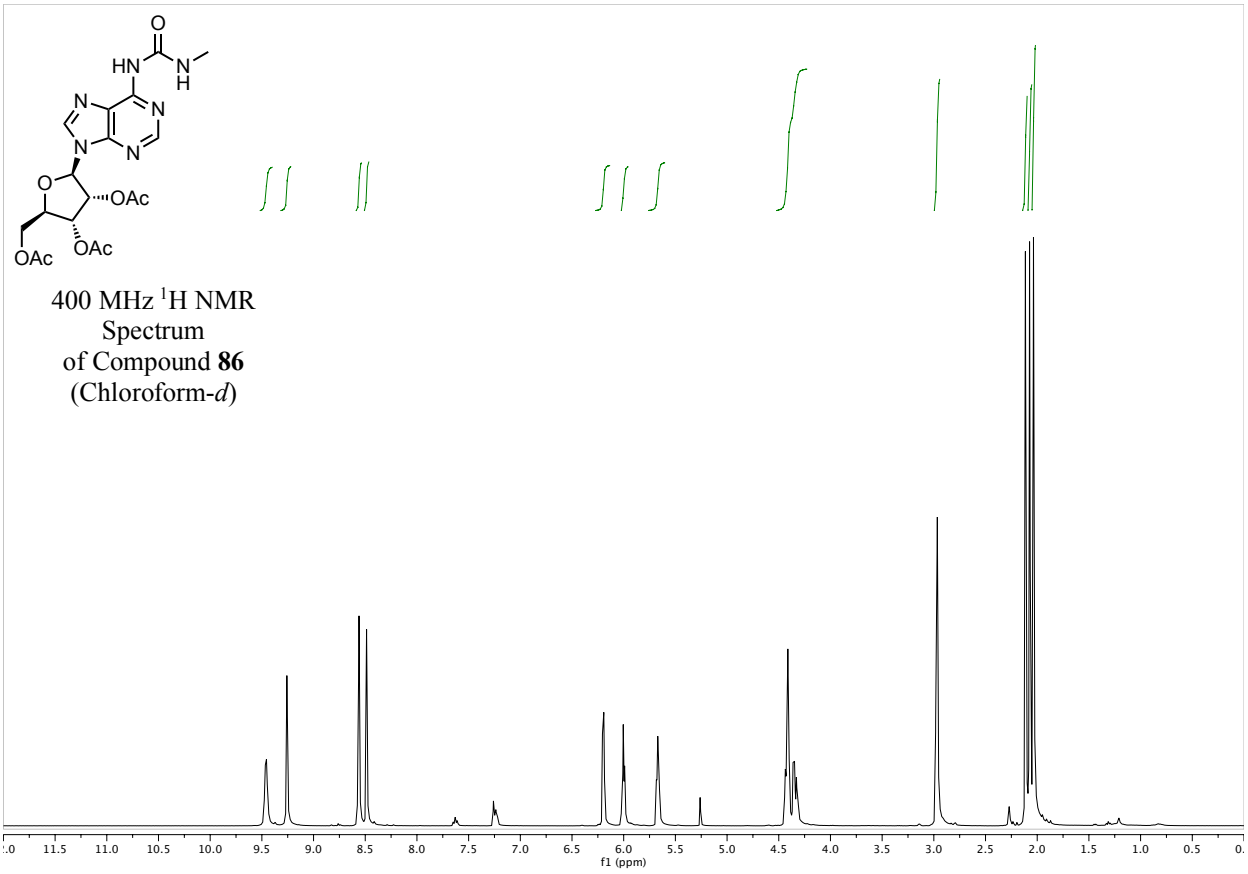
101 MHz ^{13}C NMR Spectrum of Compound **60b** (DMSO-d_6)



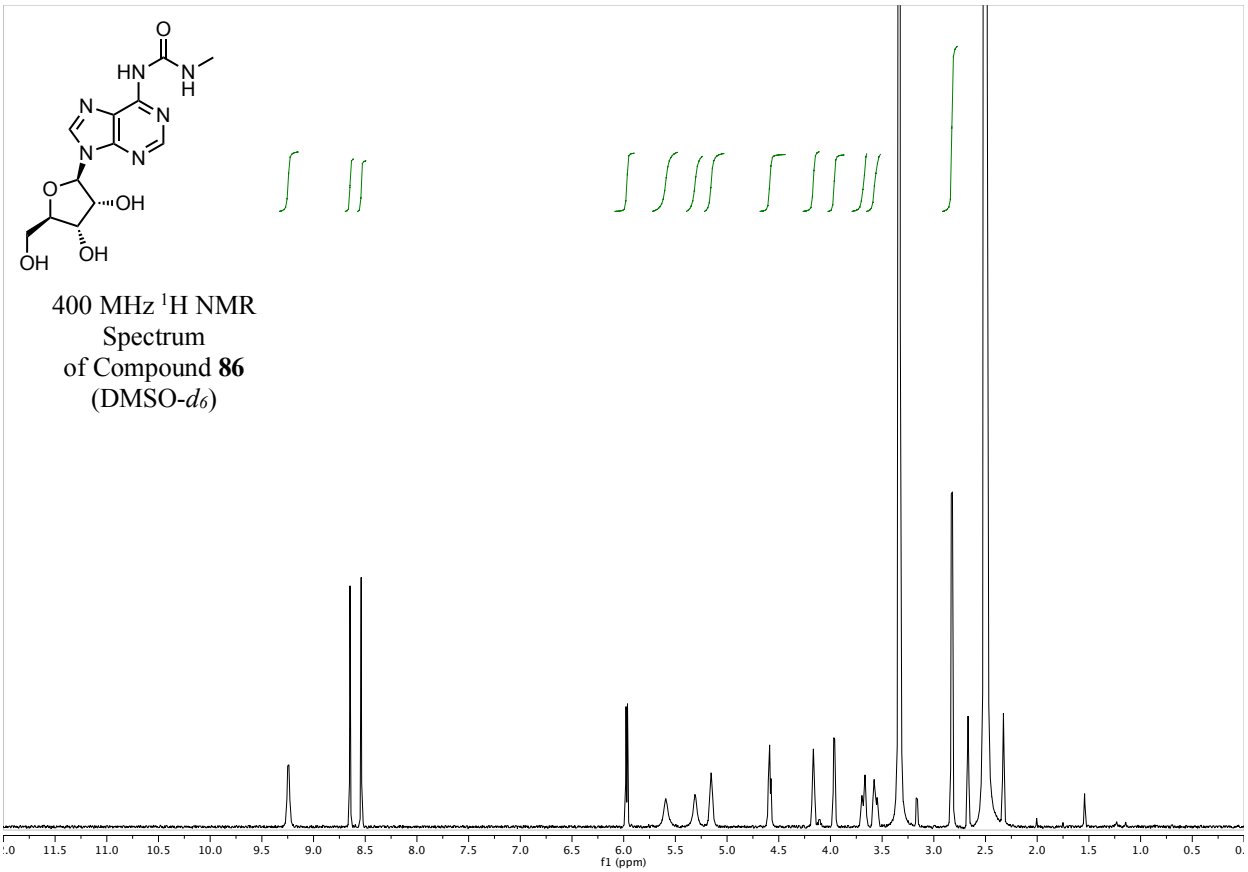
400 MHz ^{1}H NMR Spectrum of Compound **83** (DMSO-d_6)



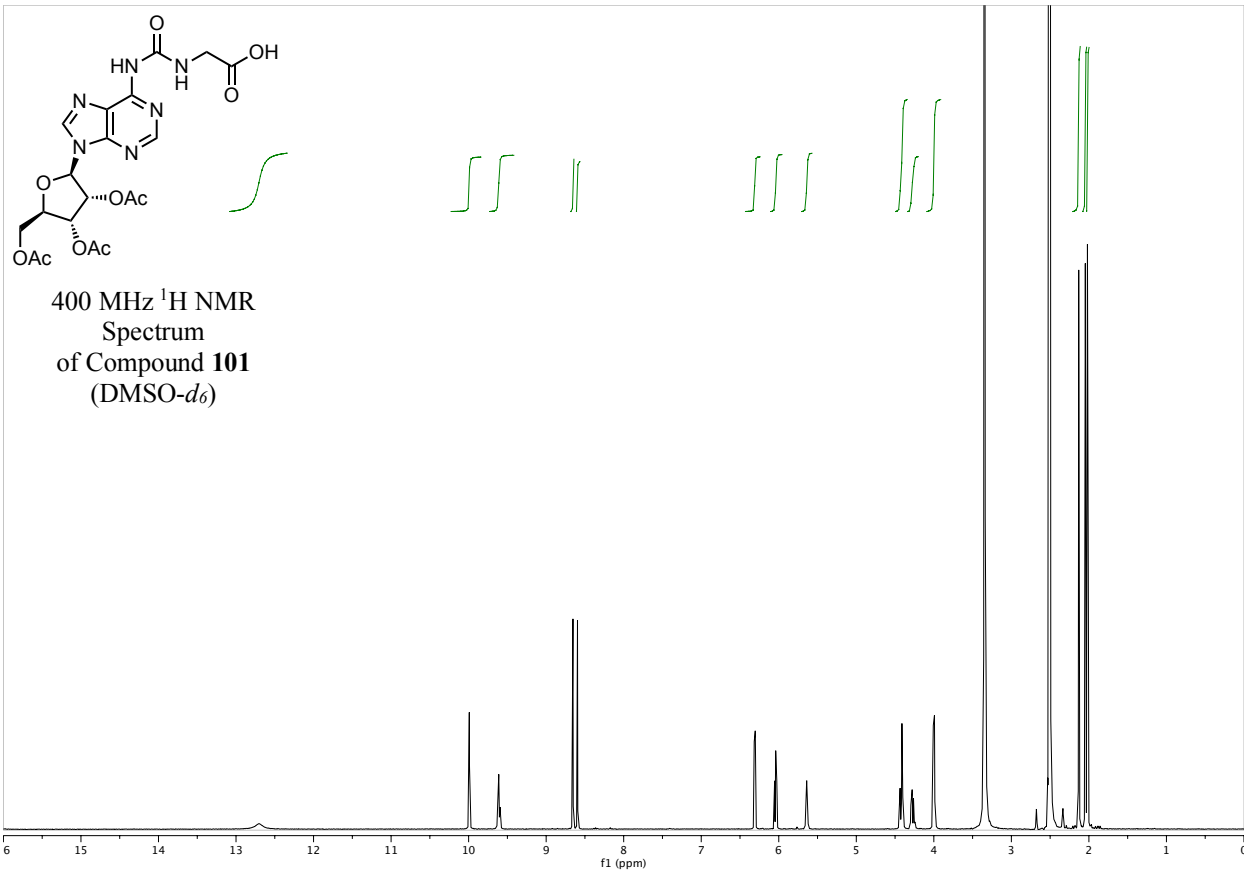
400 MHz ^{1}H NMR Spectrum of Compound **84** (Chloroform-*d*)



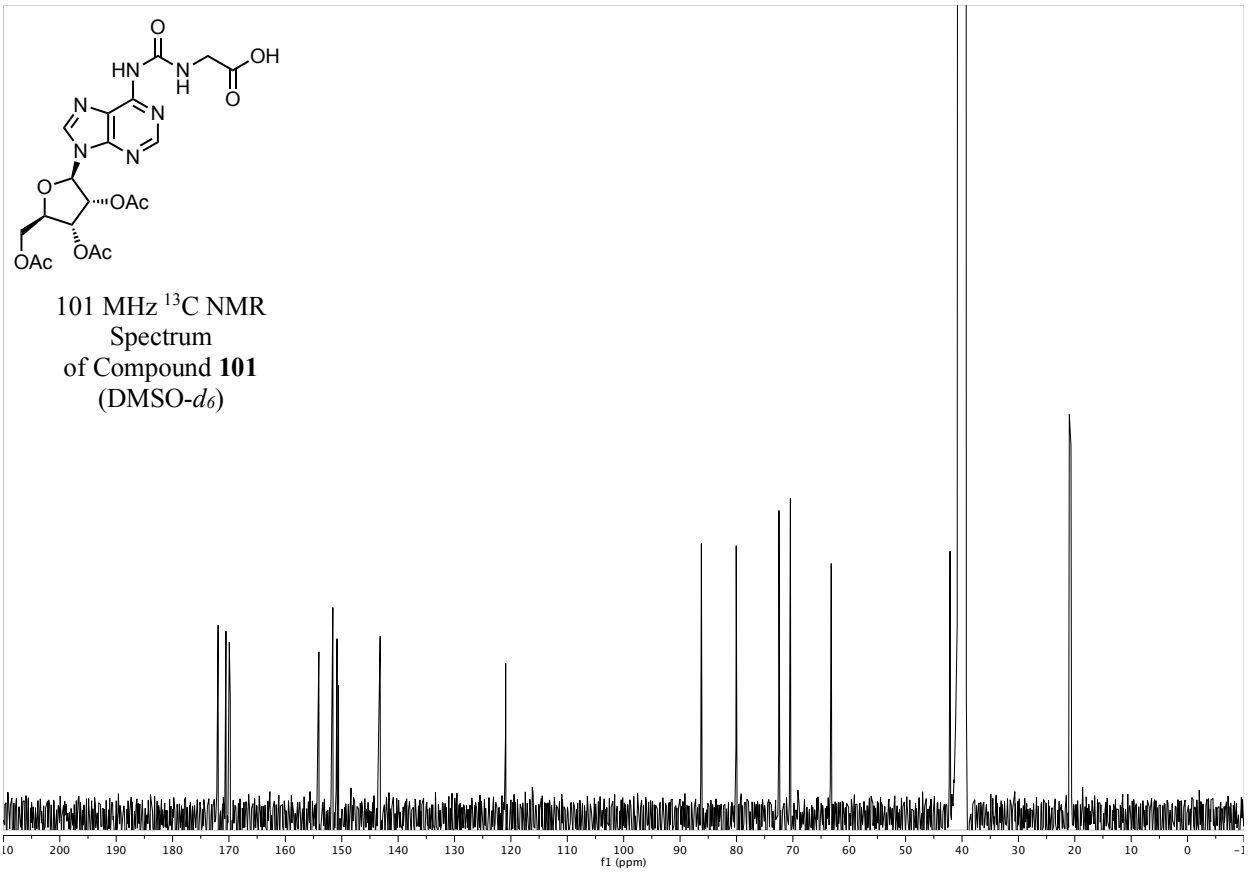
400 MHz ^{1}H NMR Spectrum of Compound **86** (Chloroform-*d*)



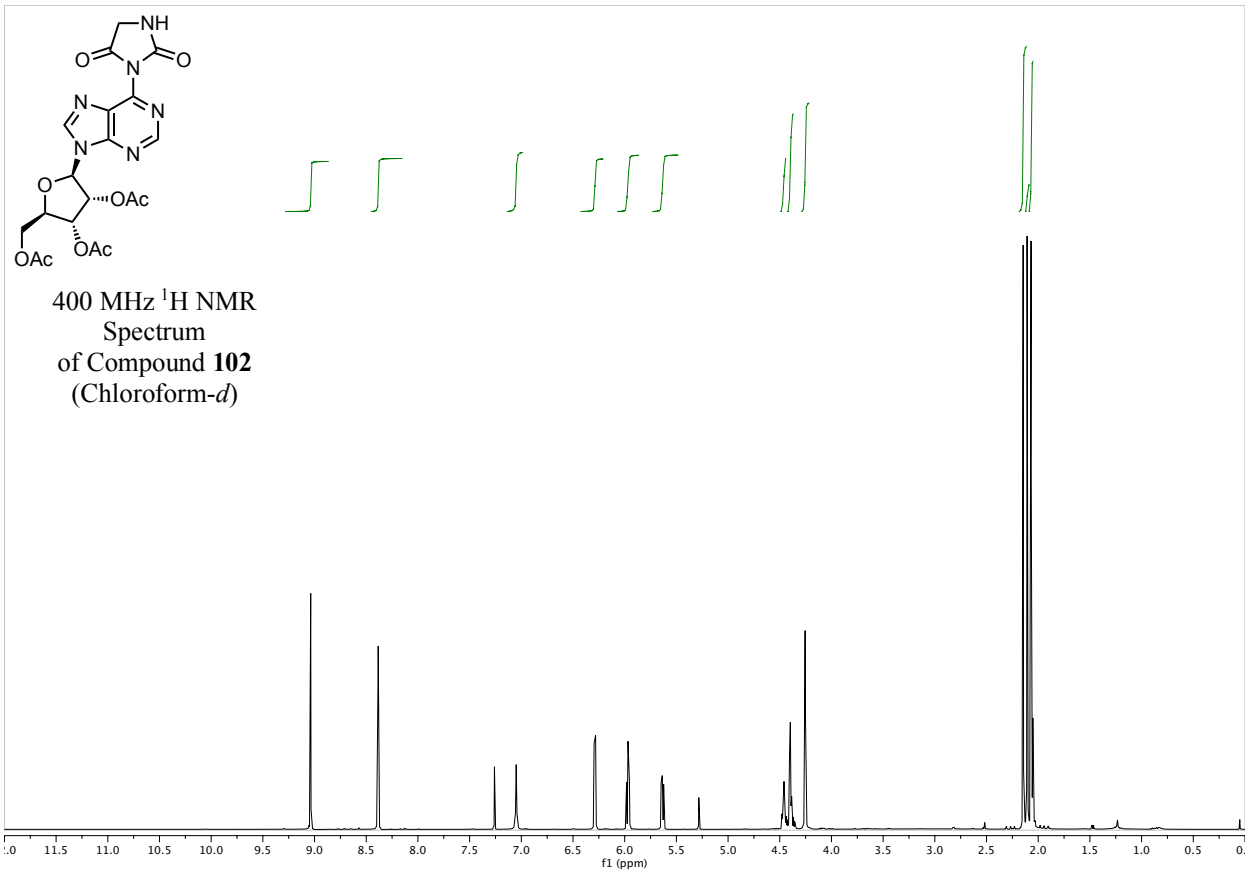
400 MHz ^{1}H NMR Spectrum of Compound **86** (DMSO-d_6)



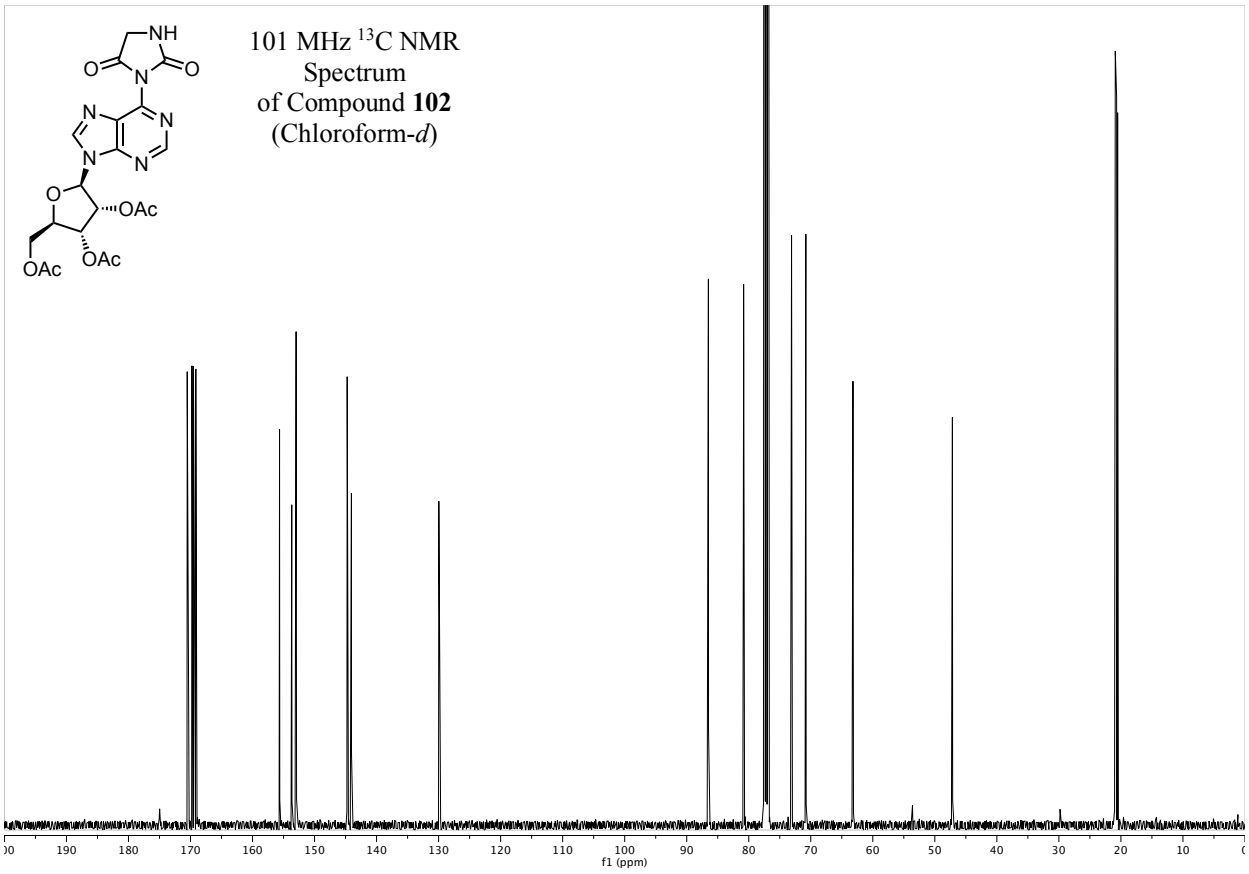
400 MHz ^{1}H NMR
Spectrum
of Compound **101**
(DMSO-*d*$_6$)



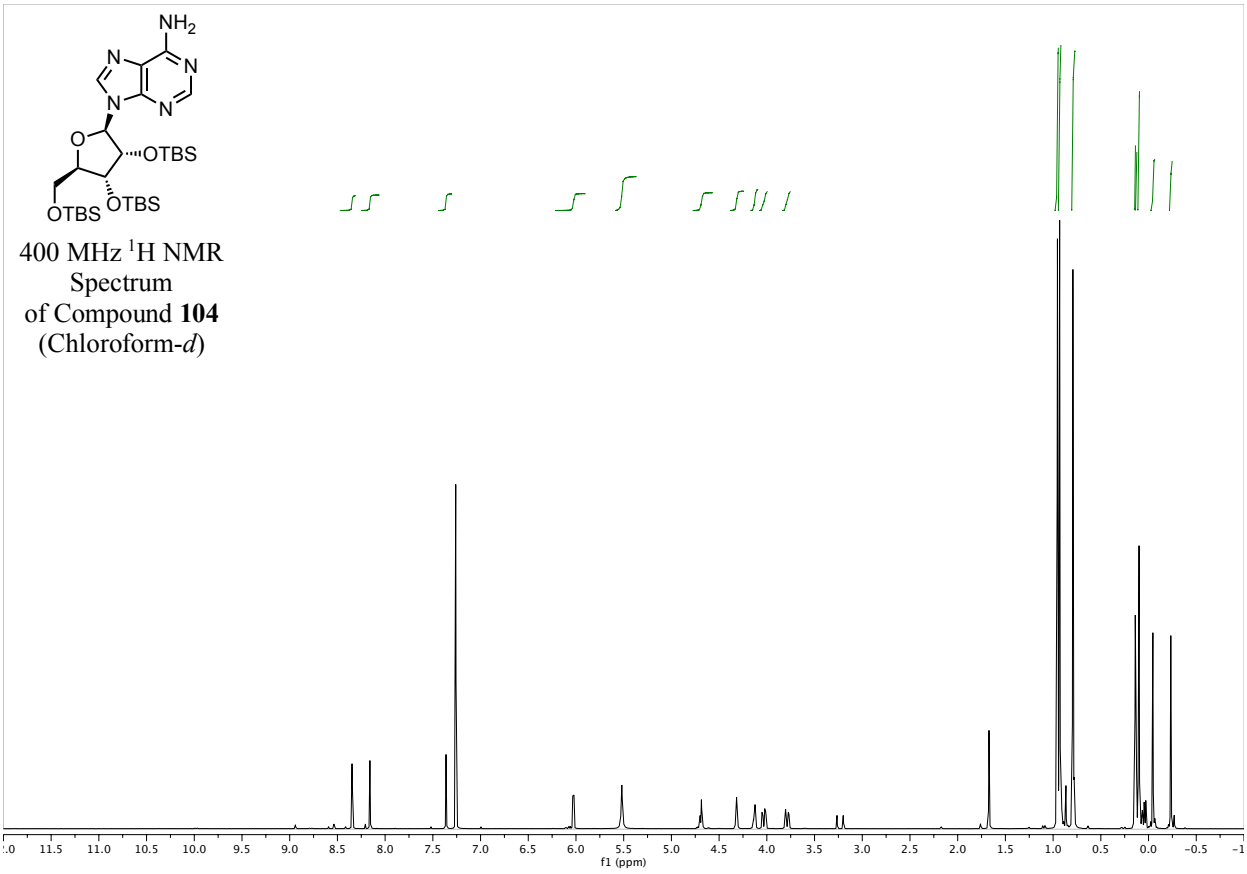
101 MHz ^{13}C NMR
Spectrum
of Compound **101**
(DMSO-*d*$_6$)



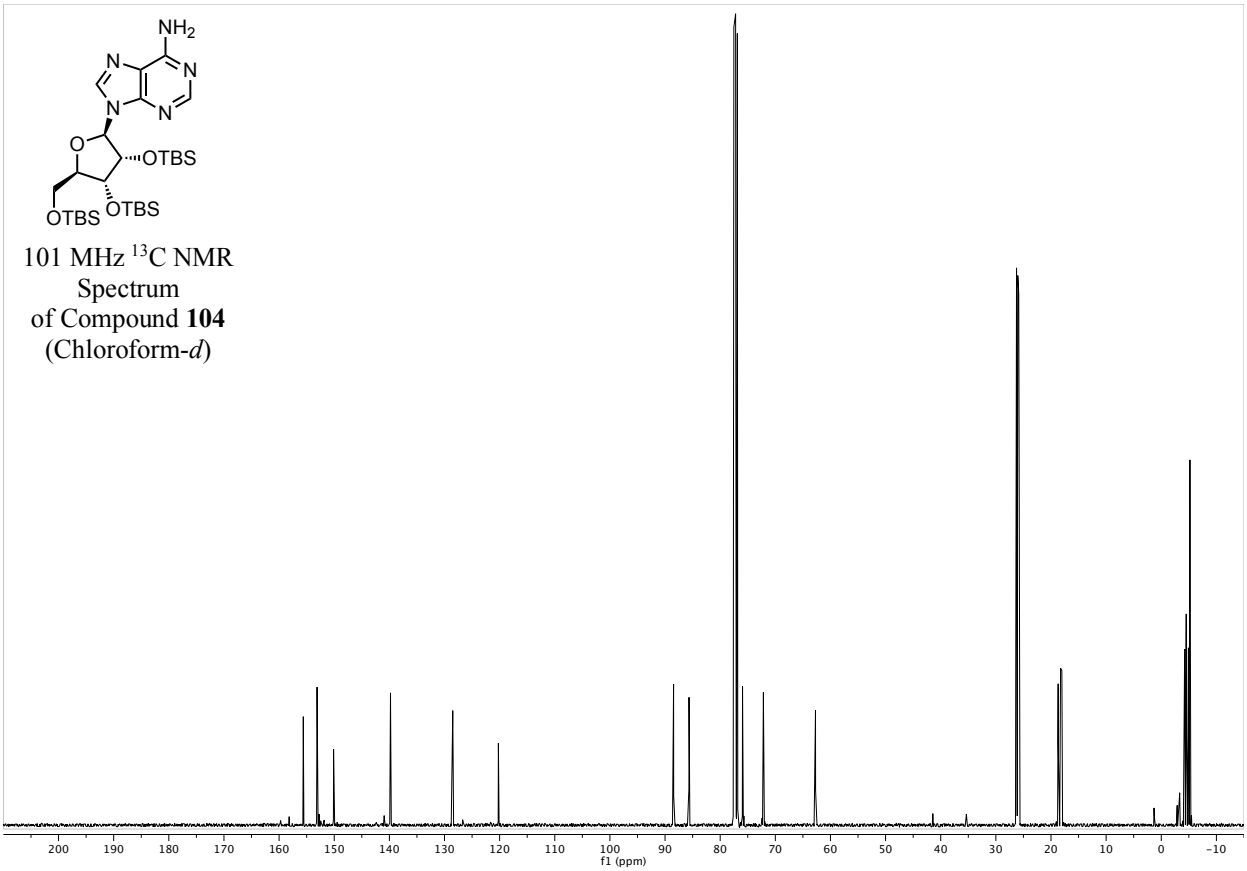
400 MHz ^{1}H NMR Spectrum of Compound **102** (Chloroform-*d*)



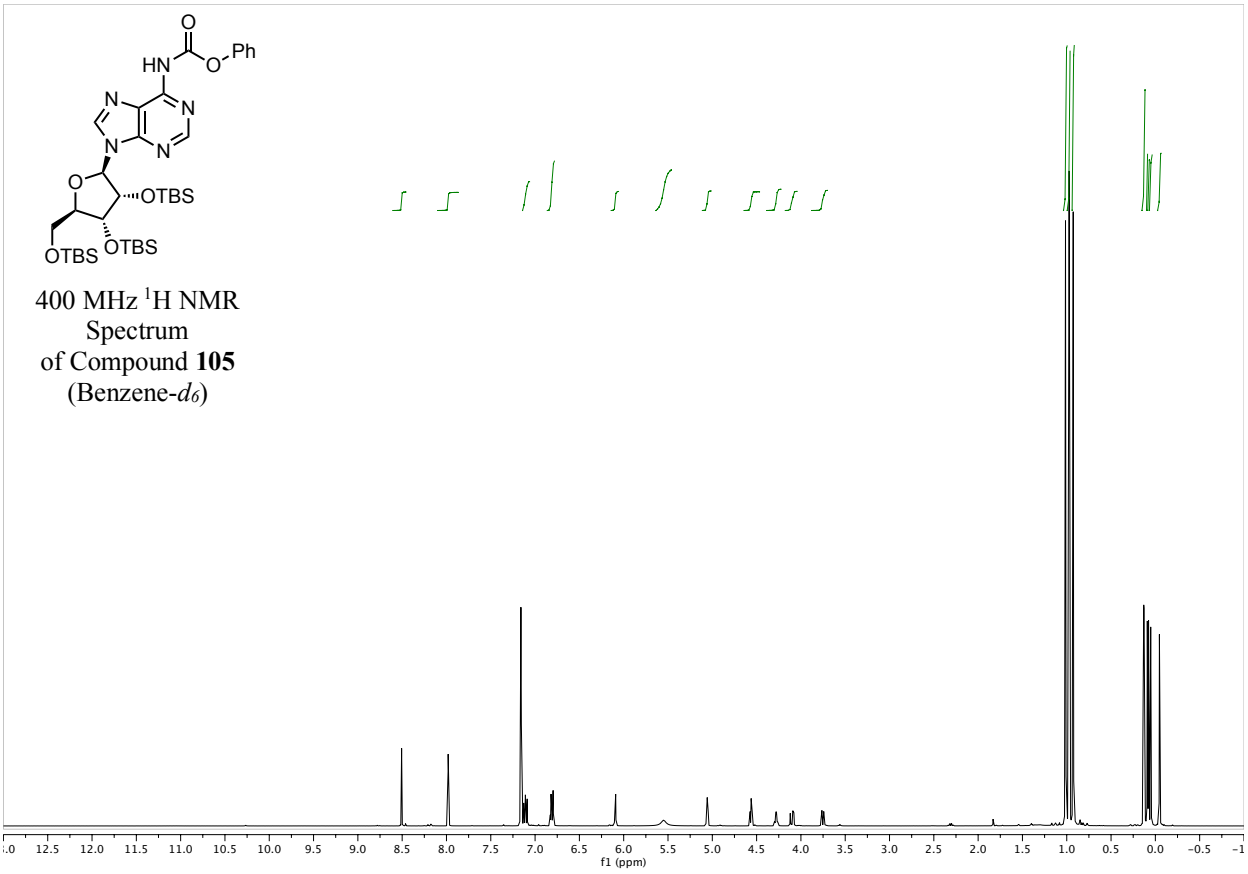
101 MHz ^{13}C NMR Spectrum of Compound **102** (Chloroform-*d*)



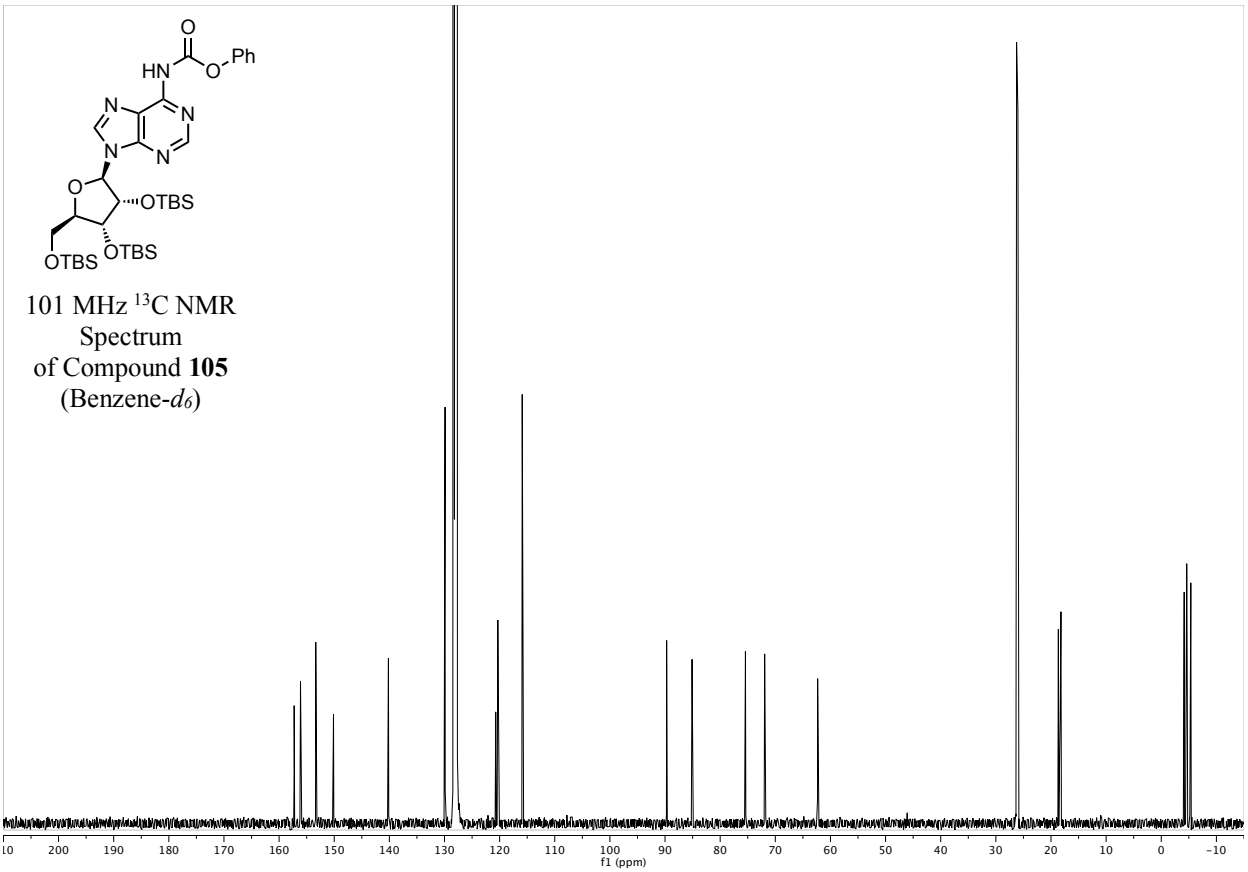
400 MHz ^{1}H NMR Spectrum of Compound **104** (Chloroform-*d*)



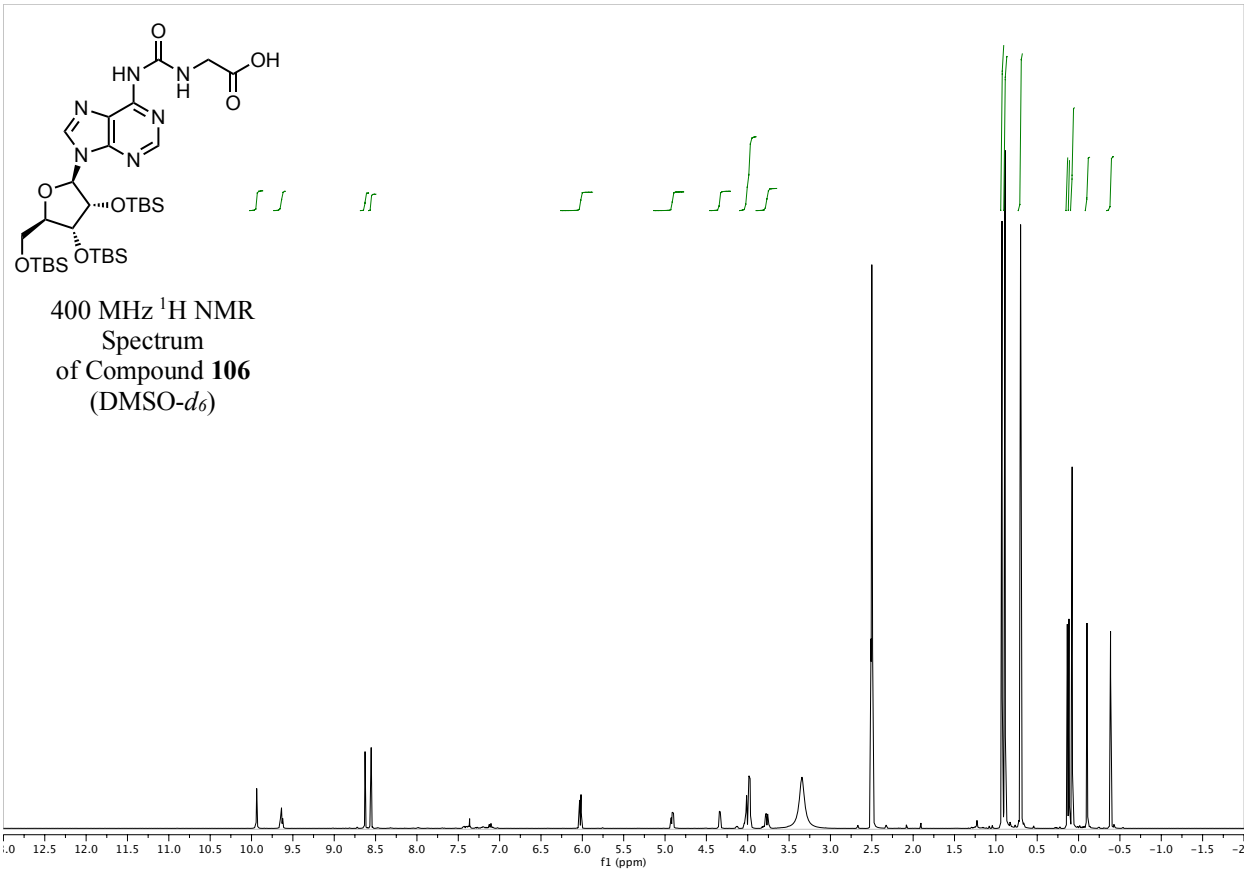
101 MHz ^{13}C NMR Spectrum of Compound **104** (Chloroform-*d*)



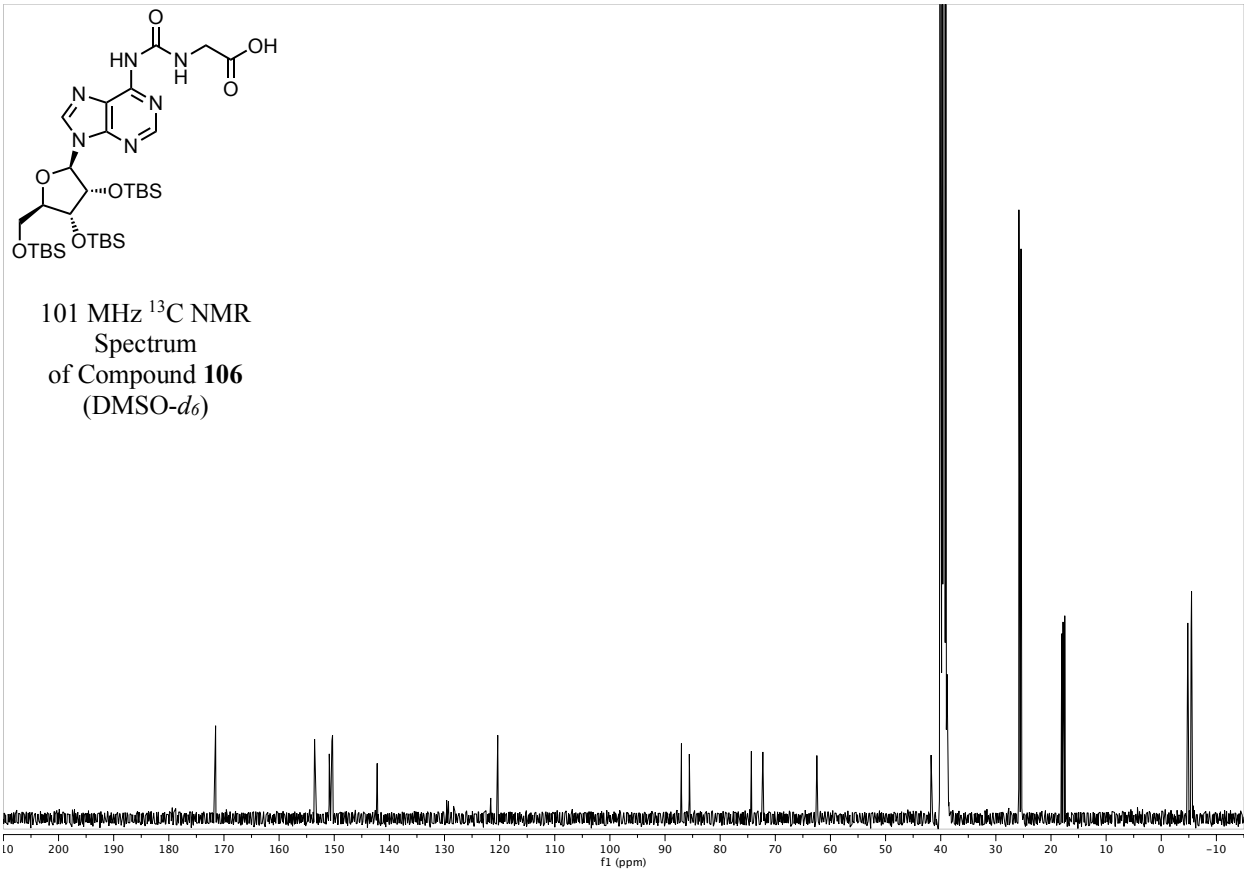
400 MHz ^{1}H NMR Spectrum of Compound **105** (Benzene-d_6)



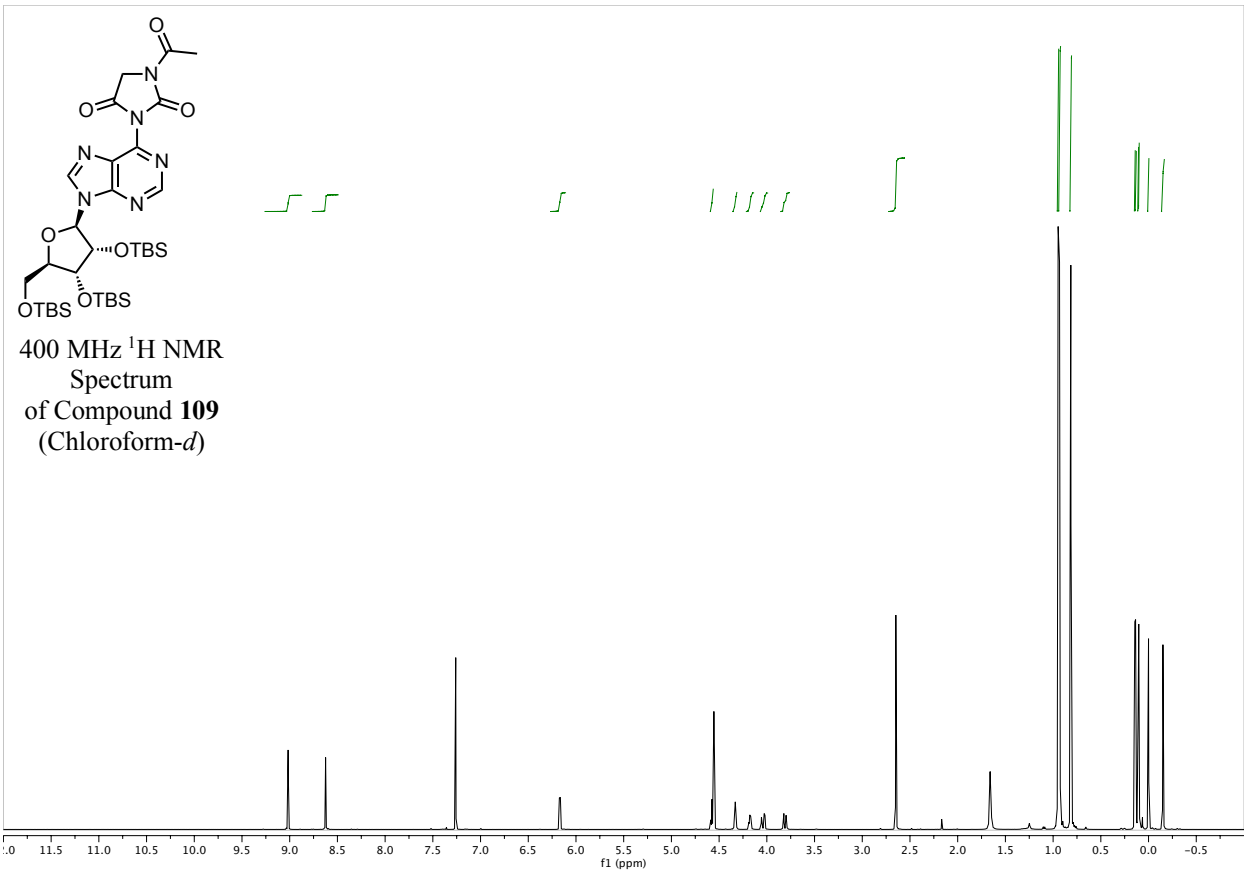
101 MHz ^{13}C NMR Spectrum of Compound **105** (Benzene-d_6)



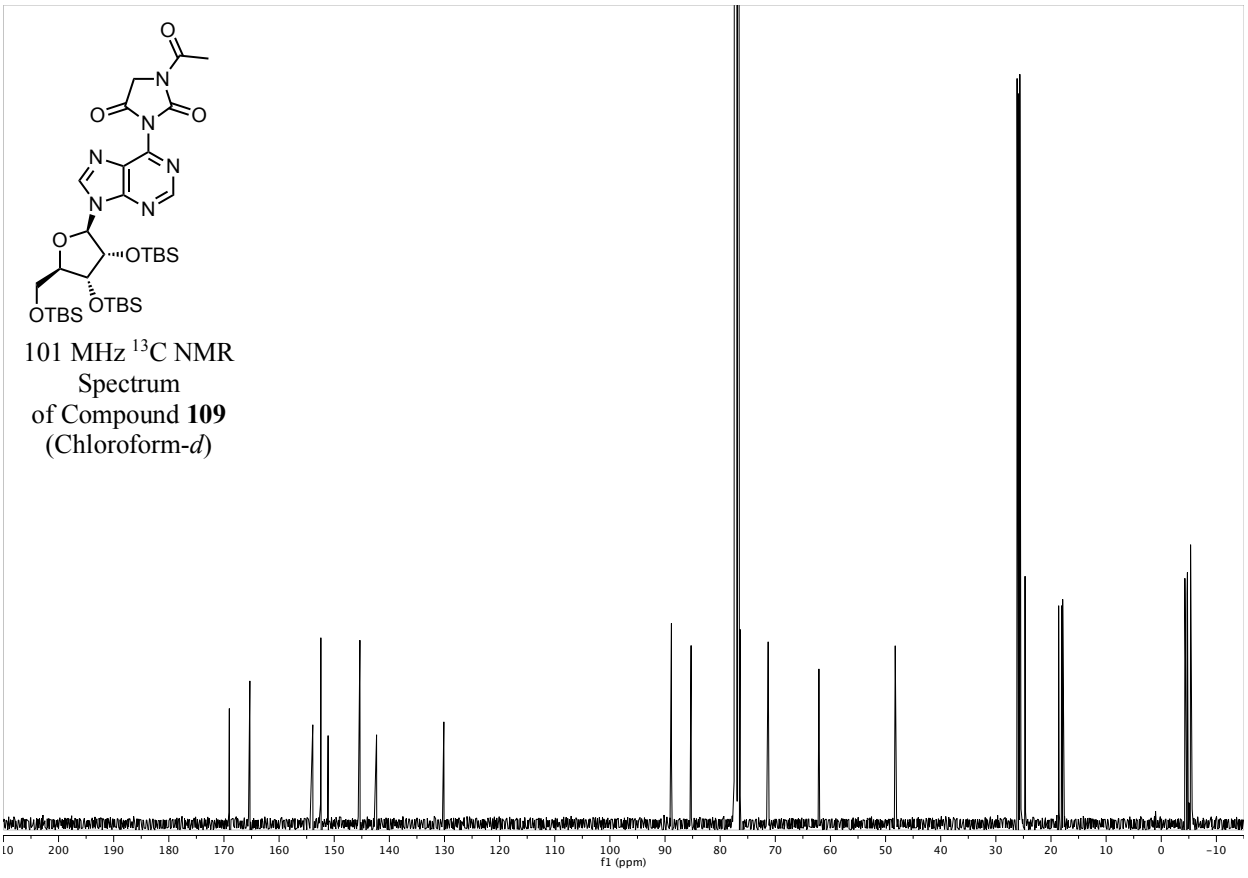
400 MHz ^{1}H NMR
Spectrum
of Compound **106**
(DMSO-d_6)



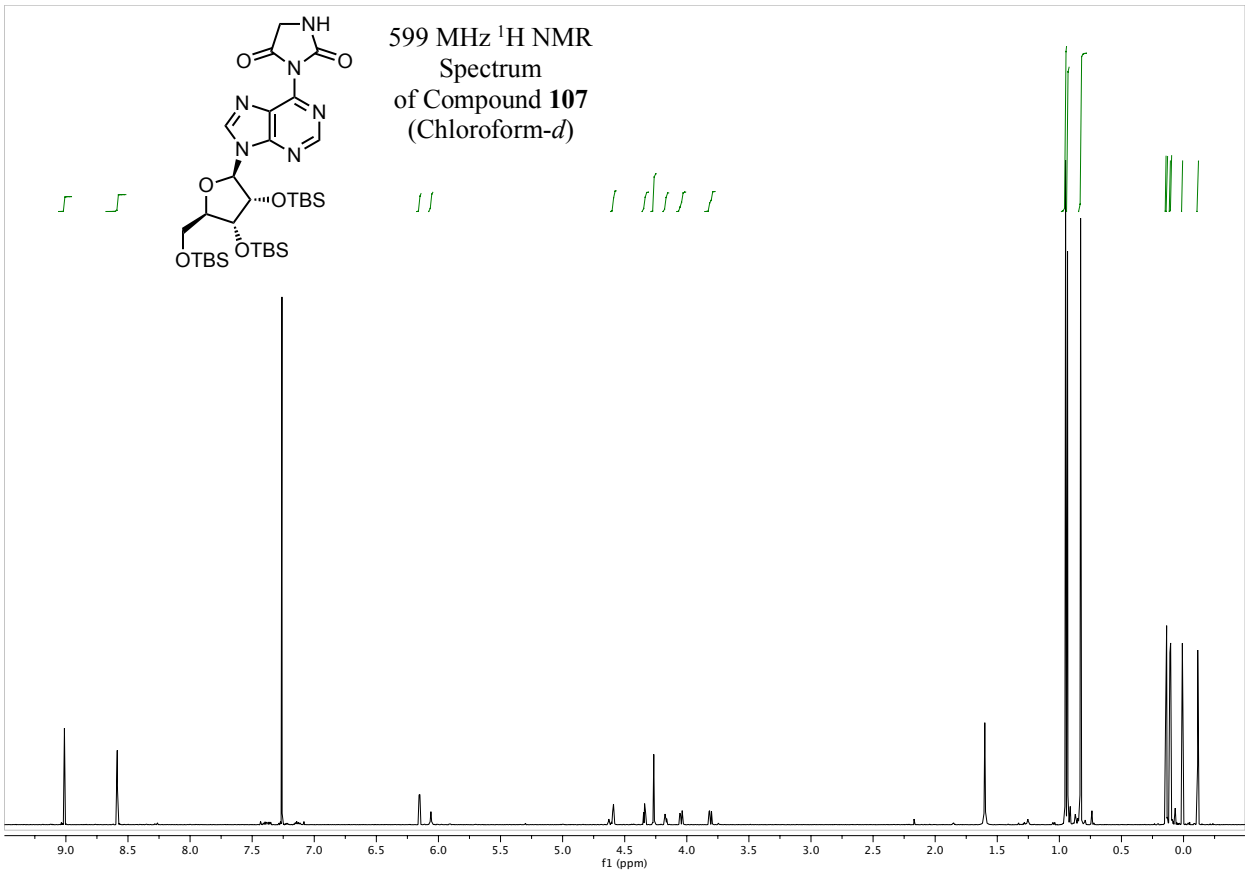
101 MHz ^{13}C NMR
Spectrum
of Compound **106**
(DMSO-d_6)



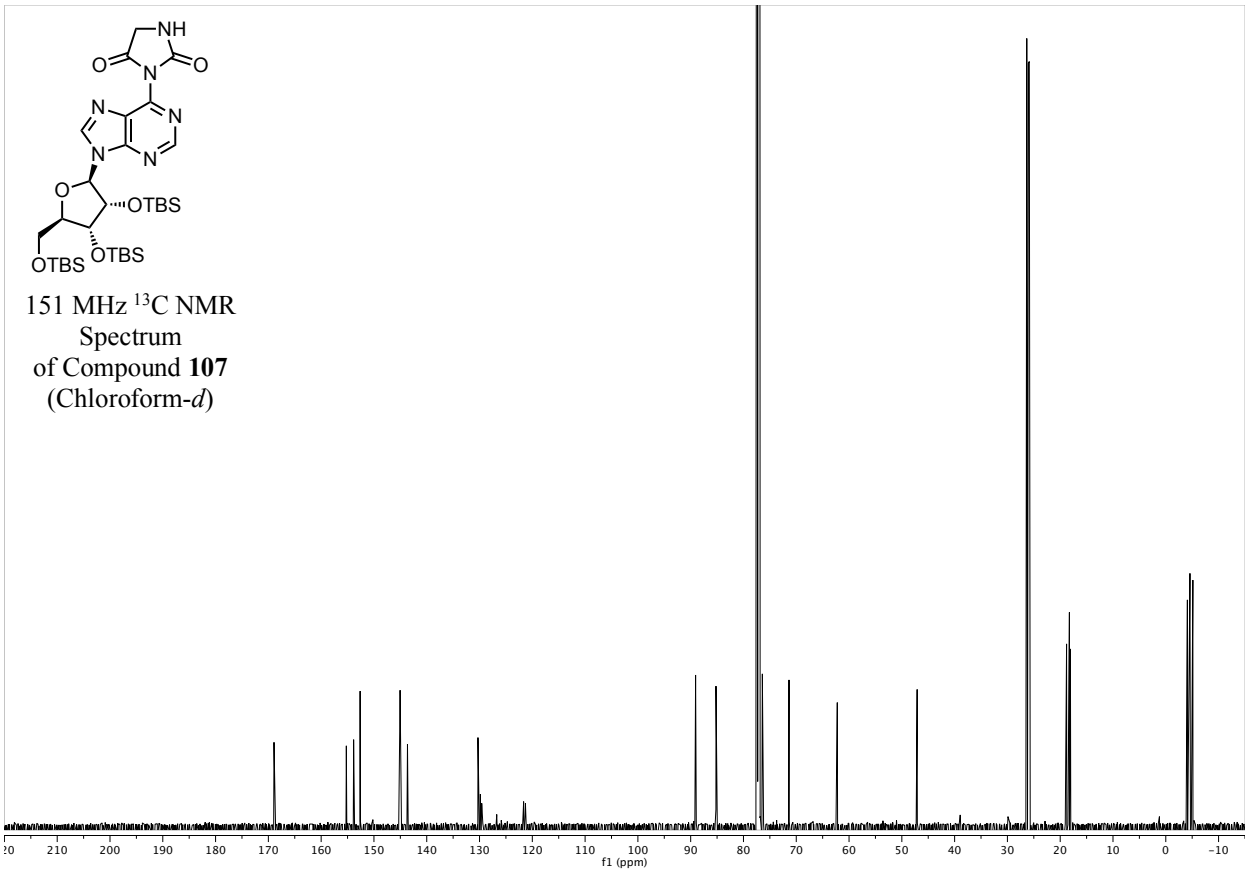
400 MHz ^{1}H NMR
Spectrum
of Compound **109**
(Chloroform-*d*)



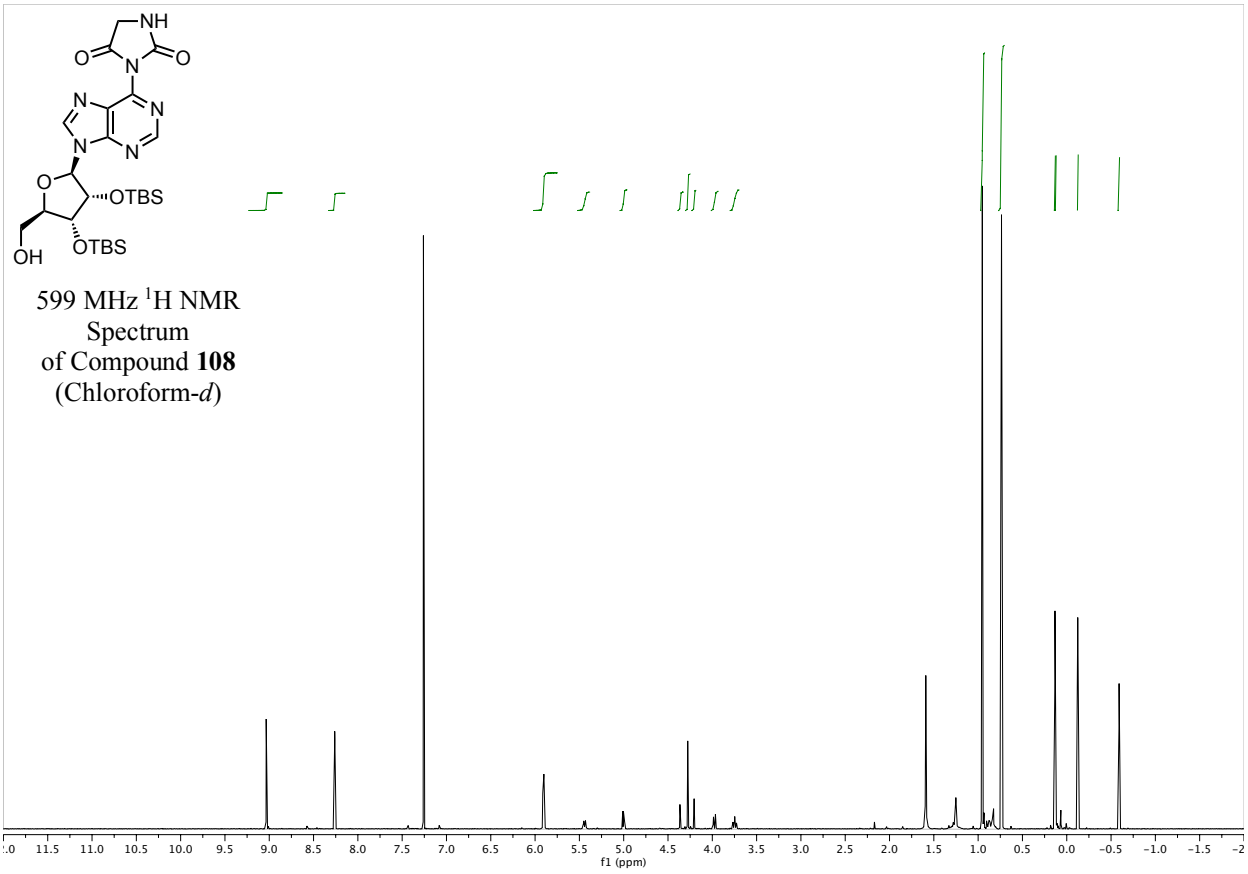
101 MHz ^{13}C NMR
Spectrum
of Compound **109**
(Chloroform-*d*)



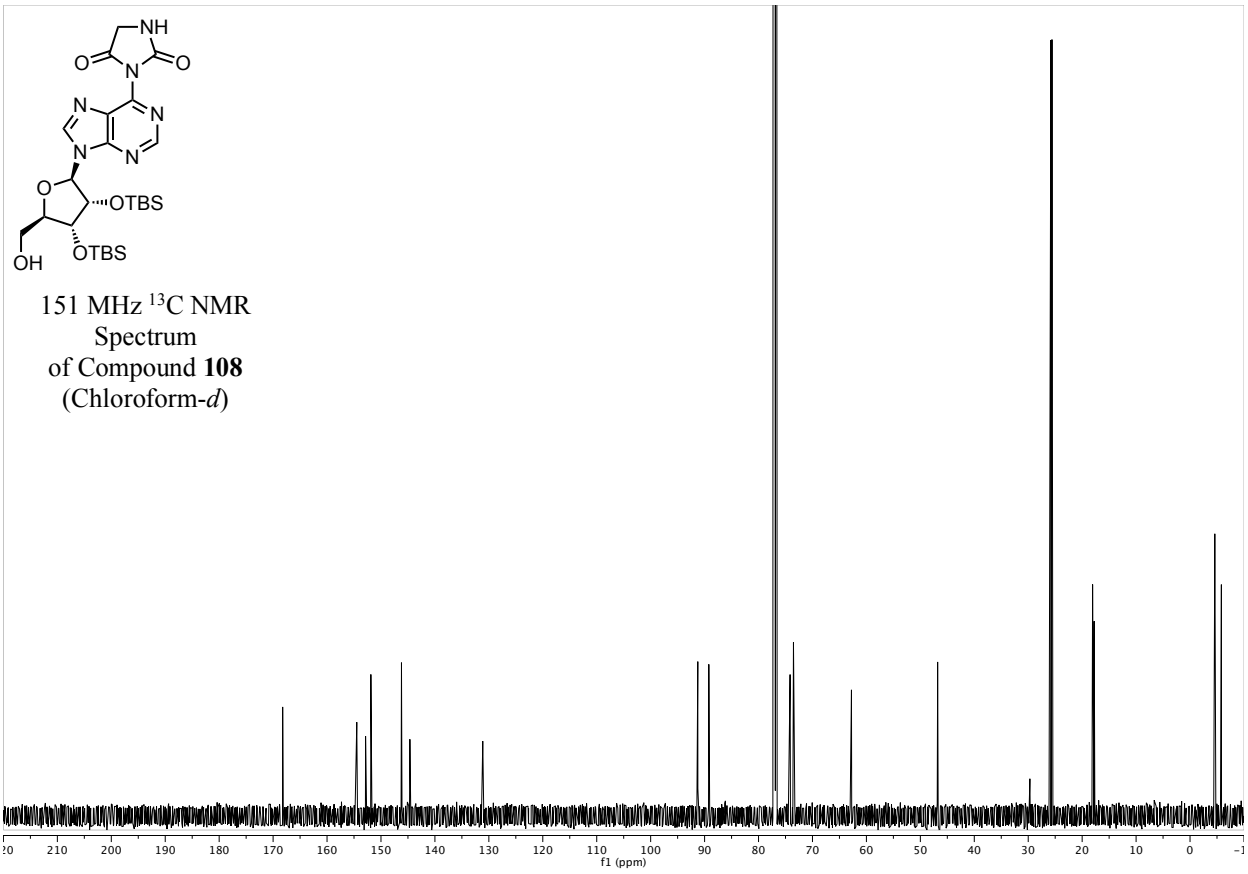
599 MHz ^{1}H NMR Spectrum of Compound **107** (Chloroform-*d*)



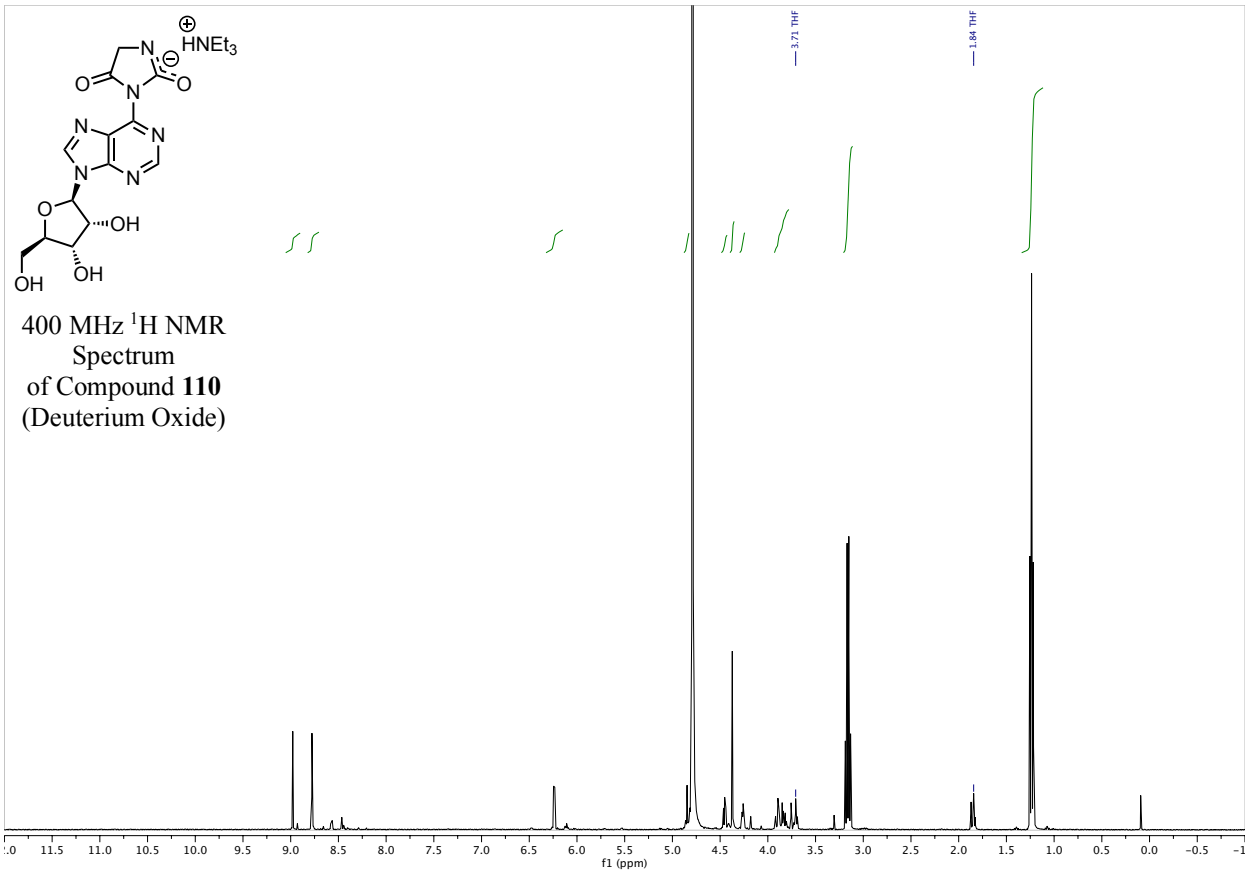
151 MHz ^{13}C NMR Spectrum of Compound **107** (Chloroform-*d*)



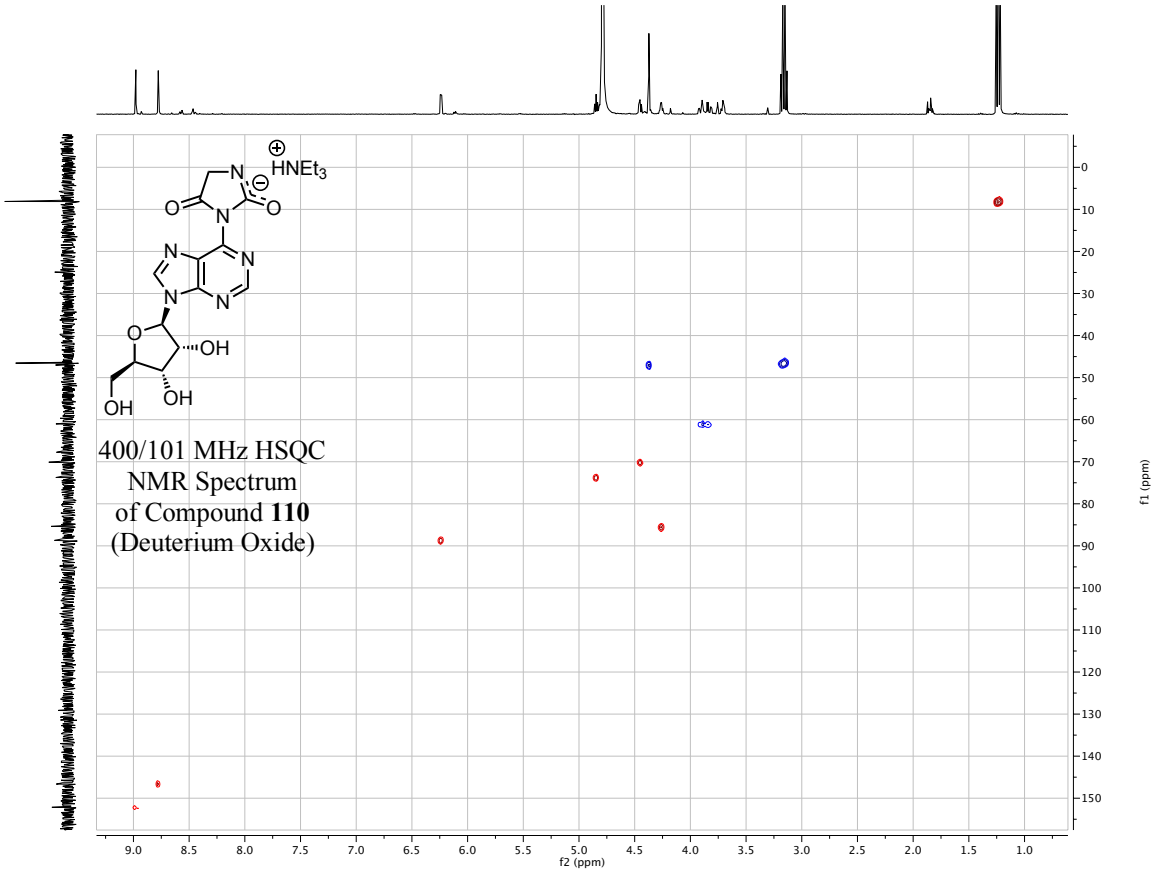
599 MHz ^{1}H NMR
Spectrum
of Compound **108**
(Chloroform-*d*)



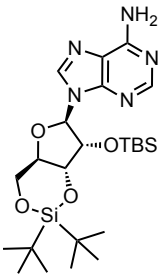
151 MHz ^{13}C NMR
Spectrum
of Compound **108**
(Chloroform-*d*)



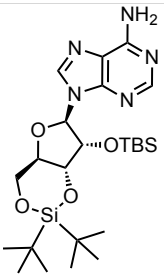
400 MHz ^{1}H NMR Spectrum of Compound **110** (Deuterium Oxide)



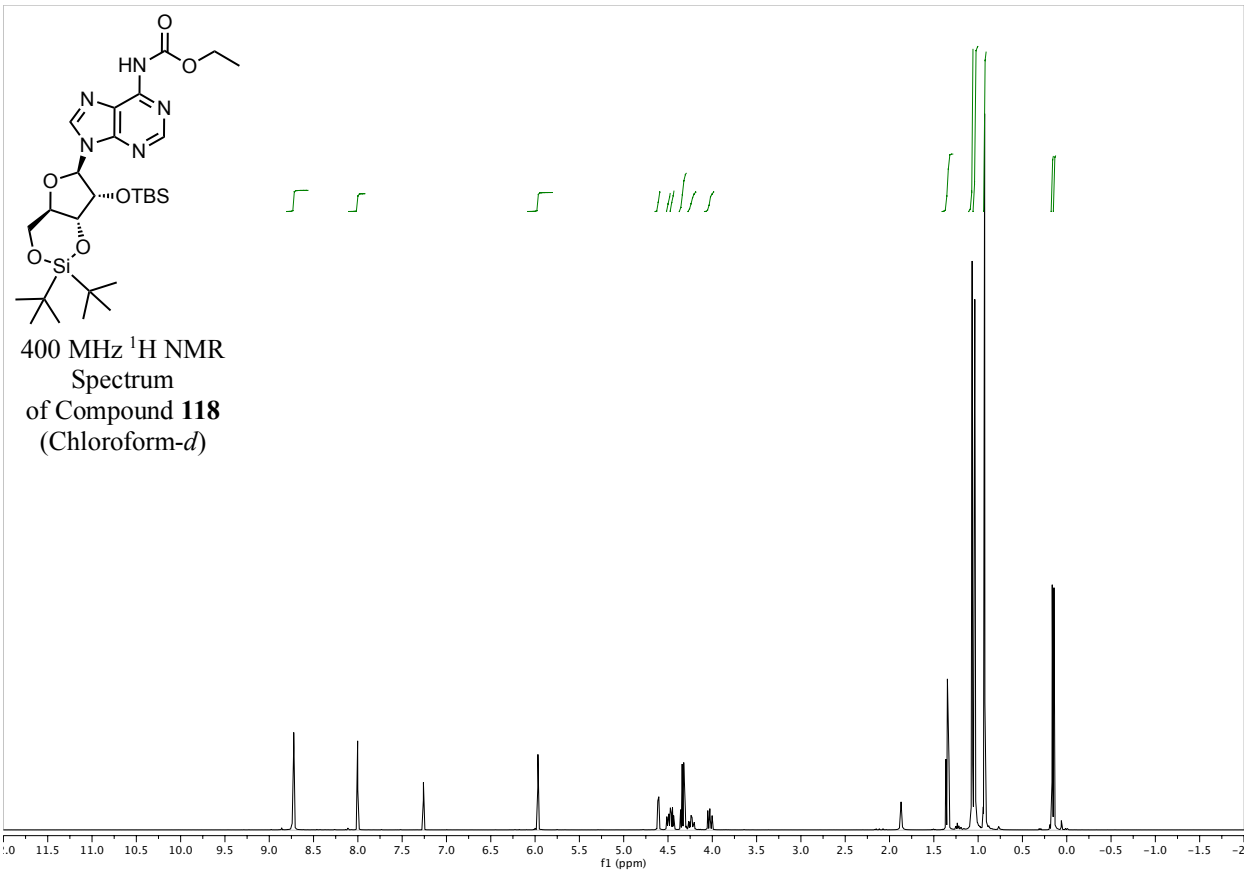
400/101 MHz HSQC NMR Spectrum of Compound **110** (Deuterium Oxide)



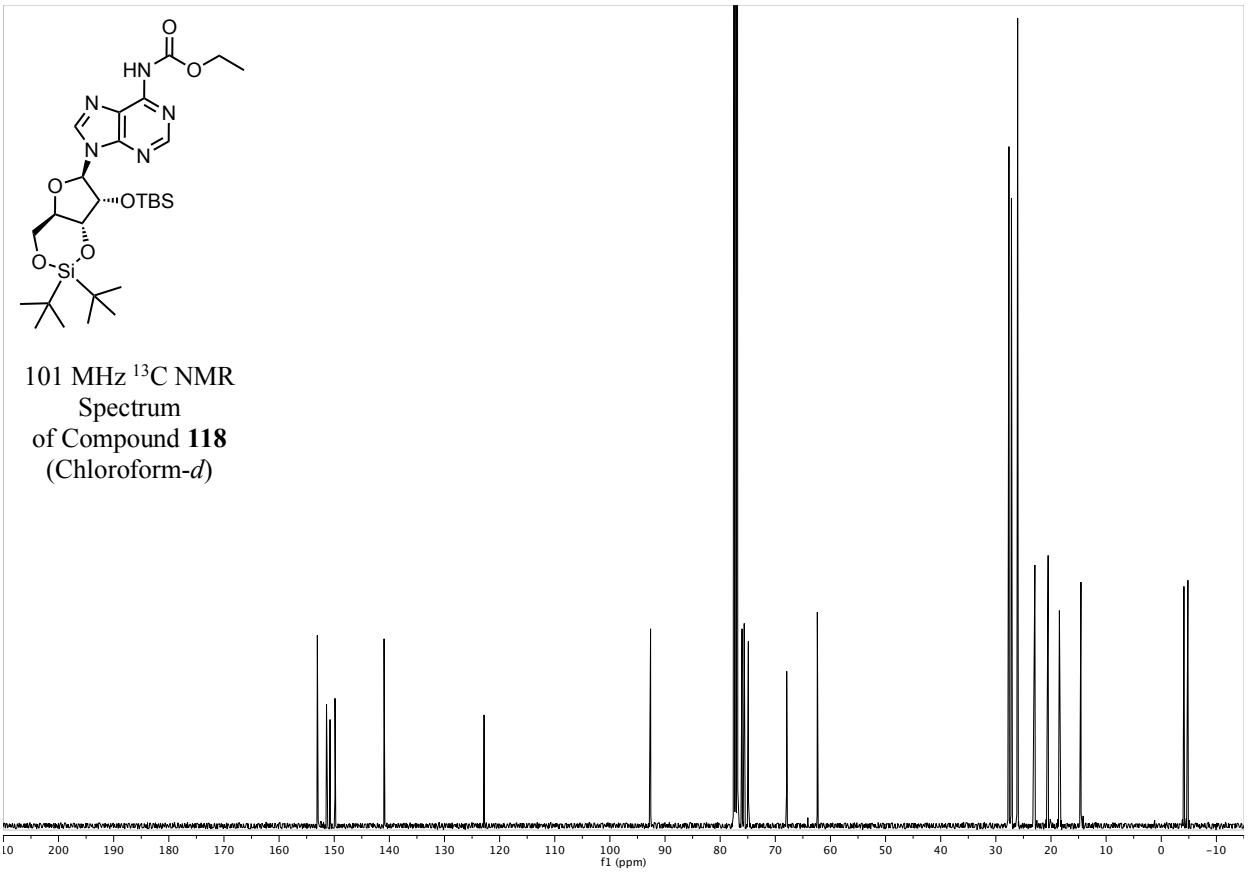
400 MHz ^{1}H NMR
Spectrum
of Compound **116**
(Chloroform-*d*)



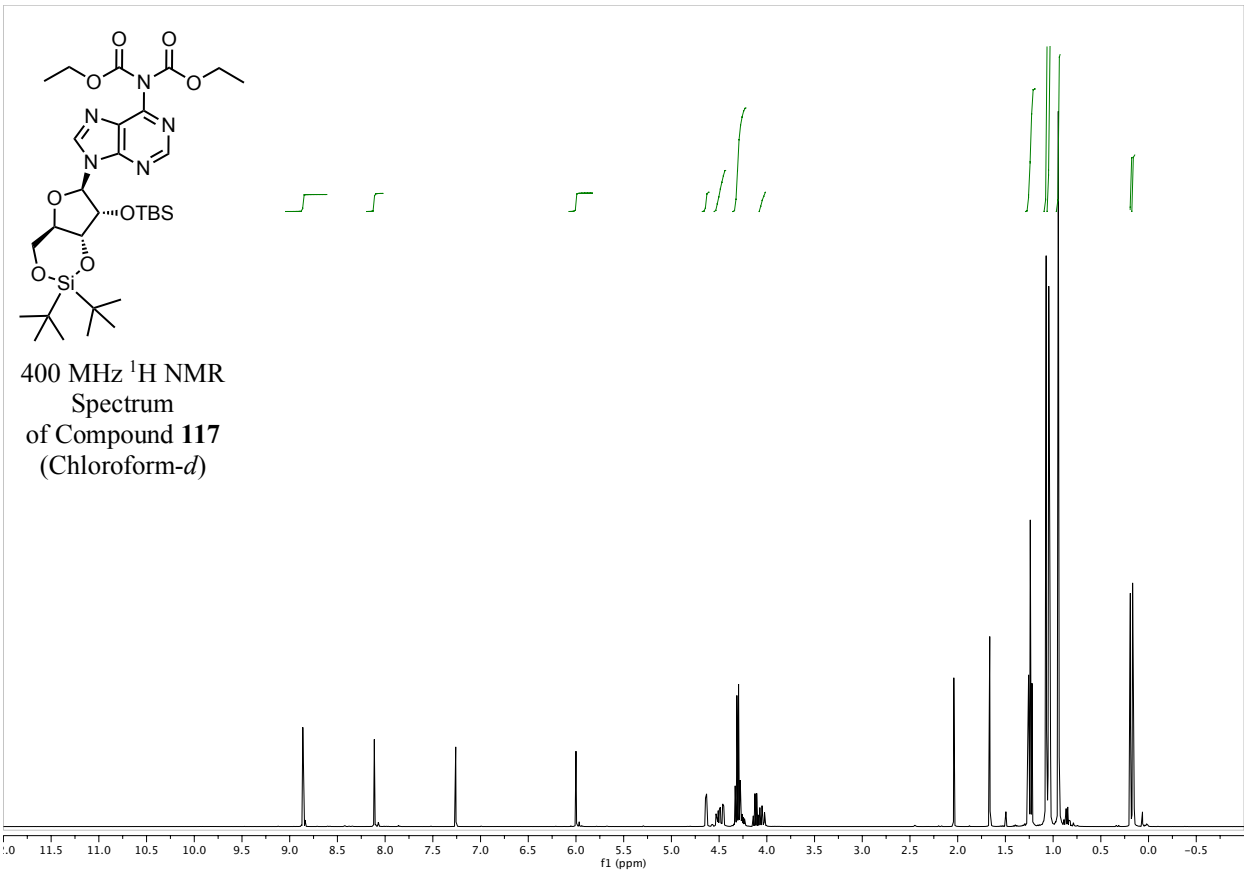
101 MHz ^{13}C NMR
Spectrum
of Compound **116**
(Chloroform-*d*)



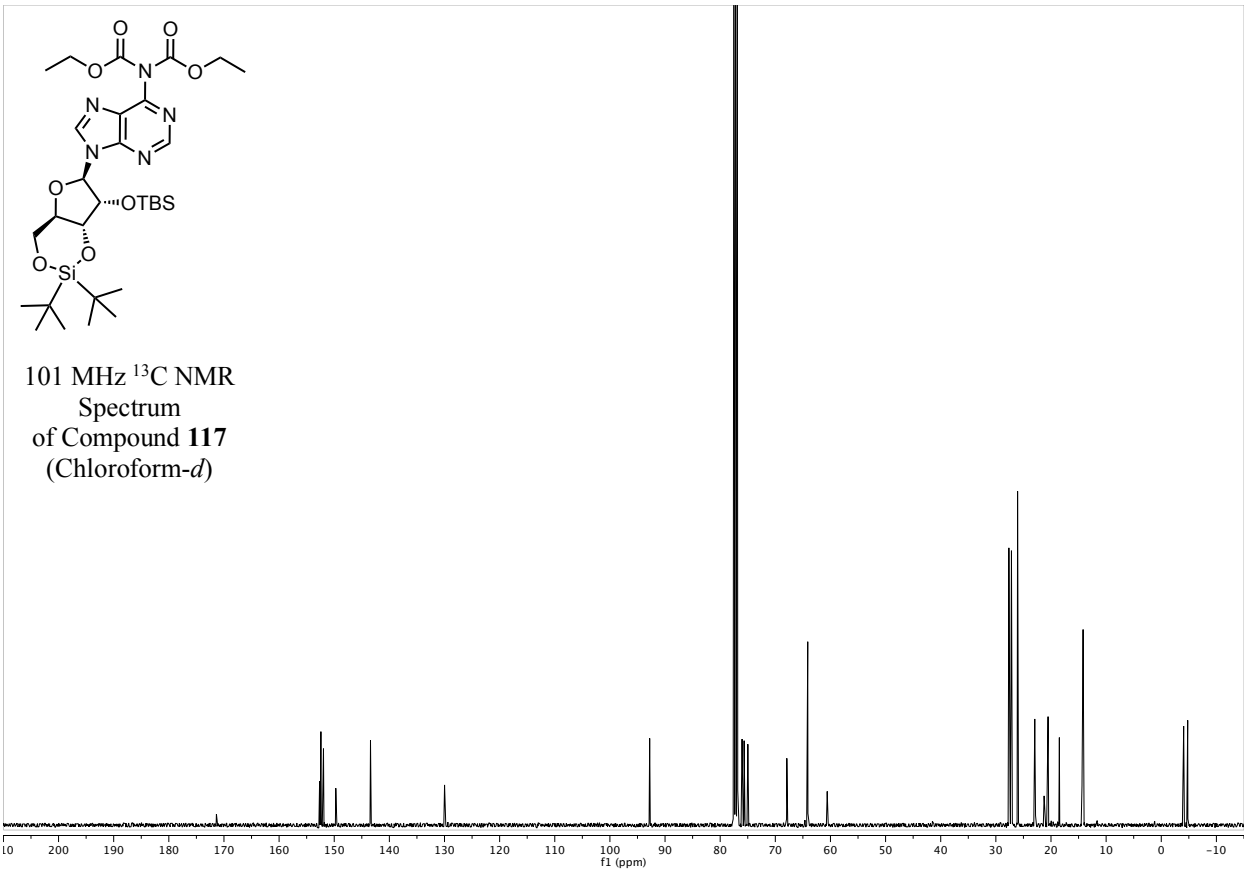
400 MHz ^{1}H NMR Spectrum of Compound **118** (Chloroform-*d*)



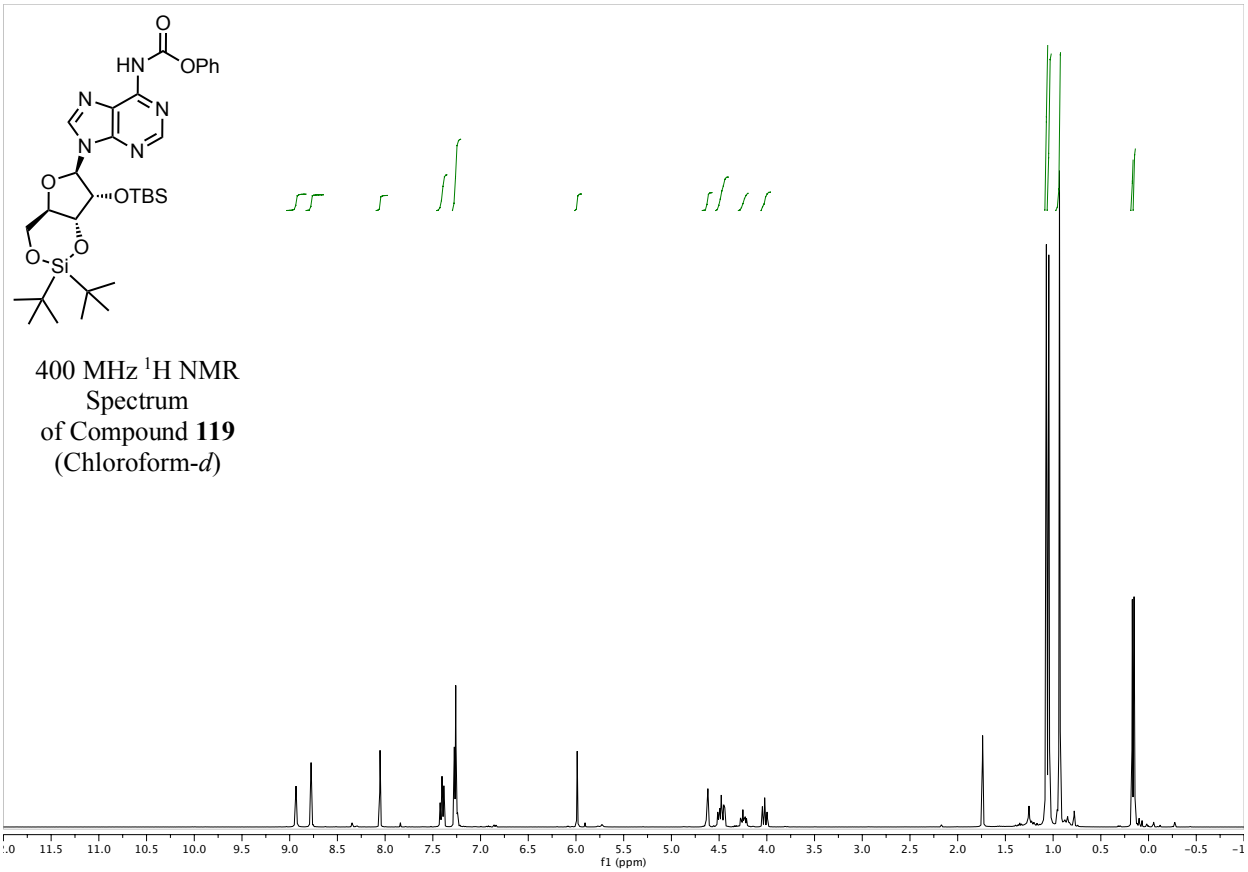
101 MHz ^{13}C NMR Spectrum of Compound **118** (Chloroform-*d*)



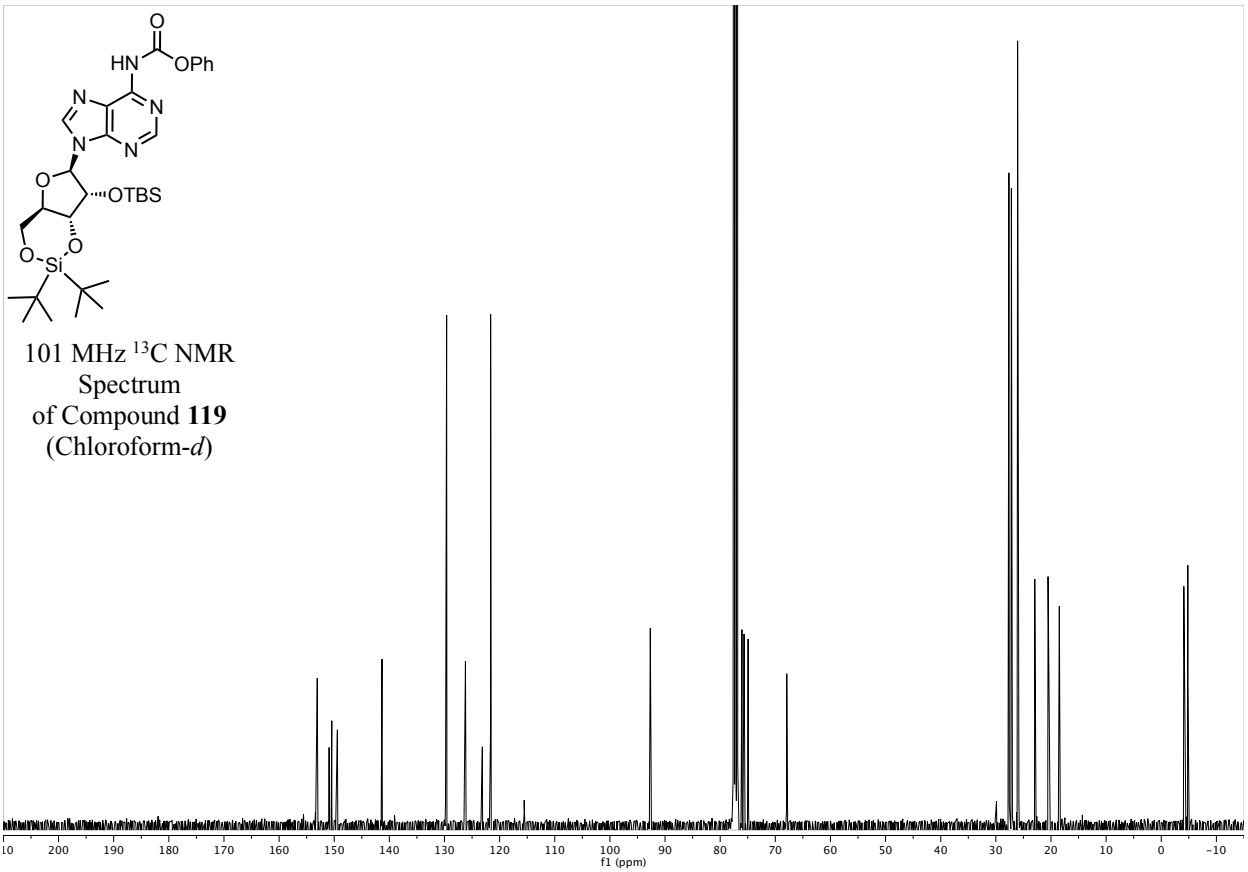
400 MHz ^{1}H NMR
Spectrum
of Compound **117**
(Chloroform-*d*)



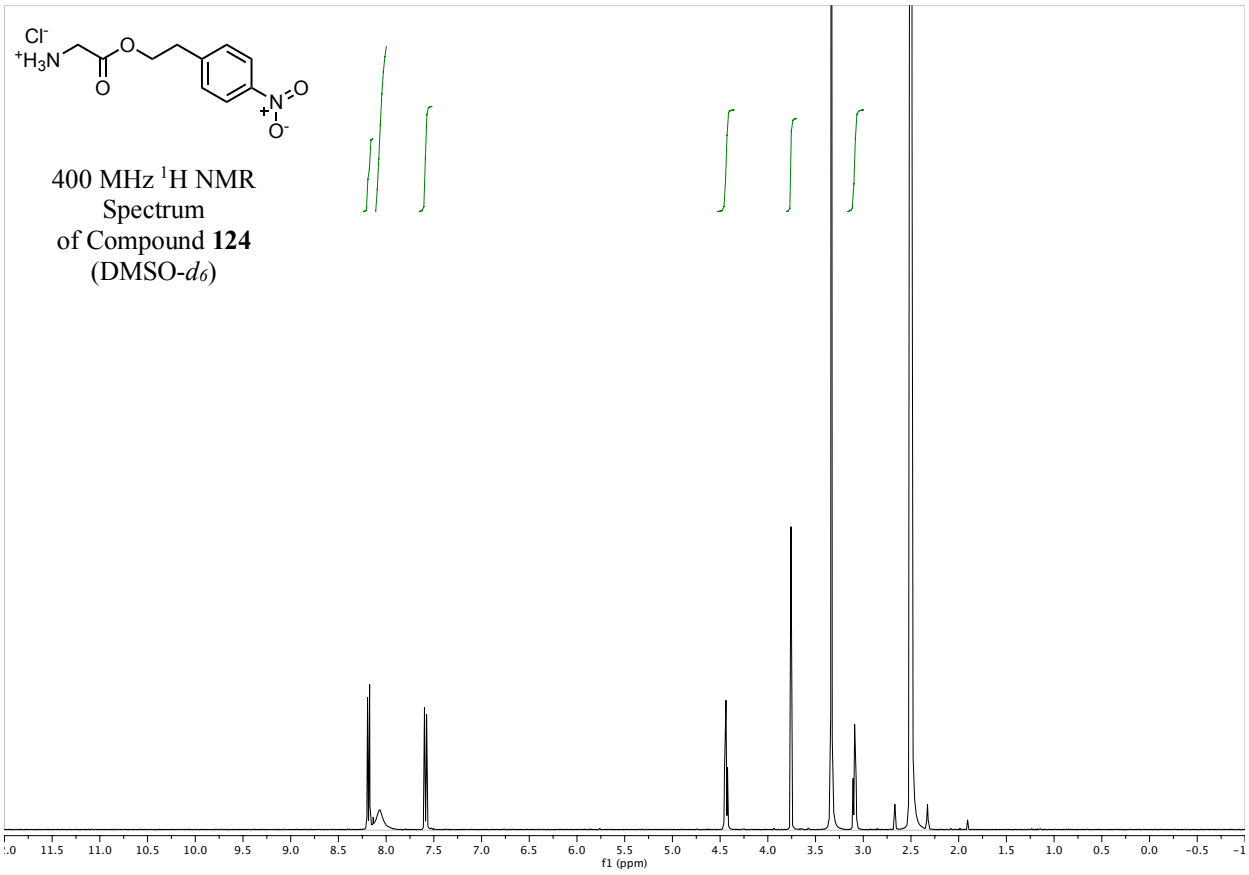
101 MHz ^{13}C NMR
Spectrum
of Compound **117**
(Chloroform-*d*)



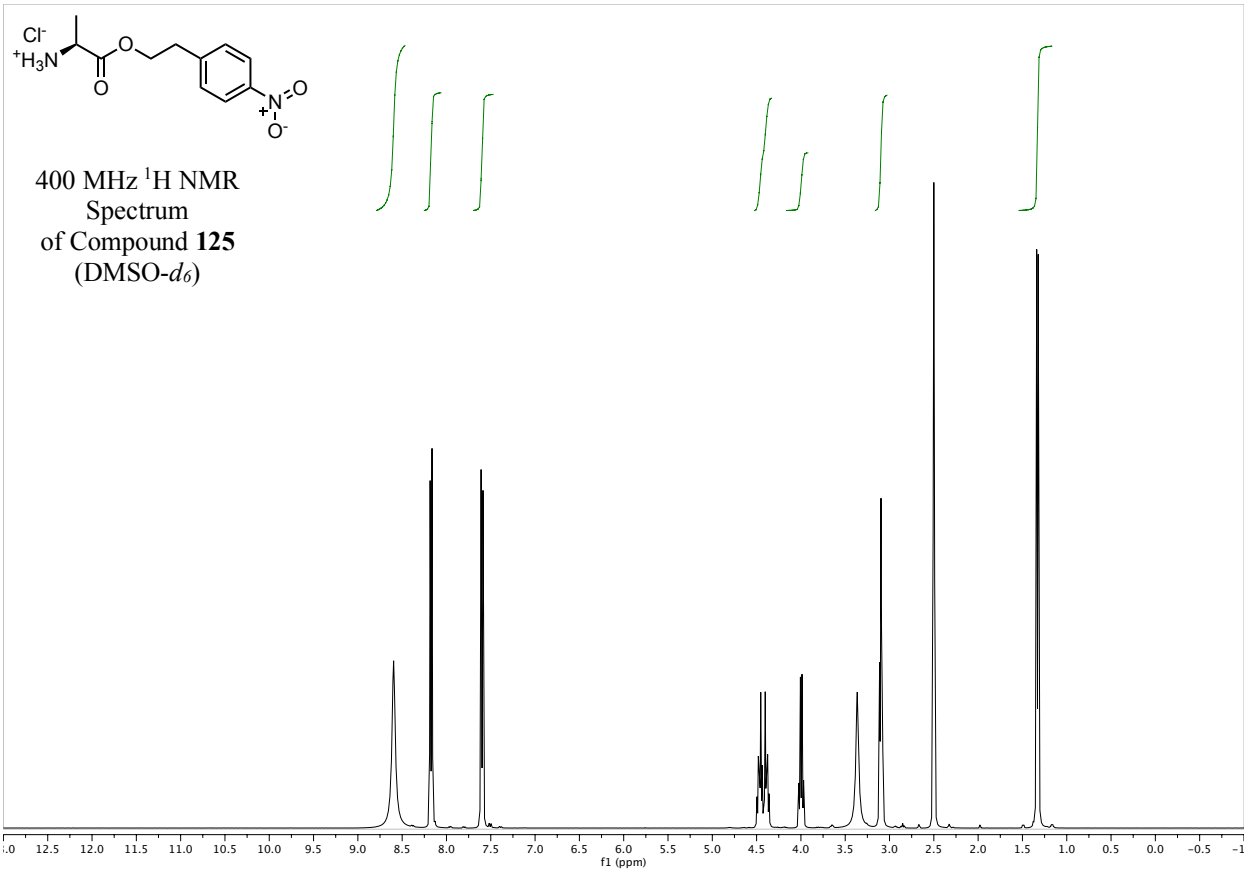
400 MHz ^{1}H NMR
Spectrum
of Compound **119**
(Chloroform-*d*)



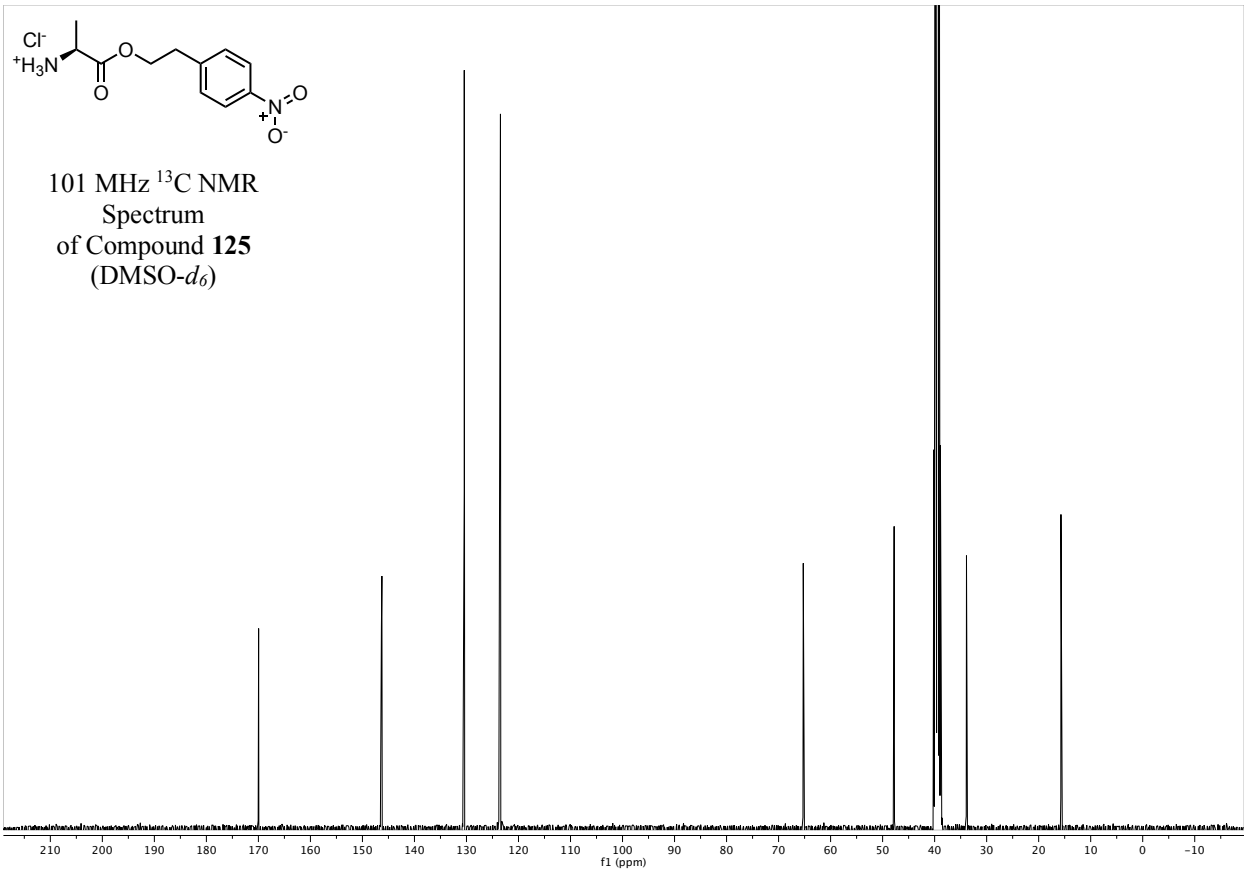
101 MHz ^{13}C NMR
Spectrum
of Compound **119**
(Chloroform-*d*)



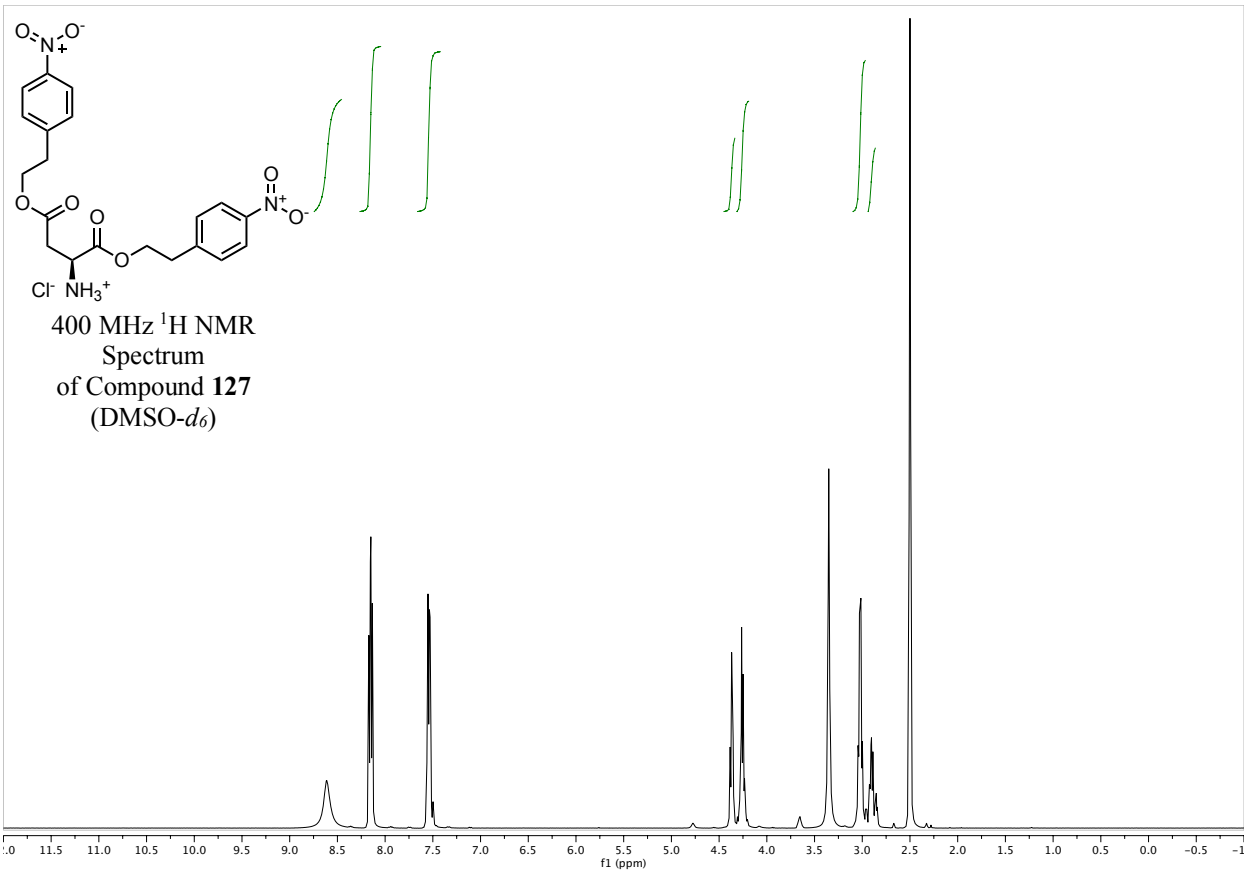
400 MHz ^{1}H NMR Spectrum of Compound **124** (DMSO-d_6)



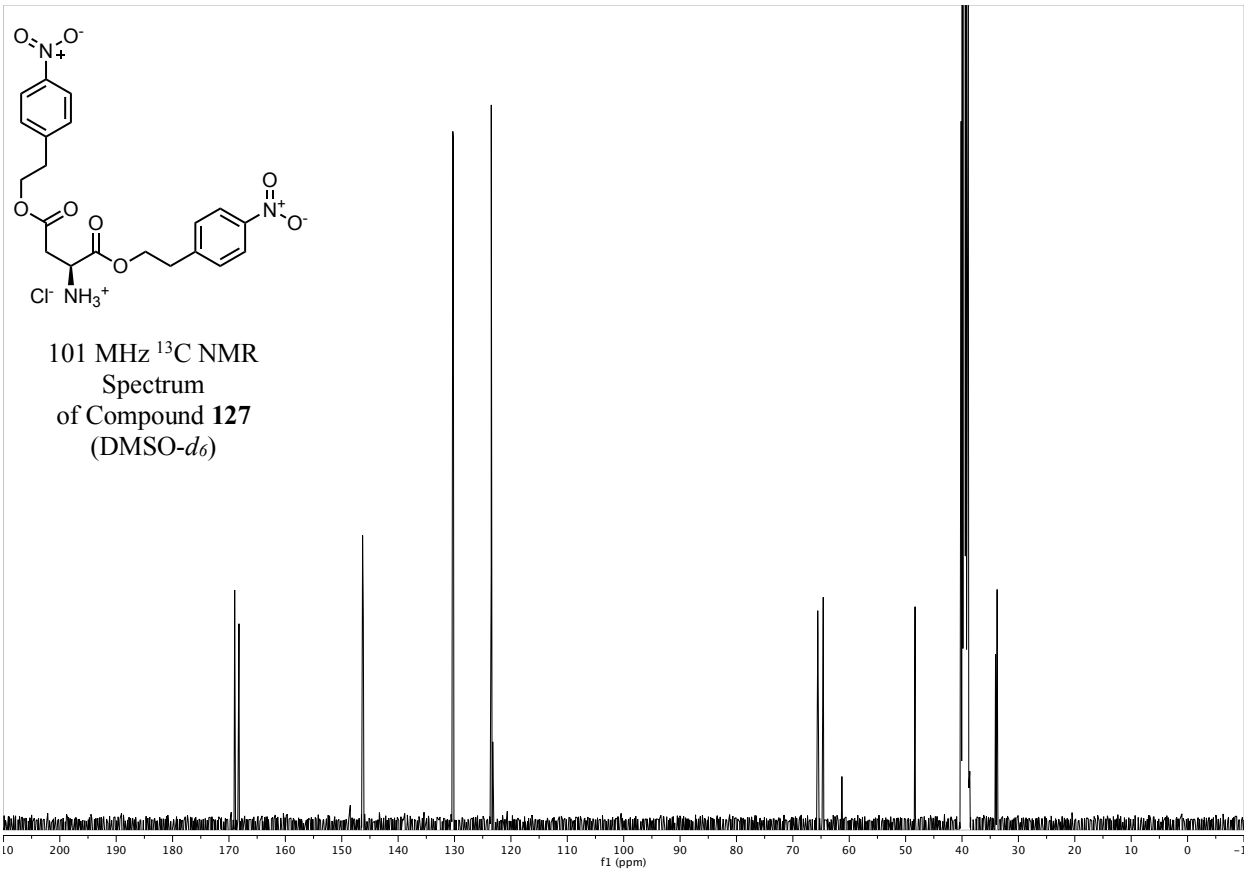
400 MHz ^{1}H NMR Spectrum of Compound **125** (DMSO-d_6)



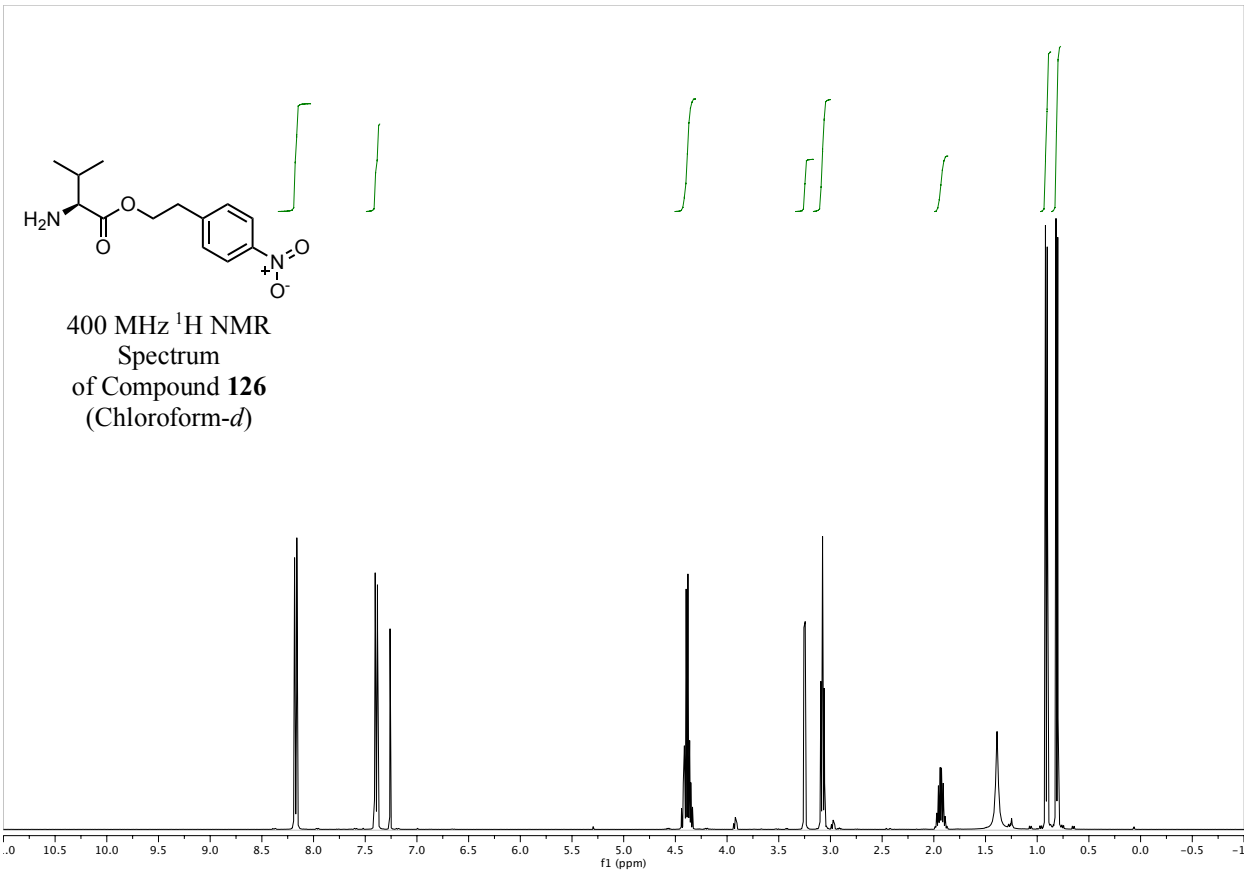
101 MHz ^{13}C NMR Spectrum of Compound **125** (DMSO-d_6)



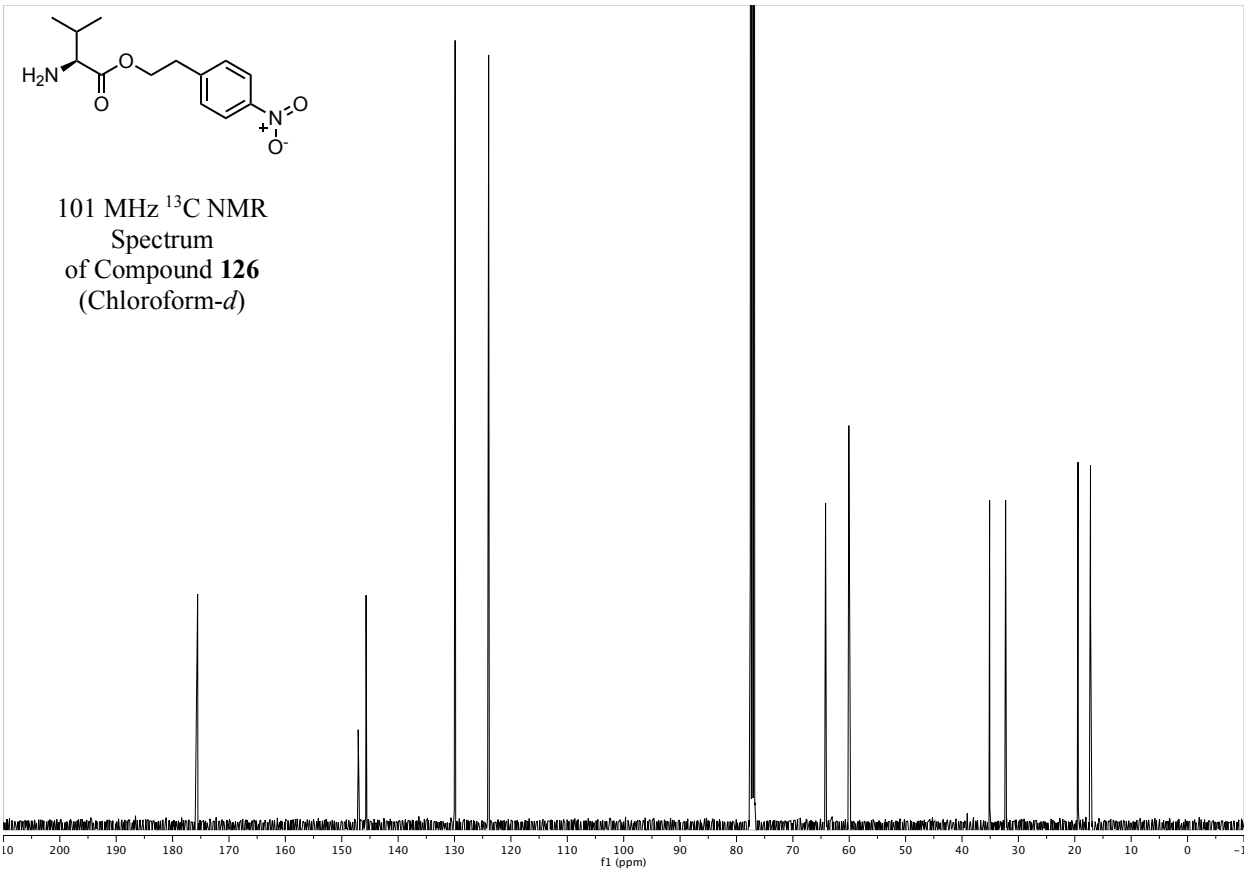
400 MHz ^{1}H NMR Spectrum of Compound **127** (DMSO-*d*$_6$)



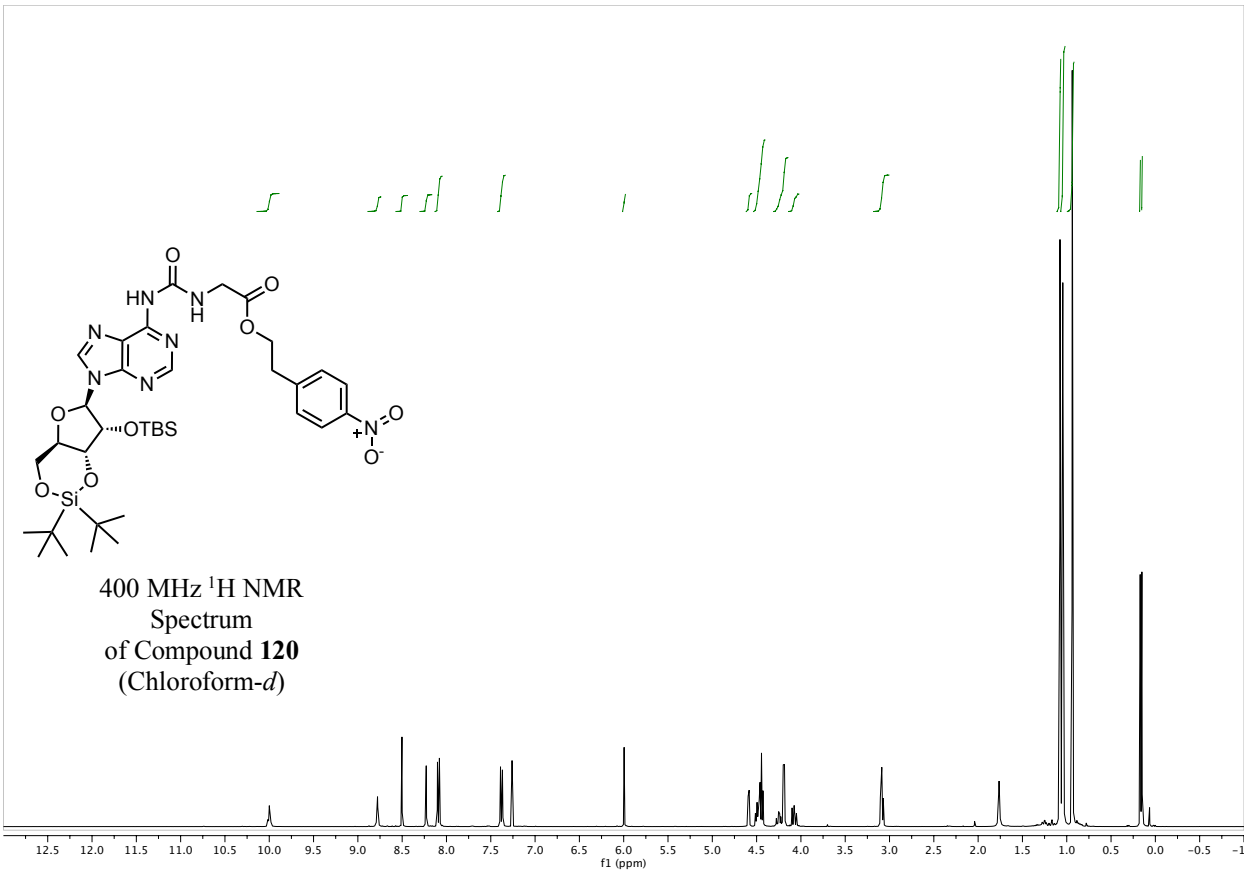
101 MHz ^{13}C NMR Spectrum of Compound **127** (DMSO-*d*$_6$)



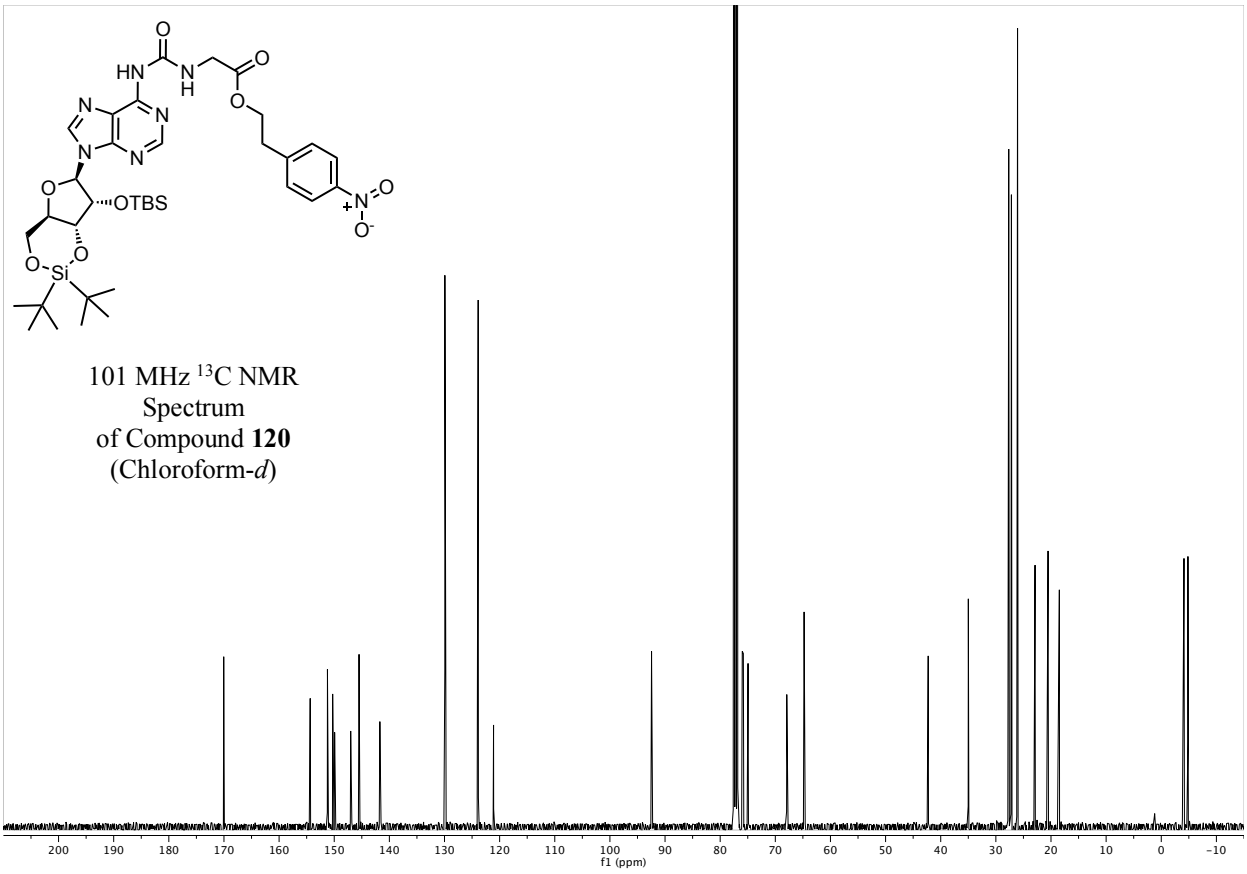
400 MHz ^{1}H NMR Spectrum of Compound **126** (Chloroform-*d*)



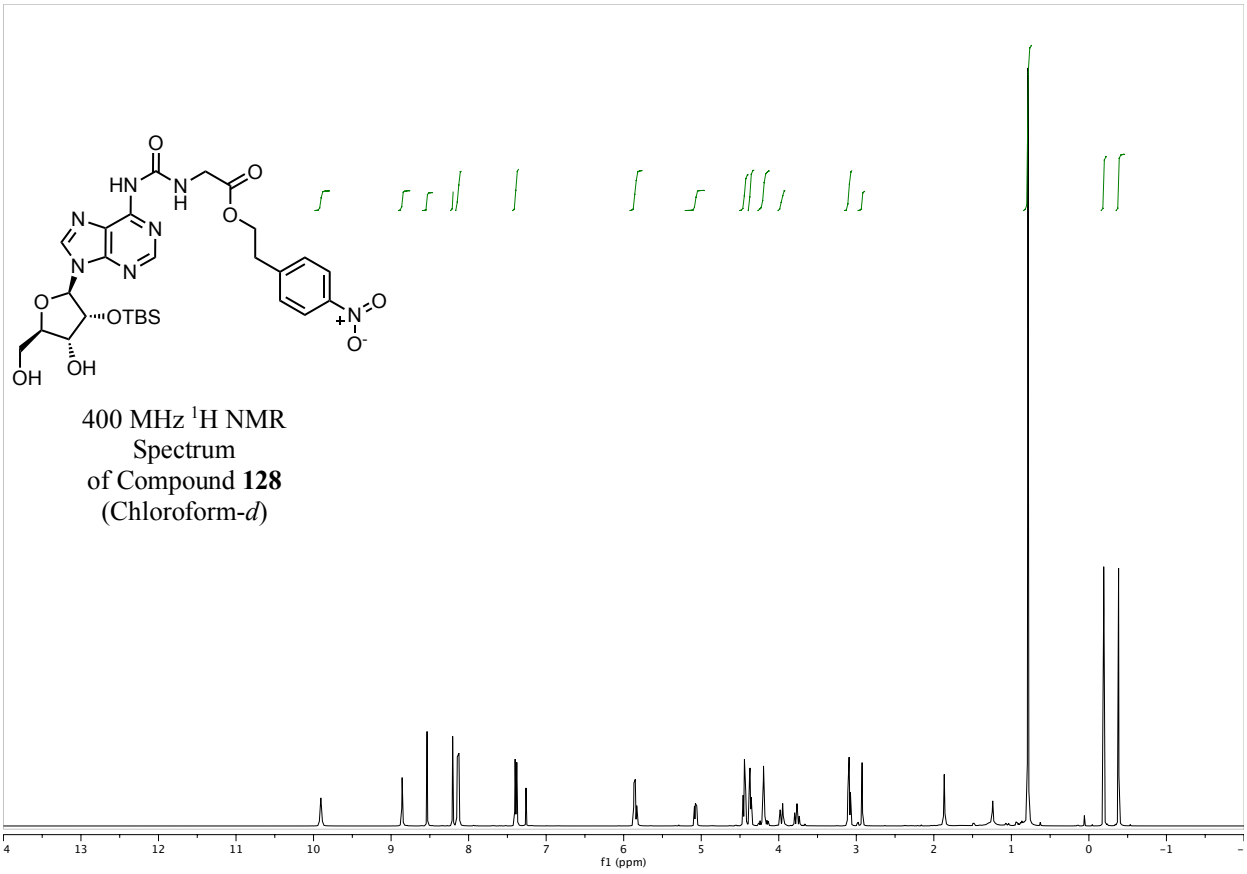
101 MHz ^{13}C NMR Spectrum of Compound **126** (Chloroform-*d*)



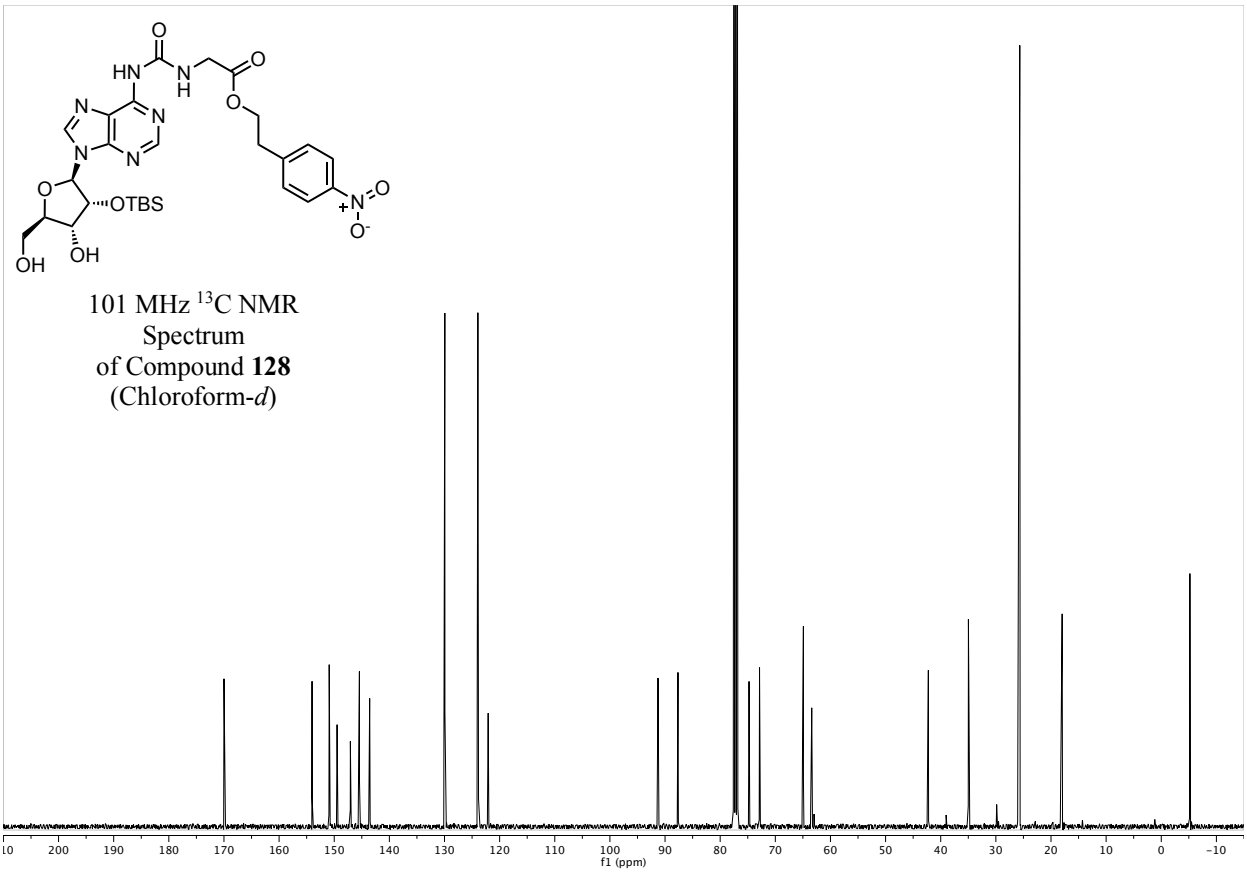
400 MHz ^{1}H NMR Spectrum of Compound **120** (Chloroform-*d*)



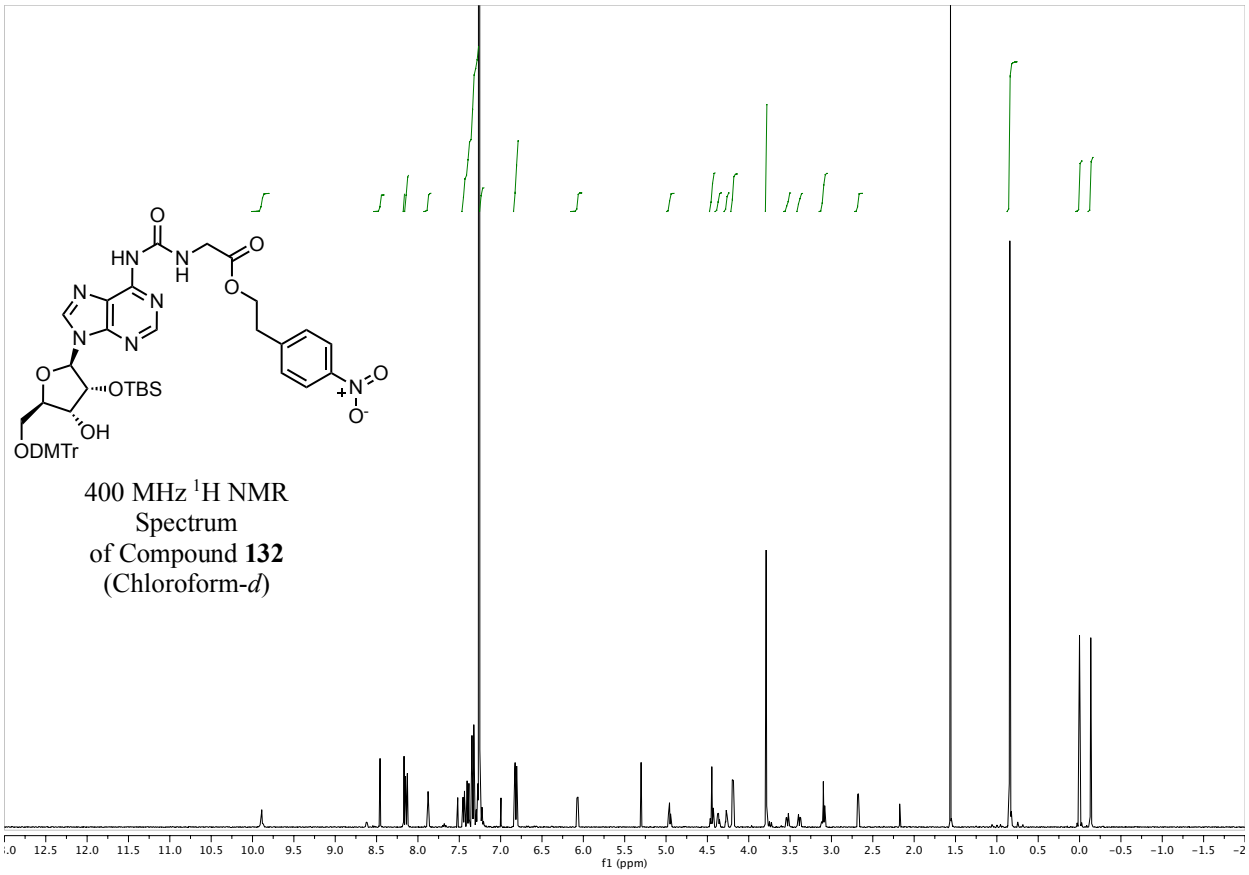
101 MHz ^{13}C NMR Spectrum of Compound **120** (Chloroform-*d*)



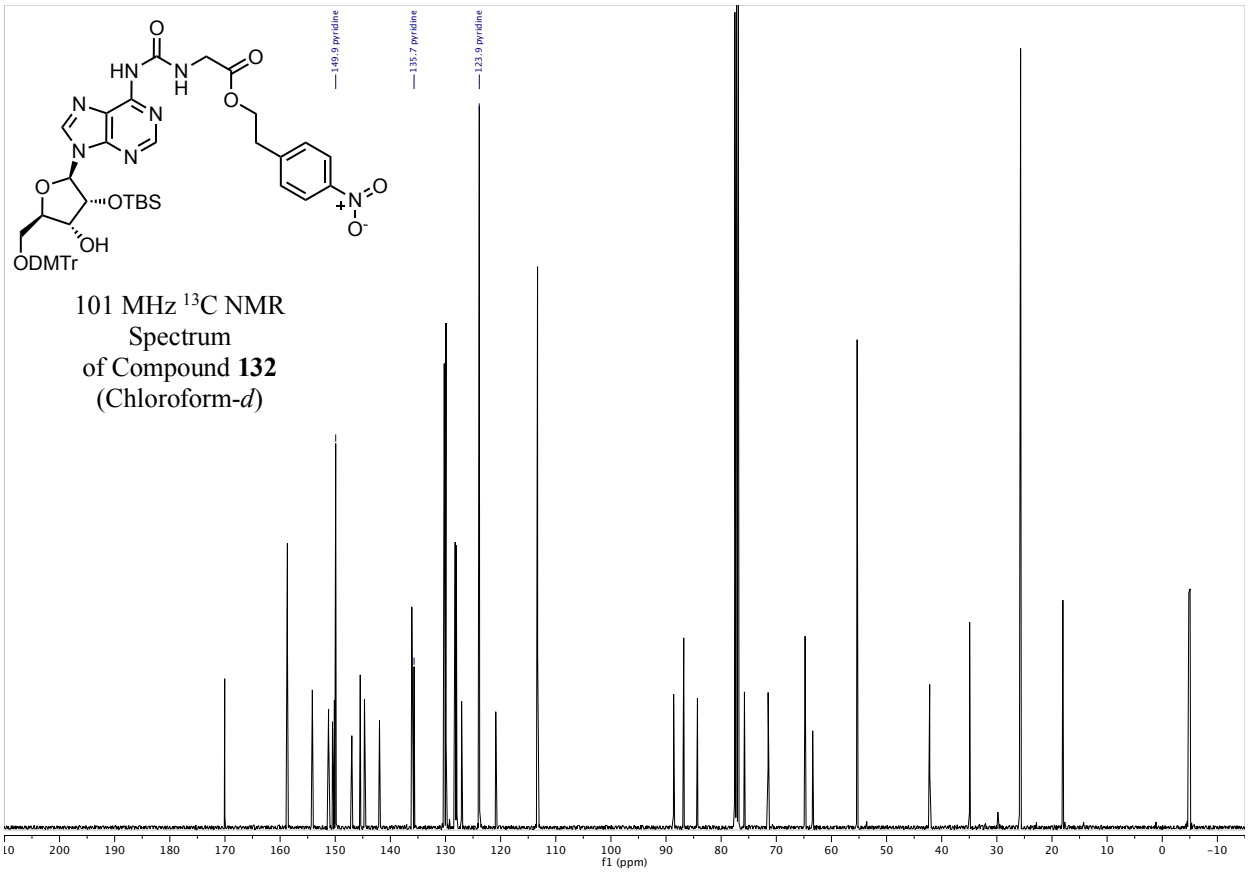
400 MHz ^{1}H NMR
Spectrum
of Compound **128**
(Chloroform-*d*)



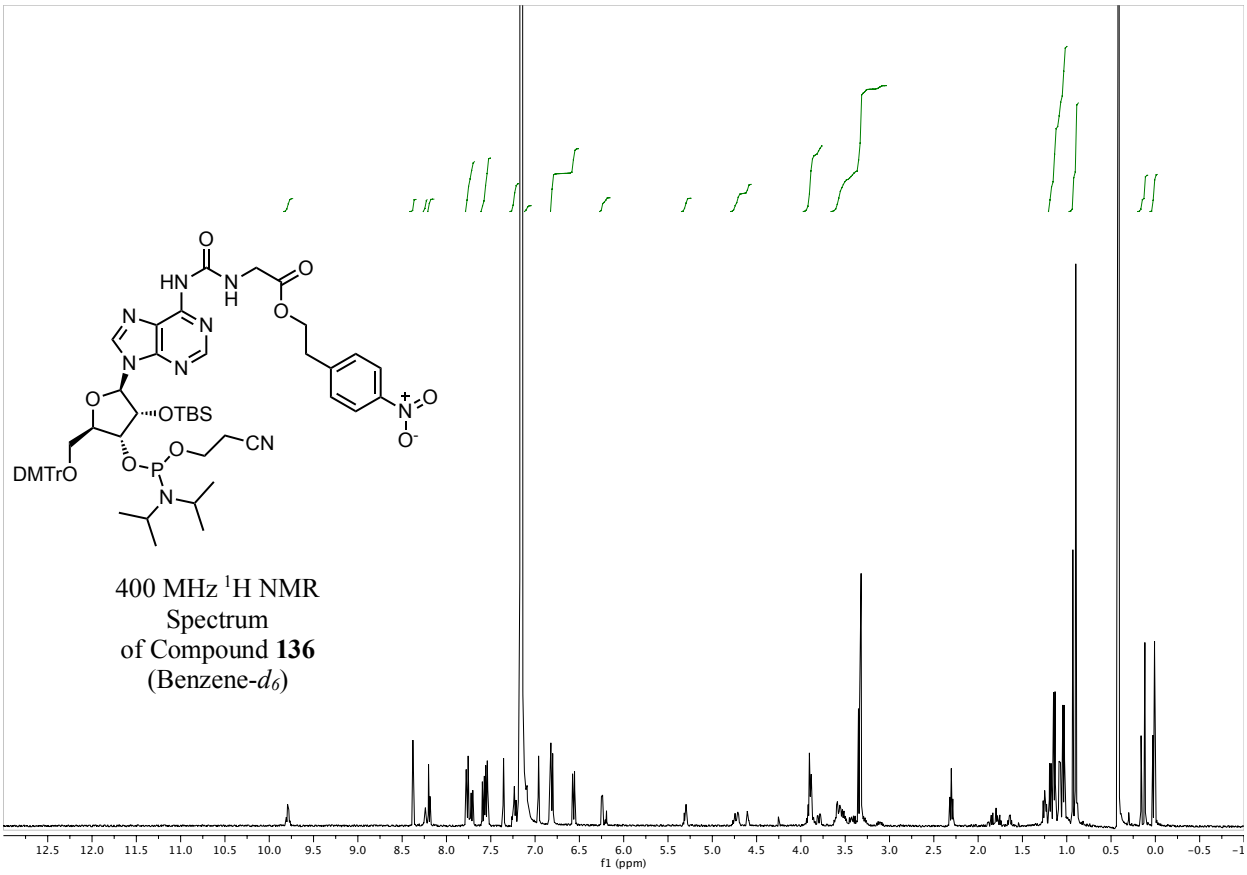
101 MHz ^{13}C NMR
Spectrum
of Compound **128**
(Chloroform-*d*)



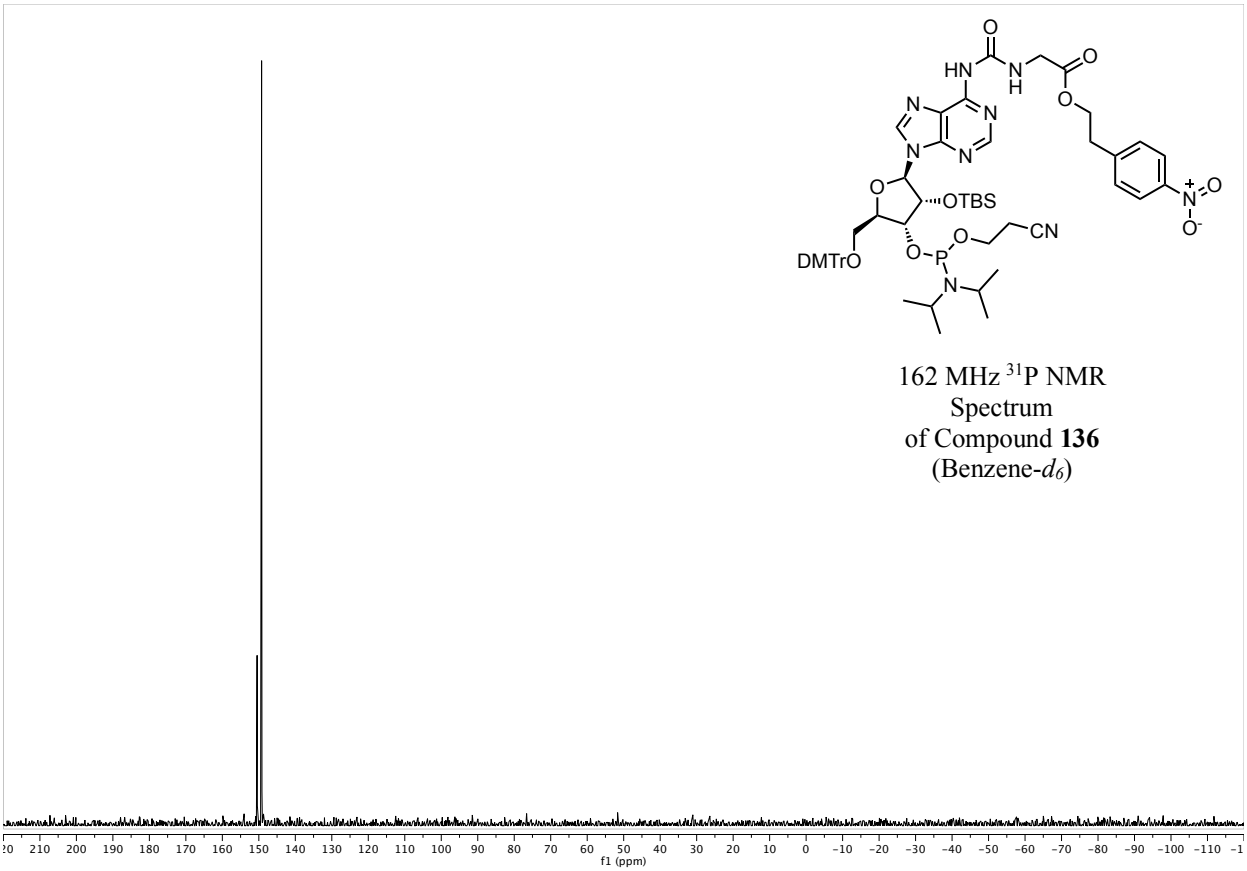
400 MHz ^{1}H NMR
Spectrum
of Compound **132**
(Chloroform-*d*)



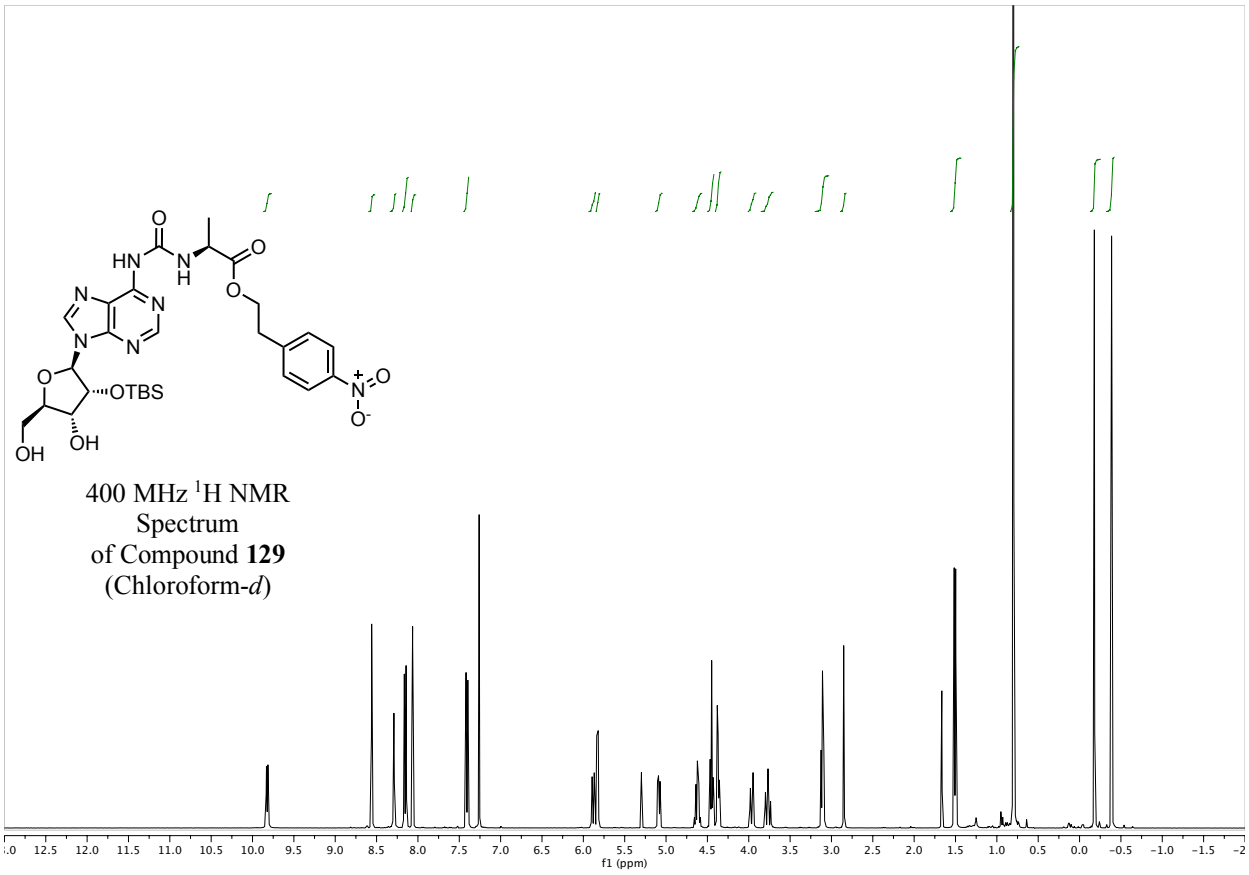
101 MHz ^{13}C NMR
Spectrum
of Compound **132**
(Chloroform-*d*)



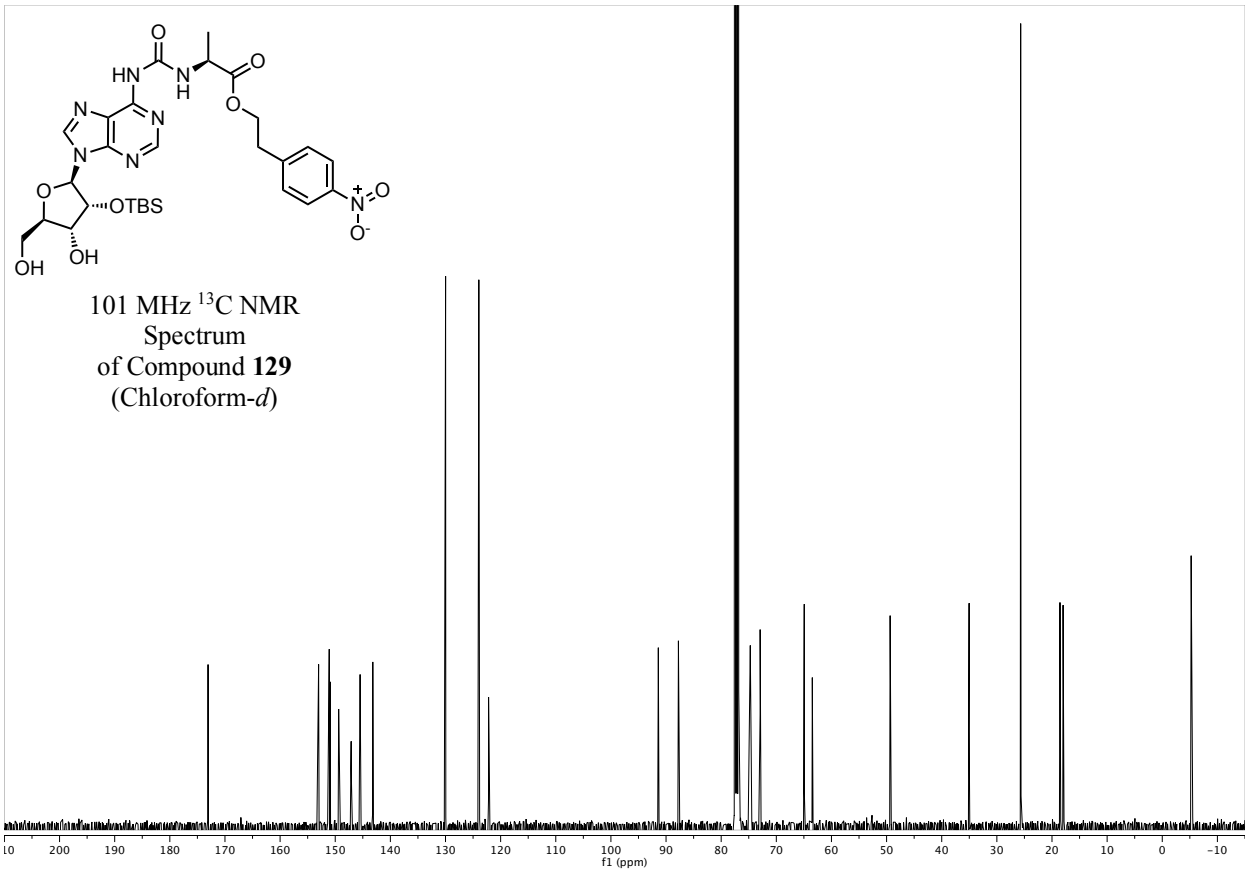
400 MHz ^{1}H NMR
Spectrum
of Compound **136**
(Benzene-d_6)



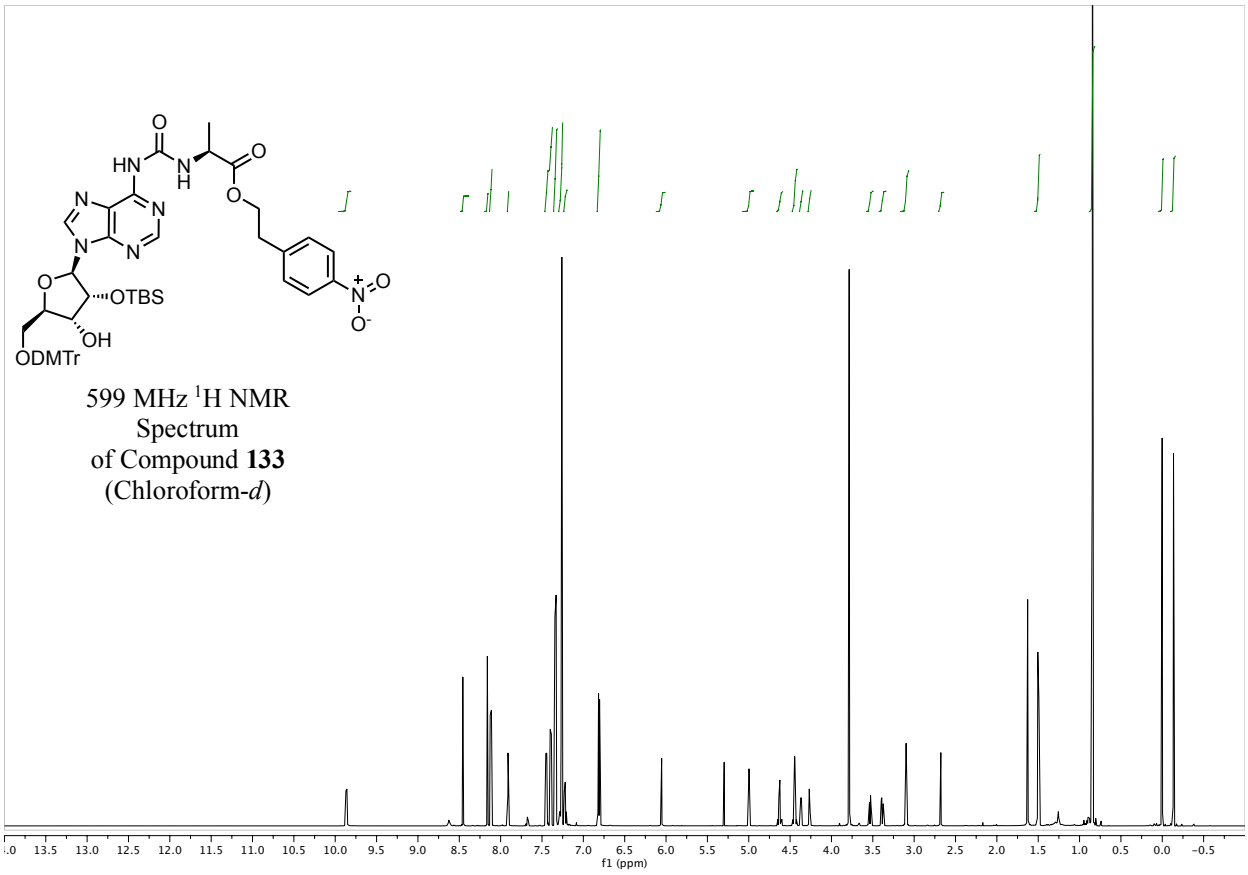
162 MHz ^{31}P NMR
Spectrum
of Compound **136**
(Benzene-d_6)



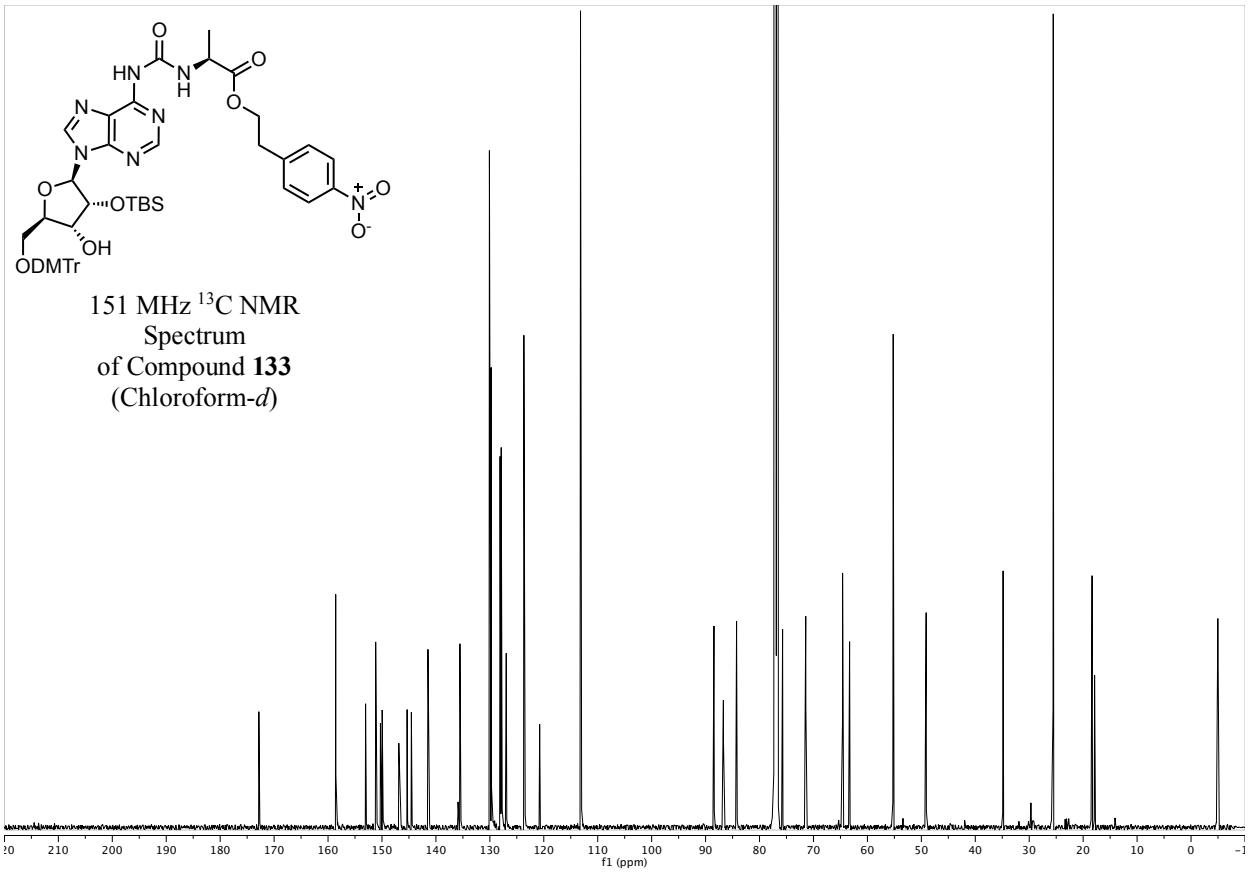
400 MHz ^{1}H NMR Spectrum of Compound **129** (Chloroform-*d*)



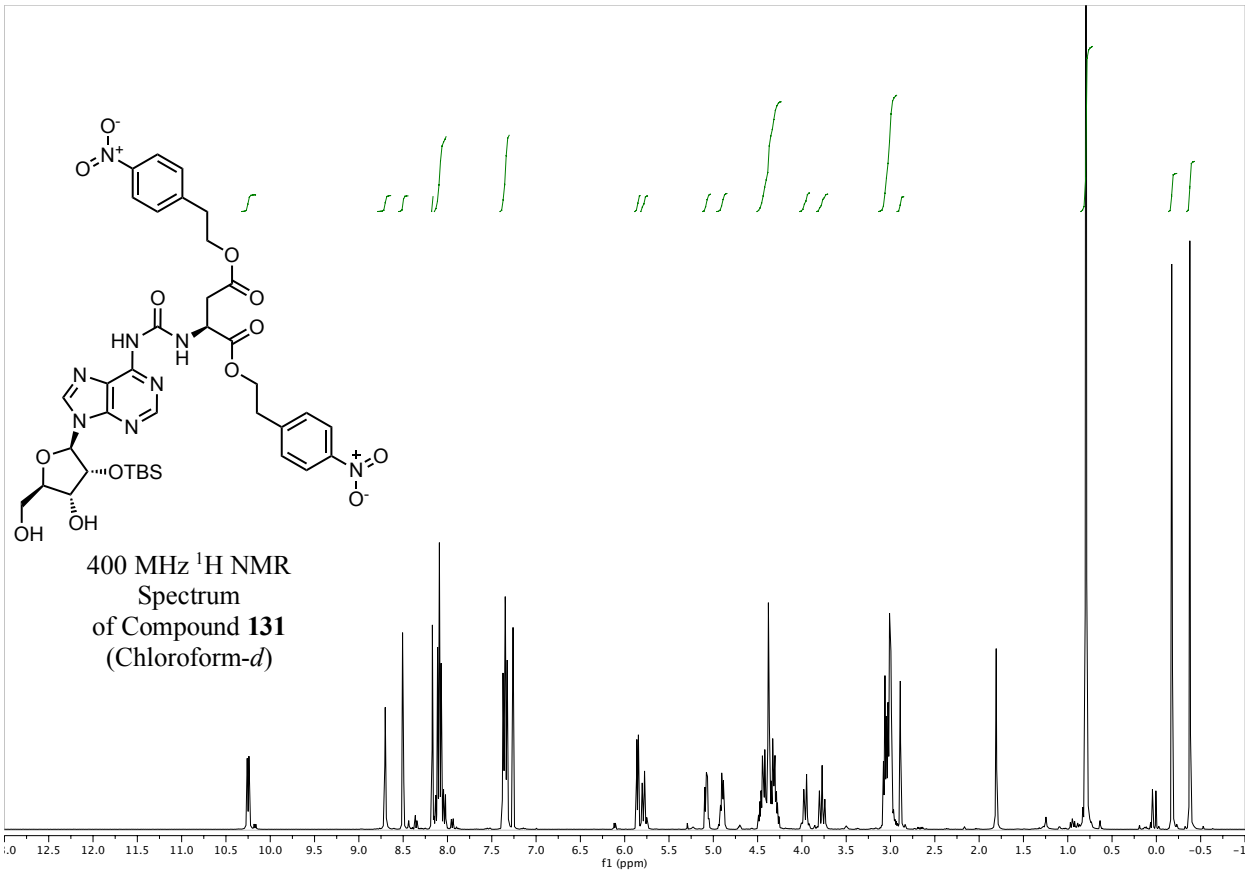
101 MHz ^{13}C NMR Spectrum of Compound **129** (Chloroform-*d*)



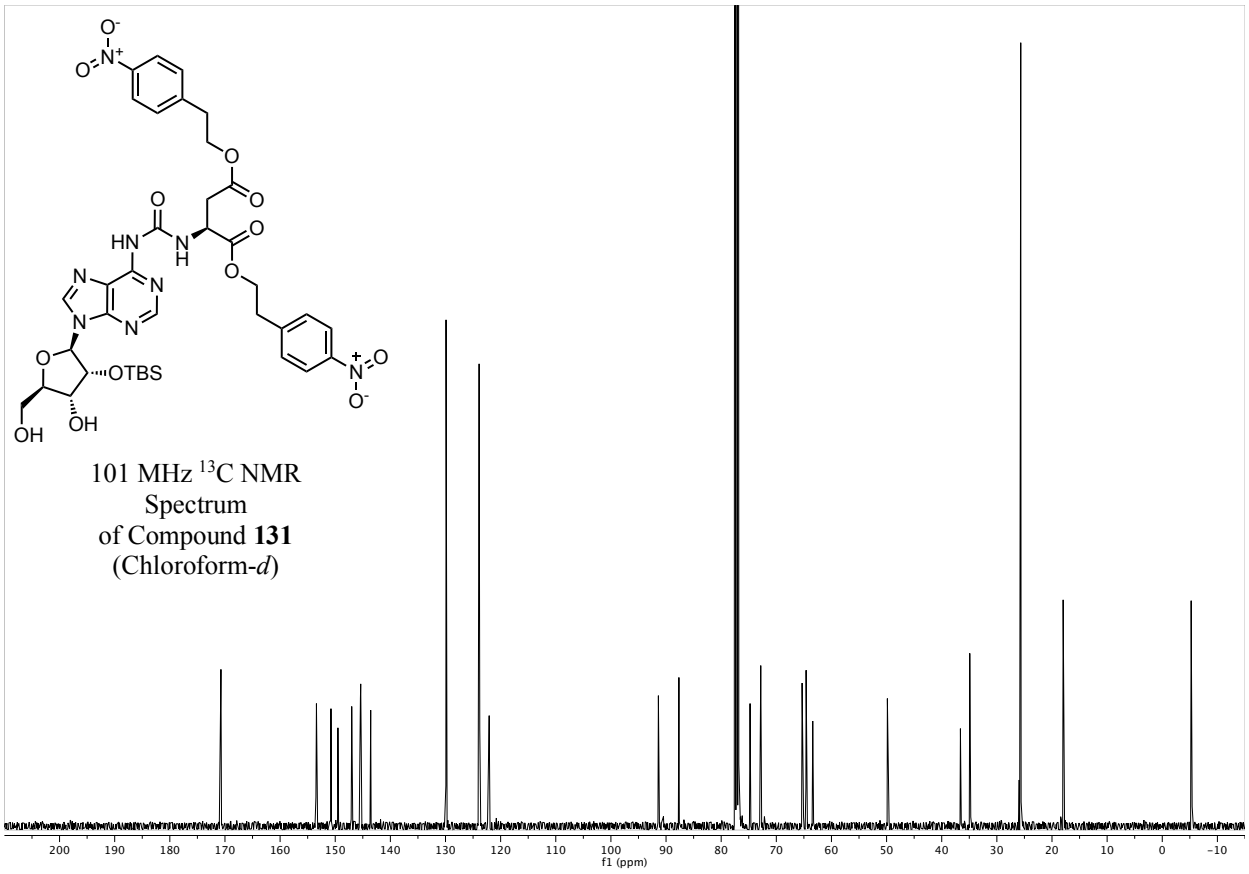
599 MHz ^{1}H NMR
Spectrum
of Compound **133**
(Chloroform-*d*)



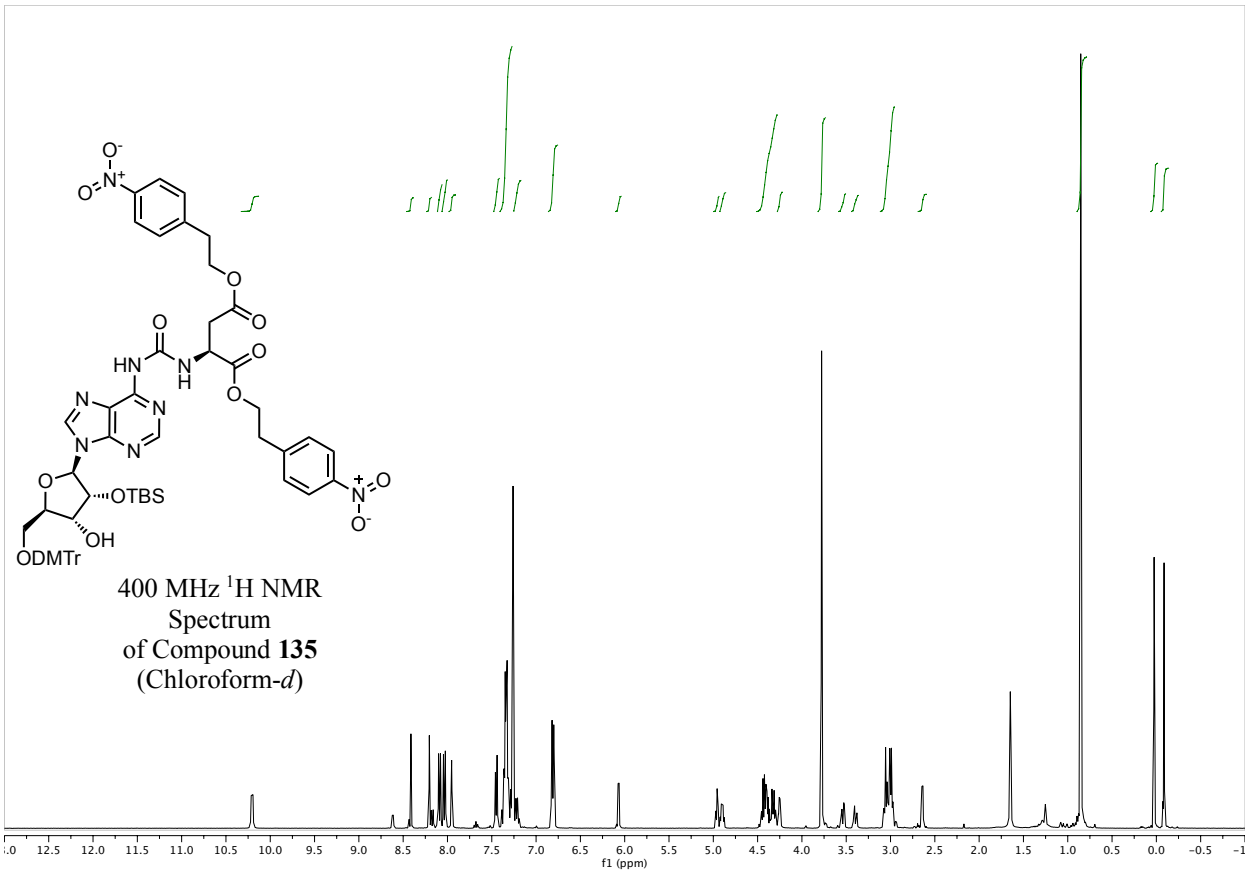
151 MHz ^{13}C NMR
Spectrum
of Compound **133**
(Chloroform-*d*)



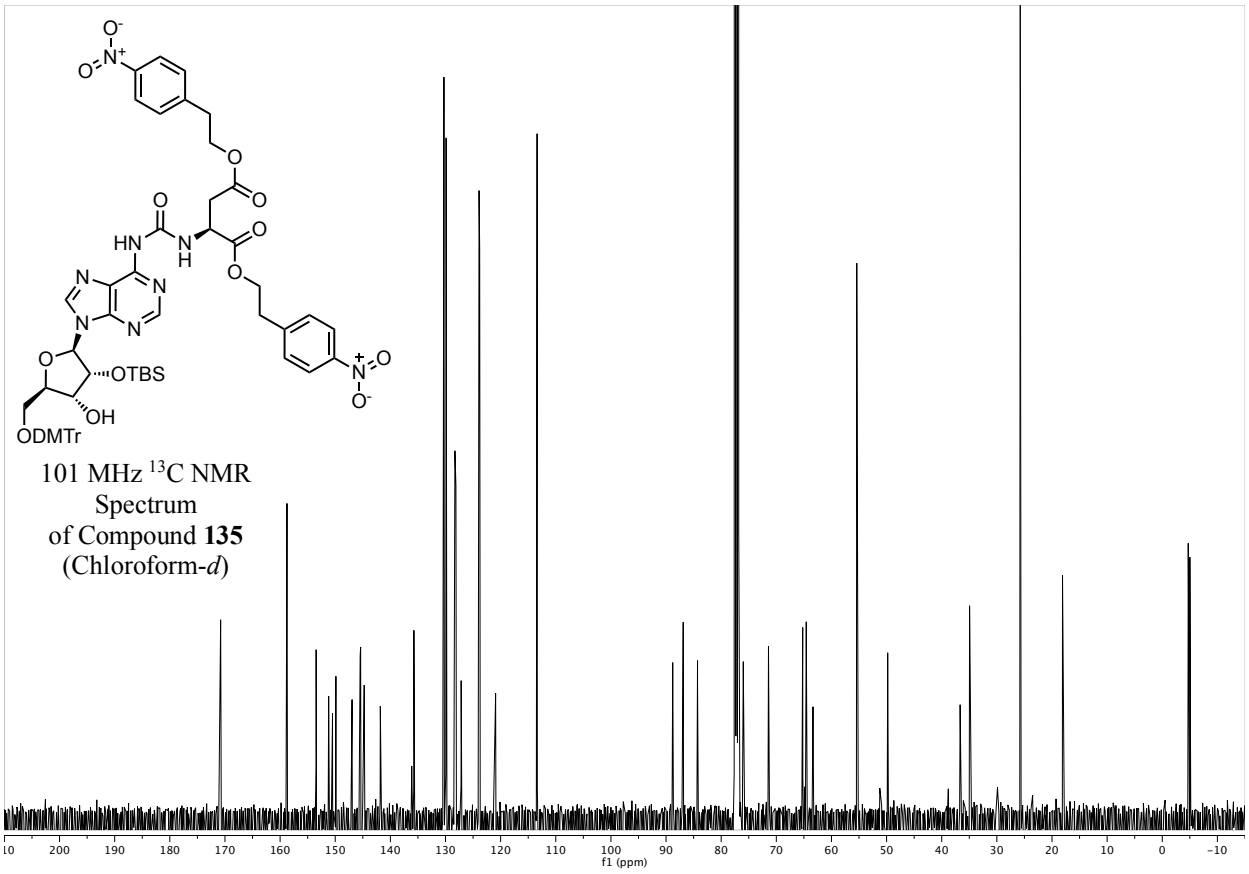
400 MHz ^{1}H NMR Spectrum of Compound **131** (Chloroform-*d*)



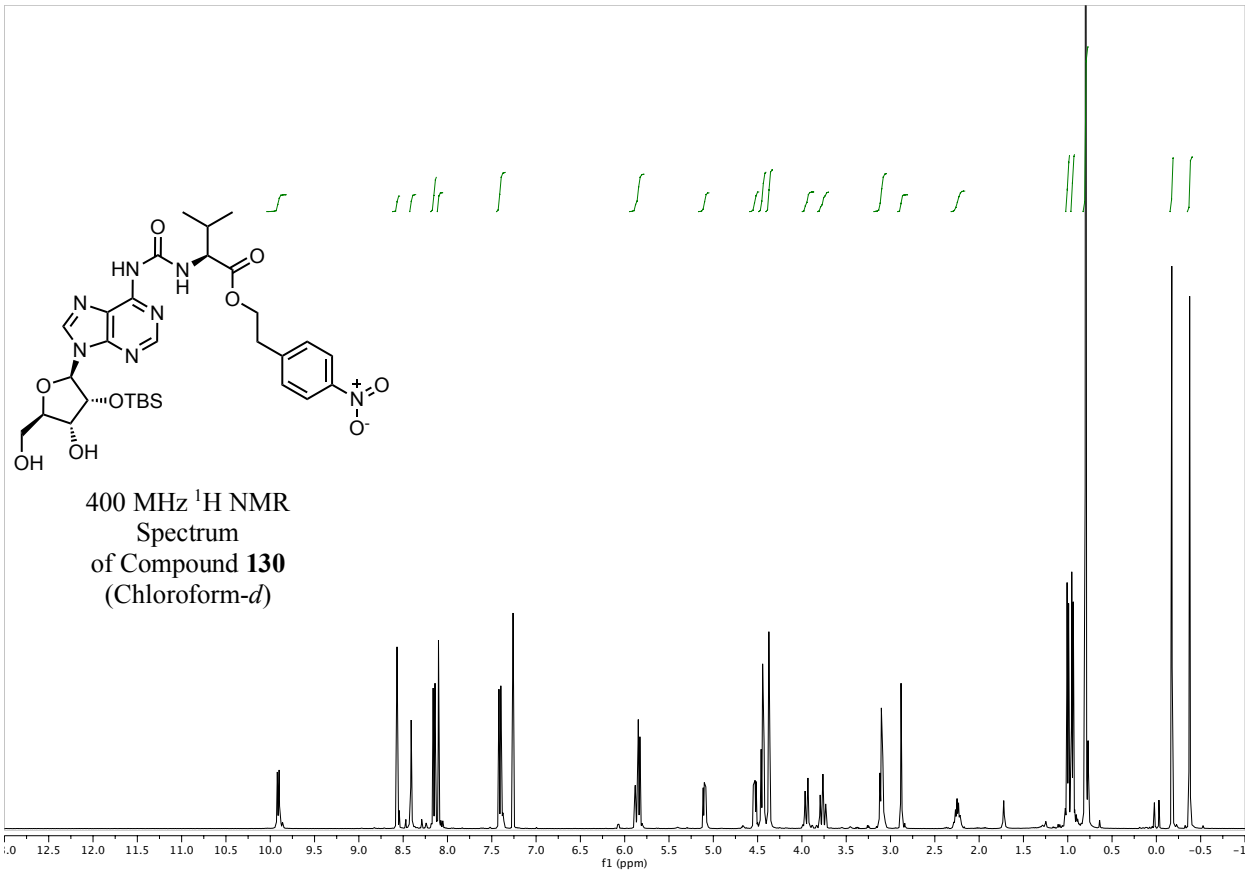
101 MHz ^{13}C NMR Spectrum of Compound **131** (Chloroform-*d*)



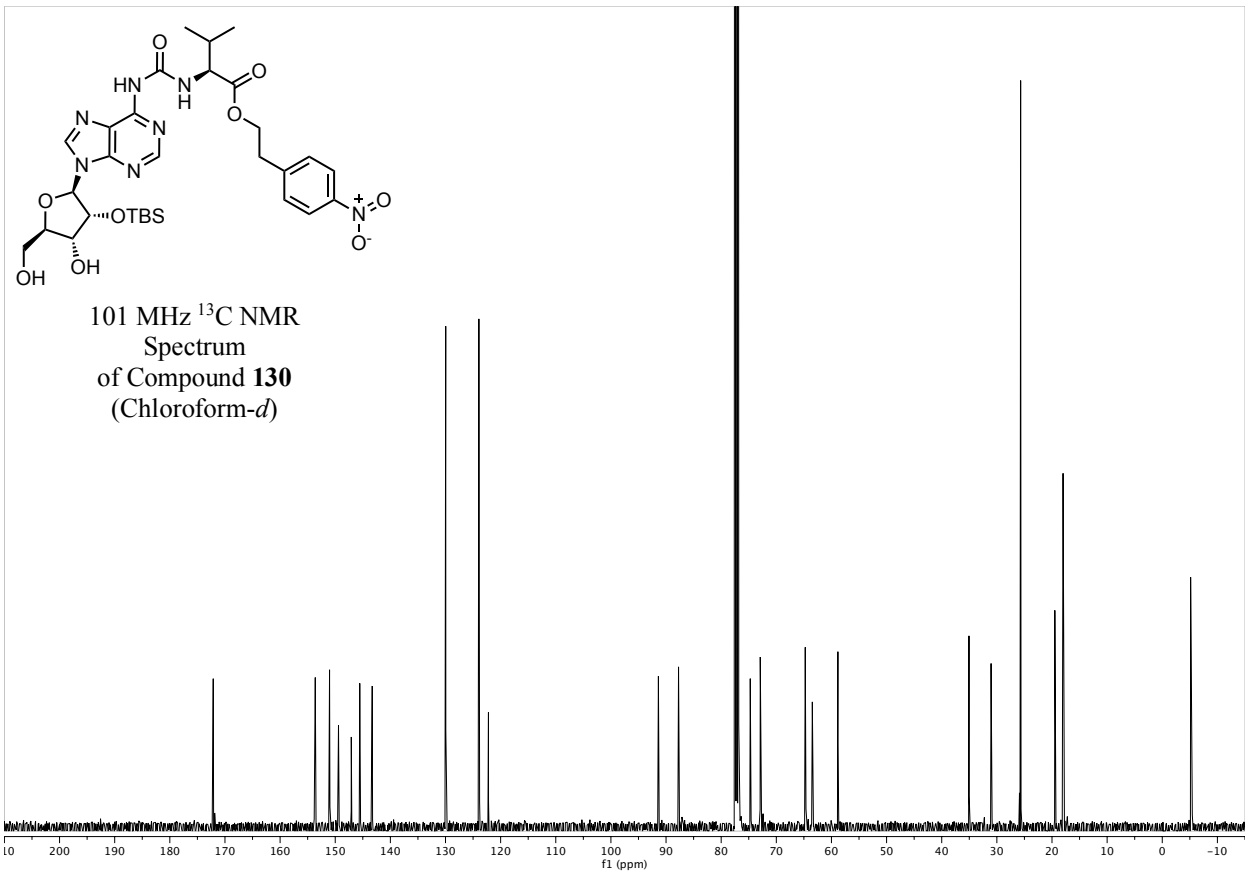
400 MHz ^{1}H NMR Spectrum of Compound **135** (Chloroform-*d*)



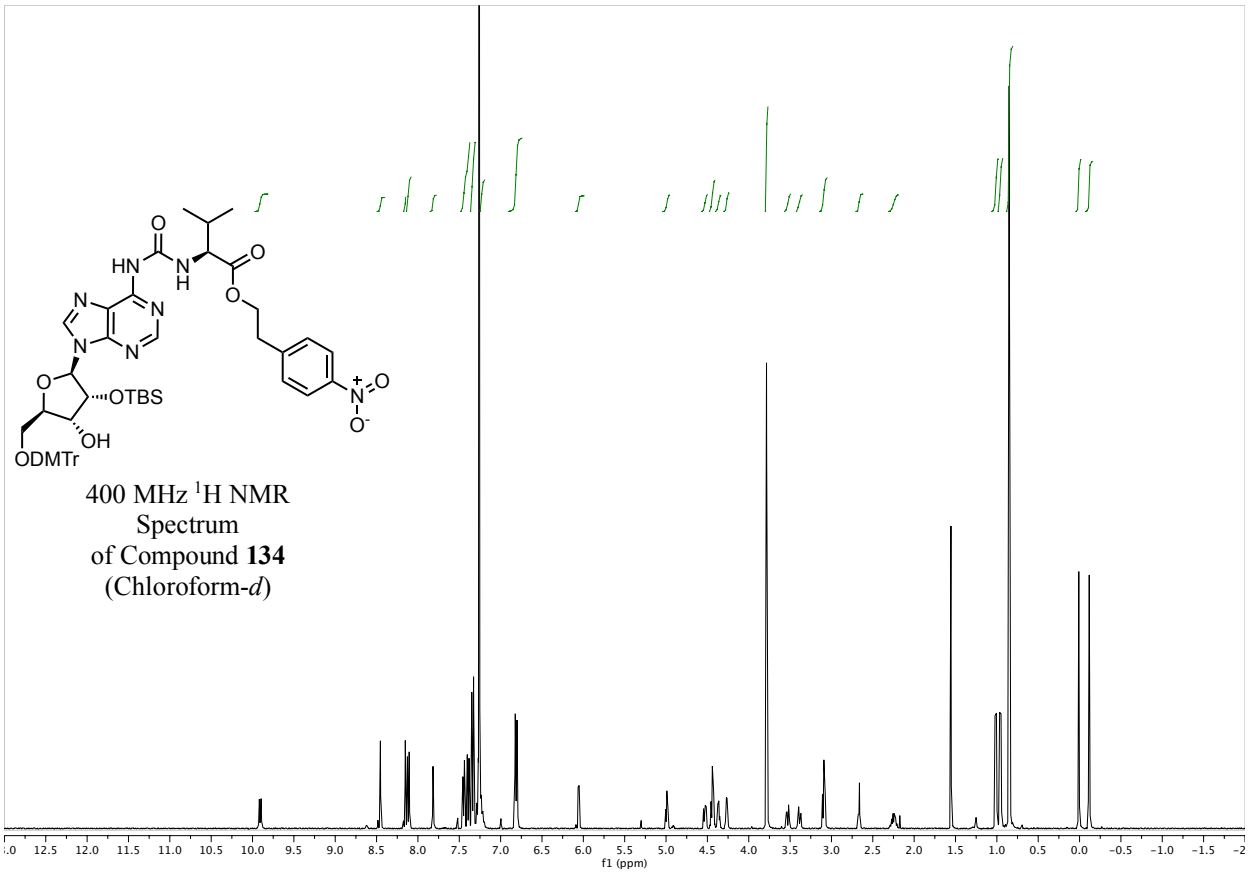
101 MHz ^{13}C NMR Spectrum of Compound **135** (Chloroform-*d*)



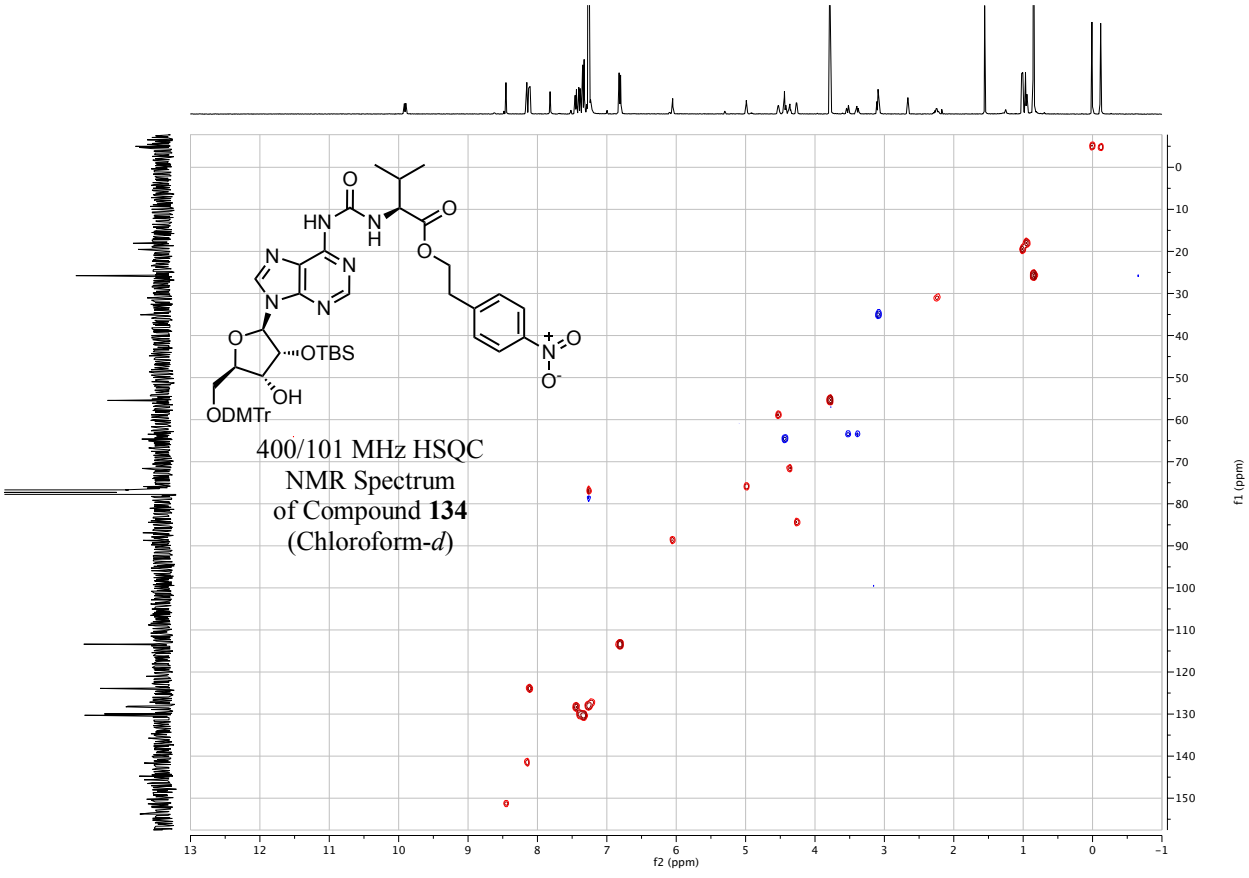
400 MHz ^{1}H NMR
Spectrum
of Compound **130**
(Chloroform-*d*)



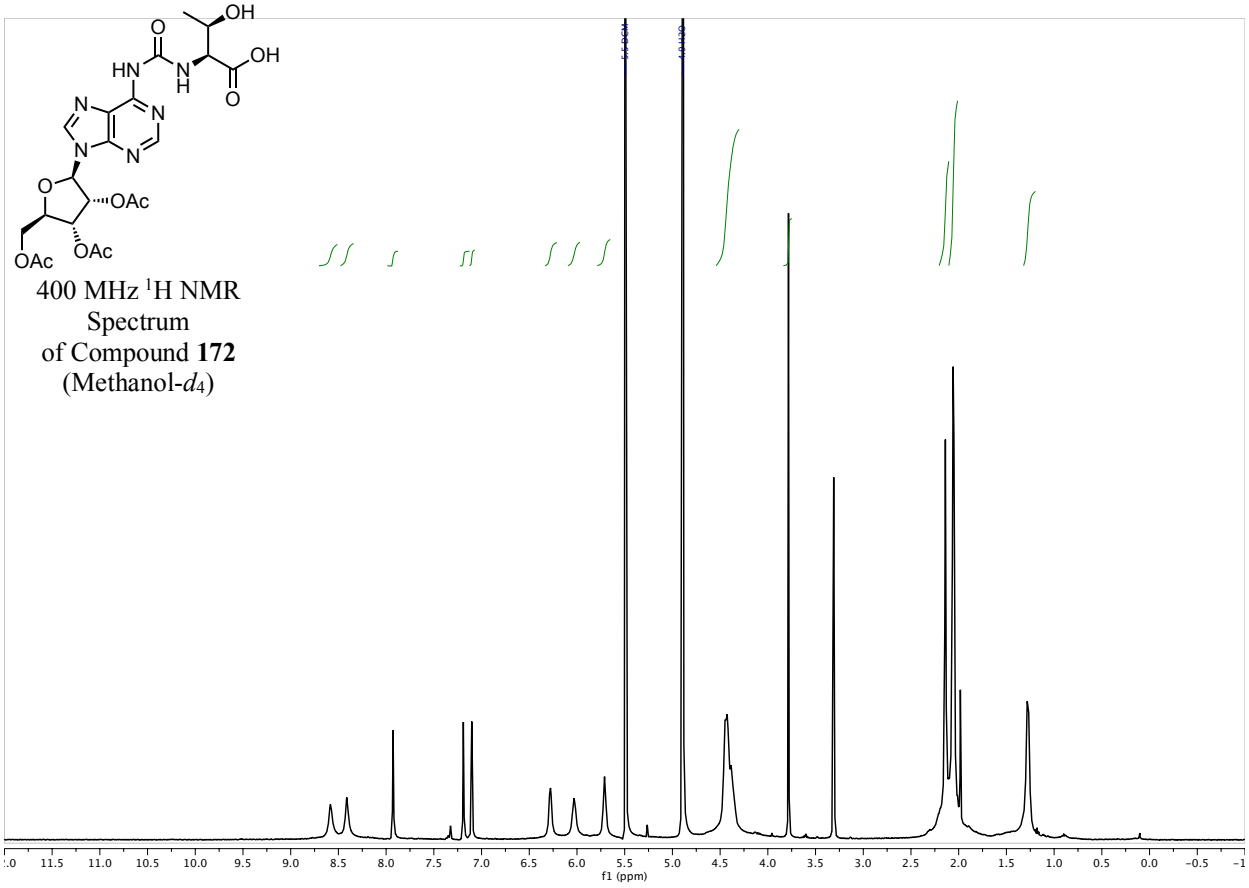
101 MHz ^{13}C NMR
Spectrum
of Compound **130**
(Chloroform-*d*)



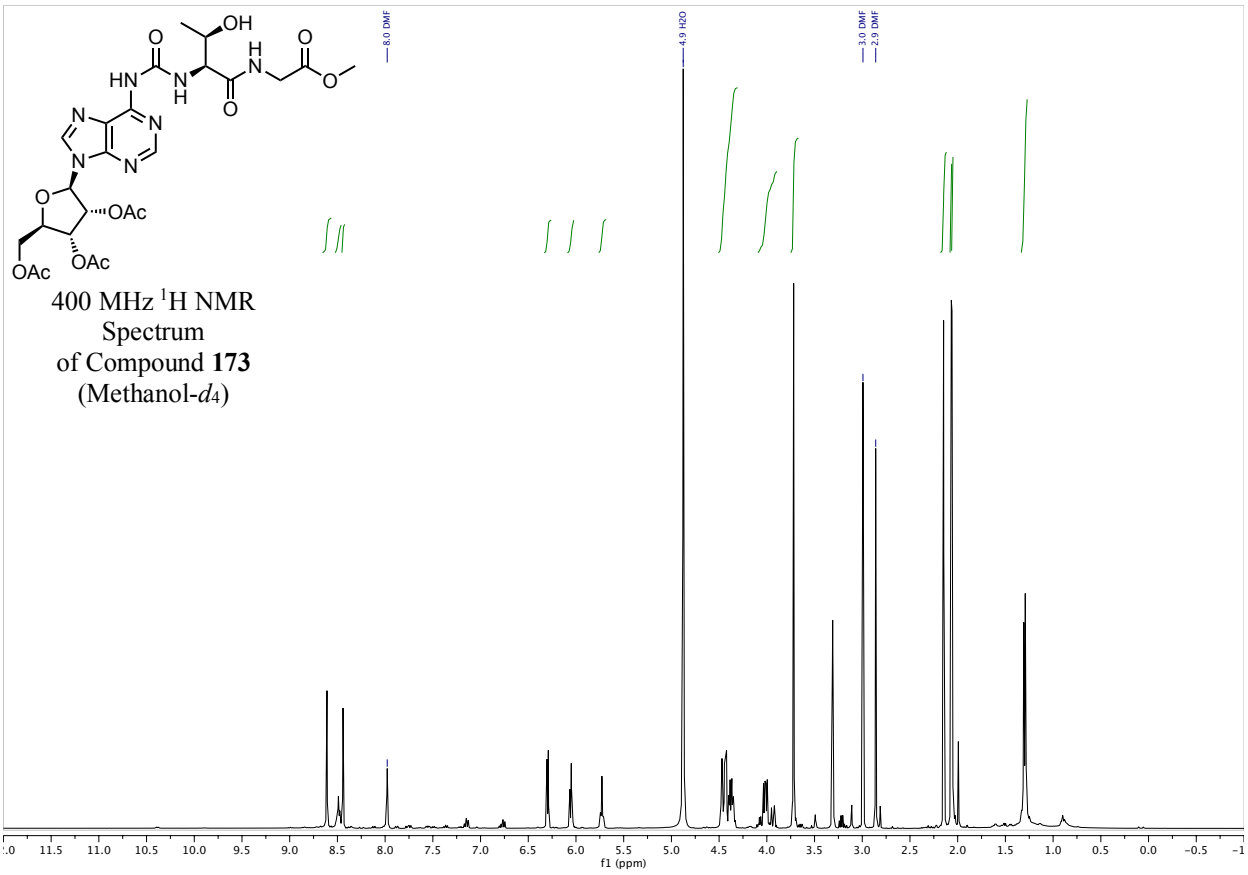
400 MHz ^{1}H NMR Spectrum of Compound **134** (Chloroform-*d*)



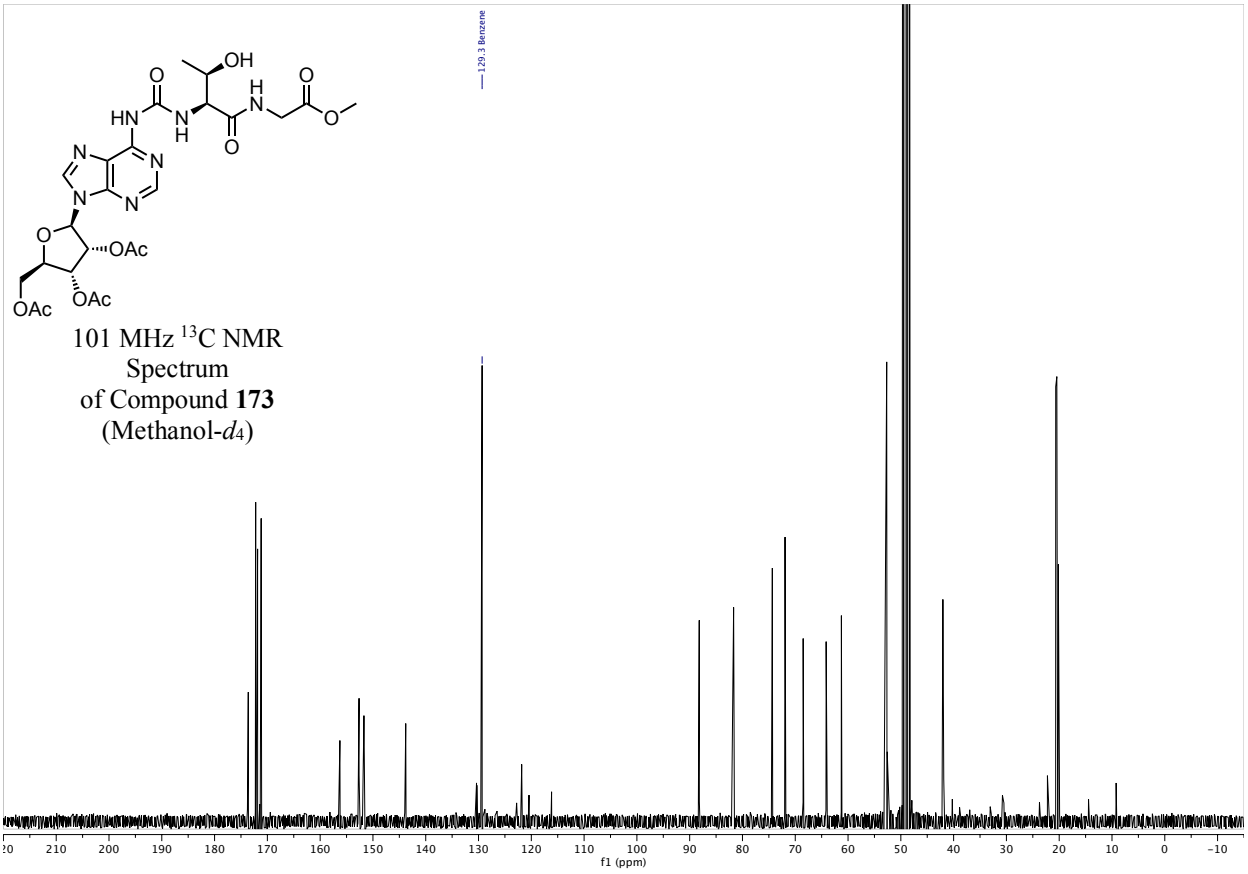
400/101 MHz HSQC NMR Spectrum of Compound **134** (Chloroform-*d*)



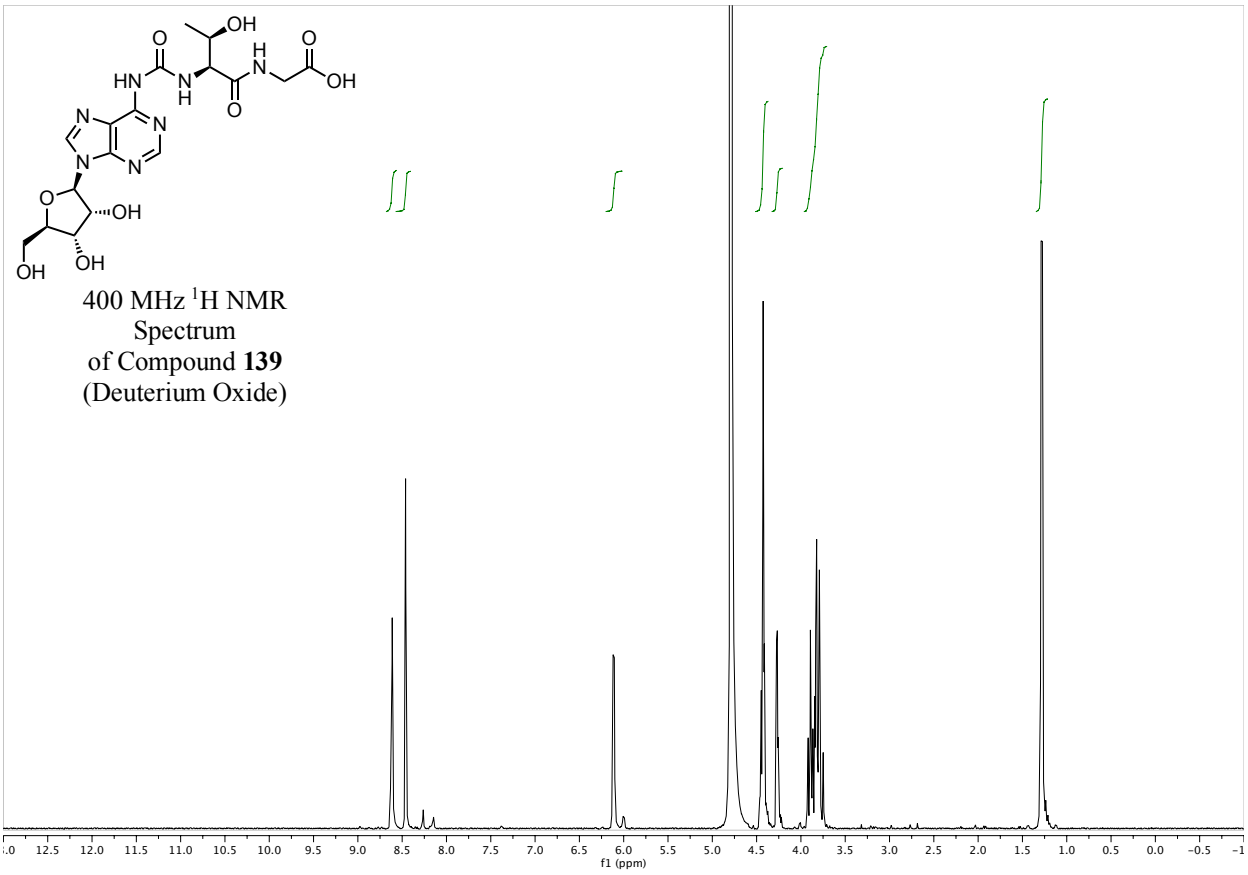
400 MHz ^{1}H NMR Spectrum of Compound **172** (Methanol-d_4)

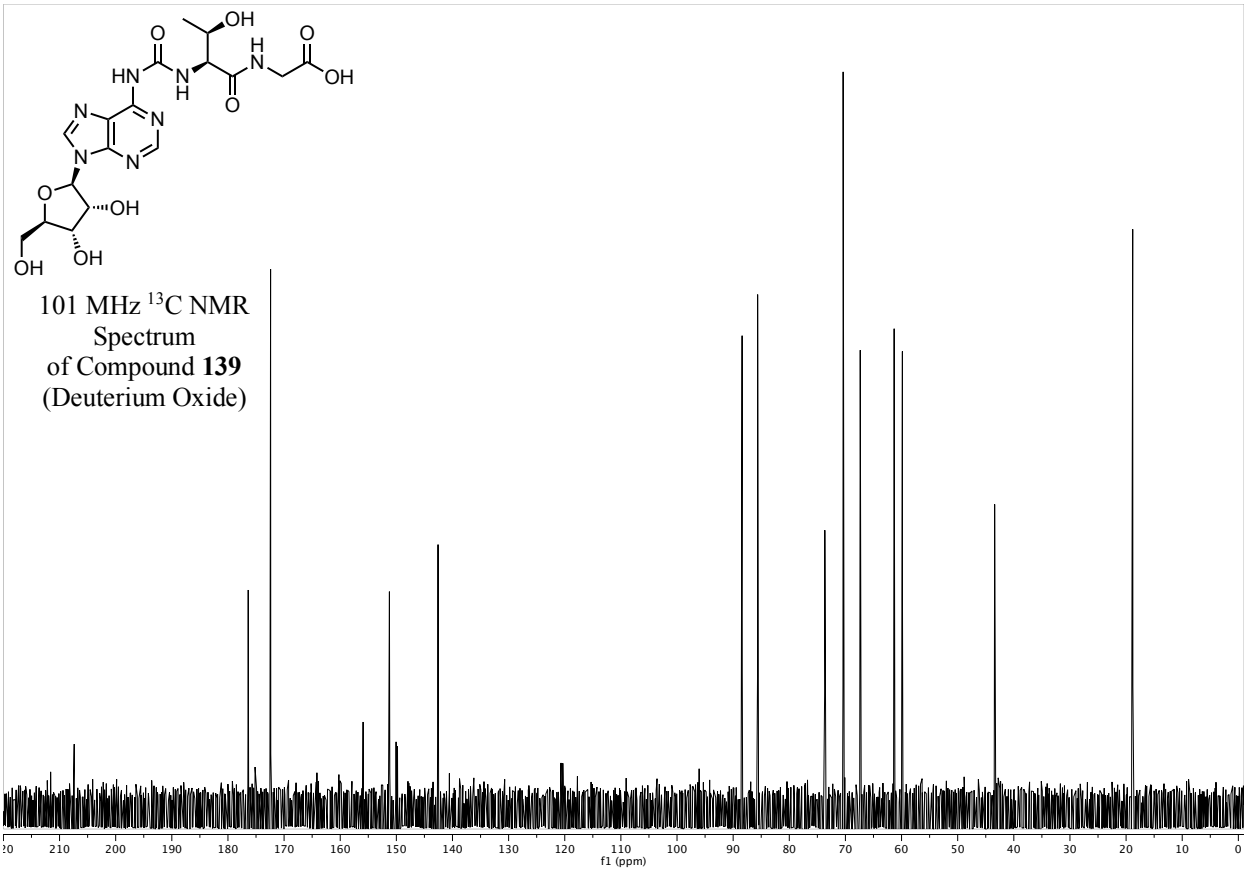


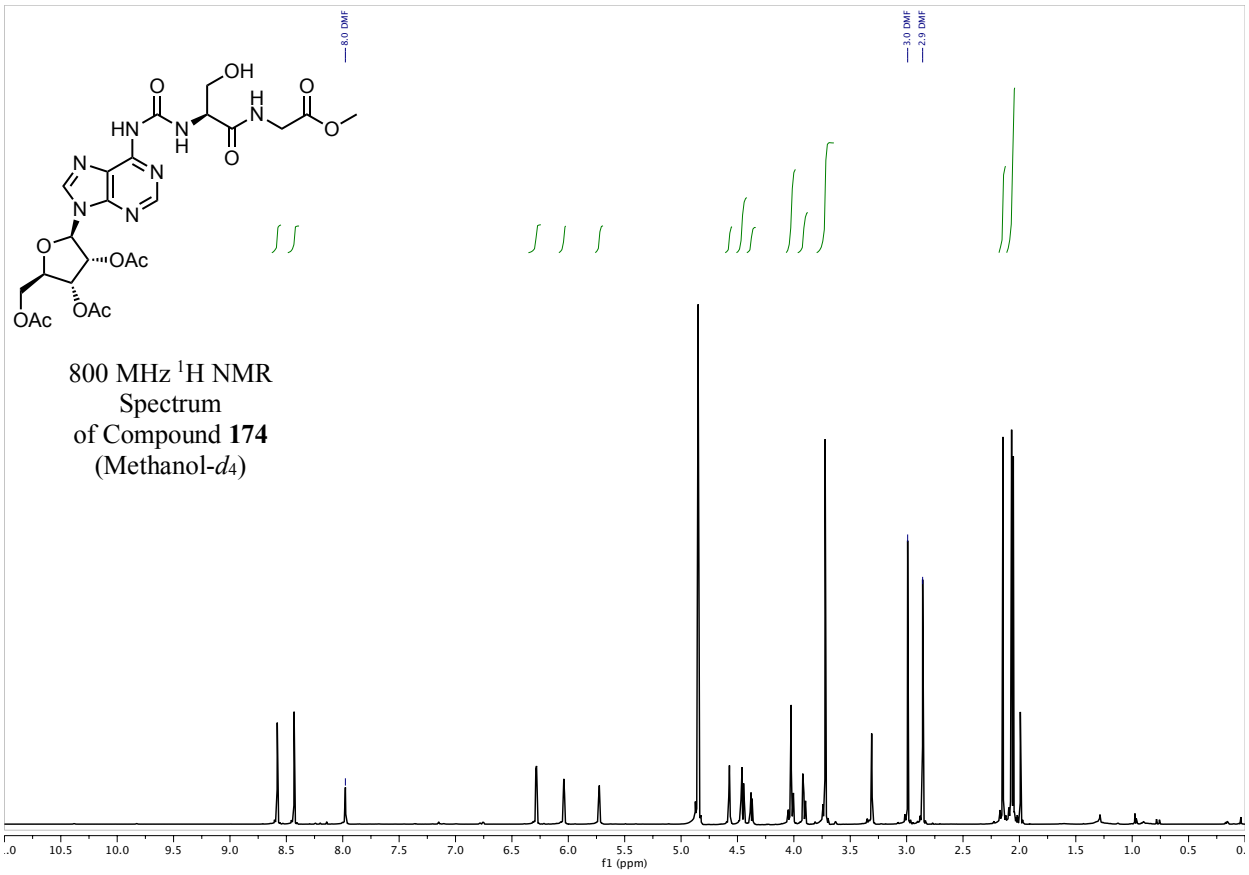
400 MHz ^{1}H NMR Spectrum of Compound **173** (Methanol-d_4)



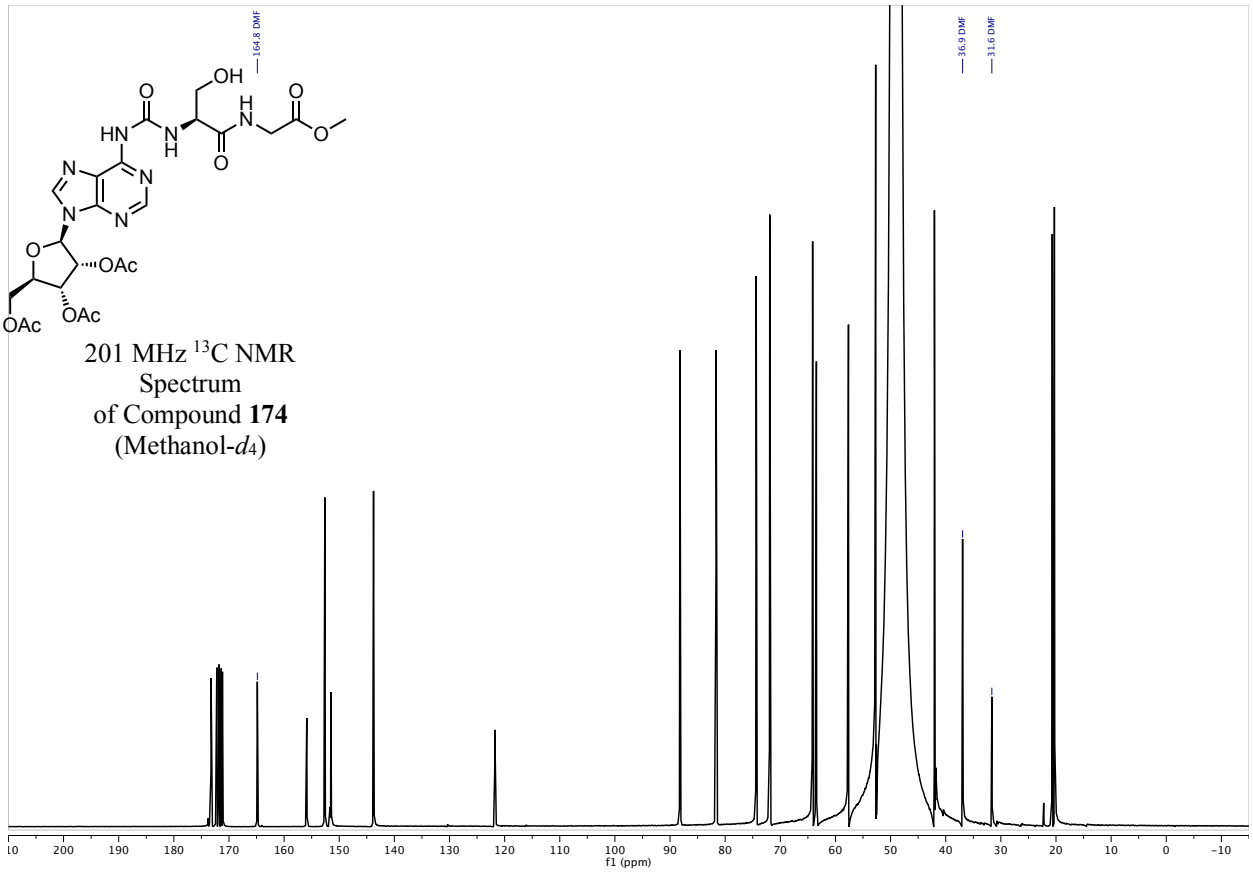
101 MHz ^{13}C NMR Spectrum of Compound **173** (Methanol-d_4)


400 MHz ^{1}H NMR
Spectrum
of Compound **139**
(Deuterium Oxide)
f1 (ppm)

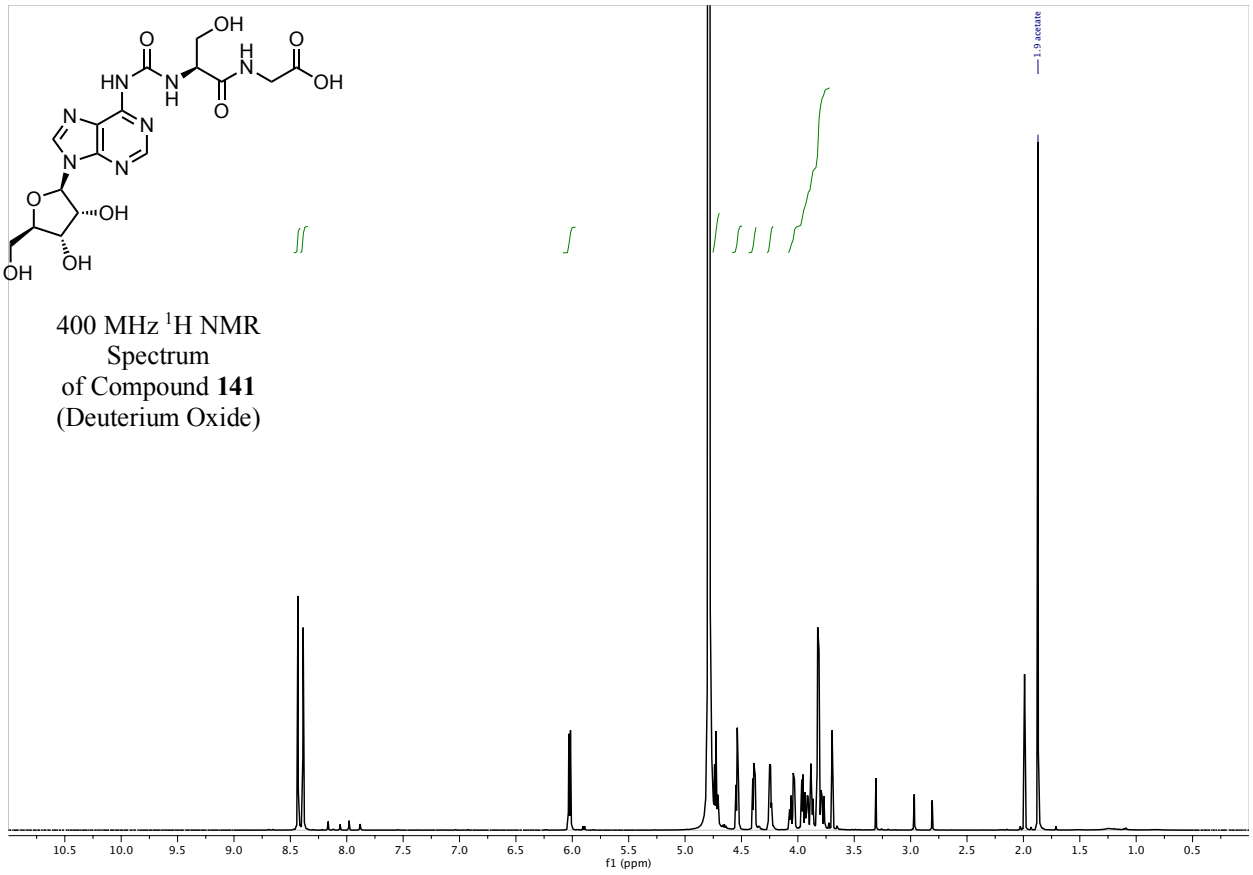

101 MHz ^{13}C NMR
Spectrum
of Compound **139**
(Deuterium Oxide)
f1 (ppm)



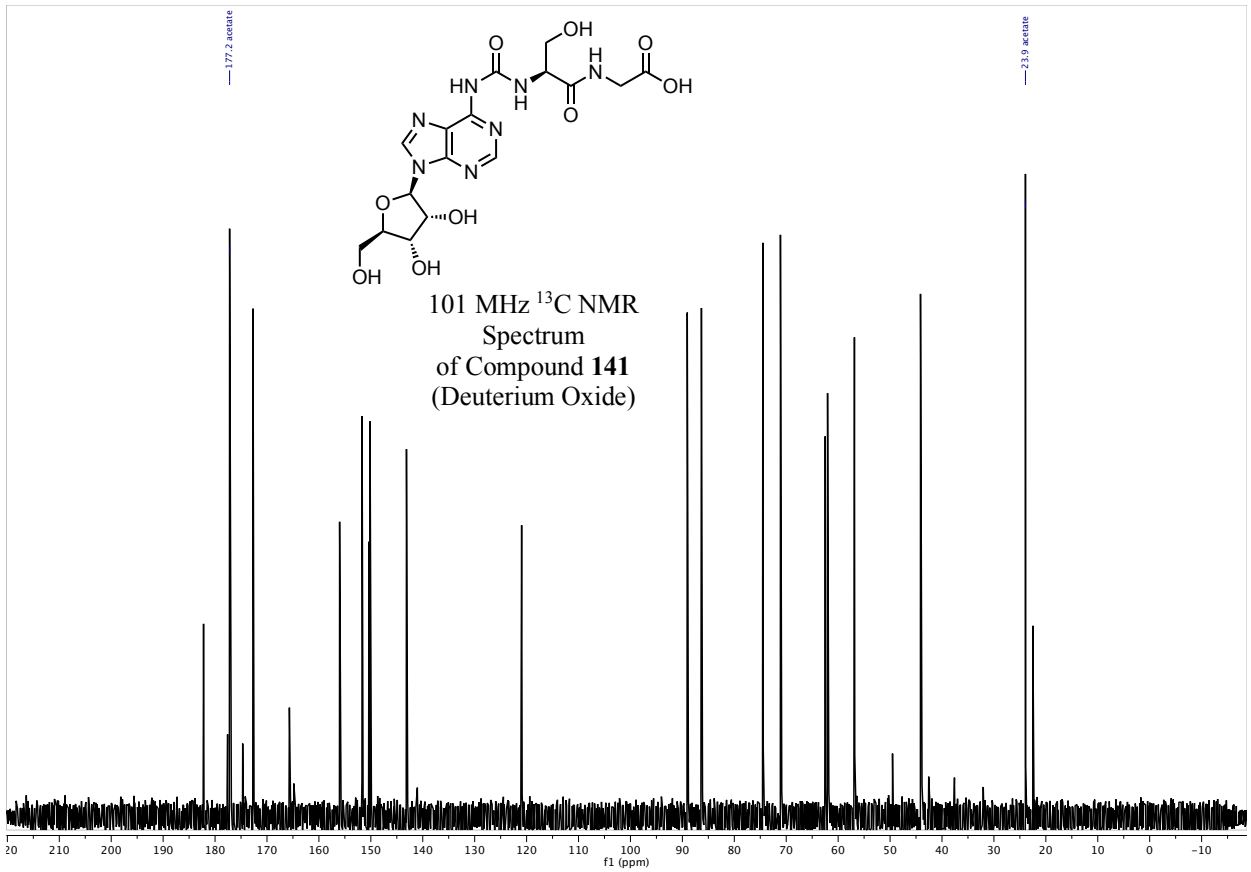
800 MHz ^{1}H NMR
Spectrum
of Compound **174**
(Methanol-d_4)



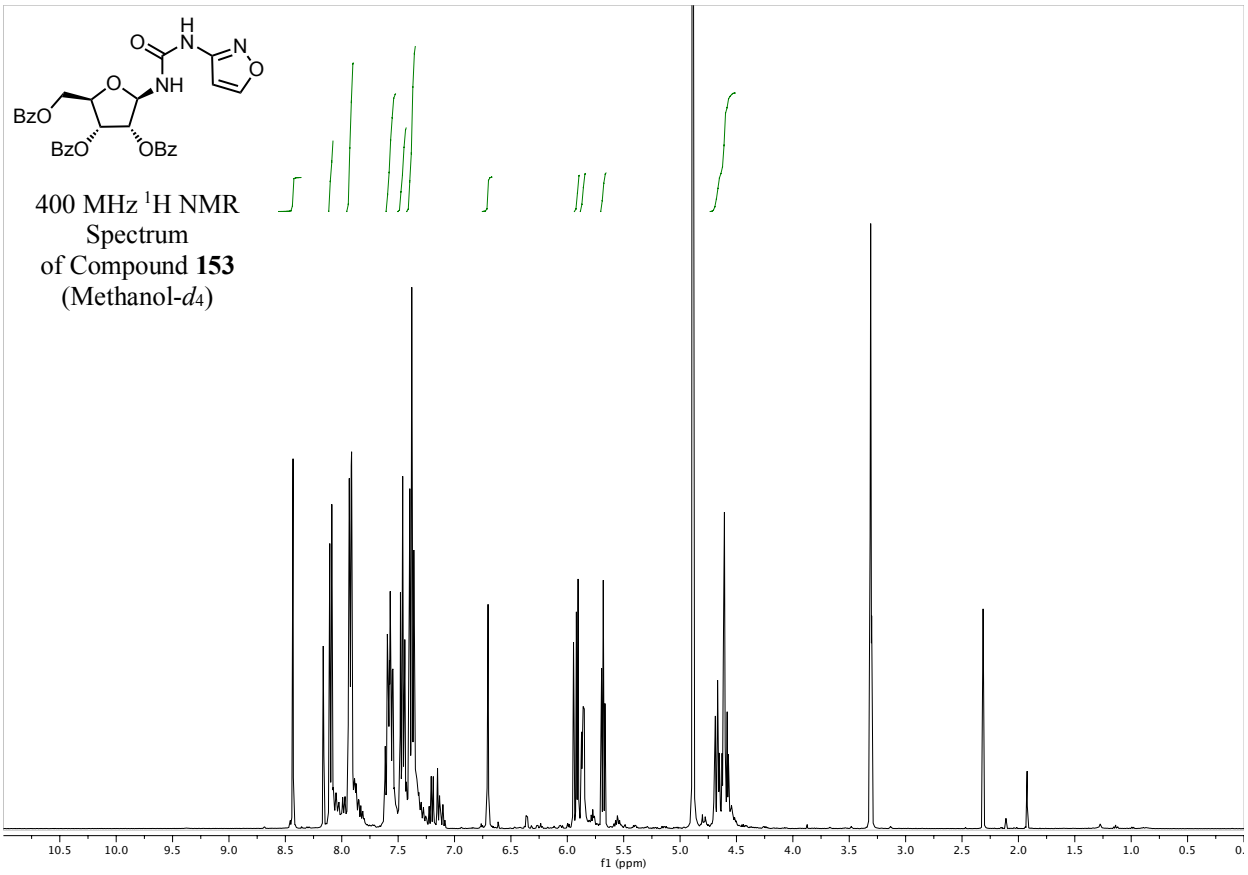
201 MHz ^{13}C NMR
Spectrum
of Compound **174**
(Methanol-d_4)



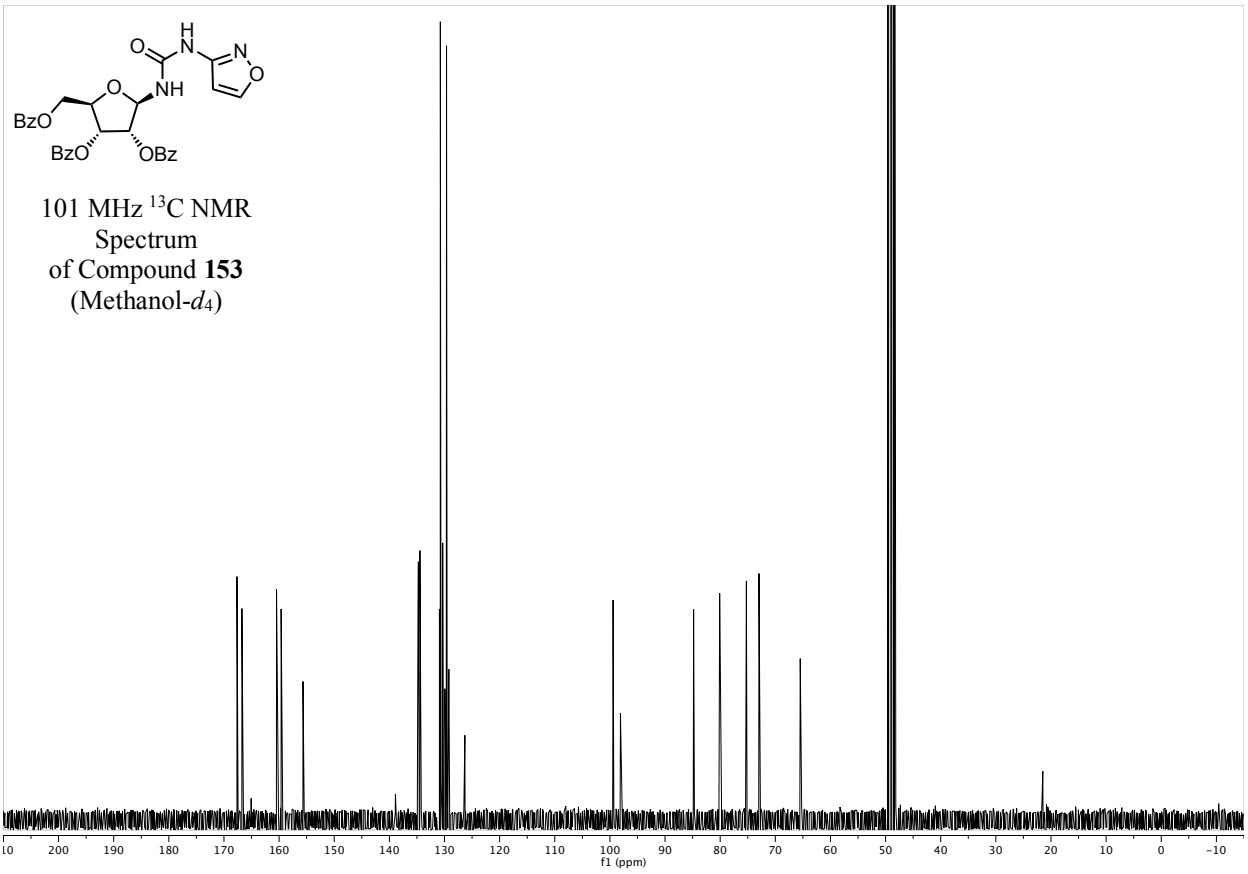
400 MHz ^{1}H NMR Spectrum of Compound **141** (Deuterium Oxide)



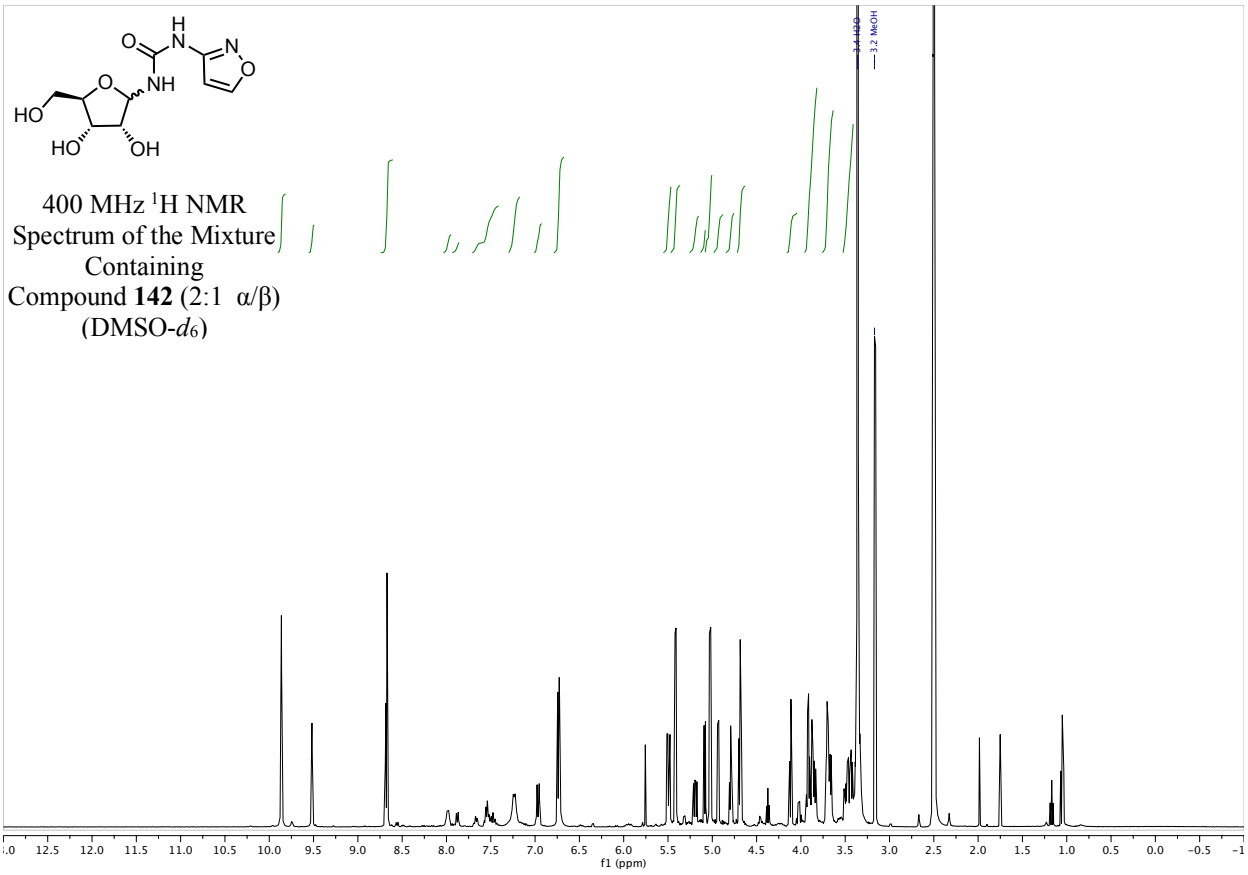
101 MHz ^{13}C NMR Spectrum of Compound **141** (Deuterium Oxide)



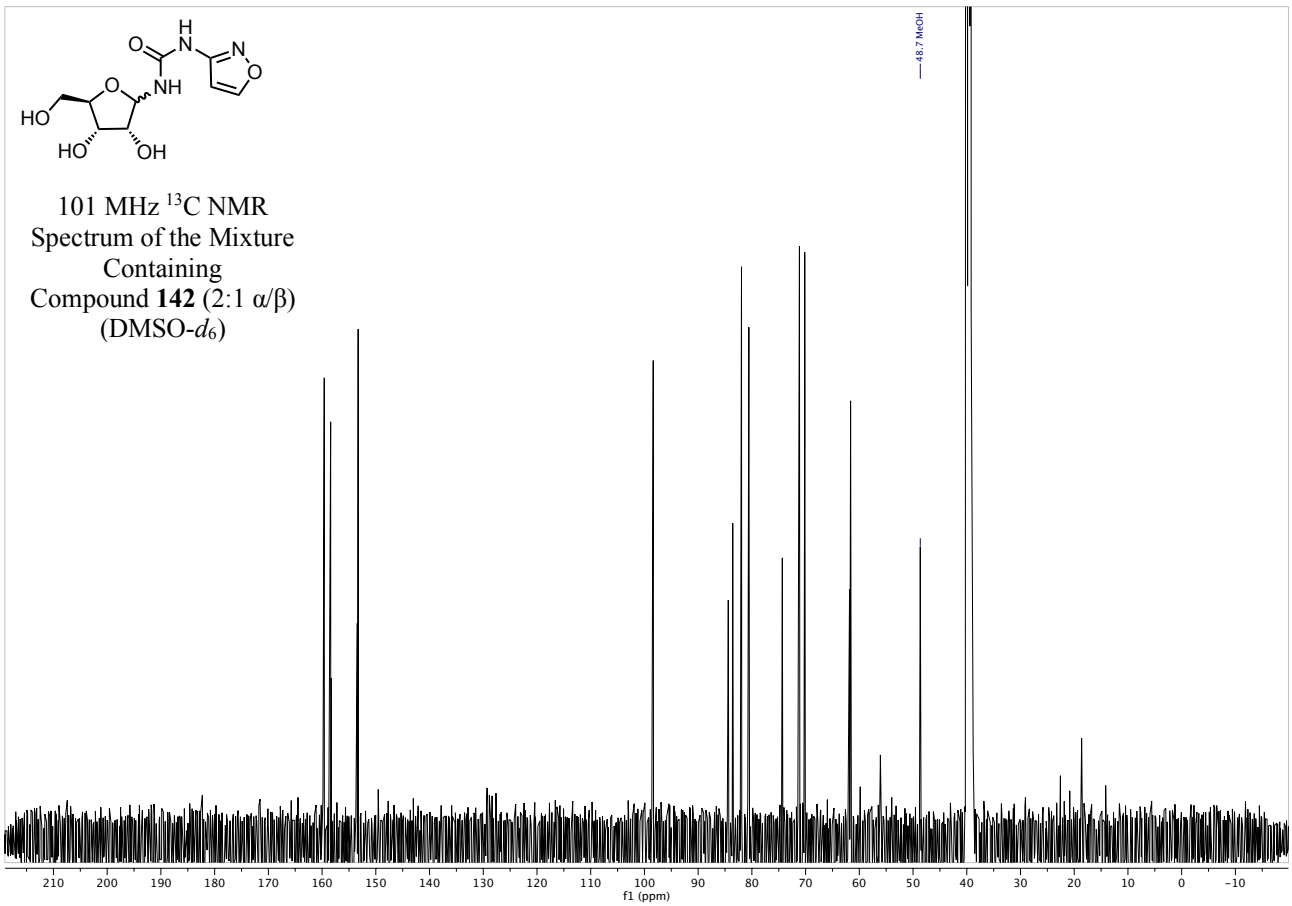
400 MHz ^{1}H NMR Spectrum of Compound **153** (Methanol-d_4)



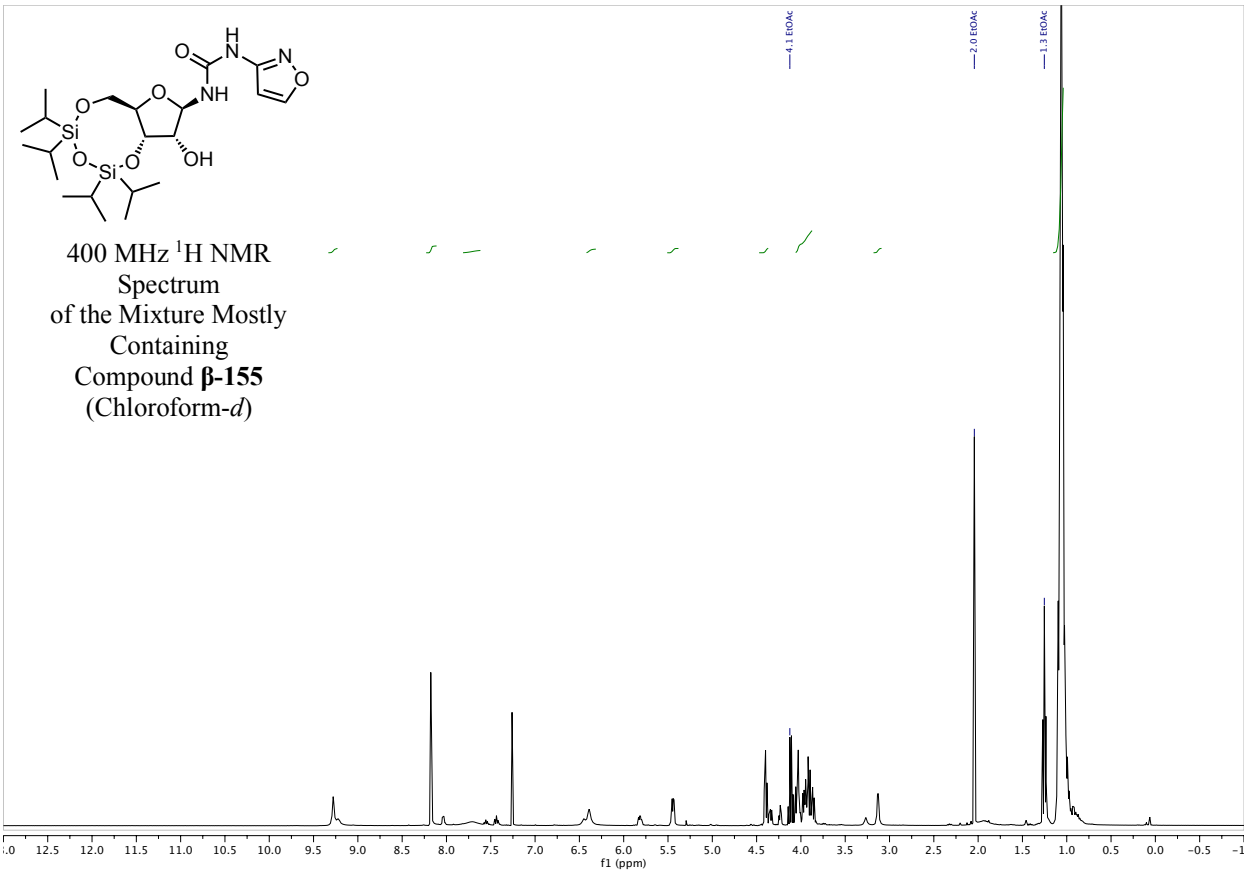
101 MHz ^{13}C NMR Spectrum of Compound **153** (Methanol-d_4)



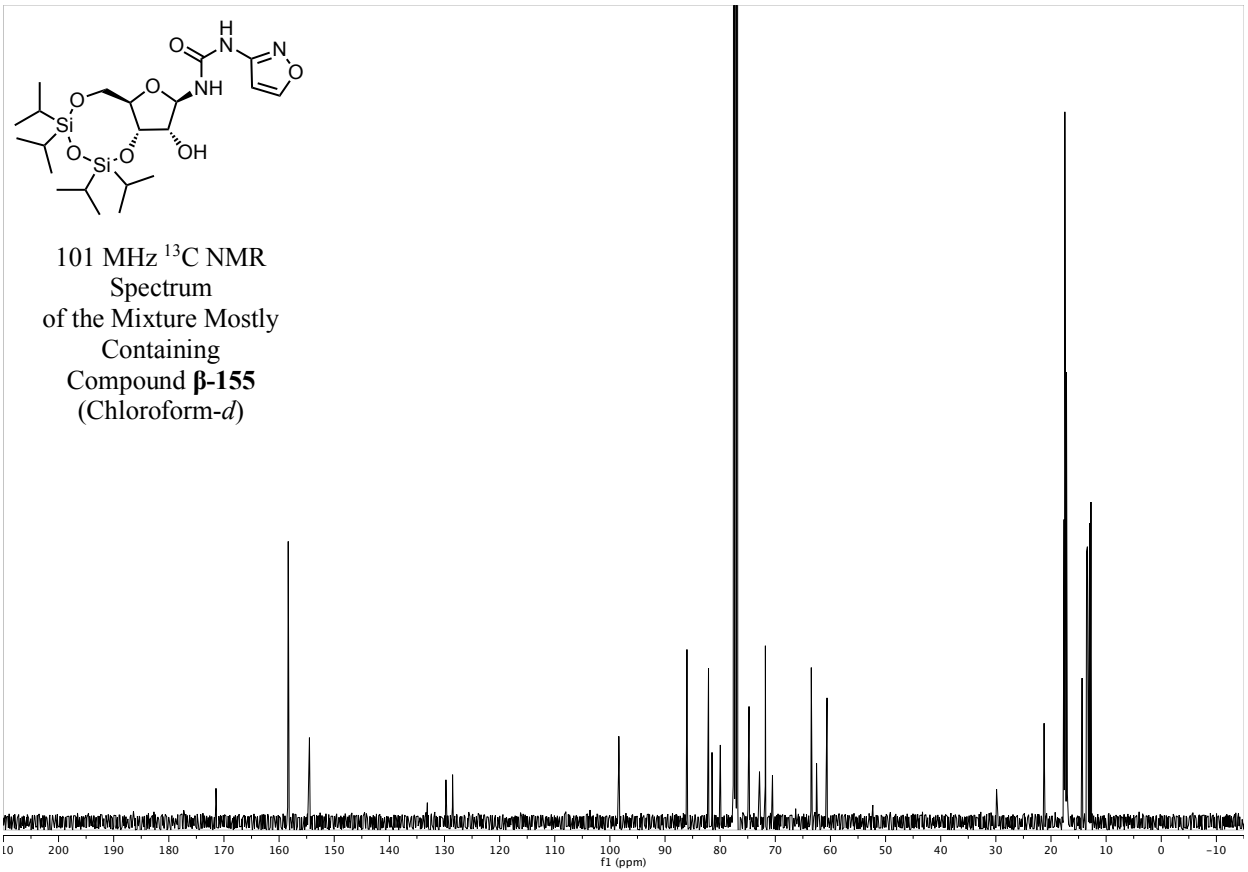
400 MHz ^{1}H NMR Spectrum of the Mixture Containing Compound **142** (2:1 α/β) (DMSO-d_6)



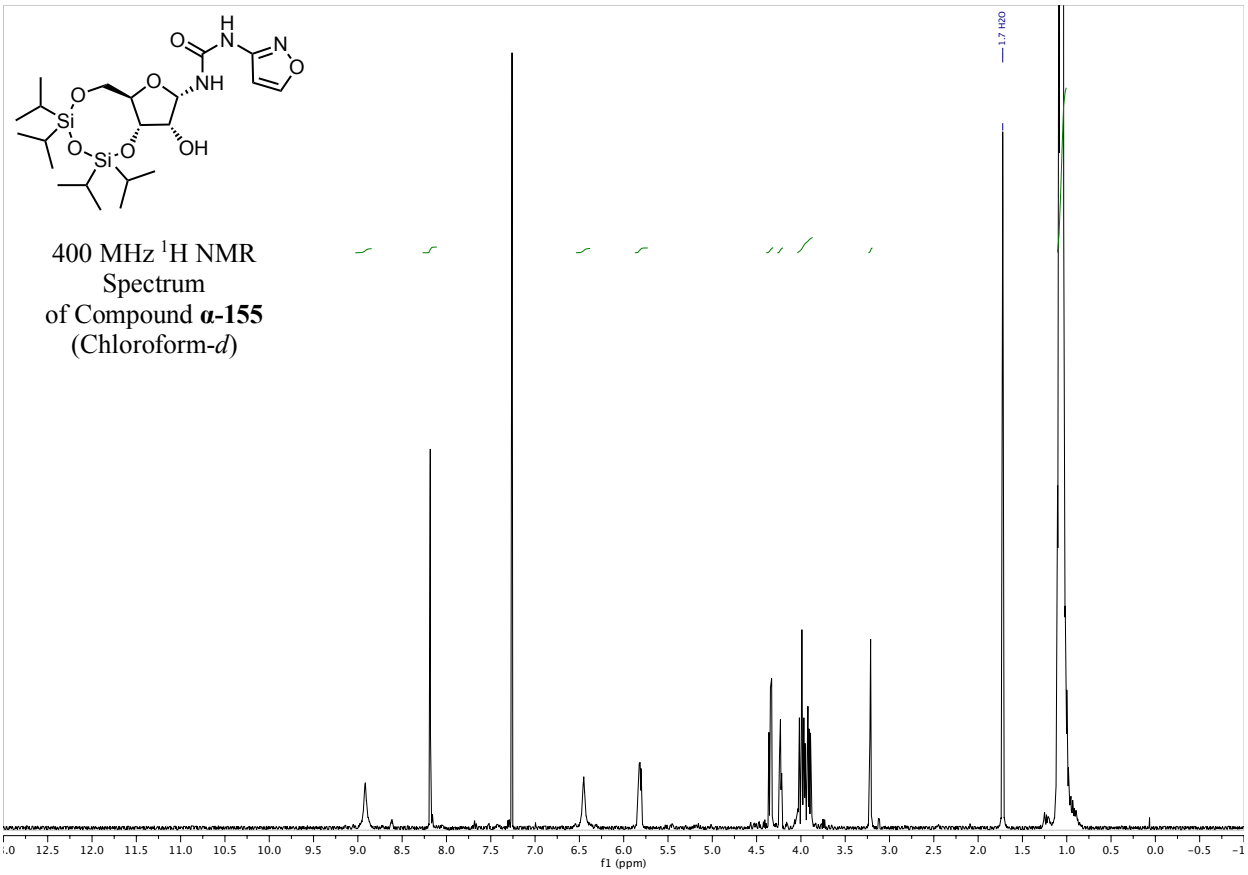
101 MHz ^{13}C NMR Spectrum of the Mixture Containing Compound **142** (2:1 α/β) (DMSO-d_6)



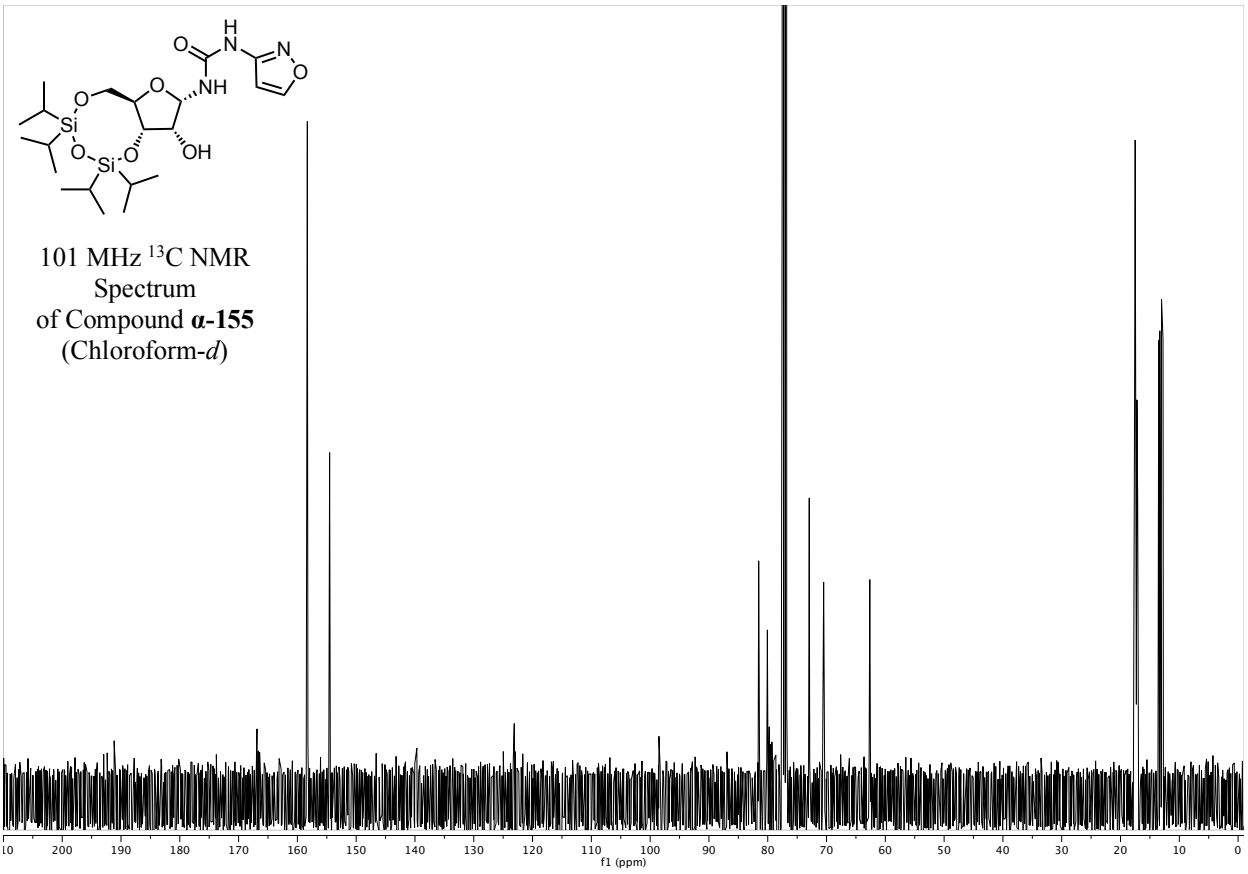
400 MHz ^{1}H NMR Spectrum of the Mixture Mostly Containing Compound **β-155** (Chloroform-*d*)



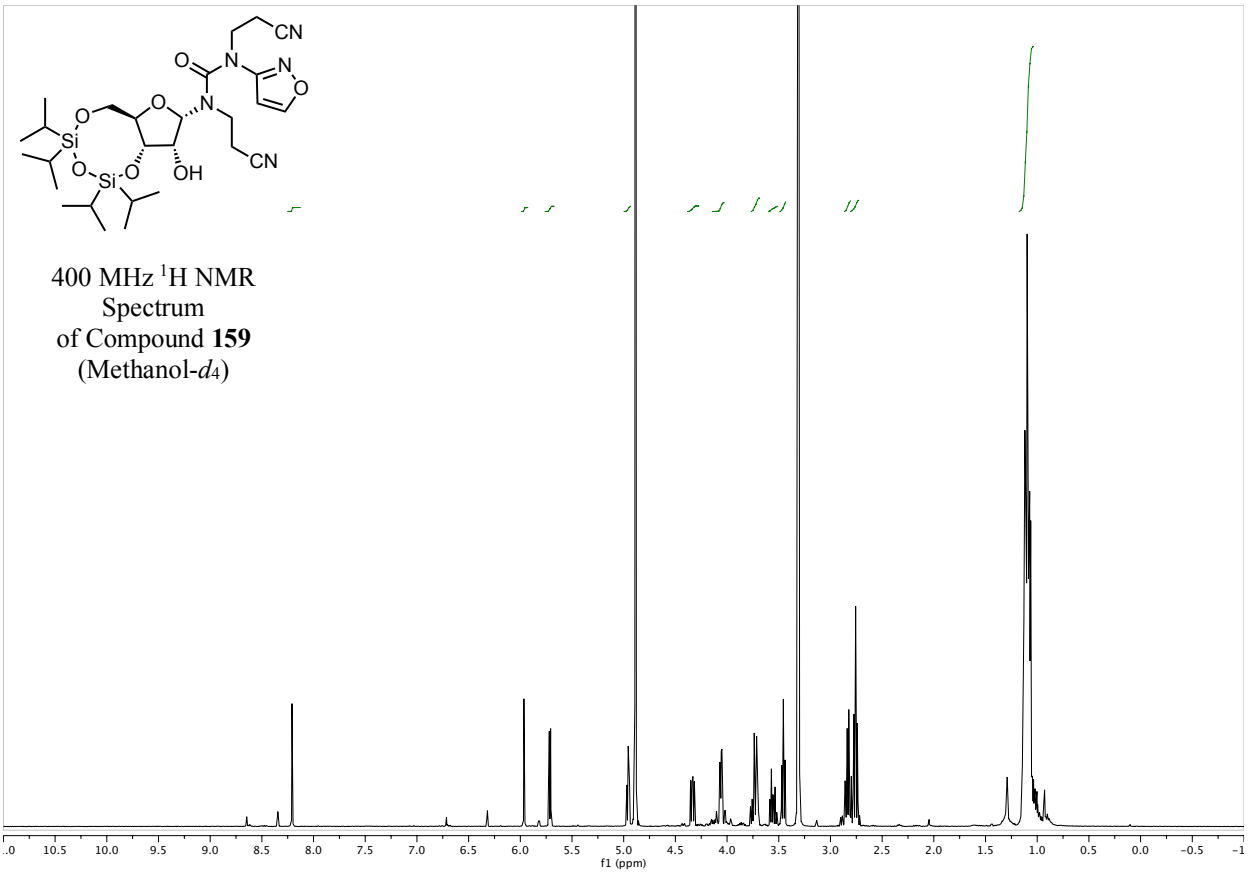
101 MHz ^{13}C NMR Spectrum of the Mixture Mostly Containing Compound **β-155** (Chloroform-*d*)



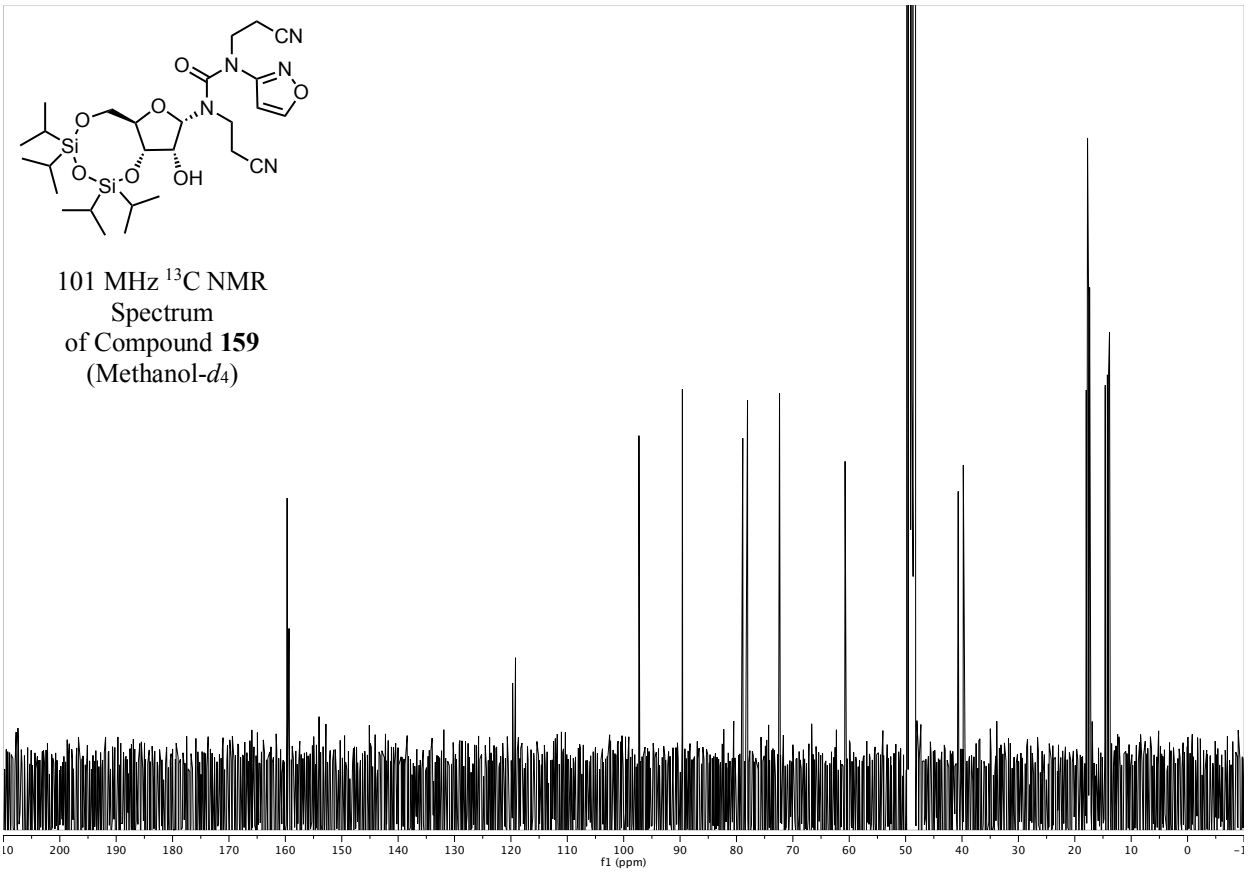
400 MHz ^{1}H NMR Spectrum of Compound **α-155** (Chloroform-*d*)



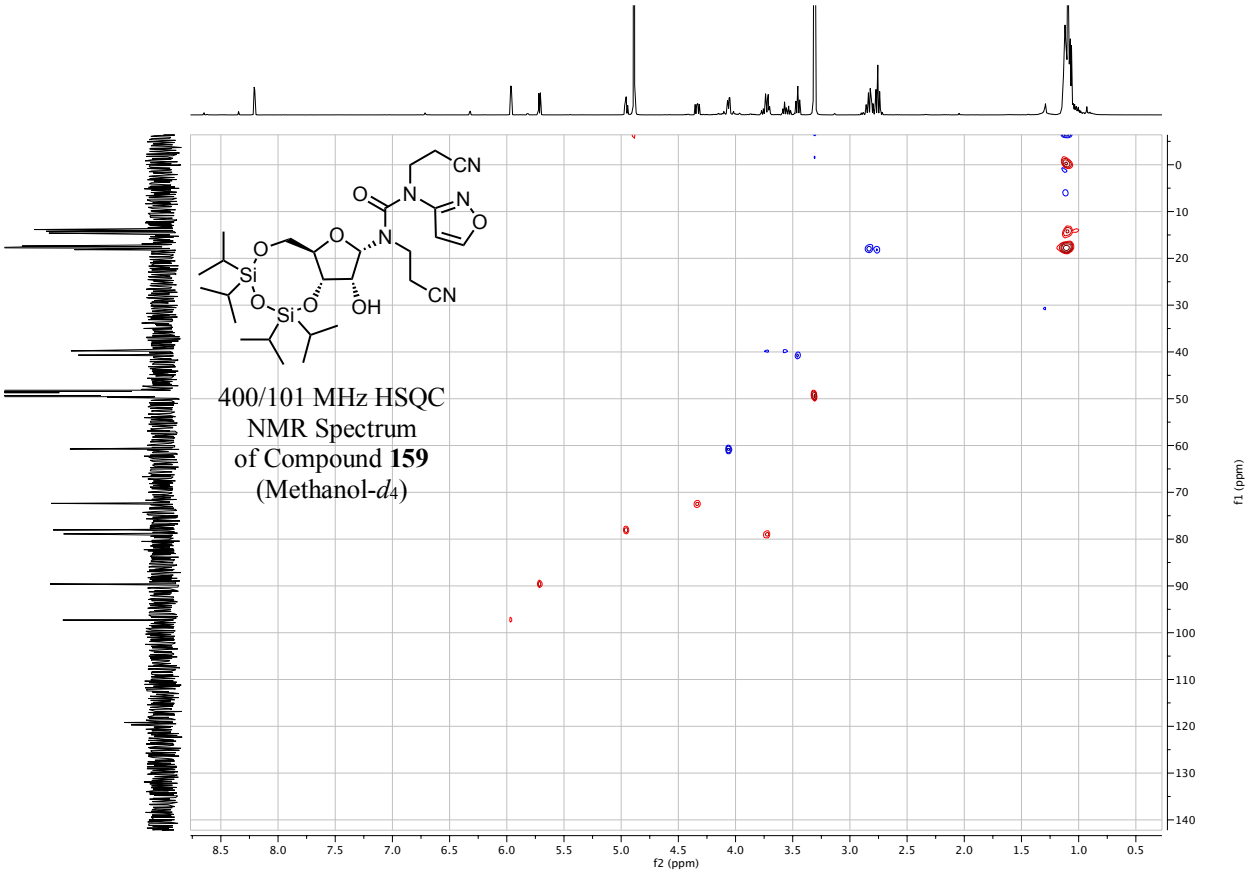
101 MHz ^{13}C NMR Spectrum of Compound **α-155** (Chloroform-*d*)



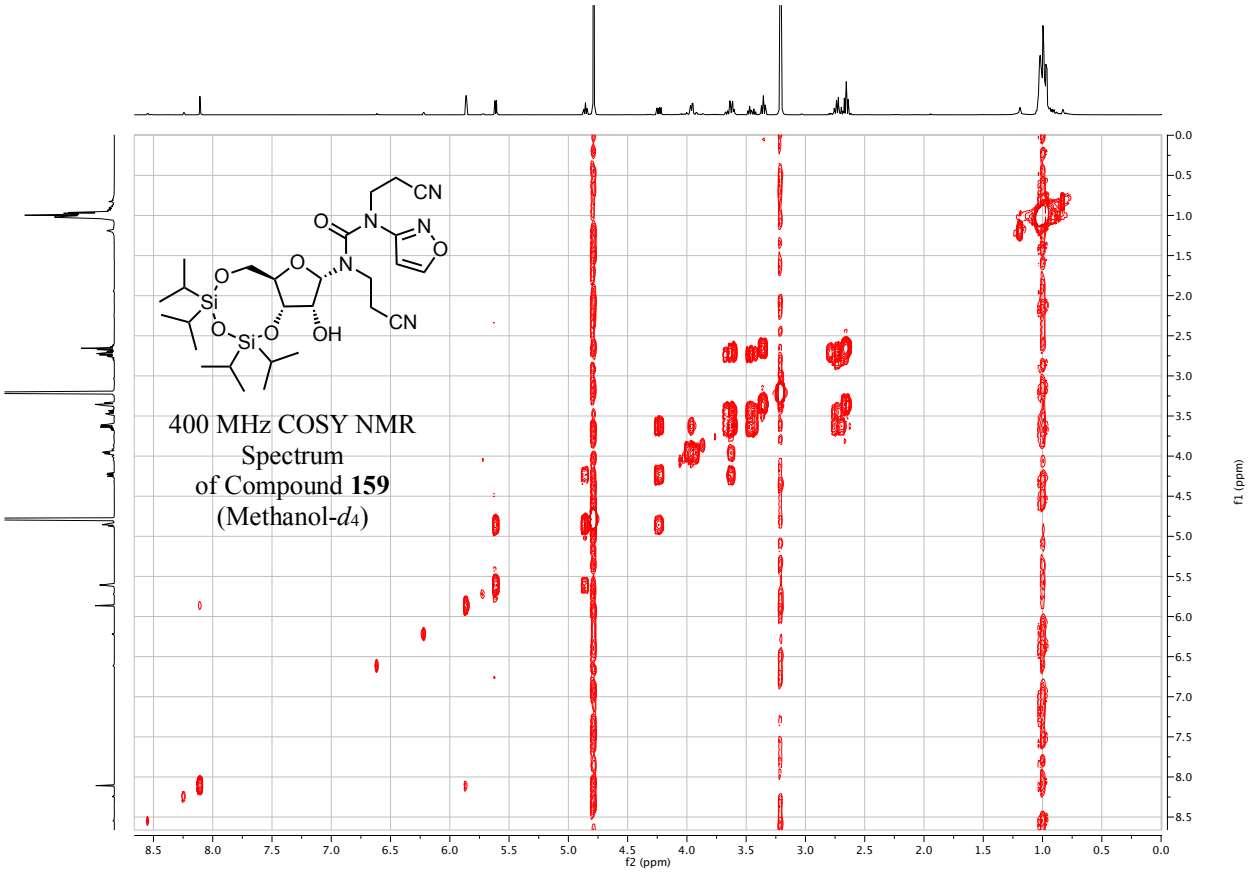
400 MHz ^{1}H NMR
Spectrum
of Compound **159**
(Methanol-d_4)



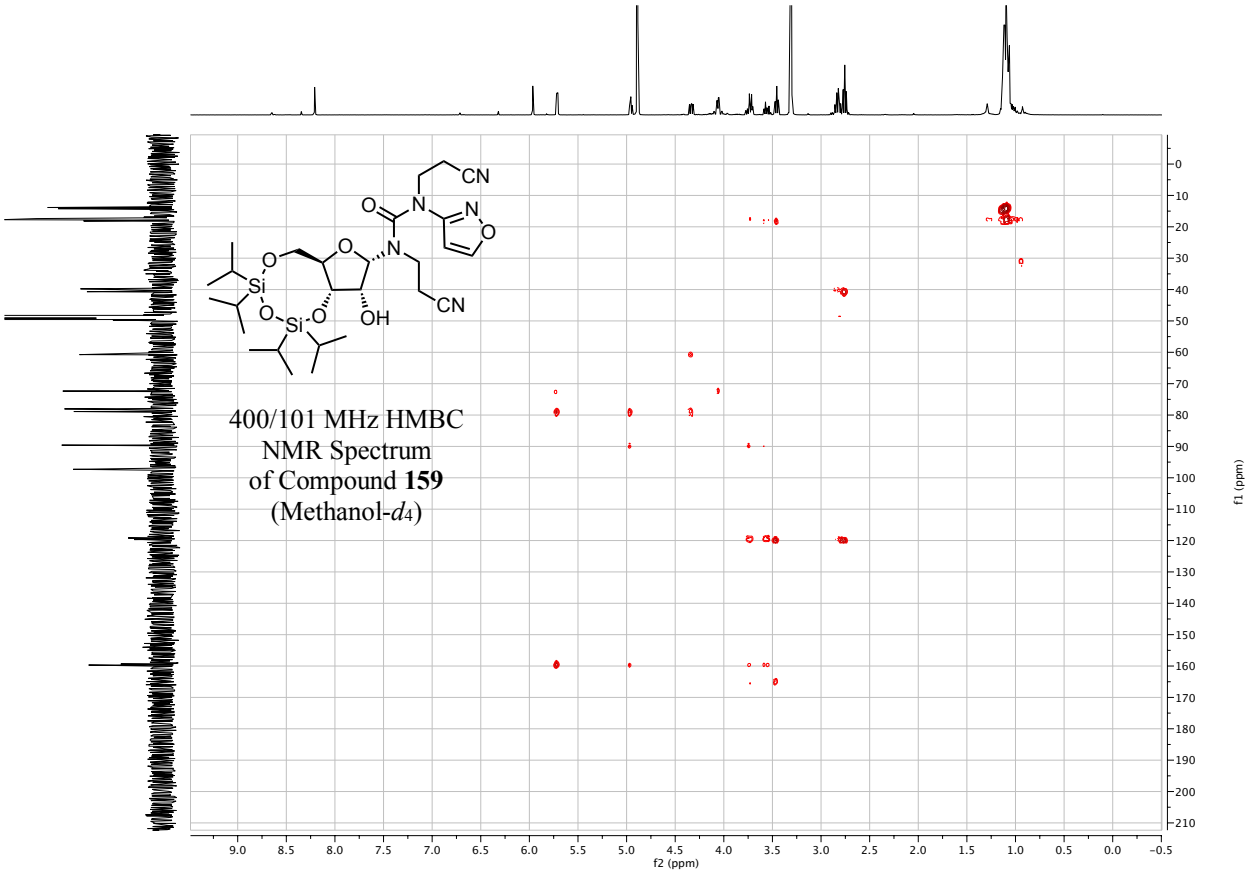
101 MHz ^{13}C NMR
Spectrum
of Compound **159**
(Methanol-d_4)

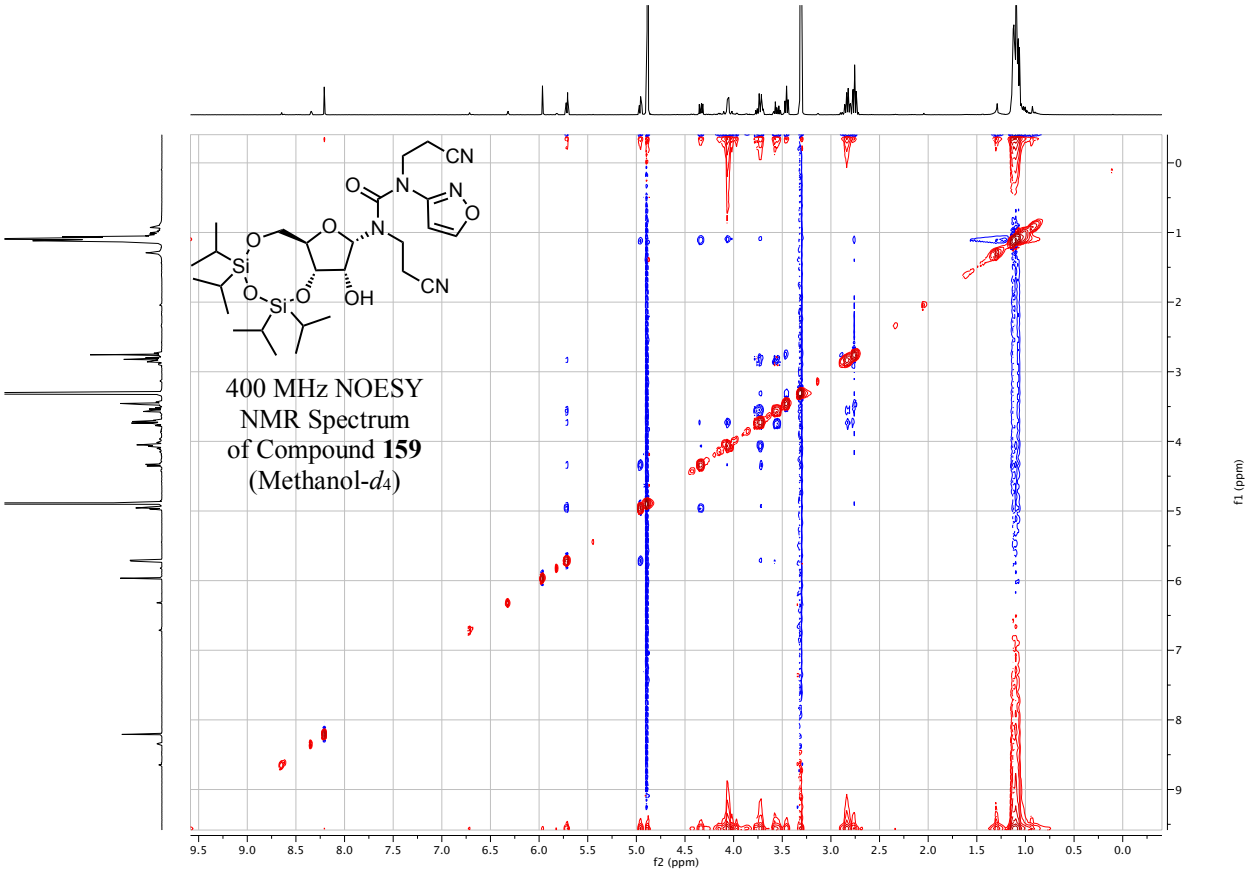


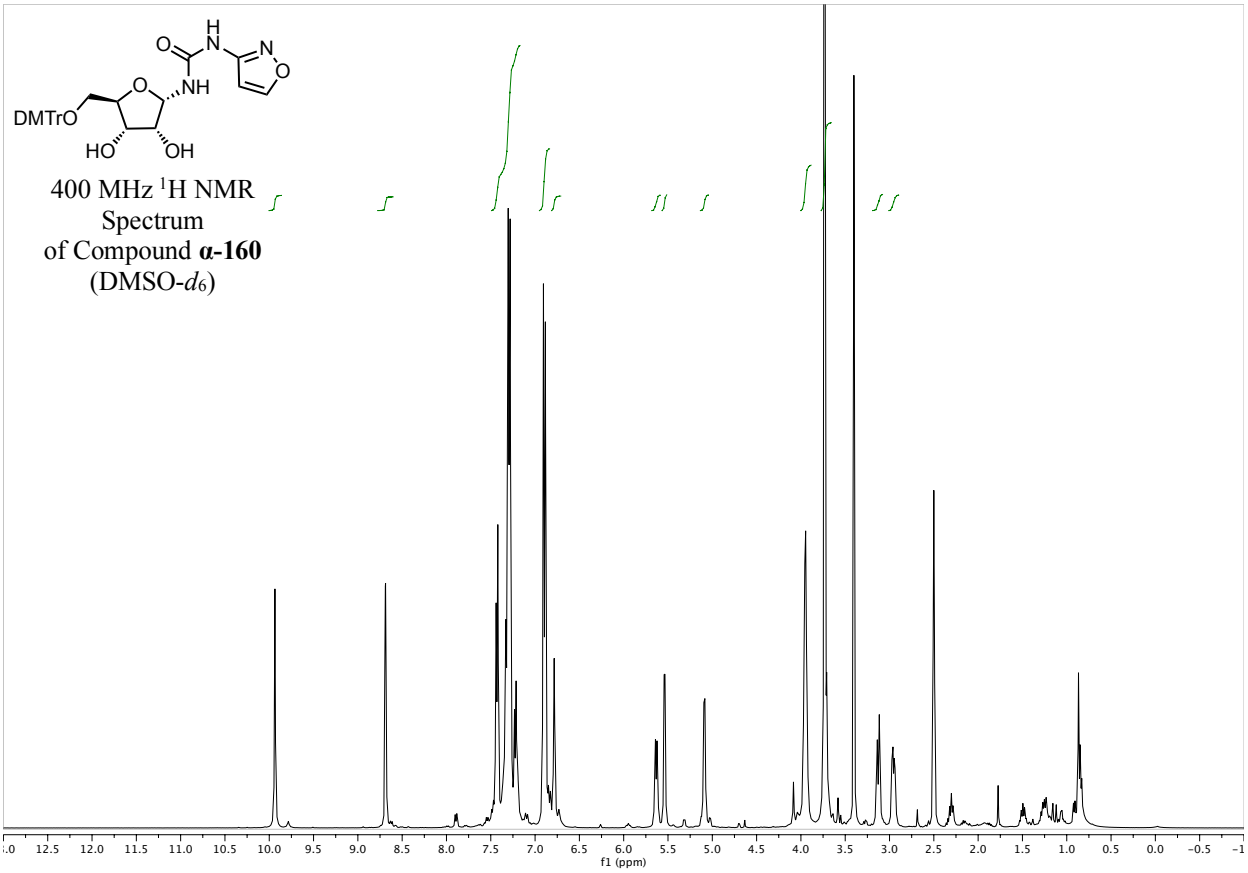
400/101 MHz HSQC NMR Spectrum of Compound **159** (Methanol-d_4)



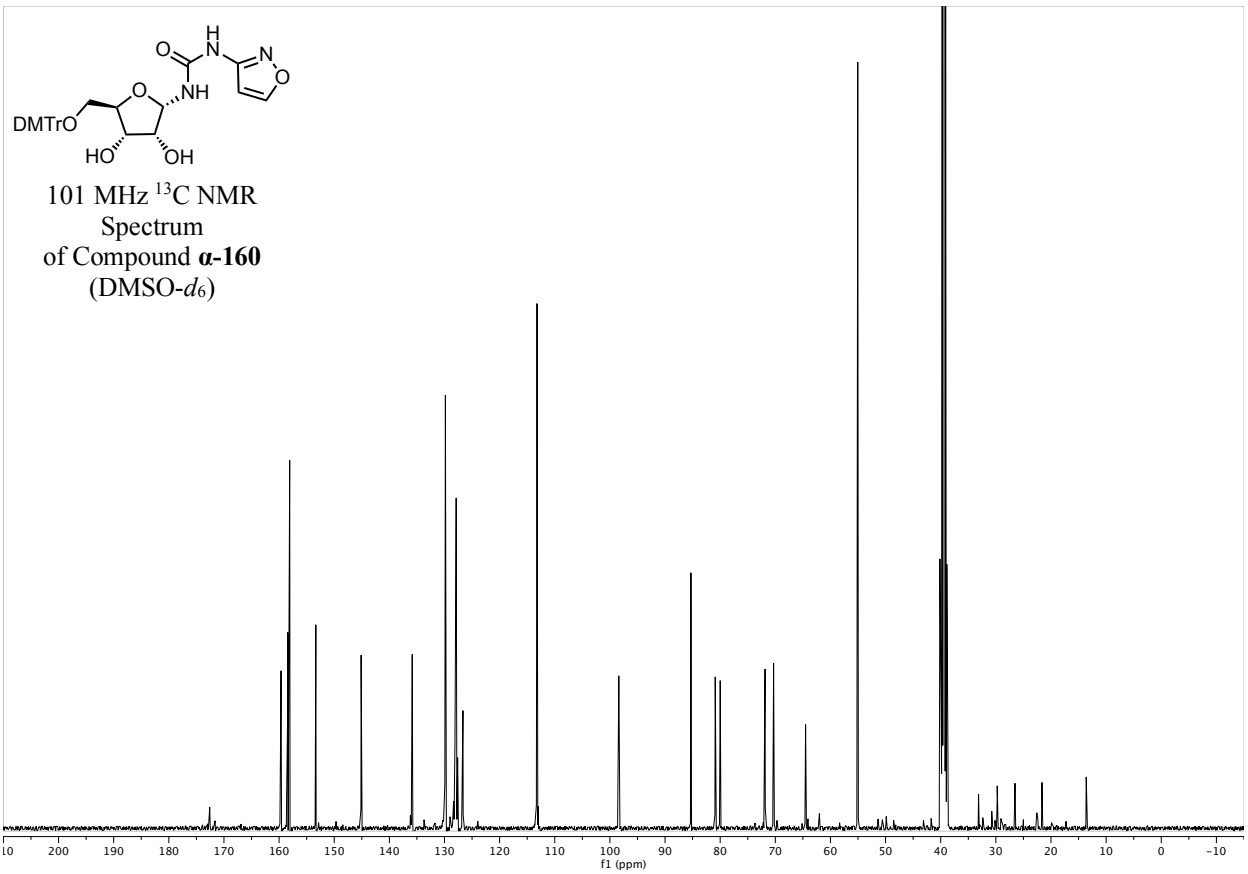
400 MHz COSY NMR Spectrum of Compound **159** (Methanol-d_4)


400/101 MHz HMBC
NMR Spectrum
of Compound **159**
(Methanol-d_4)
f2 (ppm)
f1 (ppm)

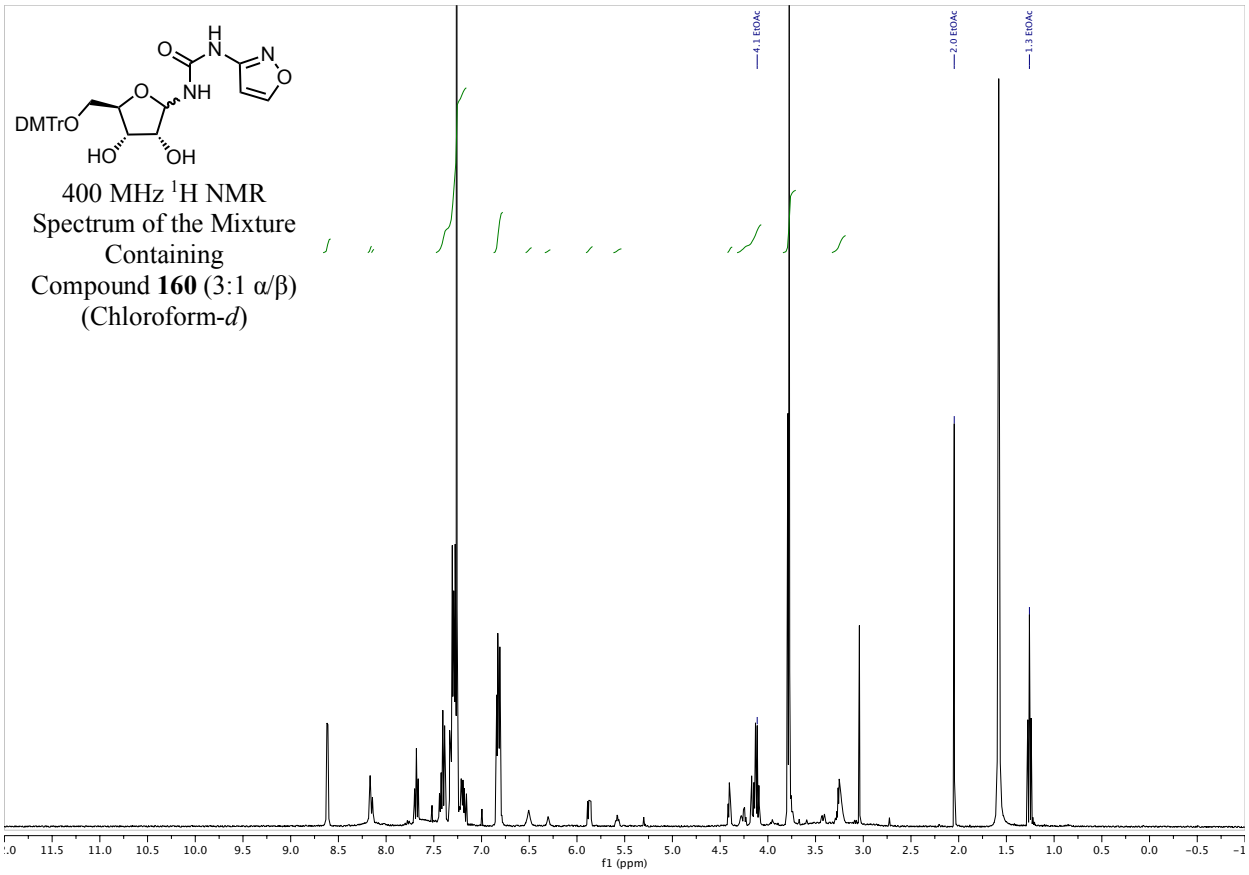

400 MHz NOESY
NMR Spectrum
of Compound **159**
(Methanol-d_4)
f2 (ppm)
f1 (ppm)



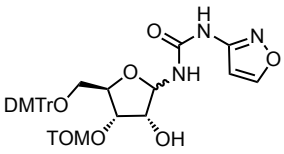
400 MHz ^{1}H NMR Spectrum of Compound **α-160** (DMSO-d_6)



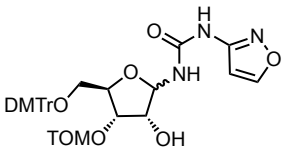
101 MHz ^{13}C NMR Spectrum of Compound **α-160** (DMSO-d_6)



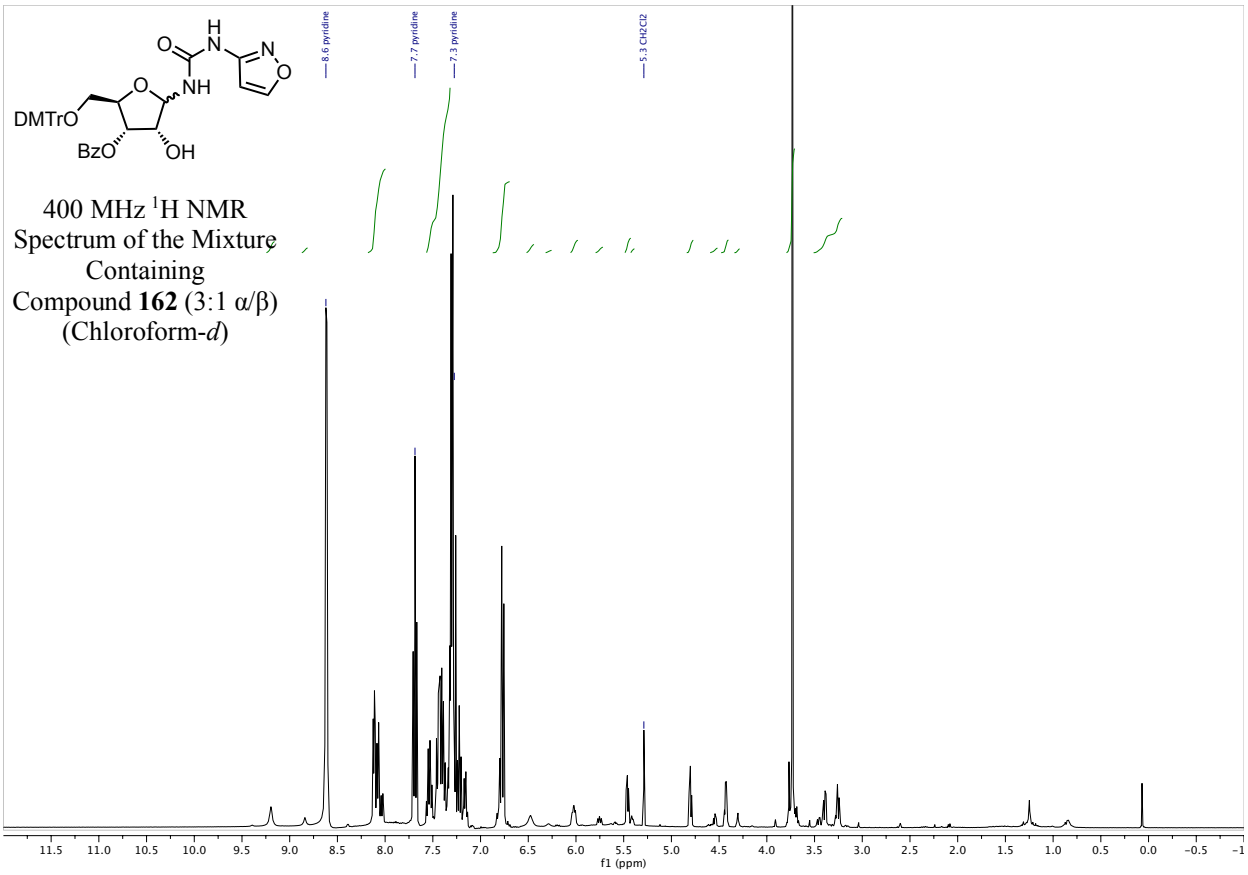
400 MHz ^{1}H NMR Spectrum of the Mixture Containing Compound **160** (3:1 α/β) (Chloroform-*d*)



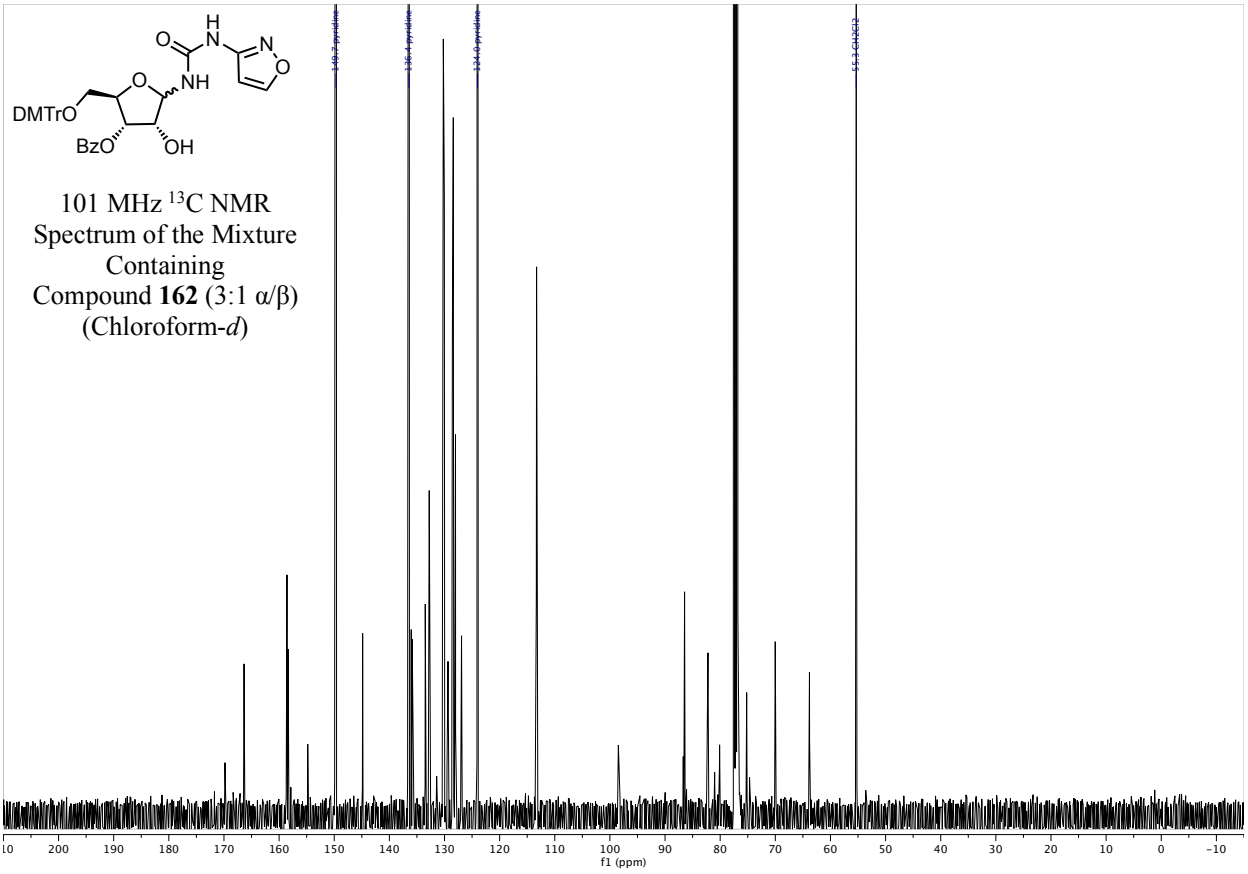
400 MHz ^{1}H NMR Spectrum of the Mixture Containing Compound **161** (2:1 α/β) (Chloroform-*d*)



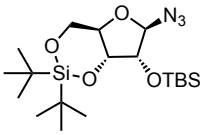
101 MHz ^{13}C NMR Spectrum of the Mixture Containing Compound **161** (2:1 α/β) (Chloroform-*d*)



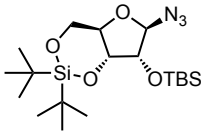
400 MHz ^{1}H NMR Spectrum of the Mixture Containing Compound **162** (3:1 α/β) (Chloroform-*d*)



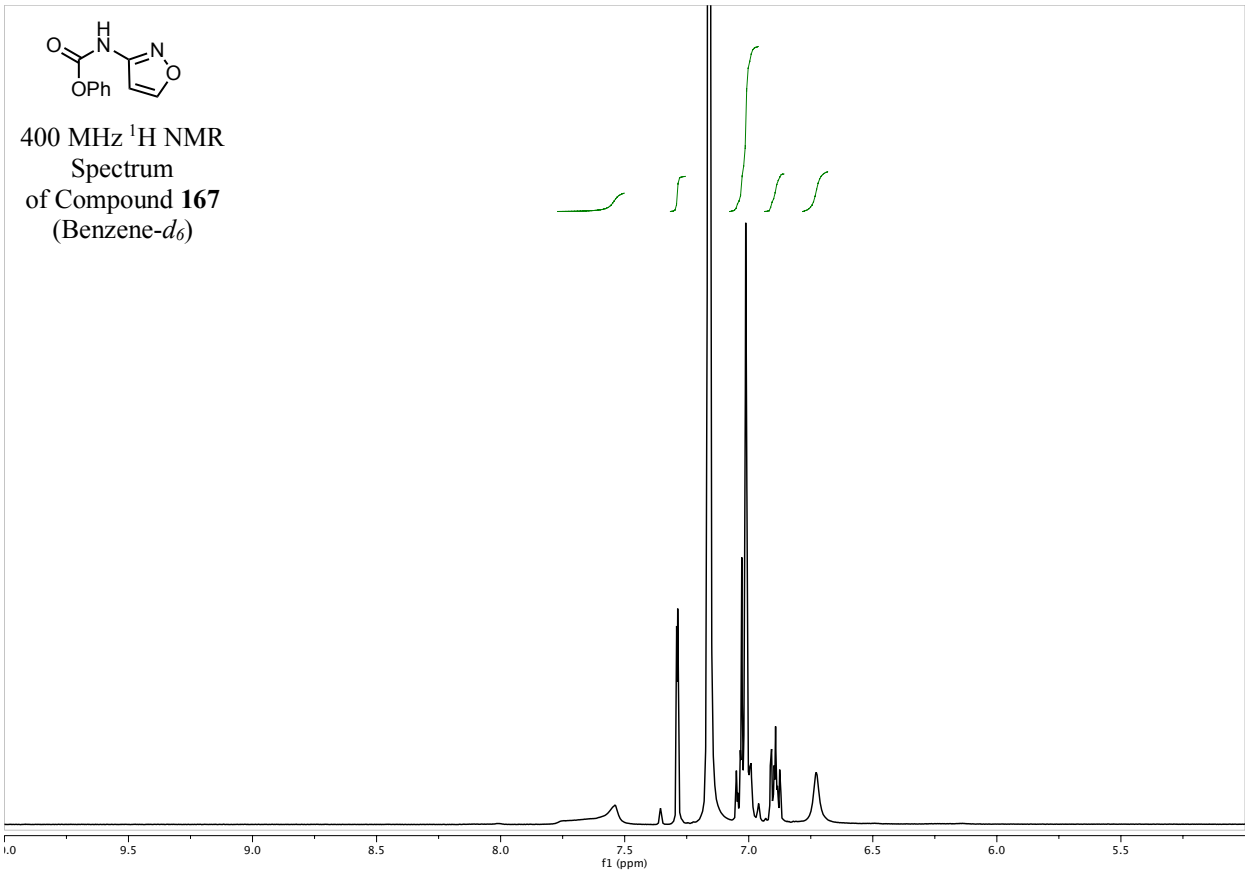
101 MHz ^{13}C NMR Spectrum of the Mixture Containing Compound **162** (3:1 α/β) (Chloroform-*d*)



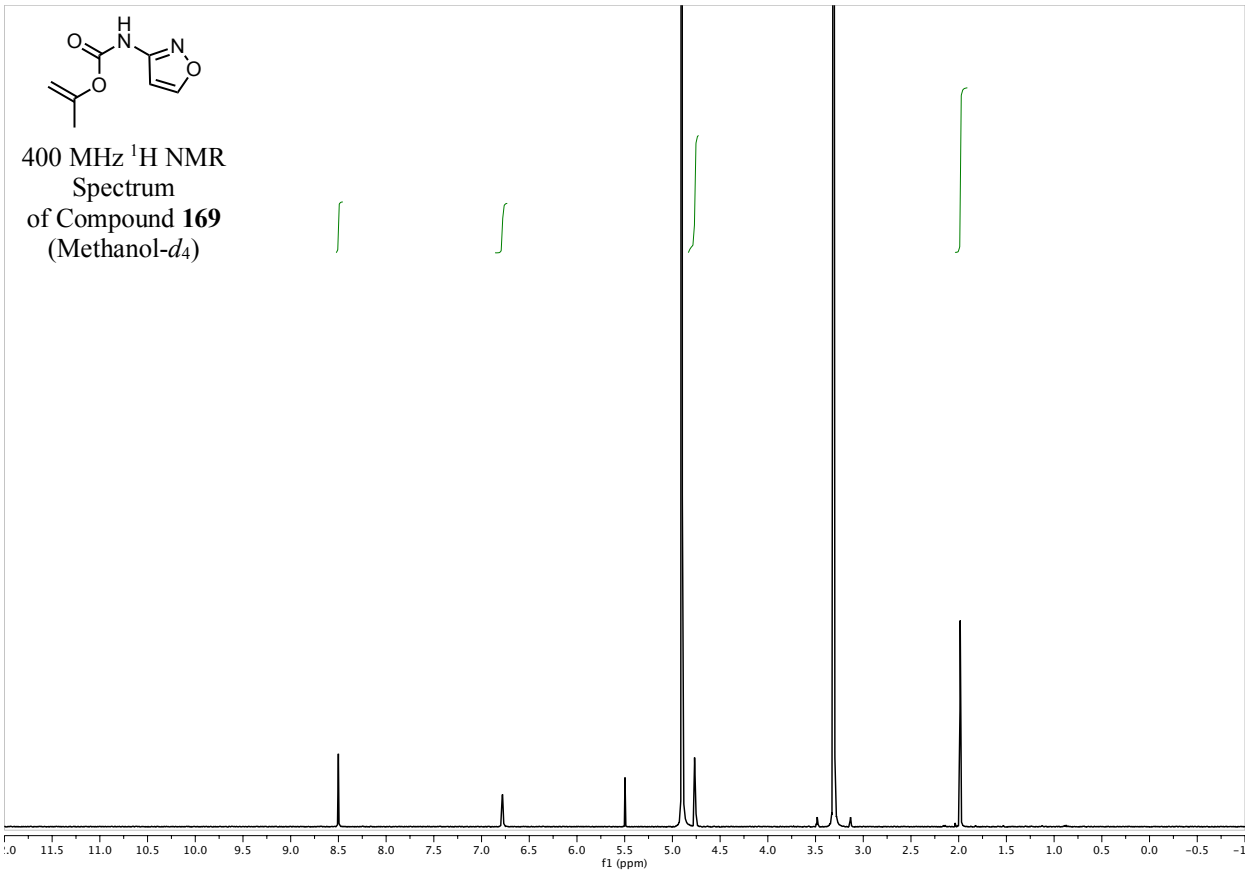
400 MHz ^{1}H NMR Spectrum of Compound **166** (Chloroform-*d*)

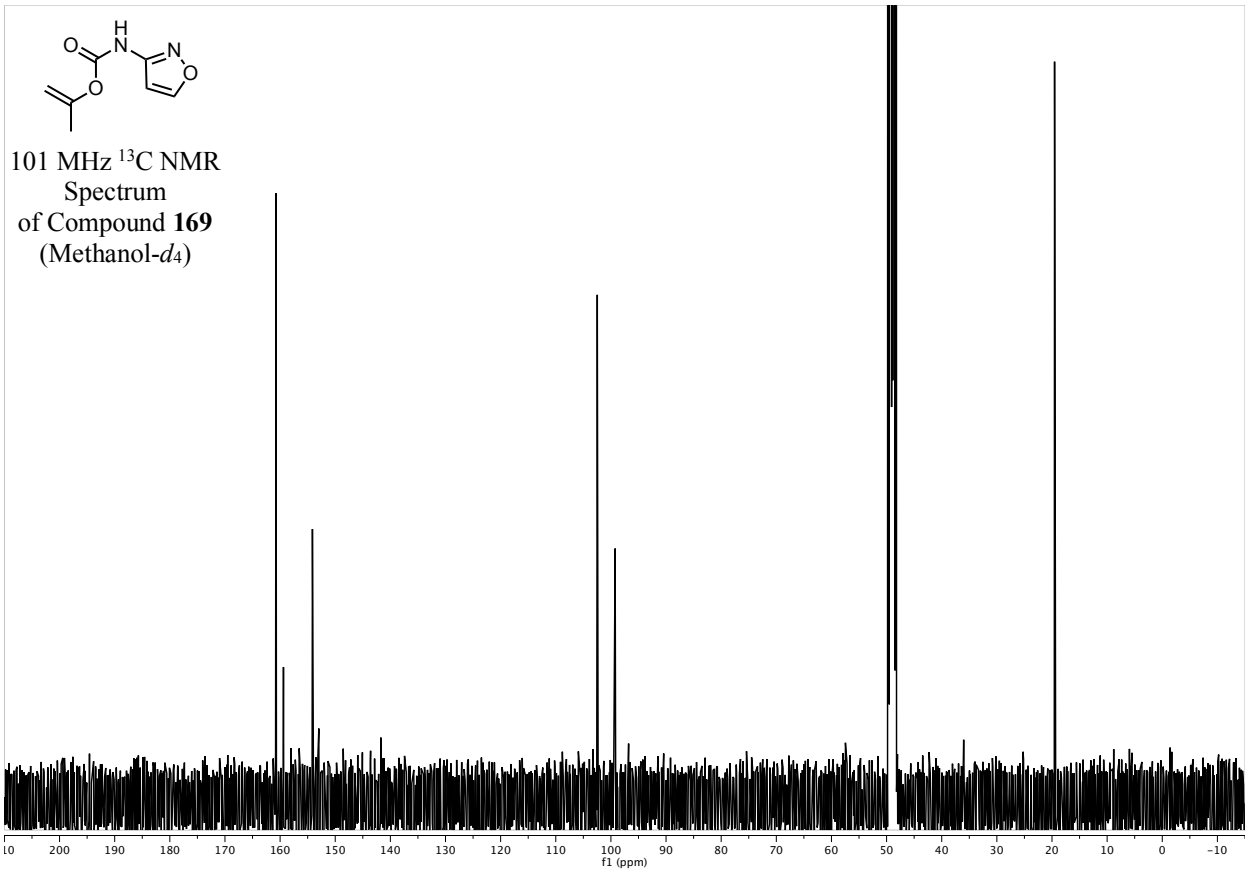


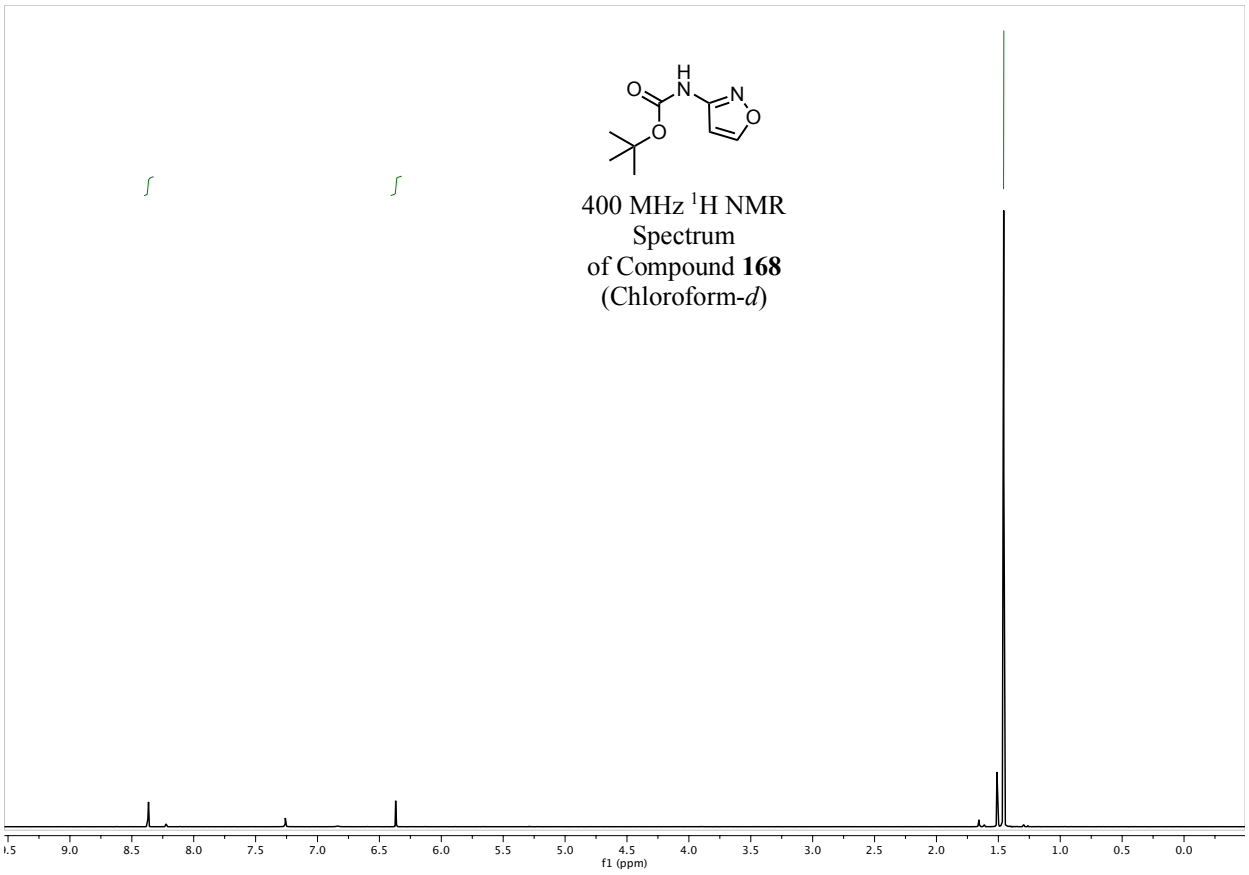
101 MHz ^{13}C NMR Spectrum of Compound **166** (Chloroform-*d*)



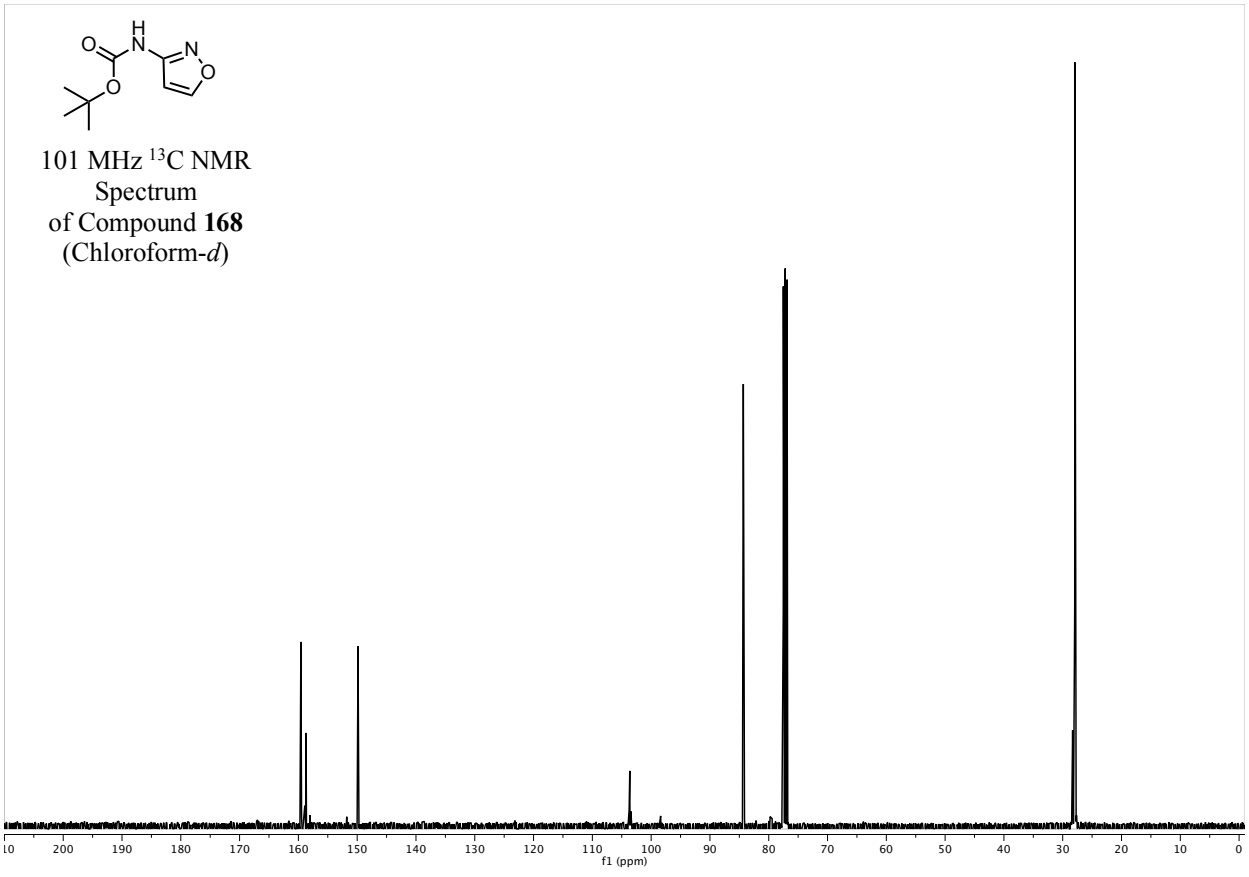
400 MHz ^{1}H NMR Spectrum of Compound **167** (Benzene-d_6)


400 MHz ^{1}H NMR
Spectrum
of Compound **169**
(Methanol-d_4)
f1 (ppm)


101 MHz ^{13}C NMR
Spectrum
of Compound **169**
(Methanol-d_4)
f1 (ppm)



400 MHz ^{1}H NMR
Spectrum
of Compound **168**
(Chloroform-*d*)



101 MHz ^{13}C NMR
Spectrum
of Compound **168**
(Chloroform-*d*)