Studies Towards Nucleic Acids at the Origins of Life

Ben W. F. Colville

UCL

Thesis Submitted to University College London for the Degree of Doctor Philosophy
I, Ben W. F. Colville confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Parts of this thesis have already been published at the time of writing:

Abstract

Nucleic acids are at the heart of extant biology and the structure of life’s original genetic polymer is still a question of debate. The RNA world theory proposes that RNA was the first nucleic acid employed as life’s genetic polymer due to its dual ability as informational storage (genotype) and primordial catalyst (phenotype). However, ribonucleotides are complex chemical structures, and simpler or more stable nucleic acids, such as threose nucleic acid (TNA) or DNA, can also carry genetic information.

In principle, nucleic acids like TNA could have played a vital role in the origins of life but the advent of any genetic polymer requires synthesis of its monomers. This work demonstrates a high-yielding, stereo-, regio- and furanosyl-selective prebiotic synthesis of threo-cytidine, an essential component of TNA. This work uses key intermediates (aminooxazolines) and reactions previously exploited in the prebiotic synthesis of the canonical pyrimidine ribonucleoside cytidine. It avoids the low yielding glycosylations that have previously been demonstrated for constructing nucleic acids and utilises and efficient photochemical anomerization that is enabled by selective anhydronucleoside thiolysis. This work also demonstrates that erythro-specific 2’,3’-cyclic phosphate synthesis provides a mechanism to photochemically select TNA cytidine and suggests that TNA may have coexisted with RNA during the emergence of life.

This thesis also investigates whether DNA be delivered simultaneously with RNA, the co-emergence of both would further probe the place and importance of DNA at the origin of life. This work expands on previous work towards DNA and examines the role of irreversible thiol addition to anhydronucleosides as a route towards DNA precursors.

Finally, the question of why the nucleobases present in extant biology (A, G, C, U/T) were chosen is addressed. This work explores the paradigm of UV stability as a selection pressure and the results contrast with current thinking that the purine nucleobases were chosen for their ability to resist degradation by UV-light.
Impact Statement

This work has a number of potential impacts for those inside and outside academia. Inside academia this work would be of great interest to those carrying out research in the field of prebiotic chemistry and on the origin of life. TNA, the focus of Chapter 2, has been continually discussed as a potential progenitor to RNA at the earliest stages of life since its discovery and the work presented here represents a significant advance towards elucidating a complete prebiotic synthesis of TNA.

Additionally, as focus has turned to RNA/DNA vaccines in recent years, greater interest is being paid to fundamental nucleic acid chemistry. TNA has been proposed a viable antisense oligonucleotide for therapeutic use and all the work presented in this thesis is focused on understanding the chemical behaviour of modified and natural nucleic acids.

Chapter 4 presents work on the UV stability of ribonucleosides, biologically ubiquitous building blocks, which sheds new light on the existing paradigm of photochemical interactions with genetic polymers. This has potential implications for the prebiotic chemistry community and those focused on studying photolesions effects on human health.

The work presented here is also of great interest to the general public. The question of abiogenesis has always captivated humanity and this thesis sheds more light on this fascinating topic. This work adds to the open-ended question of whether the biomolecules we see in biology today were the very first to arise or whether they were chosen for their innate properties.

The interdisciplinary nature of studying the origins of life has led to informal and formal collaborations with a number of other researchers internationally (Sasselov lab – Harvard University, Szabla lab – University of Edinburgh) and locally (Sutherland lab – MRC-LMB). Work from this thesis has and has been presented at international symposia (Simmons Collaboration on the Origins of Life 2019, IOoIM 2018, CRC 235 Emergence of Life Summer school 2019) and recently been accepted for publication in Angewandte Chemie.
This thesis is dedicated to A. I. McMullen, my grandfather. I never got the chance to discuss science with him and his contribution to human knowledge will continue to dwarf my own for a little while longer. I think he would be proud to see another chemist in the family.
"I must not fear.

Fear is the mind-killer.

Fear is the little-death that brings total obliteration.

I will face my fear.

I will permit it to pass over me and through me.

And when it has gone past I will turn the inner eye to see its path.

Where the fear has gone there will be nothing. Only I will remain."

_Dune_ – Frank Herbert

“So what?”

_Anon_

“What are the three most important rules of the chemist?”

This I knew from Ben.

"Label clearly. Measure twice. Eat elsewhere."

_The Name of the Wind_ – Patrick Rothfuss
Acknowledgements

A PhD belongs to one person but can never be completed alone. I could not have done so without the help and support of the many people I have been lucky enough to cross paths with. I must first thank Matt for your supervision over the course of two degrees, for your guidance and instilling in me the desire for asking big questions. Your endless enthusiasm for science is infectious and I hope to keep it with me forever. I have been very fortunate that I have been able to work on such fascinating topics and for that I thank you as well. Thanks must of course also go to Lello for taking me under your wing during my masters year and showing me how to do science and for your continual support from then on. To Daniel for your support with actual chemistry and many excellent discussions at the Buchis. To Cal and Sam for your friendship and support in the lab and since. The Powner group is a testament to Matt’s supervision and I must also thank the members of the Powner group with whom I have been fortunate enough to work alongside and share so many amazing memories, Alvaro, Benji, Jasper, Adam, Shaun, Becca, Christian, Pierre, Arif, Natalie, Kathryn, Leonardo, Jyoti, Arran, Andrew, Abid, thank you for your support and the bits of knowledge that snowball into something like a thesis over the years! Also thanks must go to Archie for a shared passion for science, spicy lunches and ball sports. I just hope we don’t bump into each other at too many more overnight interviews! Thanks to Abil for an NMR facility that made my PhD possible and in general to the other staff of the chemistry department for the help they provided and in particular Dewi Lewis, without him showing me round UCL on a Sunday I surely never would have come to UCL in the first place. On the same topic I must thank Dr. Peter Jones (CRGS) for inspiring a love for chemistry without which I would never have begun this journey.

To my parents, Brigid and Alan, you have done everything you can to help me since I can remember and my thanks go beyond words. Your constant encouragement and belief in me has always kept me going. To my brother, Tom, who despite his age has long been a role model for me and who’s appetite for science almost eclipses my own. To Karen and Mark, this thesis would have been impossible without your involvement and I am truly fortunate to have such a kind, caring and generous family.

Charlie, thank you for your friendship, your endless support and unerring positivity ever since we began at UCL. Chloe and Laura, your support of my scientific endeavours means a great deal. Thanks to Iain for your kindness and support when times were tough. My friends have kept me going and their own success inspires me continually, James, Karl, Chiara, Rachael, Izzy, George, Ella, Annie, thank you.
The Author

The author was born in Hackney, London in July 1994 and remained there until the age of 9 before relocating to Chappel, Essex. He attended Colchester Royal Grammar School for his GCSE and A-levels. His return to London to study Chemistry at UCL culminated in winning the Charles Vernon prize and a first class MSci degree (2016). He then returned to the group of Prof. Matthew Powner for his PhD studies.
Compound Numbering and Nomenclature

Atom Numbering

Atom numbering conventions for pentose and tetrose sugars, purines, pyrimidines and other relevant or generic structures.

Compound prefixes

Compound numbers of the type x-YYYY-ZZZ refer to compounds with multiple possible diastereomers of interest; x refers to the stereochemistry of the anomeric carbon and can be either α or β and is applied when this stereochemistry is of interest, yyy refers to the sugar stereochemistry and can be ribo, ara(bino), threo, erythro, xylo, lyxo and ZZZ refers to the compound number. Amino acids and their derivatives are denoted using their 3-letter codes (AA, generic amino acid) and combination of this with prefixes (Ac-) or suffixes (-CN) describe the amino acid derivative.
### List of Abbreviations

- °C  Degrees Celsius  
- 2AI  2-aminoimidazole  
- 2AO  2-aminooxazole  
- 2AT  2-aminothiazole  
- A  Adenosine (ribofuranosyl)  
- AA  Amino Acid  
- AAO  Arabinose aminozaoline  
- ADP  Adenosine diphosphate  
- AMP  Adenosine monophosphate  
- AmTP  Amidotriphosphate  
- Aq.  Aqueous  
- ATP  Adenosine triphosphate  
- b.p.  Boiling point  
- COSY  Correlation spectroscopy  
- CVP  Cyanovinylphosphate  
- D  2,6-Diaminopurine (ribofuranosyl)  
- DAP  Diamidophosphate  
- DCM  Dichloromethane  
- DKP  Diketopiperazine  
- DMAP  4-Dimethylaminopyridinet  
- EAO  Erythrose aminozaoline  
- Equiv.  Equivalents  
- G  Guanosine (ribofuranosyl)  
- Ga  Billion years (Giga annum)  
- GOE  Great Oxidation Event  
- HMBC  Heteronuclear Multiple Bond Correlation spectroscopy  
- I  Inosine (ribofuranosyl)  
- LAO  Lyxose aminozaoline  
- LHB  Late heavy bombardment  
- LUCA  Last Universal Common Ancestor  
- M.p.  Melting point  
- M/z  Mass to charge ratio  
- NERPE  Non-enzymatic (RNA) primer extension  
- NMR  Nuclear magnetic resonance  
- PAL  Present Atmospheric Level/Limit
• **P<sub>i</sub>**  Inorganic Phosphate (PO<sub>4</sub><sup>3-</sup>)
• **p-LAO**  Pyranosyl lyxose aminooxazoline
• **Ppm**  Parts per million
• **Quant.**  Quantitative
• **Quant.**  Quantitative yield (~100%)
• **RAO**  Ribose aminooxazoline
• **Ref.**  Reference
• **RP-FCC**  Reverse phase flash column chromatography
• **RT**  Room temperature
• **r-TCA**  Reverse Tricarboxylic acid (cycle)
• **TAO**  Threose aminooxazoline
• **TCA**  Tricarboxylic acid (cycle)
• **UV**  Ultraviolet
• **X**  Xanthosine (ribofuranosyl)
• **XAO**  Xylose aminooxazoline
• **Δ**  Application of heat
• **μL**  Microlitre(s)
# Table of Contents

1. Introduction .................................................................................................................. 20
  1.1 Why Study the Origins of Life and What is Life? .................................................. 20
  1.2 Biology Down and Geology Up ............................................................................. 22
    1.2.1 LUCA: A Biology Down Approach ................................................................ 22
    1.2.2 Geology Up and The Need for the Prebiotic Toolbox ................................. 23
  1.3 History of Earth and Conditions for Abiogenesis ............................................. 24
    1.3.1 Planetary formation ....................................................................................... 25
      1.3.1.1 Disks, Bombardments and Accretion ...................................................... 25
    1.3.2 The Evolution of the Early Earth’s Atmosphere ......................................... 25
      1.3.2.1 Hazes and Clouds .................................................................................. 27
      1.3.2.2 Ozone ................................................................................................... 27
    1.3.3 The Prebiotic Chemist’s Inventory ............................................................. 28
      1.3.3.1 Prebiotic Availability of Phosphate ....................................................... 28
      1.3.3.2 Simple Prebiotic Feedstocks ............................................................... 29
      1.3.3.3 Cyanosulfidic Protometabolism ........................................................... 31
  1.4 Perspectives in prebiotic chemistry ................................................................. 33
    1.4.1 Metabolism in prebiotic chemistry .............................................................. 33
      1.4.1.1 Chemiosmosis ........................................................................................ 33
      1.4.1.2 Wood-Ljungdahl Pathway ................................................................... 35
      1.4.1.3 Ancient Glycolysis ............................................................................... 38
      1.4.1.4 Tricarboxylic Acid Cycle (TCA) ............................................................. 38
      1.4.1.5 Reverse Tricarboxylic Acid Cycle (r-TCA) ........................................... 40
    1.4.2 Peptides in prebiotic chemistry ...................................................................... 43
      1.4.2.1 Protein World Theory .......................................................................... 43
      1.4.2.2 Amino Acid Synthesis ......................................................................... 43
      1.4.2.3 Making Peptide Bonds ........................................................................ 47
      1.4.2.4 Aminonitriles as Amino Acid Analogues in Peptide Synthesis ............ 49
      1.4.2.5 Catalytic Peptide Ligation .................................................................... 50
1.4.3 Lipids in Prebiotic Chemistry ................................................................. 52
1.4.3.1 Prebiotic Synthesis of Lipids and Their Precursors .................. 52
1.4.3.2 Prebiotic Chemistry Inside Membranes ........................................... 55
1.4.4 RNA World Hypothesis ........................................................................ 57
1.4.4.1 The Central Dogma of Molecular Biology ................................... 57
1.4.4.2 Formalizing the RNA World Hypothesis ...................................... 57
1.4.5 Alternative Genetic Polymers at the Origins of Life ......................... 59
1.4.5.1 DNA in the context of RNA ............................................................ 60
1.4.5.2 ANA ............................................................................................... 60
1.4.5.3 GNA ............................................................................................... 61
1.4.5.4 PNA ............................................................................................... 62
1.4.6 Oligomerisation of Genetic Polymers .................................................. 62
1.4.6.1 Non-Enzymatic Primer Extension - NERPE ................................ 63
1.4.7 A Holistic Approach to the Origins of Life ....................................... 64
1.5 Prebiotic Synthesis of Nucleic acids ....................................................... 65
1.5.1 Classic Disconnection Approach ......................................................... 65
1.5.2 Prebiotic Sugar Synthesis ..................................................................... 65
1.5.2.1 Formose Reaction ........................................................................... 65
1.5.2.2 Prebiotic Synthesis of Other Sugars ............................................. 67
1.5.3 Prebiotic Nucleobase Synthesis .......................................................... 68
1.5.4 Prebiotic Glycosylation ........................................................................ 70
1.5.4.1 Problems with Glycosylation ......................................................... 70
1.5.4.2 Glycosylations of 1',2'-Cyclic Phosphates ................................... 72
1.5.4.3 FaPy Glycosylations ....................................................................... 72
1.5.5 Alternative Routes to Nucleic Acids .................................................... 76
1.5.5.1 Alternate Routes to Pyrimidines ..................................................... 76
1.5.5.2 Powner and Sutherland Methodology ......................................... 77
1.5.5.3 Problems with the Powner and Sutherland methodology .......... 78
1.5.5.4 Alternate Routes to Purines ............................................................. 81
1.6 Prebiotic Photochemistry ......................................................................... 83
1.6.1 Prebiotic Photochemical Reduction ......................................................... 85
1.6.2 Prebiotic Photochemical Synthesis of Nucleic Acids ................................. 86
  1.6.2.1 Arabino-Nucleic Acids ................................................................. 87
  1.6.2.2 Photochemical Routes to DNA ...................................................... 88

2 Threose Nucleic Acid - TNA ........................................................................ 91
  2.1 Introduction ............................................................................................. 91
    2.1.1 Etiology of Nucleic Acid Structure .................................................. 91
      2.1.1.1 Hexose, Pento-pyranosyl and Pento-furanosyl Nucleic Acids ........... 92
      2.1.1.2 Threose Nucleic Acid (TNA) ....................................................... 92
      2.1.1.3 TNA as a Xeno-Nucleic Acid ..................................................... 94
      2.1.1.4 TNAs in Prebiotic Chemistry ..................................................... 94
    2.1.2 Previous Prebiotic Syntheses of TNA Monomers .................................. 95
    2.1.3 Formation of 2-aminooxazoles ......................................................... 99
  2.2 Results and Discussion ............................................................................ 101
    2.2.1 Tetrose-aminoazoles ......................................................................... 101
      2.2.1.1 Tetrose Oxazolidinone thiones .................................................. 105
      2.2.2 Tetrose-anhydronucleosides .......................................................... 107
    2.2.3 Thiolysis of tetrose anhydronucleosides in formamide ......................... 112
    2.2.4 Scalable thiolysis procedure ................................................................ 114
    2.2.5 Hydrolysis of Anhydros ...................................................................... 116
    2.2.6 Oxidative Recyclization of β-tetrose Thionucleosides .......................... 117
    2.2.7 Synthesis of Authentic Tetrose Nucleic Acids ..................................... 119
    2.2.8 Photochemical anomerization ............................................................ 124
      2.2.8.1 Irradiation of Tetrose α-Thiocytidines .......................................... 129
      2.2.8.2 Attempted Anomerisation of Oxygenic Tetrose β-cytidines ........... 129
      2.2.8.3 Irradiation of Threose Thiouridines .............................................. 130
    2.2.9 Peroxide Oxidation ............................................................................. 132
    2.2.10 Prebiotic Phosphorylation ............................................................... 135
    2.2.11 Phosphorylation of Erythro-Nucleosides .......................................... 139
    2.2.12 Selection of TNA cytidines ............................................................. 143
5.3 Supplementary Figures ........................................................................................................... 176

5.3.1 TNA ......................................................................................................................................... 176

5.3.1.1 One pot formation of TAO/EAO ...................................................................................... 176

5.3.1.2 Oxazolidinone thione formation ...................................................................................... 178

5.3.1.3 Reaction of TAO and EAO with cyanoacetylene .............................................................. 179

5.3.1.4 Formamide thiolysis procedure ....................................................................................... 180

5.3.1.5 Thiolysis of threo-89 and erythro-89 .............................................................................. 181

5.3.1.6 Hydrolysis of threo-89 and erythro-89 ............................................................................ 182

5.3.1.7 Reaction of β-threo-111 and β-erythro-111 with hydrogen peroxide ...................... 183

5.3.1.8 Hydrogen peroxide oxidation of α-threo-111............................................................... 184

5.3.1.9 Reaction of β-threo-111 and α-threo-111 with hydrogen peroxide ..... 185

5.3.1.10 UV-Vis Spectra .................................................................................................................. 186

5.3.1.11 Photoanomerization of β-threo-111............................................................................. 188

5.3.1.12 Photoanomerization of β-threo-111 and β-erythro-111 .............................................. 190

5.3.1.13 Photoanomerization of β-threo-111 and β-erythro-111 spiked with authentic β-threo-111 and β-erythro-111 ................................................................. 191

5.3.1.14 Photoanomerization of β-threo-111 and β-erythro-111 spiked with authentic α-threo-111 and α-erythro-111 ................................................................. 192

5.3.1.15 Irradiation of β-threo-85 .................................................................................................. 193

5.3.1.16 Irradiation of β-erythro-85 .............................................................................................. 195

5.3.1.17 Irradiation of β-threo-148 ............................................................................................... 196

5.3.1.18 Irradiation of β-threo-112 ............................................................................................... 197

5.3.1.19 Irradiation of α-threo-111 ............................................................................................... 198

5.3.1.20 Irradiation of α-erythro-111 ........................................................................................... 200

5.3.1.21 Irradiation of α-threo-111 and α-erythro-111 ............................................................... 201

5.3.1.22 Hydrolysis of α-threo-111 .............................................................................................. 202

5.3.1.23 Cyanacetylene accelerated hydrolysis ........................................................................... 203

5.3.1.24 Phosphorylation of β-erythro-111 ............................................................................... 204

5.3.1.25 Irradiation of β-threo-111 and β-erythro-151 ............................................................... 205

5.3.1.26 Irradiation of β-erythro-151 .......................................................................................... 207
5.3.1.27 Phosphorylation and irradiation of β-threo-111 and subsequent treatment by phosphatase.................................................................208
5.3.1.28 Phosphorylation of β-threo-111..........................................................210
5.3.1.29 Phosphorylation of α-threo-111..........................................................212
5.3.1.30 Phosphorylation of α-threo-111..........................................................214
5.3.1.31 Phosphorylation of α-erythro-111.........................................................216
5.3.2 DNA........................................................................................................218
5.3.2.1 Reaction of ara-102 with ethanethiol in DMF ........................................218
5.3.2.2 Reaction of ara-89 with EtSH and NaH ................................................219
5.3.2.3 Reaction of ara-126 with EtSH and NaH.................................................220
5.3.2.4 Thiolysis Screening Method 1...............................................................221
5.3.2.5 Reactions of ara-102 in DMF with EtSH/EtSNa ....................................222
5.3.2.6 Reactions of ara-102 in formamide with EtSH/EtSNa ..........................223
5.3.2.7 Reaction of ara-126 in formamide..........................................................224
5.3.2.8 Reaction of ara-89 in formamide...........................................................225
5.3.2.9 Reaction of ara-102/ara-126 with prebiotic sulfur nucleophiles ..........226
5.3.2.10 Reaction of cysteamine with ara-102 in water.....................................227
5.3.2.11 Attempted desulfurisation reaction of 157 with Raney-Nickel............228
5.3.2.12 Irradiation of 157 ................................................................................229
5.3.2.13 Irradiation of 166 ................................................................................230
5.3.3 UV stability..............................................................................................231
5.3.3.1 General Method for Investigating UV Stability ....................................231
5.3.3.2 Irradiation of A 2mM vs 117 2 mM.......................................................232
5.3.3.3 Irradiation of A 2 mM vs I 2 mM............................................................234
5.3.3.4 Irradiation of A 2 mM vs D 2 mM ..........................................................236
5.3.3.5 Irradiation of A 10 mM vs D 2 mM .........................................................238
5.3.3.6 Irradiation of A 10 mM vs I 2 mM..........................................................240
5.3.3.7 Irradiation of A 10 mM G 2 mM.............................................................242
5.3.3.8 Irradiation of A, I, G and X 2 mM .........................................................244
5.3.3.9 Irradiation of A, I, D and X 2 mM..........................................................246
5.4 Synthesised Compounds

5.4.1 TNA

5.4.1.1 2-thiooxazole

5.4.1.2 Tetrose-aminoazolines

5.4.1.3 Synthesis of propiolamide

5.4.1.4 Synthesis of cyanoacetylene

5.4.1.5 Tetrofuranosyl-2,2'-anhydrocytidine hydrochloride

5.4.1.6 Two step synthesis of β-tetro-thiocytidines

5.4.1.7 Characterisation data for β-threo-111

5.4.1.8 Characterisation data for β-erythro-111

5.4.1.9 Synthesis of tetrose β-cytidines

5.4.1.10 Synthetic route for authentic standard of α-threo-111

5.4.1.11 Characterisation data for β-threo-148

5.4.1.12 Characterisation data for β-erythro-148

5.4.1.13 Synthesis of rac-Tetrofuranosyl-2,2'-anhydouridine

5.4.1.14 Synthesis of β-threo-112

5.4.1.15 Synthetic route for authentic standard of α-threo-111

5.4.1.16 Synthesis of Calcium L-threonate

5.4.1.17 Synthesis of L-threonolactone

5.4.1.18 Synthesis of 2',3'-di-O-acetyl-β-cytidines

5.4.1.19 Silylation of 2-thiocytosine

5.4.1.20 Synthesis of 2',3'-di-O-benzoyl-2-thio-thiocytidine

5.4.1.21 Synthesis of authentic standard of α-threo-thiocytidine

5.4.1.22 Synthetic route for authentic standard of α-erythro-111

5.4.1.23 Synthesis of 2',3'-di-O-benzoyl-erythronolactone

5.4.1.24 Synthesis of 2',3'-di-O-benzoyl-erythrothiocytidine

5.4.1.25 Synthesis of α-erythro-thiocytidine

5.4.1.26 Peroxide oxidation of α-threo-111

5.4.1.27 Peroxide oxidation of α-erythro-111
5.4.1.28 Synthesis of β-erythro-thiocytidine-2′,3′-cyclic phosphate β-erythro-151
......................................................................................................................308

5.4.2 DNA Chapter ..............................................................................................311
5.4.2.1 2′,8-cyclo-adenosine .............................................................................311
5.4.2.2 Synthesis of 2′-deoxy-2′-ethylmercapto-8-oxy-adenosine ..........312
5.4.2.3 Synthesis of 2′-deoxy-2′-ethylmercapto-ribocytidine .......................314
5.4.2.4 Synthesis of 2′-deoxy-2′-ethylmercapto-ribouridine .................315

References ............................................................................................................316
# 1 Introduction

The focus of this thesis is the prebiotic chemistry of nucleic acids and what follows in this chapter is a general introduction to prebiotic chemistry to contextualise the work that is presented in later chapters.

Prebiotic chemistry is the process of figuring out the instruction manual to make life’s first cells from a given “toolbox” of prebiotic compounds inferred from geology. It therefore encompasses the discovery of chemistry that can be used to probe these questions as well as investigating what selection processes could have acted on these biomolecules on the early Earth.

Research on the origins of life stumbled blindly for many years, not knowing how or even where to look for answers. More recently, these questions have now become things that can be studied. Due to many technological developments there is now the ability to analyse previously opaque and complex reaction mixtures (high frequency NMR, powerful chromatographic techniques), the ability to accurately date rocks (allowing inference of a much clearer picture of the primordial Earth’s geology), as well as having a better understanding of planetary dynamics and atmospheric formation. These are all key parts in this inherently interdisciplinary field of origins of life (abiogenesis) research. Presented in this section is an exploration of: what life is in the context of this research, the geochemical prebiotic approach presented throughout the rest of this thesis, an overview of the inferred geological conditions at the origins of life and finally a foray into several theories on the origins of life are presented to give vital context to discussions in later chapters.

## 1.1 Why Study the Origins of Life and What is Life?

The origin of life is an unanswered question of special interest throughout the human condition with scientific, philosophical and religious rationale seeking to explain it. Currently no one theory has achieved a broad consensus leaving many diverse, interesting questions yet to be answered.

Examples of such questions may be:

- Where did life start? What environments were conducive to its beginning?
- What came first? Metabolism, genetics or something else entirely?
- What role did DNA play at the origin of life, if any?
Studying the origins of life on Earth would well equip humanity with the tools to search for life (molecular biosignatures, habitable areas) on other planets.

NASA defines life as “a self-sustaining chemical system capable of Darwinian evolution”. This implies that as a chemical system it should be possible to rediscover the chemical pathways towards life. This thesis considers the origins of life, abiogenesis, a fundamentally chemical problem to solve.

Solving this problem has posed a challenge. There are many factors but chief among them, is the sheer time scale of the events that have led to extant life. For most of human history the belief was that the timescale of events that had shaped the planet was on the order of thousands of years but it was the realisation that the timescale was over billions of years that has caused many of the difficulties in definitively revealing the mechanisms of the origin of life. The reason for this is that the long timescales give time for evidence trapped in the rock record to be eroded by the metamorphic nature of the planet’s tectonics. Due to this there is very little empirical evidence from the Hadean (4.6-4.0 Ga) and Archean (4.0-2.5 Ga) eras that can directly inform on how life did start (Figure 1). The earliest fossils of bacteria have been dated to 3.5 Ga. This reveals that life was established, albeit at a very simple level compared to extant life, as rapidly as 1 billion years after the earth was formed.

![Figure 1 Timeline from the formation of Earth to present day with geological eras overlaid.](image)

So how can life be studied if there is no empirical evidence to be examined now and how can, in the case of this thesis, chemistry be used in answering these questions?
This has been formalised in recent years as solving the question of abiogenesis, the original origin of life or living organisms from inanimate matter. Two major routes have been devised to investigate abiogenesis, the Geology Up approach and the Biology Down approach.

1.2 Biology Down and Geology Up

Two common approaches to attempt to understand the origins of life are the Biology Down and the Geology Up approaches. These broad terms encompass many individual theories and approaches within them to solving the problem of the origin of life. A brief discussion of both terms follows to provide context for later discussions of particular theories.

1.2.1 LUCA: A Biology Down Approach

The Biology Down approach advocates the use of phylogenetics to trace back extant life to a common origin point in genetic history. This is known as the Last Universal Common Ancestor (LUCA) and is seen as an intermediate point between abiotic chemistry and ancient (yet still highly complex) microbial life but not, however, the first living organism. LUCA likely had all four key components of a minimal functional protocell: a genome, catalysis, metabolism and a membrane (Figure 2).

To find a more specific makeup of LUCA previous work has investigated the most conserved pathways (identical/similar genetic sequences that appear in species over generations) in biology and seeing which are the most ‘ancient’. From these insights it is proposed by Weiss et al. that LUCA was “anaerobic, CO$_2$-fixing, H$_2$-dependent with a Wood-Ljungdahl pathway, N$_2$-fixing and thermophilic”. These complex metabolic pathways preclude LUCA from being the first living cell and imply that a period of evolution must have occurred to arrive at LUCA.

Unfortunately, this appears to offer an incomplete picture as these investigations can only reveal potential information about what early life may have relied on for its function (i.e. genetics, catalysis, metabolism) but not how any of these systems themselves came about. In effect, this approach only informs the synthetic targets for prebiotic chemistry and implies a requirement for a Geology Up approach to investigate which reactions could have been possible given a geochemical scenario and which would have led to these essential components.
Before biology there must have been a chemical system that created it. This then means that the origins of life can be considered as a problem of organic synthesis. However, without the constraints of geology to inform the scope of potential chemical space available it is difficult to assess the plausibility of any approach or theory. The Geology Up approach uses the rock record that still exists to create and constrain a “prebiotic toolbox” with which prebiotic chemists can work with.

The work in this thesis is primarily focused on the prebiotic chemistry of nucleic acids and as such it is constructive to understand that this approach is the one considered here. Therefore, the need for a handle on the conditions of the early Earth to construct a “prebiotic toolbox” is necessary before the discussion of any individual theories within the Geology Up approach.

1.2.2 Geology Up and The Need for the Prebiotic Toolbox
1.3 History of Earth and Conditions for Abiogenesis

Abiogenesis was a purely chemical process and so some element of this must involve the construction of more complex molecules from generationally and structurally simpler molecules. To think about the reactions that can bring about complexity it is necessary to examine the prebiotic chemist’s “toolbox”. This will consist of the potentially prebiotic feedstock molecules based on the predicted conditions of the early Earth.

Contemporary chemical synthesis makes use of complex reagents, advanced protecting group strategies as well as a plethora of solvents and sophisticated purification technologies. These handles enable the optimisation of a huge array of different transformations and enable generation of complex products in high purity and yields. To a modern chemist prebiotic chemistry may seem, in contrast, constrained to the point of futility. The scope of the conditions, potential feedstock molecules and purification opportunities are all reduced in comparison to modern chemistry. Therefore, prebiotic chemistry has plausibility guidelines built into it; all the conditions and reagents used must be consistent with what can be gleaned from geochemistry and other predictions about the early Earth.

The history of Earth (Figure 1) may provide insights into the environments available for the origins of life and how old life is. Recent research, particularly geology, has sought to constrain the upper and lower limits of important small molecules (phosphate, HCN, etc.) in order to not only glean insight into how life started on earth but to use in the search for extra-terrestrial life. Despite ample chemicals available on Earth currently, life has not been observed to form de novo from “inanimate materials”. This suggests that the conditions were different on the early Earth and had different feedstocks available. The following section involves a brief discussion on the formation of Earth and its atmosphere and the resulting “prebiotic inventory” available. Much of this thesis focuses on UV-light chemistry and accordingly an understanding of the early Earth’s atmosphere is important to constrain photochemical conditions.

UV-light has been shown to be very useful in the context of many prebiotic reactions making it unlikely to not have played some part in the origin of life, it is also important in many parts of this thesis. At the origins of life, much like today, the effect of UV-light on processes taking place on or near the surface of the Earth are largely governed by the wavelengths attenuated by the atmosphere. An understanding of the early Earth’s atmosphere is therefore crucial to inform which types of photochemistry would be possible.
1.3.1 Planetary formation

1.3.1.1 Disks, Bombardments and Accretion

Earth is believed to have been formed at the same time as other planets in the solar system around 4.5 Ga through the accretion process of protoplanetary disks and collisions with other planetesimals. After the moon-forming event late in Earth’s formation and other bombardments (dubbed the “late veneer”) the mass of Earth had reached ~80% of its eventual total mass during the Hadean era (4.6-4.0 Ga). The Hadean is characterised as a mostly molten, inhospitable and volatile time in the history of Earth with constant volcanic outgassing due to frequent asteroid collisions. It is believed to have been the time period in which the protoplanetary disk of volatile-rich material that formed the earth had solidified enough for the silicate-rich mantle to separate from the ferrous core, enabling various dating of rocks with more lithophilic (affinity for silicates) or siderophilic (affinity for iron) tendencies. The Hadean has been dubbed the geologic “dark ages” as the upheaval and tectonics has erased much evidence, its geologic record lacks well preserved sedimentary or volcanic rocks.

1.3.2 The Evolution of the Early Earth’s Atmosphere

After the Hadean and the late heavy bombardment (LHB, 4.1-3.8 Ga, a particularly violent period of asteroid impacts) the early Earth entered the Archean era (4.0-2.5 Ga) where a stable hydrosphere and atmosphere could form. Before this any asteroid of size 400-500 km or larger could, on impact, vaporise entire oceans and cause instability in the hydrosphere needing as much 1000 years for the precipitation to reform the ocean. This may have precluded the survival of any early life due to the critical role of water in many prebiotic regimes leading to the theory that life predominantly originated after the LHB (although some disagree).

Most models of the origin of life are based on aqueous prebiotic chemistry that is used to build higher complexity molecular building blocks. In addition, there is little geological evidence to suggest periods of glaciation during the Archean. It is therefore assumed that the Archean atmosphere must have been capable of retaining enough heat to keep the surface temperature above 0 °C ensuring surficial liquid water.

The lack of rare gases (Xe gives the clearest example) in comparison to cosmic abundances preclude a primary captured atmosphere from having been retained, this is almost certainly due to hydrodynamic escape of accreted hydrogen early in the Earth’s lifetime. The atmosphere must be a secondary atmosphere, one formed not by accretion during planetary formation but instead one formed from accumulation of meteoritic impacts, volcanism and other terrestrial processes.
Additionally, the young sun is modelled to have been weaker (25-30% lower luminosity 4 Ga\textsuperscript{14,15}) than it is currently as well as emitting more strongly in the middle and far UV (< 300 nm\textsuperscript{16}), commonly referred to as the “faint young sun paradox”. Indeed some models have predicted that the early Earth could have had average surficial temperatures as low as -1 °C due to its reduced thermal output\textsuperscript{17}.

The atmosphere has serious implications for the amount of UV-light hitting the surface (surface actinic flux) of any terrestrial system this chemistry may be based in due to the attenuation of incident solar flux by atmospheric gas molecules. Is it possible to arrive at a model of an atmosphere that would have kept surficial liquid water and constrain potential photochemistry? This has been thoroughly reviewed elsewhere but a brief discussion here is useful to contextualise the work in later chapters\textsuperscript{18,19}.

Sagan and Mullen when first discussing this problem found that their calculations of the surface temperatures required atmospheric greenhouseing due to Earth’s high albedo and putative atmospheric compositions. These could be aligned with surficial liquid water by the predicted presence of ammonia (NH\textsubscript{3}), albeit in low quantities, alongside molecular nitrogen (N\textsubscript{2}) and carbon dioxide (CO\textsubscript{2}) that are believed to have been the major components of the Earth’s early atmosphere\textsuperscript{17}. This would also align with early ideas of a reducing atmosphere\textsuperscript{20,21}. However several subsequent studies have argued that photochemical conversion of NH\textsubscript{3} to N\textsubscript{2} in the atmosphere would have rendered this incompatible with other models (Kasting does argue that the amount of NH\textsubscript{3} may be suitable for incorporation into prebiotic reactions)\textsuperscript{22,23}. This could have happened incredibly quickly on a geological timescale, maybe as quickly as 40 years, if the mixing ratio was less than 10\textsuperscript{-4}. This photochemical process occurs across a relatively broad photochemical band up to 230 nm and the reactions are summarised in below (Table 1). Methane (CH\textsubscript{4}) has also been proposed as a viable method to ensure adequate temperatures on the early Earth\textsuperscript{24}. A weakly-reducing atmosphere, one composed of bulk oxidized gases (N\textsubscript{2} and CO\textsubscript{2}) but containing lower levels of reducing gases (CO, H\textsubscript{2} and CH\textsubscript{4})\textsuperscript{19}, can also be reconciled from impacts during the addition of the “late veneer” a series of meteoritic impacts after the Moon forming events\textsuperscript{25}. Zahnle et al. show that this could deliver reducing potential in the form of metallic iron that was then oxidised by H\textsubscript{2}O and CO\textsubscript{2} giving H\textsubscript{2}, CO and CH\textsubscript{4}. Big impacts are needed to do anything meaningful, as lots of smaller impacts do considerably less.
1.3.2.1 Hazes and Clouds
An interesting caveat also proposed by Sagan concluded that if suitable levels of CH₄ were present, an organic haze (c.f. Titan), could have been formed that would have protected NH₃ from destruction<sup>26</sup>. This process can occur when the ratio of CH₄:CO₂ exceeds ~0.1 when a CO₂/N₂/CH₄ mixture is irradiated by UV-light, leading to CH₄ radical polymerisation into particulates<sup>27</sup>. This in turn would have an impact on the amount of UV-light hitting the surface (surface actinic flux) and reduce the amount available for photochemical reactions. Kasting showed that when the optical properties of the haze and size of the particles in the haze were taken into account that the cooling rendered by such a haze would actually have created a negative feedback loop which would have caused a powerful "antigreenhouse effect" and a cessation of CH₄ production<sup>28</sup>. Later chapters investigate UV-light photochemistry on the early Earth. Accordingly, of consideration is the influence of clouds and water depth on UV-light attenuation. The much higher flux of UV-light during the Archean would necessitate a 30 m water column to be equivalent to conditions on Earth’s surface today<sup>29</sup>. Clouds have also been shown to attenuate UV-light but the effect is hard to quantify<sup>30</sup>. A large asteroid could have vaporised entire oceans and caused long lasting cloud cover (~1000 years)<sup>10</sup>. Therefore, this atmospheric water could have attenuated much of the UV-light even for a (geological) relatively short time period.

\[ \text{NH}_3 + h\nu \rightarrow \text{NH}_2 + \text{H} \]  \hspace{1cm} \text{R1}
\[ \text{NH}_2 + \text{NH}_2 + \text{M} \rightarrow \text{N}_2\text{H}_4 + \text{M} \]  \hspace{1cm} \text{R2}
\[ \text{N}_2\text{H}_4 + \text{H} \rightarrow \text{N}_2\text{H}_3 + \text{H}_2 \]  \hspace{1cm} \text{R3}
\[ \text{N}_2\text{H}_4 + h\nu \rightarrow \text{N}_2\text{H}_3 + \text{H} \]  \hspace{1cm} \text{R4}
\[ \text{N}_2\text{H}_3 + \text{N}_2\text{H}_3 \rightarrow \text{N}_2\text{H}_4 + \text{N}_2 + \text{H}_2 \]  \hspace{1cm} \text{R5}

*Table 1* Photolysis destruction of ammonia<sup>28</sup>.

1.3.2.2 Ozone
Oxygen (O₂) absorbs UV-light (160-200 nm) and can fragment to give two O radicals which can attack another molecule of O₂ to form ozone (O₃) which instead absorbs UV-light (100-280 nm). If sufficient O₃ could be produced it would attenuate wavelengths with \( \lambda < 280 \text{ nm} \) which have been shown to be particularly harmful to biological systems. O₂/O₃ is present in appreciable quantities in the extant atmosphere so it must have emerged at some point in the evolution of the earth’s atmosphere and biologically effective ozone layer screening could have been established at 1% to 10% of present
atmospheric levels (PAL) of $O_2$. However, an abiotic production of $O_2$ presents challenges. The amount of $O_2$ from photodissociation of $H_2O$ and subsequent escape of $H$ was small\textsuperscript{12,31} which is why levels of $O_2$ during the Archean are thought to have been low ($<10^6$ PAL\textsuperscript{32}). Indeed, oxygen is mostly believed to have been a consequence of biotic processes and is signified by the Great Oxidation Event (GOE, Figure 1) a dramatic rise in oxygen levels thought to have taken place 2.4 Ga, after the emergence of cellular life, as photosynthetic cyanobacteria multiplied\textsuperscript{33,34}. This changed the atmosphere of the planet which is mirrored by the rock record (in particular banded iron formations), consequently this would have also led to the extinction of many forms of life unused to these oxidising conditions\textsuperscript{35}.

In conclusion and despite ongoing contention particularly around accurate numerical constraints\textsuperscript{19}, most models assert that a secondary atmosphere of $N_2/CO_2$ in an 80:20 mix with traces of other chemicals ($CH_4/NH_3$) was the dominant atmosphere during the Archean\textsuperscript{36,37}. This would have effectively attenuated wavelengths $\lambda < 200$ nm giving an available window for photochemistry occurring upon absorbing UV-light with wavelengths between 200 nm and 300 nm\textsuperscript{38}.

1.3.3 The Prebiotic Chemist’s Inventory

With a model for understanding the early Earth’s formation by accretion and subsequent atmosphere, a “prebiotic toolbox” can now be constructed. As discussed, the early Earth’s atmosphere during the Archean was probably low in $O_2$ and so a mildly reducing atmosphere ($N_2$, $CO_2$, $CO$, $CH_4$) would have predominated. Additionally, the temperature is believed to have been in the range 0-50 °C and pH of the hydrosphere believed to have been in the range pH 5-7\textsuperscript{10,39}. This has led to the conclusion that simple, reduced carbon and nitrogen compounds are believed to have been present in the highest quantities (Figure 3 shows a selection most relevant to this work). A short overview of several compounds highly relevant to work presented in later chapters are explored here.

1.3.3.1 Prebiotic Availability of Phosphate

Phosphate’s ($PO_4^{3-}$) necessity at the origin of life has been debated\textsuperscript{40} but its’ ubiquity in modern biology (nucleotides, phospholipids, energy storage (ATP)) and many prebiotic syntheses are highly indicative of its presence\textsuperscript{41–44}. It was believed to have been unavailable in the requisite high millimolar to molar concentrations due to its tendency to form insoluble minerals in the presence of calcium (to form apatite) and also with acidic $Fe^{3+}/Al^{3+}$. Recent work has revealed the possibility of carbonate-rich ($CO_3^{2-}$) lakes instead sequestering calcium in carbonate minerals allowing phosphate to accumulate
1.3.3.2 Simple Prebiotic Feedstocks

Most of the chemical matter available on the early Earth would have been from its initial accretion but some material was certainly delivered by subsequent meteoritic impacts. For example, hydrogen cyanide (HCN), predominantly formed from the reactions of atmospheres containing C, H, O and N by photochemistry, lightning, meteor impacts, or volcanism\(^2,47,48\), has also been proposed to survive impact if delivered meteortically\(^49\). HCN chemistry is the basis for most of the work presented here and has shown to be a vital synthon for a great deal of prebiotic chemistry. Once generated atmospherically there must have been a sequestration mechanism to enable aqueous chemistry to occur. Ferrocyanide (Fe(CN)\(_6\)\(^{4-}\)) has been proposed as a prebiotically plausible method due to ferrous iron’s (Fe\(^{2+}\)) affinity for cyanide\(^47,50\). Formation of ferrocyanide (and its simple salts with Mg\(^{2+}\)/Na\(^+\) and Ca\(^{2+}\)/K\(^+\)) also protects it from destruction by UV-light as its relative insolubility would have caused it sediment on lake floors\(^2\). This would then be released by thermal metamorphosis (from geochemical fluctuations) when lake beds dried out to yield concentrated cyanide salts (NaCN/KCN) although the temperatures required are high (up to 700 °C)\(^2\). Other geochemically constrained species from minerals have been thoroughly reviewed by Hazen and others and some relevant ones are indicated here (Figure 3)\(^51\).

\[ \text{Figure 3} \text{ Prebiotically “plausible” chemical species, ones shown here are a representative sample and many more are probable.} \]
Pinto et al. have reported that formaldehyde 1 (the first reduction product of HCN) formation can also be produced in the atmosphere by the photolysis of CO$_2$ and H$_2$O when the atmospheric levels of CO$_2$ are high$^{31}$. While water is assumed to have been the major prebiotic solvent others have also suggested the first hydration product of HCN, formamide (HCONH$_2$), could have been formed in adequate quantities$^{52}$.

Wöhler demonstrated, in what is arguably the first organic chemistry described (in 1828), that a biological molecule, urea 2, can be synthesized from an inorganic compound, ammonium cyanate (Scheme 1a)$^{53}$. Subsequently urea 2 has been implicated in many prebiotic phosphorylation reactions among others (vide infra, 2.2.10).

![Scheme 1](image)

**Scheme 1** a) Pioneering work of Wohler’s urea 2 synthesis$^{53}$; b) Overview of the Miller-Urey reactions and some of products and their eventual uses elucidated by others$^{54}$.

In another early and pioneering prebiotic experiment, Miller and Urey performed their now iconic spark discharge reaction on a mixture of CH$_4$, NH$_3$, H$_2$ and “steaming” H$_2$O to simulate a lightning strike initiating a chemical reaction$^{55}$. Analysis revealed HCN and a mixture of Strecker products (vide infra, 1.4.2.2) including the amino acids: glycine Gly, alanine Ala and aspartic acid Asp (Scheme 1b) and subsequent reanalysis of closely related experiments has also yielded a broader range of amino acids (including methionine Met), although many are non-proteinogenic$^{56,57}$. 
1.3.3.3 Cyanosulfidic Protometabolism

Building on the work of Miller and Urey, Patel et al. demonstrated (Scheme 2) they could selectively generate 12 proteinogenic amino acids, as well as the lipid precursor glycerol and most importantly simple aldehydic sugars glycolaldehyde 3 and glyceraldehyde 4. Hydrogen Cyanide was the only feedstock molecule and the key to the synthesis is Killiani-Fischer type homologation of HCN involving multiple photochemical reductions of nitrile (C≡N) bonds using stoichiometric hydrogen sulfide (H₂S) as a reducing agent and UV-light.

This process has also been shown to work with bisulfite (HSO₃⁻)⁵⁸. The use of cyanide as opposed to formaldehyde 1 avoids the problem of utilizing umpolung chemistry of formaldehyde to form glycolaldehyde 3 which before relied on the formose reaction (vide infra, 1.5.2.1)⁵⁹. Instead, it makes use of the cyanohydrin glycolonitrile 6 and uses it to fix carbon as a non-volatile and stable form. The other advantage of this process is that it releases ammonia into the system, alleviating concerns that nitrogenous compounds have a weaker basis in the "prebiotic toolbox" (vide supra,1.3.2). Note the α-
deoxygenation of glycolaldehyde 3 as an alternative fate to give acetaldehyde 8 the precursor to alanine Ala and threonine Thr via cyanide homologation to lactaldehyde. A similar process from glyceraldehyde 4 to give acetone 10 occurs and allowed them to obtain valine Val and leucine Leu. Glycerol 5, a key component in lipid synthesis, is also produced from the reduction of dihydroxyacetone (DHA) 9 the equilibration product in water of 4.

A common problem with prebiotic reactions is the incompatibility of species reactive enough to be useful also undergoing reactions with enough selectivity to avoid polymerisation, destruction or derivatization by other species. A clear example of this is the brown tar observed when the autocatalytic homologation of formaldehyde 1 is left unchecked (vide infra, 1.5.2.1). The most striking aspect of Patel et al.’s work is the relatively tight distribution of products, no oligomerization to intractable tars, no egregious side reactions to squander material. 3 and 4 have been shown to be key synthons in the synthesis of nucleic acids and their use is described in detail (vide infra, 1.5.5.2).

Recently work by Rimmer et al. has used rate constants derived from the “light” (reactions needing UV-light) and “dark” (no UV-light requirement) reactions of this cyanosulfidic protometabolism to ascertain an “abiogenesis zone” defined as: “the zone in which a yield of 50% for the photochemical products is obtained, adopting the current UV activity as representative of the UV activity during the stellar lifetime and assuming a young Earth atmosphere”37. This gives the ability to characterise exoplanets orbiting stars active enough to yield a surface actinic flux capable of pushing these reactions over this threshold which is of great interest in detection of potential extra-terrestrial life.

With an understanding of the geochemical and atmospheric effects on the early Earth and the subsequent plausibly available synthons an exploration of prebiotic chemistry can now be discussed.
1.4 Perspectives in prebiotic chemistry

Due to the interdisciplinary nature of the origins of life as a field of research there are many overlapping and conflicting theories that seek to explain the origins of life. Here several of them will be discussed as they are relevant to the work in later chapters. The theories discussed here are based on the idea that life originated on Earth and do not include the ideas of Panspermia, the delivery of the preconstructed components essential to life or life itself from other planets to Earth. Many theories imply that one particular type of biomolecule was first or most important and therefore this section finishes with a brief discussion on new ideas of a more holistic “systems chemistry” approach incorporating several previously siloed theories.

1.4.1 Metabolism in prebiotic chemistry

Metabolism is the process by which chemical compounds are broken down (catabolism, Krebs cycle (TCA), glycolysis) and built back up (anabolism, Calvin cycle, r-TCA cycle). This provides chemical energy for life’s need to sustain an out of equilibrium state. Extant biology harnesses chemical energy to drive reactions against thermodynamic control and uses chemical energy to create higher order metabolites vital for its function.

A “Metabolism First” scenario envisages that non-enzymatic chemical cycles and pathways would have been established first; providing the full suite of necessary biomolecules before any other part of a minimal protocell (Figure 2) subsequently followed by fusion into proto-cellular structures. This idea of a “Metabolism First” scenario for the origin of life remains attractive to this day. If a purely chemical system could be established that could provide all the necessary compounds that life might need then the inference is that “life” would naturally develop out of this. Therefore, developing chemical reaction networks deliver these required compounds is the focus of research for this scenario.

1.4.1.1 Chemiosmosis

Much of this work is inspired by observations of chemoautotrophs, bacteria that harness chemical energy instead of light to fix CO₂ and turn it into higher order metabolites. They are often found in deep sea alkaline hydrothermal vents (i.e. Lost City in the Mid-Atlantic Ridge). Wächterhäuser was one of the first to suggest that hydrothermal vents could have been a suitable place for the origin of life. He argued that chemoautotrophs at the origin of life must have first solved the problem of carbon fixation⁶⁰. The “Iron-Sulfur world” was his description of how this process could have occurred⁶¹. He postulated that iron sulfide (FeS) could have been oxidised by hydrogen sulfide (H₂S) to pyrite (FeS₂) thereby reducing CO₂ and CO (from the atmosphere) to simple organic molecules.
formaldehyde 1 and formate 14. The “pioneer organism”, said to be “a composite structure with an inorganic substructure and an organic superstructure” would have to rely on simple ligand-accelerated auto-catalytic cycles to drive its function and survival\textsuperscript{62}. Despite popularity, a lack of convincing experimental proof for the “Iron-Sulfur” world and self-sustaining autocatalytic cycles on the whole has affected its widespread adoption\textsuperscript{63,64}. Autocatalysis is proposed as a requirement for an emergent metabolism but will not be covered here\textsuperscript{65}.

\[ \text{Scheme 3 Wächterhäuser's proposed reduction reaction to generate formaldehyde 1 and formate 14 via the oxidation of FeS to pyrite (FeS}_2\text{)}^{60}. \]

Interestingly, recent work has however elucidated a prebiotic synthesis of Fe-S clusters. Fe-S clusters have been proposed as ancient cofactors due to their continued incorporation into extant biology (ferredoxins, NADH dehydrogenase etc.), leading to convincing arguments that Fe-S clusters may be privileged or pre-determined species. Bonfio et al. demonstrated that UV-light (\textit{vide infra}, 1.6.1) can drive their synthesis implying that a deep sea origin may be less viable if taking into account water’s attenuation of UV-light as a function of depth\textsuperscript{66}.

Deep sea alkaline hydrothermal vents have also been implicated in pH-gradient-driven (Chemiosmosis) scenarios that align with a “Metabolism first” approach. This process of Chemiosmosis, vital in extant life (e.g. ATP-synthase is driven by a proton gradient), has been suggested as a solution to the problem of sustaining a far from equilibrium system akin to extant life.

Despite many contributions to the literature from the groups of Russe\textsuperscript{l}\textsuperscript{67}, Martin\textsuperscript{68}, Lane\textsuperscript{69} and others\textsuperscript{70} experimental evidence for pH gradients being harnessed abiotically for metabolic reactions remains elusive as well as any higher-order metabolites made in such a way. Jackson has produced a thorough analysis of many of these claims\textsuperscript{71}. He highlights a two-fold problem with this concept. Firstly, no experimental evidence for the correct type of membranes required for such a process have been discovered; secondly,
abiotically harnessed chemiosmotic pressure has not been demonstrated to drive metabolic reactions.

The membranes required for such an origin scenario are not found at the site of the Lost City\textsuperscript{72,73}. They are proposed to be strong, thin membranes allowing a pH gradient to exist across it while also providing structural support. Extant biology makes use of transmembrane proteins that fully span the membrane to facilitate the movement of protons; no such transmembrane structures have been reported.

Molecular “machines” for pH driven chemiosmosis are proposed to be the driving force for metabolic reactions most commonly identified as the reduction CO\textsubscript{2} to 14 by H\textsubscript{2} (in a similar way to \textit{Wächterhäuser}). However, no such “machines” able to couple this process have been demonstrated and such molecular complexity seems unlikely to have arisen under such conditions.

There is a broad trend in “metabolism first” origins of life research to attempt replication of enzymatic reactions abiotically. The following section explores two directions (Scheme 4) that research has moved in towards the central metabolic process of the citric acid cycle.

\begin{center}
\textbf{Scheme 4 Two Routes to central metabolism} a) Wood-Ljungdahl pathway followed by the reverse citric acid (r-TCA) cycle  
\textit{b) Triose Glycolysis followed by the citric acid (TCA) cycle}
\end{center}

\textbf{1.4.1.2 Wood-Ljungdahl Pathway}

The Wood-Ljungdahl pathway (reverse acetyl-CoA pathway) is utilised by bacteria and archaea (acetogens and methanogens) to fix CO\textsubscript{2} using H\textsubscript{2} as the electron source followed by subsequent incorporation of the acetate 15 and pyruvate 16 generated into biosynthesis as Acetyl CoA\textsuperscript{74}. LUCA is proposed to have utilised this process generating interest in a prebiotic equivalent\textsuperscript{1}.
Varma *et al*. have recently demonstrated (Scheme 5a) that native iron (Fe$^0$) powder can generate several compounds in the Wood-Ljungdahl pathway from CO$_2$.$^{75}$ Previous work has demonstrated that this was possible$^{77}$ but this was the first time it had been demonstrated in an arguably plausibly prebiotic way. Formate 14 (0.41 mM), acetate 15 (0.18 mM), and methanol (0.12 mM) were all observed; of special interest pyruvate 16 (0.03 mM) was only observed in low amounts despite its importance in metabolism. Powdered “native iron” (Fe$^0$) is produced in very small amounts by serpentinization (exothermic, H$_2$-generating oxidation of ultramafic rock), and transiently in the mantle so a reliable source for this process is up for debate.

Recently, this work was extended by a team lead by Martin, Tüysüz and Moran they have shown (Scheme 5b) that three iron minerals (greigite (Fe$_3$S$_4$), magnetite (Fe$_3$O$_4$) and awaruite (Ni$_3$Fe)) that have been found in hydrothermal vents can assist the formation of 14 from gaseous CO$_2$ and H$_2$.$^{76}$ A simulated prebiotic atmosphere (80/20 H$_2$/CO$_2$, 25 bar) at high temperatures (up to 100 °C) under alkaline conditions and metal ions (1 mmol) generated a product mixture containing: 14 (332 mM), 15 (0.56 mM), 16 (0.01 mM), methanol (0.12 mM) and methane (CH$_4$). External energy input (in the form of chemiosmosis) not being required for the fixation of CO$_2$ to organic carbon species is touted as the major advantage and allows a move away from chemiosmosis as the driving force of prebiotic metabolism, however there is still a need for energy in the form

---

\[ \text{Scheme 5 a) Varma et al. utilise Fe}^0 \text{ to form a variety of organic compounds from CO}_2^{75} \text{ b) follow up by Preiner et al. using minerals from serpentinization to achieve similar but higher yielding reactivity}^{76} \]
high temperatures. The iron minerals used are more plausible, forming naturally from serpentinization in hydrothermal systems. The addition of hydrogen as a reductant is clearly beneficial for the initial reduction of CO\textsubscript{2} to 14 (0.41 mM vs 332 mM). However, this route lacks the active catalyst surface that Fe\textsuperscript{0} provided in the earlier work potentially causing the low yields of 16, a key part of extant metabolism.

Scheme 6 a) Coggins et al.’s prebiotic synthesis\textsuperscript{78} of intermediates in triose glycolysis (red box) b) Enzymatic triose glycolysis pathway leading to the citric acid cycle, where intermediates are shared with Coggins et al.’s synthesis they are indicated (dashed arrows); DHA-P 23 also leads into the lipid pathway.
1.4.1.3 Ancient Glycolysis

The citric acid cycle consumes acetyl-CoA from glycolysis to provide stable chemical energy in the form of ATP (using electron-transport chain) making it a vital source of energy for extant life. Phosphate-containing molecules such as ATP are ubiquitous as energy storage sources in all domains of life. Metabolic pathways in extant life centre on glycolysis as one of the most vital metabolic cycles, therefore finding a prebiotically analogous route for these reactions is of clear interest in understanding the role of early metabolism.

Recently, Coggins et al. have shown a route to phosphoenol pyruvate 22, the highest-energy phosphate in extant organisms and a vital intermediate in glycolysis and other metabolic pathways\(^7\). They showed (Scheme 6a; red box) that starting with the phosphorylation (pyrophosphate or diamidophosphate) of glycolaldehyde 3, a simple 2 carbon sugar and the first product of the contentious formose reaction (\textit{vide infra}, 1.5.2.1), they could produce key triose-glycolysis intermediates with prebiotic reactions. Glyceraldehyde-2-phosphate 19 was delivered by reaction of Gly-2-P 17 with formaldehyde 1 which then underwent E1cB elimination to phosphoenolpyruvaldehyde 21 in good yield (2 steps, 49%). Enzyme catalysis is required for this step in modern biology as glyceraldehyde 2-phosphate 18 is de-activated to elimination with respect to glyc-2-phosphate 19. To explore a scenario-agnostic approach their subsequent oxidation step to deliver 22 was demonstrated with range of oxidants: ferricyanide, chlorites and MnO\(_2\); many with quantitative conversions. This demonstrates that a linear route to important metabolic compounds is possible with prebiotic chemistry.

1.4.1.4 Tricarboxylic Acid Cycle (TCA)

Work has been done on replicating the citric acid cycle non-enzymatically. Initial work by Waddell showed that high temperatures (140 °C, Scheme 7; blue arrow) and light (visible and UV, both abundant on the early Earth, Scheme 7; pink arrows) can produce a variety of compounds in the TCA cycle\(^7\) and also some via the r-TCA cycle. Interestingly, many reactions lead to succinate 30 formation potentially suggesting it as a thermodynamic minimum for the cycle. They also observed maleic acid 35 the \textit{cis} isomer of fumarate 31 indicating that this may be occurring via a radical process.

Later work by Ralser and co-workers has shown (Scheme 7; black arrows), despite some contentious discussion\(^8\), that the full set of prebiotic TCA cycle products (CoA bonded intermediates omitted) were generated in a simulated “Archean-Sediment metal mixture” consisting of ammonium persulfate ((NH\(_4\))\(_2\)S\(_2\)O\(_8\)) which was activated by ferrous sulfide (FeS)\(^8\) to generate sulfate radicals. They report a highly selective (very few non-TCA
compounds formed), high yielding (> 90% network recovery for carbon) pathway of 24 reactions that also include parts of the glyoxylate cycle (or glyoxylate shunt, *vide infra* 1.4.1.5). More work is need on either harnessing this process or its role in regulating, via catabolism, other potential autocatalytic reactions are needed for a better understanding of the role the TCA cycle could have played in an origins scenario.

Scheme 7 Overview of prebiotic pathways in the oxidative TCA cycle. Sulfate radical enabled steps (black arrows) for the oxidative TCA cycle and glyoxylate shunt pathway. Photochemical (pink arrows) and thermal (blue arrow) conversions of TCA intermediates.
1.4.1.5 Reverse Tricarboxylic Acid Cycle (r-TCA)

Current thinking posits that an ancient Wood-Ljungdahl pathway could feed into the reverse citric acid (TCA) cycle. This would allow a continuous stream of fixed carbon to be available for anabolic processes. An attractive option, much work has sought to elucidate role of the r-TCA cycle at the origins of life. Zhang and Martin investigated (presumably building on the work of Waddell) potential photochemistry pathways of the r-TCA cycle using UV-light (200-410 nm) and colloidal ZnS. They observed the photochemical reductions of oxaloacetate 33 to malate 32 (70%, quantum yield 3.9%), fumarate 31 to succinate 30 (100%, quantum yield 1.9%) as well as the carboxylation of α-ketoglutarate 29 to oxalosuccinate 28 albeit in very low yield (2.5%). Presumably irradiation of 32 under these conditions lead to 35 as indicated by the work of Wadell but this was not reported; it is a dehydration and perhaps not relevant to their discussion of reductions. Moran and co-workers have shown that products from the r-TCA cycle are accessible when incubated with Fe⁰, Zn²⁺, Cr³⁺ and CO₂ in highly acidic (1 M HCl) conditions at elevated temperatures (< 140 °C). Interestingly none of the C-C bond formation steps were accessible and the reversible hydration of 26 to 25 also proved difficult, only occurring with the inclusion of Cr³⁺ and instead hydrating to 27 with Zn²⁺. All 3 metals and prebiotically questionable micelles were needed to drive multistep sequences of 33 to 30 (41% of remaining material) and 28 to 25 (2% of remaining material). The prebiotic plausibility of native iron (Fe⁰) is up for discussion. However, Muchowska et al. have updated this work by demonstrating that ferrous (Fe²⁺/Fe³⁺) water (pH ~ 4.5 5.5) is also capable of performing some of these same transformations of the r-TCA cycle (and some of the TCA cycle). They employ aldol reactions of glyoxylate 34 and pyruvate 16 or α-ketogluarate 29 to generate, for pyruvate 16, maloyl formate 41 which then undergoes several different reactions. Key amongst them are those attributed to oxidative decarboxylation allowing access to malate 32, succinate 30 and fumarate 31. These are more akin to reactions of the oxidative TCA cycle but the overall pathway, not cycle, is in the reverse direction. Once again access to citrate 25 proves challenging, suggesting it was only through later fusion of glycolysis that 25 became a major part of this pathway. The amount of material remaining dramatically reduces with time with much of the “fixed carbon” being recycled to CO₂ via decarboxylative processes. It also reveals that malonate 36, an “off cycle” product, is the major “fixed carbon” species after 48 hours in this process. This raises questions about the utility of this pathway if no solution to the low yields are suggested.
Malonate 36 was previously implicated in work from Springsteen et al., where a mixture of pyruvate 16 and glyoxylate 34 displayed similarities to a metabolic cycle when alternating additions of 34 and hydrogen peroxide (H₂O₂) were added to regenerate 36 from oxaloacetate 33. Building on this, they recently reported a non-enzymatic α-keto acid alternative to the r-TCA cycle. Species without an α-keto acid moiety in the r-TCA cycle are weakly acid at the α-carbon making them less reactive, particularly towards enolate reactions. Accordingly, they show that the reaction between 16 and 34, 21 h at 50 °C, pH 7 in 0.5 M phosphate buffer, produces the α-keto acid maloyl formate 41 (corresponding to r-TCA species malate 32). Maloyl formate 41 then reacts further; remarkably 34 is able to act as both carbon source and reductant to deliver fumaroyl formate 40, α-ketoglutarate 29, isocitroyl formate 39 and aconityl formate 38 with a total one-pot yield of 57%. Oxaloacetate 33 was not formed in this sequence but was demonstrated to link the cycle as dimerization gave citroyl formate 37 while reduction by 34 gave once again 41.

Muchowska et al. asserted that ferrous water was required for the synthesis pathway to α-ketoglutarate 29. However, when Springsteen et al repeated this reaction under their conditions (pH 4.5 and 70 °C) but without iron the reaction of pyruvate 16 and glyoxylate 34 still proceeds to 29. This suggests that the reductive steps are actually reliant on glyoxylate instead; iron is only required for the oxidative decarboxylations of the α-keto acids to their r-TCA equivalents, something that can be achieved with metals or H₂O₂. This may also even be undesirable due to their suppressed reactivity and tendency of the iron network to prevent product accumulation.
Scheme 8 Overview of current progress toward a prebiotic r-TCA pathway: Trent Stubbs et al. α-keto acid variant (inner blue circle) of the r-TCA cycle (outer circle) with the parts of the r-TCA that correspond to the α-keto acid’s oxidative decarboxylation also indicated (orange arrows). The work of Moran and co-workers (green arrows) and steps accessible by UV-light (pink arrows) are also highlighted. Reagents and conditions: (i) Fe⁰/Cr³⁺/Zn²⁺, 140 °C 1M HCl; (ii) Fe²⁺/Fe³⁺, pH 4.5 70 °C; (iii) H₂O₂.

Metabolism is a key part of extant life and presumably early life. How the metabolic pathways came to exist is still unknown. It seems plausible that the cycles seen in biology today were improved upon by incorporation of enzymatic control and catalysis but this does not preclude the existence of an alternate set of reactions that were supplanted at a later date (although this is harder to prove). Going forward the field looks to focus on elucidating C-C bond forming reactions of the r-TCA and should also flesh out the mechanism of sustaining metabolism be that with geochemically rooted autocatalytic anabolic reactions (Wood-Ljungdahl pathway leading to r-TCA type cycle) and catabolic back reactions to ensure an out of equilibrium state is maintained. Also of interest is the provision of basic feedstocks by metabolic pathways for other domains of early life: feeding protein synthesis via amino acid precursors and the synthesis of cofactors.
1.4.2 Peptides in prebiotic chemistry

Peptides are crucial in extant biology for a host of applications. Mainly, they provide catalytic function normally via their folded tertiary enzyme structure. It is therefore important to understand the role of peptides at the origins of life. This involves understanding the prebiotic synthesis of amino acids, the reasons behind the canonical amino acids in the extant genetic code, amino acid oligomerisation and the potential interactions between RNA and peptides. Some have advocated for peptides playing the leading role at the origins of life with catalytic function of peptides providing all the necessary components for the development of all the necessary parts of protocellular function (Figure 2). This is known as the “Protein world” theory. This section will briefly explore this theory before reviewing the attempts towards polypeptides after first contextualising the synthesis of amino acids in a prebiotic setting.

1.4.2.1 Protein World Theory

The “Protein world” theory is the theory that peptides or proteins predated genetic polymers at the origin of life. Their ability to act as catalysts would have allowed a full suite of protometabolic functions to be established supposedly without the need for genetics. The key steps for this then follow: synthesis of a suite of amino acids (AAs) with sufficient chemical diversity; oligomerisation of AAs to peptides; random non-coded peptides display emergent beneficial properties and self-amplify themselves by an unknown mechanism.

Ikehara has proposed that a peptide system assembled from only the amino acids glycine, alanine, aspartic acid, and valine (GADV) could have spontaneously formed and displayed emergent function simply by virtue of existence. Kurland also writes in support of this but both fail to reconcile their considerable assumptions with empirical proof. They assume: large pools of AAs, emergent catalytic function (but based on a reduced alphabet: GADV) and simultaneous protection from hydrolysis. These reports do little to advance the field and work should instead focus on solving the prebiotic synthesis of amino acids and efficient peptide formation.

1.4.2.2 Amino Acid Synthesis

One of the key reactions in the prebiotic synthesis of amino acids is the Strecker reaction (Scheme 9a). This now well understood reaction involves the attack of ammonia on an aldehyde followed by trapping of the resultant iminium ion with cyanide. Subsequent hydrolysis of the nitrile leads to amino acids. The functionalisation of the aldehyde defines the side chain for the eventual amino acid. It is also now thought that amino acid
synthesis from Miller-Urey type processes (vide supra, 1.3.3.2) probably involved the Strecker reaction.

Scheme 9 Convergent routes to amino acid precursors (AAs): a) mechanism of the Strecker reaction (pink); b) mechanism of the Bucherer-Bergs reaction (green) c) Phosphoro-Strecker reaction of aldehydes and divergent routes from resulting phosphoaminonitriles d) sluggish Phosphoro-Strecker reaction with ketones ($R \neq H$) to give non-natural di-substituted amino acids.
A potentially improved method is the Bucherer-Berger reaction (Scheme 9b) where CO₂ aids the conversion of aminonitriles to amino acids by reacting with the N-terminus amine to give a carbamic acid. Subsequent cyclization to an N-carboxyanhydride (NCA) and ring opening to the isocyanate allows for hydantoin formation. This can undergo a more rapid two step hydrolysis via the N-carbamoylated amino acid to give an amino acid⁹¹.

The Phosphoro-Strecker reaction has also been recently explored (Scheme 9c) in which NH₃ (pKₐH = 9.2) is replaced with diamidophosphate (DAP) allowing efficient conversion of aldehydes to phosphoroaminonitriles at neutral pH rather than basic pH (9-10) required for efficient conversion with NH₃, although the synthesis of DAP does require high pH⁹². Importantly, the reaction is highly selective for aldehydes over ketones (Scheme 9d) and this sluggish reactivity suppresses the formation of α,α-disubstituted amino acids and the non-biological peptides that they would become. Aldehydic phosphoroaminonitriles can be thiolysed to thioamides in high yields, undergo acid hydrolysis to their respective aminonitriles and undergo alkaline hydrolysis to N-phosphoro amino acids.

Patel et al. have demonstrated that as part of their cyano-sulfidic protometabolism (Scheme 2) they can form many amino acids as eventual products of cyanide homologation⁴⁷,⁵⁸. They report the synthesis of 12 amino acids with aminonitriles once again the key intermediates and emphasise a tight product distribution. Their stepwise sequential sequence of Strecker reactions coupled with reductions of nitriles with H₂S or bisulfite delivers amino acids from HCN along with precursors to both nucleic acids and lipids revealing a system that is compatible with 3 of the 4 parts of the model protocell (Figure 2).

Recent work from both Moran and Springsteen and Krishnamurthy also found that the core metabolites of an α-keto acid variant of the r-TCA cycle (vide supra, 1.4.1.5) could be trapped by amines (or hydroxylamine for Moran) to also give amino acids⁸⁵,⁸⁷. Springsteen and Krishnamurthy found (Scheme 10b) that when pyruvate 16, α-ketoglutrate 29 and glycine Gly, as the amine source, were incubated at 80 °C in an aqueous solution (pH 5) containing potassium aluminium sulfate (AlK(SO₄)₂) they gave respectively alanine Ala (19%, 24 h) and glutamate Glu (44%, 4 h)⁸⁷. Moran and co-workers had previously shown (Scheme 10a) similar reactivity but necessitated the use of metallic iron (Fe⁰) and hydroxylamine to convert glyoxylate 34 to Gly (63%); pyruvate 16 to Ala (2%); oxaloacetate 33 to aspartic acid Asp (17%) and 29 to Glu (9%)⁸⁵. This also shows that fundamentally interconnected precursors can play an important role in prebiotic chemistry.
Scheme 10 Reductive aminations of α-keto acids to amino acids: a) Recent work by Moran and co-workers utilising hydroxylamine (NH$_2$OH) and metallic iron (Fe$^0$) for the conversion of α-keto acids to amino acids$^{85}$; b) Trent-Stubbs et al.’s amino acid synthesis also from α-keto acids utilising glycine as the amine source$^{87}$.

One notable absence in the work discussed is that of cysteine Cys, which provides the majority of the sulfur in extant biology. Cysteine is not able to be formed from any of the aforementioned processes as the aldehyde (β-mercaptoacetaldehyde, BMA) required for Strecker reactions is not amenable to this process. It indeed undergoes reaction with ammonia and cyanide to give α-aminonitrile 42 but as β-aminothiols are highly reactive and 42 is both an aminonitrile and a β-aminothiol it forms intractable oligomerized products (polythiazolines). This has led to Cys being thought of as a later invention despite its implication in important parts of extant biology (Fe-S clusters, CoA and other and prebiotic chemistry (vide infra, 1.6.1).

Recently, work in the Powner group has revealed that Cys may be a secondary product of serine Ser instead$^{93}$. BMA’s close analogue glycolaldehyde 3 (implicated in nucleic acid synthesis, vide infra, 1.5.5.2) undergoes the Strecker reaction in excellent yield (>90%) to give serine-aminonitrile Ser-CN (Scheme 11).
This work demonstrated that serine nitrile Ser-CN can be simultaneously N- and O-acetylated (91% yield) (or O-phosphorylated) to 43 and subsequent incubation at pH 8 (4 days, RT) and loss of thioacetate leads to N-Acetyl-dehydrolalanine 44 in high yield (85%). Dehydrolalanine (Dha) is also a common peptide modification in bio-conjugation chemistry (see work by Bernades and Davis⁹⁴) but non-acetylated Dha-CN was synthesised by Eschenmoser and was found to be highly unstable⁹⁵. This can then be trapped with H₂S to give N-Ac-cysteine aminonitrile 48 which can be elaborated into amino amide 46 and thioamide 45. This work could plausibly be extended to harness the reactivity of Dha-CN as a synthetic handle and other hard to synthesise AAs could be made this way such as tryptophan (recently observed in the deep ocean⁹⁶) which is formally the addition of indole to Dha-CN followed by hydrolysis.

**1.4.2.3 Making Peptide Bonds**

Once a pool of amino acids (or their precursors) has been synthesised the next step is the construction of peptides that are then able to fold into 3-dimensional active structures that infer emergent properties such as catalysis in extant biology. The prebiotic oligomerisation of amino acids into peptides is therefore critical to solve en route to understanding the emergence of functional peptides involvement at the origins of life. Many different routes have been proposed for this process; overall it involves the formation of an amide bond in a condensation reaction. The aqueous nature of nearly all origins scenarios therefore appear to work against any reaction that produces water
necessitating methods to circumvent or overcome this thermodynamic barrier. Some of the simplest methods involve simply heating amino acids (to extreme temperatures)\(^97\) or wet-dry cycling with the addition of salts\(^98,99\) or minerals\(^100\). However, many of these methods are unselective (many form branched chains or DKPs, \textit{vide infra} 1.4.2.2) and/or low yielding and so do little in the way of persuading that they could be used to assemble long oligomers.

An alternate methodology has been the use of condensing agents to activate the AA’s \(N\)-terminus for \(C-N\) direction chain growth, opposite to the way biology operates (\(N-C\) direction). A multitude of different species have been proposed and covering them all is beyond the scope of this chapter (see recent review by Leman and co-workers for a comprehensive overview\(^101\)). Some of the most notable: \(\text{CO}_2\), \(\text{COS}\)^102 and \(\text{CS}_2\)^103. proceed in a similar manner to the Bucherer-Bergs reaction in that they react with the AA’s free amine. \(\text{CO}_2\) is believed to have been a major part of the early Earth’s atmosphere (\textit{vide supra\textsuperscript{,} 1.3.2}) and \(\text{COS}/\text{CS}_2\) are products of volcanic outgassing. These activations result in cyclization to an \(N\)-carboxyanhydride (NCA), a versatile and well-explored intermediate\(^104\) (Scheme 12). This can undergo several fates (Scheme 13a): hydrolysis back to the AA, thiolysis to the thioacid\(^105\), phosphorylation to an aminoacyl phosphate\(^106\)/adenylates\(^107\) or, most importantly, trapping with another AA’s amine group to form an amide bond and extend the incoming AA by one monomer\(^102,103\).

\[\text{AA} + \text{COS} \rightarrow \text{S} - \text{C} - \text{N} - \text{H} - \text{A} - \text{CO} \rightarrow \text{NCA}\]

\[\text{AA} + \text{CO}_2 \rightarrow \text{C} - \text{N} - \text{C} - \text{X} - \text{H} - \text{A} - \text{CO} \rightarrow \text{NCA}\]

\[\text{AA} + \text{CS}_2 \rightarrow \text{S} - \text{C} - \text{N} - \text{H} - \text{A} - \text{S} - \text{O} \rightarrow \text{NCA}\]

\textit{Scheme 12} Chemical activation pathways of AAs to \(N\)-carboxyanhydrides (NCAs) or 2-thiono-5-oxazolidones with \(\text{CO}_2\), \(\text{COS}\) or \(\text{CS}_2\).
Low yields, lack of prebiotic availability of activating agents and side reactions of both the products and the AAs side chains have hamstrung the widespread acceptance of NCAs as the origin of peptide formation. The most deleterious side reactions that occur, for a dipeptide (or longer), is formation of highly stable diketopiperazines (Scheme 13b). Uncapped N-termini can attack acid/amide bonds in the next residue causing cyclization and loss of 2 monomer units from any growing chain. NCAs avoid this step during the synthesis of dipeptides but are far from immune unless the resulting dipeptide is intercepted immediately after formation.

![Scheme 13](image)

Scheme 13 a) NCAs as a central reactive intermediate in peptide synthesis; b) dipeptide “bite-back” to form diketopiperazine (DKP)

### 1.4.2.4 Aminonitriles as Amino Acid Analogues in Peptide Synthesis

To avoid the problem of DKP formation, recent work (Scheme 14a) has sought to exploit the nascent reactivity of α-aminonitriles (AA-CN)\textsuperscript{108}. These impart greater reactivity, over AA, at a lower pH due to the amine’s pKa \(= 5.4\) for AA-CN. They show near quantitative ferricyanide-mediated acetylation of AA-CN with thioacetate\textsuperscript{109} to yield \(N\)-acetyl-aminonitriles (Ac-AA-CN), \(N\)-acetylation is a common pathway in extant biology. This not only avoids DKP formation (Scheme 13b), via a blocked N-terminus, but also activates the nitrile moiety. Activated \(N\)-acetyl-aminonitriles can then undergo an efficient thiolysis to thioamides (quantitative conversion, but not high yielding for AA-CN or other nitriles like acetonitrile) followed by hydrolysis to its corresponding thioacid (Ac-AA-SH). This hydrolysis requires strong base and heating for an extended period of time but it is
nonetheless high yielding. The thioacid was then susceptible to biomimetic (N-C direction) chain growth. A wide variety of prebiotically plausible activating agents (ferricyanide, cyanoacetylene 12, CuCl_2) were able to stoichiometrically activate the thioacid which is rapidly trapped by the next aminonitrile forming an amide bond. All 20 proteinogenic α-aminonitriles were tolerated as substrates (with activated N-Ac-Gly-SH) and the method is suitable for making oligomers, indeed Canavelli et al. also report the synthesis of a Gly_5-pentamer using this method. High α-aminonitrile selectivity was also observed when tested in competition reactions against β-alanine nitrile and only glycine amide (Gly-NH_2) at pH 9 was able to out compete Gly-CN. Across a broad pH range (pH 5-9) the major product is still N-Ac-Gly-Gly-CN. As nucleotides were also not reactive and well tolerated this demonstrates a pathway that is compatible with a more holistic systems chemistry approach to peptide bond formation.

### 1.4.2.5 Catalytic Peptide Ligation

Catalytic peptide ligation has been an elusive target in origins research but recently Foden et al., building on their synthesis of N-acetyl-cysteine (vide supra, 1.4.2.2), have recognised that N-Acetyl-cysteine (Ac-Cys-OH) can enable regio- and chemo-selective organocatalytic proteinogenic α-peptide ligation (Scheme 14b)\(^9\). Once again this leverages the remarkable reactivity of α-aminonitriles. While the full mechanism has not been fully elucidated, cysteine’s thiol is proposed to reversibly add to the nitrile in a biomimetic fashion (reminiscent of non-ribosomal peptide synthesis, growth in an N-C direction) activating it to a transient thioimidate which can then be trapped by an incident amino acid’s (or similar) amine moiety to give the amidine. This methodology is both highly general and highly selective; it tolerates all amino acids (as amine donors) but exhibits only selectively thiol addition for Ac-α-AA-CN over Ac-β-AA-CN. Ac-Cys-OH also does not promote a great deal of aminonitrile hydrolysis, aided perhaps by the kinetic stability of AA-CN (when Ac-Gly-CN is incubated at pH 7, 60 °C they observed only 6% hydrolysis after 24 h). They also discovered this catalytic ability extends to many cysteinyl peptides and even primitive thiols (Scheme 14c) drastically increasing the likelihood of this being possible for a variety of origins scenarios.

Amino acids and the peptides they form are vital for extant life. Work towards understanding the origins of amino acids and peptides has progressed considerably but there are still many challenges remaining unsolved. The most auspicious of these being the connection between nucleic acids and the genetic information they carry and the peptides they code for. Homochirality also remains an unsolved problem but not just for peptides. Lately there has been a shift towards a more holistic approach to origins
scenarios and there is optimism that this will enable the ability to answer broader questions in the field.

**Scheme 14** Prebiotic routes to peptide synthesis via α-aminonitriles: a) Canavelli et al.’s ligation of aminonitriles to N-acetyl peptide nitriles by iterative, sequential thiolysis, hydrolysis and oxidation; b) Foden et al.’s thiol catalytic peptide ligation; c) a selection of high yielding thiol catalysts used by Foden et al.; note their similar structure.
1.4.3 Lipids in Prebiotic Chemistry

Lipids, when self-assembled into membranes, provide extant life both encapsulation (concentration of biomolecules) and separation (sequestration of a system from outside influences). The importance of these properties at the origin of life cannot be overstated. The need to retain important molecules after their prebiotic synthesis before they are lost to near-infinite dilution in an ocean environment or prevent potentially parasitic species from afflicting a replicating genetic system is paramount en route to life. What follows in this section is an overview of current progress towards the synthesis of lipids in prebiotic chemistry and their assembly into membranes. There have been proponents of a “Lipid world” theory but this will not be discussed here.

Fundamentally lipids are amphiphilic molecule consisting of a polar (hydrophilic) head group and one or two non-polar (hydrophobic) tails. Phospholipids are ubiquitous in extant biology and consist of a polar phosphorylated head group and long alkyl chain tail group(s). Therefore, phosphorylation and assembly of long alkyl chains under prebiotic conditions are vital in addressing the synthesis of lipids. Alternatively, a fatty acid (glycerol connected by an ester bonds to 3 longer-chain acids) can also act as lipids.

1.4.3.1 Prebiotic Synthesis of Lipids and Their Precursors

Early work towards this stemmed from the Fischer-Tropsch process (Scheme 15a), the reaction to CO and H₂ at high temperatures (500 °C), which has also been shown to operate under prebiotic conditions and its Kölbel-Engelhardt variant (Scheme 15a) which instead of utilising H₂, only present in low amounts in the early Earth’s atmosphere (vide supra, 1.3.2), uses H₂O. These processes can produce aliphatic hydrocarbons and other oxygenated species such as alkanols up to 18 carbons long. When mixtures of short, medium and long chain alkanols were subjected to phosphorylation conditions (ammonium phosphate, urea 2, 100 °C), compatible with nucleotide synthesis (vide infra, 1.5.5.2), only the medium and long chain alkanol phosphorylation products were detected (Scheme 15b) indicating some selection for longer chain aliphatics during prebiotic phosphorylation.
Recently, work on the acylation of glycerol-3-phosphocholine \textsuperscript{49} by \textit{N}-acetylimidazoles\textsuperscript{116} has also shown (Scheme 16a) that there is selectivity for longer chain lipids as very little di-acylation was observed when \textit{N}-acetyl-imidazole \textsuperscript{50} was used (max 40\%, pH 7/8) but when longer chain acyl imidazoles (butyryl, hexanoyl, octanoyl and decanoyl) were used near quantitative yields (> 95\%, MeCN/H\textsubscript{2}O, pH 7) were observed. The polar head group investigated here can be derived from DHA phosphate \textsuperscript{23} and has been observed by Patel \textit{et al.}\textsuperscript{47} (\textit{vide supra,} 1.4.1.3 and Scheme 6) with the other product of its formation glycerol 2-phosphate has been shown to, when acylated, self-assemble into vesicles\textsuperscript{117}. Gibard \textit{et al.} also investigated the reactivity of glycerol \textsuperscript{5} under prebiotic phosphorylation conditions (\textit{vide infra,} 2.2.10)\textsuperscript{118}. They observed the phosphate-mediated esterification of glycerol \textsuperscript{5} and nonanoic acid \textsuperscript{51} to give cyclo-phospholipid products (Scheme 16b) which spontaneously self-assembled into micelles and vesicle-like bilayers when mixed with water (pH \approx 8.5). UV-light driven synthesis of amphiphiles from \alpha-keto acids and alkanols has also been demonstrated\textsuperscript{119,120}.

\textbf{Scheme 15} a) Fischer-Tropsch and Köbel-Engelhardt variant; b) selective phosphorylation of long chain alkanols
Scheme 16 a) Fernández-García et al.'s selective acylation of 49 with longer chain acylating agent (N-acetyl-imidazole 50); b) Gibard et al.'s phosphate mediated esterification of glycerol and nonanoic acid 51 the product of which then self-assembles into micelles.

Natural phospholipids had been unobtainable targets due to their synthesis requiring high energy thioester intermediates and the need for membrane-bound enzymes. Now, evidence for a possible early origin of natural diacylphospholipids has been demonstrated in water. Liu et al. utilised a thioester mediated transacylation of 52 to produce 54 which lead to self-assembly into a membrane (Scheme 17). The quaternary ammonium moiety provides vital solubility in water. A level of pre-organization of 53 also enabled the reaction as shown by reduced reactivity when a detergent was added, or the reaction carried out below the critical micelle concentrations of 52 and 53.
Scheme 17 Thioester mediated transacylation leading to natural phospholipids that self-assemble into membranes\textsuperscript{121}.

1.4.3.2 Prebiotic Chemistry Inside Membranes

In a prebiotic milieu the co-operativity of lipids with nucleic acids and peptides and metabolic cycles is vital. Many prebiotic processes have now been shown to be compatible with the membrane environments. Amino acids, ribose and nucleic acids have all been demonstrated to bind to and stabilise vesicles composed of fatty acids (decanoic acid) against the disruptive nature of divalent cations (Mg\textsuperscript{2+} in particular) and salts (NaCl)\textsuperscript{122,123}. Encapsulated catalysts, such as RNA Polymerase, are able to function inside vesicles when substrates are added externally\textsuperscript{124}. Recently, activated monomers have been shown to be able to cross membranes for use in non-enzymatic primer extension\textsuperscript{125} and when coupled with the membrane-stabilizing effect of the addition of citrate \textit{25} (the final product of r-TCA cycle, \textit{vide supra}, 1.4.1.5), to chelate Mg\textsuperscript{2+}, this can allow the templated copying of RNA inside model protocells and decrease the amount of RNA hydrolysis (also promoted by divalent cations especially Mg\textsuperscript{2+})\textsuperscript{126}. A model system has also demonstrated the ability to preferentially form peptide bonds inside fatty acid vesicles which in turn enhance the growth of the model protocell\textsuperscript{127}. Recent work on the active species in non-enzymatic RNA primer extension (NERPE, \textit{vide infra}, 1.4.6.1) has revealed an imidazolium bridged dinucleotide\textsuperscript{128} as the key structure but as yet no work has been demonstrated on the permeability of such a species to primitive membranes. Fatty acid membranes have also been shown to undergo cycles of growth and division under prebiotically plausible conditions\textsuperscript{129,130}. Formation of stable vesicles from fatty acids is also compatible with the recently demonstrated light-releasable activation strategy\textsuperscript{131–133} (methyl isocyanide \textit{55} promotes the formation of imidazolium activated nucleotide monophosphates) and leads to cooperativity\textsuperscript{134}.
The importance of lipids and the membranes they form clearly have far reaching effects on prebiotic chemistry and therefore a more complete understanding of their interactions with evolving prebiotic systems is needed to fully understand their place at the origins of life.
1.4.4 RNA World Hypothesis

1.4.4.1 The Central Dogma of Molecular Biology

Crick’s central dogma of molecular biology states that the flow of genetic information is generally downstream. DNA is transcribed to RNA and RNA codons are translated into the correct amino acids in the ribosome to make proteins. There are “special transfers” but not all of these have been demonstrated, RNA→RNA has been shown but not DNA→Protein. Interestingly the “unknown transfers” still are yet to be discovered 50 years later. This implied at the time that the transfer of information was immutably downstream to proteins and seeded the idea of the RNA World hypothesis.

![Diagram of the Central Dogma of Molecular Biology]

**Figure 4** Crick’s central dogma of molecular biology, note the downstream flow of information with the exception of the DNA replication. The dashed arrow represents the recently discovered ability of RNA ribozymes to replicate RNA from either template strands or ligate small oligomers of RNA, as this does require a process of in vitro evolution it is added to this scheme for completeness but is not considered part of Crick’s original work.

1.4.4.2 Formalizing the RNA World Hypothesis

It was thought unlikely that a synchronous appearance of proteins and nucleic acids could have occurred. Therefore, this posed the problem of which came first the nucleotide or the protein? Enzymes are needed to make more proteins and replicate genetic polymers. Nucleotides are needed to provide the information as to which proteins to make but enzymes are required to construct said proteins. Nucleotides must also be copied and replicated by enzymes in order to evolve.

With the discovery of ribozymes, (catalytic strands of RNA), by Cech and Altman, it was postulated that RNA could act as both genotype (the informational polymer) as well as
phenotype (a catalyst for reactions). These *E. Coli* ribonuclease-p ribozymes were found to be able to ligate and splice strands of RNA, possibly removing the need for proteinogenic processes at the very earliest stage of life.

The ideas of Woese, Orgel and Crick in the late 1960s that RNA could also act catalytically were formalised by Gilbert with his idea of an “RNA World” coined in his paper of 1986. It is discussed and proposed as “[A] world, containing only RNA molecules that serve to catalyse the synthesis of themselves.” Reader and Joyce, attempting to add more evidence to the RNA world hypothesis, discovered that a constitutionally simple ribozyme, made up of only diaminopurine D and uridine, can catalyse the template-directed joining of two strands of RNA (one 5′-triphosphate and 3′-hydroxyl). Joyce’s later work on ribozymes resulted in the discovery of one that can cleave amide bonds and one such sequence that achieved self-replication with exponential growth, however this requires complex oligonucleotide substrates but shows that ribozymes can be efficient catalysts. Also, through their process of *in vitro* evolution Joyce et al. achieved amplification of RNA by an RNA polymerase ribozyme. The evolved ribozyme can synthesise (using an RNA template) aptamers, ribozymes and small amounts of tRNA showing that the replication of a genetic polymer and its conversion into functional molecules is possible without the presence of proteins.

Ribozymes, at the origin of life, would allow for the establishment of a system in which RNA can evolve by random recombination and mutation, and eventually achieve a range of catalytic functions aided by co-enzymes, which have been proposed to be molecular fossils of the RNA world. After this, ribozymes could start to synthesise proteins and enzyme catalysis takes over rendering ribozymes slowly less useful. Evidence for this may be gleaned from the structure of the ribosome which has been shown not to contain any protein motifs close (within 20 Å) to the site of peptide bond formation, a possible vestige of ribozyme existence. This revelation led to the RNA world gaining widespread acceptance and is currently one of the most popular theories on the origin of life.

However, the RNA world hypothesis is not without its critics and as Kurland believes “[the] RNA world is an expression of the infatuation of molecular biologists with base pairing in nucleic acids played out in a one-dimensional space with no reference to time or energy.” Since the early concepts of Orgel and others the idea of the RNA World has been the subject of much debate with a plethora of reviews that cover all aspects of it from its plausibility to finer details such as non-enzymatic polymerisation. The work in later chapters is involved in probing the RNA world hypothesis in some way but
if the RNA World hypothesis is to be true then there are four problems proposed by Orgel that need to be solved\(^\text{154}\):

1. The non-enzymatic synthesis of nucleotides

2. The non-enzymatic polymerisation of nucleotides to give random strands of RNA

3. The non-enzymatic copying and/or replication of RNA

4. The establishment of an RNA system, including catalysts, that is self-sustaining and able to grow in a pre-biotic environment.

Work in Chapters 2 (TNA) and 3 (DNA) of this thesis focuses on the first of these points and work in Chapter 4 (nucleic acid UV stability) is tied in with the fourth. Points 2 and 3 are covered here (\textit{vide infra}, 1.4.6) for completeness but are beyond the scope of this work.

1.4.5 Alternative Genetic Polymers at the Origins of Life

One paradigm relevant here is that of the "my grandfather’s axe" paradox\(^\text{153}\). In this context it implies that RNA was not the first genetic polymer to emerge and that it has instead emerged from a process of chemical evolution from something else. However, if all the pieces that encompass it change (i.e. if the first genetic polymer used different sugars, bases and backbone-connective linkages) is it still the same genetic polymer that was started with? The arguments for this may be synthetic (perhaps alternate bases and sugars were easier to synthesise prebiotically), maybe the alternate system gave a better function as a polymer (more effective catalysis) or even was just more stable. For reasons that are still being elucidated these selection pressures have yielded RNA as a dominant genetic polymer so there must be something about RNA that is privileged. A great deal can be learnt about this by exploring alternate genetic polymers and comparing aspects such as prebiotic synthesis, base pairing strength/fidelity and stability (e.g. aqueous hydrolysis or UV-light). In fact, several other alternate genetic polymers, with backbones based on RNA, have been proposed and investigated for their base pairing properties in the context of the origins of life. These will be discussed in some detail in Chapter 2. However, work on the relevant modification are discussed here.
1.4.5.1 DNA in the context of RNA

The question of if DNA could have been the first genetic polymer or was concomitant with prebiotic RNA synthesis will be examined in more detail (vide infra, 1.6.2 and Chapter 3) but the oligomeric properties of DNA are still of interest. DNA has been shown to exhibit catalytic properties when subjected to in-vitro evolution in the form of DNAzymes\textsuperscript{155,156} and the clear advantages of transferring information to DNA from RNA are apparent from their relative rates of hydrolysis\textsuperscript{157}. The idea that a combination of chemical stability as well as decreased copying fidelity of RNA (and from DNA to RNA) suggest that DNA was driven in a genetic takeover from RNA to DNA has also been explored\textsuperscript{158}.

1.4.5.2 ANA

Arabinose nucleic acid (ANA) (Figure 5) is conceptually the 2'-epimer of RNA. Recent studies by Roberts et al. have revealed an efficient, divergent and prebiotic synthesis of ANAs that has re-surfaced ANA as a potential alternate genetic polymer (vide infra, 1.6.2.1)\textsuperscript{159}. ANA can form homoduplexes in a manner similar to DNA homoduplexes (B-helix) and form heteroduplexes, albeit quite weakly, with RNA and DNA\textsuperscript{160}. It has been proposed as a candidate xeno-nucleic acid (XNA) for antisense treatment due to its far greater stability to hydrolysis (and digestion by RNase H)\textsuperscript{161}. Pinheiro and Holliger have also shown it is capable of heredity and evolution\textsuperscript{162} and catalysis (RNA endonucleases)\textsuperscript{163}. Efforts to reconcile the synthesis of ANA oligonucleotides have revealed that it is a poor substrate for nonenzymatic primer extension (NERPE, vide infra, 1.4.6.1) due to its much lower efficiency (~14 times slower for addition of arabino-G and ~30 times slower for addition of arabino-A compared to ribo-A and ribo-G) and
that once an ANA nucleotide is added to the primer further primer extension is significantly hampered\textsuperscript{164}.

\textbf{Scheme 18} a) Phosphorylation of GNA monomers lead to 2',3'-cyclic phosphates which inactive it to oligomerisation b) ribose-2',3'-cyclic phosphates can still be oligomerised due to 5'-hydroxyl\textsuperscript{165}. $B = \text{nucleobase (A, G, C, U)}$

\textbf{1.4.5.3 GNA}

Glycol nucleic acid (Figure 5), a recently invented base pairing system by \textit{Meggers and co-workers} is based on an acyclic three-carbon propylene glycol backbone\textsuperscript{166}. Despite acyclic backbones being widely assumed to be incompatible with the required “preorganization” for duplex formation and destabilising backbones that they are part\textsuperscript{167} it is interesting that GNA can base pair with itself in a helical Watson-Crick-Franklin mode ($T_M = 63 \, ^\circ\text{C}$). \textit{Meggers} also observed that the enantiomers, (R)-GNA and (S)-GNA, did not cross pair with each other and neither did they form stable antiparallel crosspairs with DNA, indicating that should this system have prebiotic relevance it would be an ancestor to only RNA. (S)-GNA was found to form heteroduplexes with RNA due to (S)-GNA being structurally similar to TNA (\textit{vide infra}, Chapter 2), conceptually it is TNA after removing a CH\textsubscript{2}O unit from the tetrahydrofuran ring. However, it has also been shown that GNA and TNA do not crosspair, presumably due to TNA’s DNA-like B-form helix\textsuperscript{168}. This
reveals that although conceptually simple and similar they almost certainly were not consecutive polymers en route to RNA and DNA. Phosphorylation may also pose a significant problem due to the (more) reactive primary 2′-alcohol promoting cyclization to the 2′,3′-cyclic phosphate (Scheme 18a) rather than oligomerising. Compare this to ribose-2′,3′-cyclic phosphates (Scheme 18b) which can still oligomerise due to 5′-hydroxyl\(^{165}\).

### 1.4.5.4 PNA

Peptide Nucleic acid (Figure 5), has been proposed as an alternative genetic polymer\(^{169−171}\). It is based on an achiral, non-ionic, amide backbone joined to nucleobases. PNA can form duplexes with RNA and DNA and even can template the oligomerisation of RNA as well as being be oligomerised on RNA templates\(^ {170}\). Some thioester-modified PNAs have even shown self-assembling properties\(^ {172}\). While the ability to translate, select and amplify PNAs has been demonstrated (albeit for in-vitro evolution experiments), a prebiotic synthesis is yet to be fully elucidated\(^ {173}\). The future prospects of PNA due to its dual ability as peptide and nucleic acid are intriguing\(^ {174}\).

### 1.4.6 Oligomerisation of Genetic Polymers

Orgel argued that one key part of the RNA world theory would be the non-enzymatic oligomerisation of monomers and the subsequent non-enzymatic copying of the resulting strands. That would then facilitate a system that could undergo Darwinian evolution. However, despite many advances there is no known and accepted way to polymerise long chains of ribonucleotides (into RNA) once they have been made or copy them with high fidelity to ensure heritable genetic transfer. There has been considerable work carried out seeking to address the issues not all of which can discussed here.

Modern enzymes can enable a totally regioselective synthesis of 3′→5′ linked RNA and while prebiotic chemistry has yet to fully address this problem of potential linkage isomerism, iterative rounds of degradation and repair can be used to correct 2′→5′-linkages into 3′→5′-linkages\(^ {175}\) and some work has also shown that they may not even be as detrimental as first thought as there is not a completely adverse effect on structure or catalytic ability of RNAs\(^ {176}\).

Additionally, the tendency for shorter oligomers to outcompete longer chain ones, simply based on concentration has been proposed as a problem. However, the work of Braun and co-workers have shown that thermophoresis, enabled by the flow of a solution of oligomers over an area of steep thermal gradient, has the ability to increase the concentrations of longer chain polymers which is a key problem with the polymerisation
of oligonucleotides. Strand separation is also a vital pre-replication stage for non-enzymatic synthesis and the energy for this process must come from somewhere geochemically plausible. Previous work has shown that it can be based on pH cycles of and wetting and drying and even salt.

1.4.6.1 Non-Enzymatic Primer Extension - NERPE

An unsolved problem in origins of life research is the way in which complex biomolecules were activated and assembled from their constituent monomers at the origin of life. Extant biology uses large, complex enzymes to take these monomers and join them together to build the biopolymers it needs (DNA/RNA, proteins). Non-enzymatic primer extension seeks to answer part of this question by utilizing a template strand and extending a pre-organized base-paired complimentary strand using chemically activated substrates. Extensive early work by Orgel and co-workers revealed that while biology can make good use of ribonucleotide triphosphates (e.g. ATP) as activated monomers they are not reactive enough at the origins of life, before the advent of complex enzymes. Instead they found that using 2-methyl-imidazole activated 5′-monophosphates could enable this Non-Enzymatic Primer Extension (NERPE).

Advances to this chemistry made by Szostak and co-workers have led to optimised conditions for this to take place (Mg$^{2+}$, ~pH 8). Without the inclusion of Mg$^{2+}$ the rate is sluggish, and the membrane destabilizing effects of Mg$^{2+}$ can be mitigated with the addition of citrate (the final product of the r-TCA cycle, vide supra, 1.4.1.5). Replacing 2-methylimidazole with 2AI provided a 10-100 fold rate enhancement over that of MeI when using 2AI. 2AI has been shown to form under conditions that also give 2AO. It was previously believed that the mechanism for this chemistry involved the primer’s terminal 3′-OH attacking the incoming mononucleotide activated phosphate but the isolation of a 2-aminoimidazolium-bridged dinucleotide intermediate has revealed that the ability of this to base pair and enable subsequent attack on the activated phosphorous dramatically enhance rate and fidelity. This key intermediate has also recently been shown to be compatible with prebiotic activation chemistry.

With particular relevance to the work in later chapters the ability of alternate sugar backbones and nucleobases to undergo efficient NERPE has been thoroughly investigated. DNA activated monomers undergo NERPE but are still 20 fold slower than RNA. Similarly, Arabinosides, even with preformed bridged dinucleotide, proved to be even slower (~100 fold) when investigated on an RNA template and in effect act as a chain terminator as the primer is almost incapable of being extended once one arabino-nucleotide has been added.
Subsequent investigations of nucleobase heterogeneity has revealed that 8-oxo-purines, recently synthesised by Stairs et al. (vide infra, 1.5.5.4) all perform poorly as substrates for nonenzymatic RNA copying, independently of whether they are present as activated mononucleotides at the end of the primer and when in the template lead to dramatic reduction in fidelity. However, inosine demonstrated efficient copying with high fidelity and taken together with the recent synthesis of inosines by Sutherland (vide infra, 1.6.2.2) have led to significant interest in inosine as a potential alternative nucleobase as a prebiotic synthesis of guanosine still remains elusive. Unnatural linkages such as pyrophosphates or amino acids are tolerated by this chemistry which is of particular note for the work explored in Chapter 2.

Due to better mechanistic understanding and tangible improvements to the rate and fidelity of NERPE the picture that emerges is that this process could have acted as chemical selection for RNA as the variety of alternatives (DNA, ANA, TNA (vide infra, Chapter 2) and alternate nucleobases) tested all display inferior rates and fidelities of copying apart from inosine and 2-thiocyotosine which is fascinating given the recent routes to both (vide infra, 1.6.2). Many are still able to be copied, even if at a lesser extent. This leads to a scenario in which the composition of a heterogeneous pool of prebiotic nucleic acids has great impact on the genetic polymers that are assembled from it and such the chemical synthesis of these monomers bears thorough investigation.

1.4.7 A Holistic Approach to the Origins of Life

As discussed, there are many compatible and conflicting theories on the origins of life. Recently however previously siloed theories have instead been considered in a more holistic nature, the demonstration of cooperative reactivity and linked synthetic routes to many prebiotic targets is becoming more prevalent. For example: it seems unlikely that only proteins emerged when many syntheses of other prebiotic molecules that also lead to amino acids have been demonstrated. More emphasis is now being put on what selection pressures could have resulted in extant biology from prebiotic chemistry.
1.5 Prebiotic Synthesis of Nucleic acids

For the RNA world theory to hold true, an abiotic synthesis that yields the required monomers is a strict requirement (vide supra, 1.4.4.2). Later chapters (2 and 3) are concerned with the prebiotic synthesis of nucleic acid from their prebiotic precursors. To contextualise that work, the following section explores work towards the prebiotic synthesis of nucleic acids.

1.5.1 Classic Disconnection Approach

The classical retrosynthesis of nucleic acids resulted in a synthesis that relied on direct attachment of the three components together using nucleophilic substitution, i.e. attach the sugar to the phosphate and attach the base to the sugar via a Hilbert-Johnson-type reaction. Here a brief overview of the prebiotic synthesis of the constituent parts (sugar, and nucleobase) for this approach will be examined.

![Scheme 19 Classical disconnection of pyrimidine nucleosides to ribofuranosyl sugar and base appears simple but experimental results show that the forward reaction is disfavoured](image)

1.5.2 Prebiotic Sugar Synthesis

1.5.2.1 Formose Reaction

Butlerow’s 1861 “formose” reaction has long been thought of as a viable prebiotic synthesis of sugars needed at the origins of life. They found that if formaldehyde 1 (0.1 M) was taken in a CaOH_2 solution a wide variety of sugar aldol products were formed including ribose (as well as other tetrose, pentoses, hexoses and their degradation products). The reaction begins with formaldehyde being converted into glycolaldehyde 3 (Scheme 20), giving a good precedent for this as a prebiotic feedstock molecule. This, albeit slow, step is believed to occur via metal-assisted umpolung chemistry of the
formaldehyde anion formed. This then adds to another molecule of 1 to give glycolaldehyde 3 and the now easier to enolise 3 reacts at a faster rate, defined as autocatalytic. The mechanism for the initial step (1 to 3) of this process is still not fully understood and while work by Breslow has helped to elucidate parts work is still ongoing fully elucidate this 160 years later.197

Scheme 20 Proposed mechanism of the formose reaction although the initial step is still not fully understood.97

Glycolaldehyde 3 then goes on to react again to give C₃ and higher sugars. Analysis after derivatization by Decker et al. (Figure 6) reveals even a cursory inspection shows a huge array of compounds (over 30) and interestingly threose (C₄ sugar, peaks 5 and 11) is present in fairly high quantities while ribose (C₅ sugar, peak 14) is present in a lower proportion.198

Despite the formose reaction being vital to many hypotheses of prebiotic chemists, especially those with glycosylation as the key step, ribose is formed in very low quantities (< 1%) and this is not improved even with the addition of initiator additives and catalysts.199 Additionally, the kinetics of the reaction are such that interception of ribose before its continued reaction to a “tar” is highly challenging. Formaldehyde 1 from the atmosphere is also probably too low a concentration for this to be an efficient process as it is need in 0.1 M concentration and models suggest formaldehyde production would be on the scale of 10⁻³ M31. Ribose 56 is not indefinitely stable to the conditions of its formation even if it is removed from a localized formose reaction, as it has an estimated t₁/₂ at pH 7 of 44 years at 0 °C, just 73 mins at 100 °C and was found to be <3 h at pH 10.2 and 55 °C.200,201 This has prompted work towards the stabilisation of 56. There are indications that borate minerals can stabilise the reaction of 3 and 4 by complexing the two hydroxyls of 4 but borates availability during the early Archean (~4 Ga) is still
in question\textsuperscript{204}. Lead (Pb\textsuperscript{2+}) has also been employed in a similar way\textsuperscript{205}. Amino acids, in the form of zinc-proline complexes, can increase the yield to 19\% but still generates a complex mixture of tetroses, pentoses and other higher order sugars\textsuperscript{206}.

The lack of pentose selectivity in these reactions almost certainly implies mixtures in any prebiotic processes if the formose reaction is invoked as the carbon source. Work to rectify this has been sparse but recently reports that the mineral hydroxyapatite can improve the reaction of 1 and 3 and when instead DHA 9 (main equilibration product of 4 in aqueous solutions) is used a mixture of products was formed consisting of ribose 56 (20\%), and other sugars (ketoses: ribulose (8\%) and xylulose (11\%)) as well as the decomposition product glycolic acid (40\%)\textsuperscript{207}. The fact that all the intermediates are sensitive to reaction with nitrogenous substances appears to be another blow.

\textbf{Figure 6} Decker et al.'s gas chromatogram of the formose reaction after derivatization as n-butoxime trifluoroacetates\textsuperscript{198}. Threose 57 (peaks 5 and 11, pink), erythrose 58 (peaks 4 and 7, blue) and ribose 56 (peak 14, green) are indicated by circles.

\subsection*{1.5.2.2 Prebiotic Synthesis of Other Sugars}
Routes to deoxyribose sugars have been of considerable interest to the field. This is in part due to the RNA world theory implying that either DNA arose later than RNA or that it was an invention of an RNA world. By providing a concomitant synthesis of DNA and RNA this would force a re-evaluation of these ideas. Toward this end work by Sutherland and co-workers has shown that pentose sugars can undergo efficient photoreduction to deoxyribose sugars\textsuperscript{208}. A mixture of ribose 56 and arabinose 59 sugars are likely products from a formose reaction and when irradiated (254 nm) for 6h in the presence of HCN, thiocyanate (KSCN) and sulfide (H\textsubscript{2}S) gave deoxyribose in 52 and 43 \% yield
from ribose and arabinose respectively. This solvated electron reduction of cyanohydrins chemistry will be discussed in more detail later (vide infra, 1.6.1).

**Trapp and co-workers** recently showed a mechanochemical synthesis of C₄ and larger sugars starting from glycolaldehyde 3 or glyceraldehyde 4. When 3 was subjected to oscillatory ball milling (30 Hz, 90 min) with various prebiotically plausible minerals (brucite, montmorillonite and portlandite) as catalysts (20 mol%) they observed tetrose and hexose formation in fair yields ((42.91 ± 0.99)% and (12.16 ± 0.35)% respectively). They also observed small amounts (<0.2 mmol) of pentoses when 3 and 4 were subjected to similar conditions but instead with CaOH₂ (20 mol%). **Clarke and co-workers** have also demonstrated that amino acids can promote the synthesis of tetrose and deoxyribose sugars.

All told formose and aldol-type reactions seem unlikely to be a plausible route to sugars but its invocation in prebiotic synthesis remains a mainstay even today (vide infra, 1.5.4.1).

### 1.5.3 Prebiotic Nucleobase Synthesis

Many routes have been established for the prebiotic synthesis of purine nucleobases and for some pyrimidine nucleobases. The belief that a simple condensation between sugar and base would yield the desired nucleoside led many groups to developing synthetic routes to nucleobases. The earliest of these to be useful was **Oró** who showed that it was possible to construct the purine adenine by the polymerisation of HCN (Scheme 21a). By heating ammonium cyanide (NH₄CN) solution at 70 °C for a number of days he achieved a very low yield of about 0.5%. **Schwartz et al.** improved this polymerisation by adding trace amounts of glycolonitrile. **Ferris, Orgel and Sanchez** elucidated the mechanism of this process. Surprisingly, they showed that the tetramer of HCN, diaminomaleonitrile (DAMN), which is stable, can react either with UV-light (77-82% after irradiating at 350 nm for 2 h) or with formamidine to give 4-aminoimidazole-5-carbonitrile (AICN), a subsequent hydrolysis gave 4-aminoimidazole-5-carboxamide (AICA). 62 can react with another molecule of HCN to furnish the canonical nucleobase adenine, while a cyanate reaction with 63 allowed guanine to be obtained. Other non-canonical nucleobases: 2,6-diaminopurine, hypoxanthine and xanthine can also be formed. HCN readily polymerises under conditions such as these and the full gamut of these pathways has recently been thoroughly reviewed.
Oro et al.’s polymerisation of HCN to give adenine 65 and guanine 67 as well as non-canonical nucleobases xanthine 68, hypoxanthine 66 and diaminopurine 64; b) Ferris’s reaction of cyanoacetylene 12 with 2 molecules of cyanate 70 to yield cytosine 71.

Formamide has also been thoroughly investigated as an alternate to HCN due to the belief that is more facile to concentrate than its readily polymerised precursor. Saladino, Di Mauro and co-workers have extensively investigated and reviewed this area and have shown that formamide can yield nucleobases under a range of conditions. Heating (160 °C) or irradiating mixtures of formamide and inorganic catalysts (montmorillonite, Murchison meteorite etc) has yielded all canonical nucleosides (including T) and AICA 63 amongst others and is thought mostly to occur via dehydration back to cyanide and then the subsequent chemistry already discussed.

While a number of fairly plausible pyrimidine syntheses have been proposed they suffer from the same problems as that of Oro’s work on purines in that the yields are poor and
reaction conditions are geochemically restricted\textsuperscript{223}. Cytosine \textbf{71} is hypothetically the easiest to form and has been demonstrated by Ferris \textit{et al.} (Scheme 21b) by reacting cyanoacetylene \textbf{12} with cyanate \textbf{70}. The cyanaovinyl urea formed can be converted to cytosine \textbf{71} in 19\% yield (100 °C, 1 day) and deamination of \textbf{71} to uracil \textbf{73} has been shown to be very slow but possible (in neutral aqueous solution, t\textsubscript{1/2} = 300 years at 30 °C)\textsuperscript{224,225}. Robertson and Miller have also shown that cyanoacetaldehyde \textbf{72}, when reacted with highly concentrated urea \textbf{2} (20 M) can also yield cytosine \textbf{71} in good yield (53\%)\textsuperscript{226}.

The prebiotic formation of nucleobases has a long history and still an active area of study. It certainly has more convincing evidence to suggest that nucleobases could be formed on the early Earth, especially compared to sugars. The conjoining of pre-formed sugars and nucleobases by glycosylation will now be discussed.

\textbf{1.5.4 Prebiotic Glycosylation}

With an understanding of the prebiotic synthesis of sugars and purine and pyrimidine nucleobases attempts to glycosylate said sugars can now be examined. It is important to keep in mind that the problem of an ample supply of sugars in the furanosyl ribose configuration is still lacking. The reactions invoking pure ribose as the glycosyl donor must be viewed through this lens.

\textbf{1.5.4.1 Problems with Glycosylation}

Work by Fuller \textit{et al.} failed to achieve yields of greater than 4\% and 9\% for the synthesis of ribo-purines β-adenosine \textbf{A} and β-guanosine \textbf{G} respectively using dry heating (100 °C, 2 days) of the nucleobase adenine \textbf{65} and ribose \textbf{56}\textsuperscript{227,228}. There are several proposed reasons for these low yields. Firstly, adenine \textbf{65} is only reactive through the correct nitrogen site (N\textsuperscript{0}) when in its minor tautomer, with 3 other sites (N\textsuperscript{1}, N\textsuperscript{3}, N\textsuperscript{7}) in the ring, and the exocyclic nitrogen (C\textsuperscript{6}), more reactive when in the major tautomer\textsuperscript{229}. To compound the problem, not only does ribose prefer to sit in its pyranose form rather than its furanose form, which would conceptually give access to its oxocarbenium ion, but is also reactive to nitrogenous species in its open chain aldehyde form, readily diverting material from the desired reaction. Nucleophilicity is quenched when acid is used to activate the reaction and the final nail in the coffin is that the equilibrium position of the reaction actually favours the reagents as heating nucleosides in the dry state leads yields their component parts; sugar and base.
Scheme 22 Fuller et al.’s work towards adenosine A and guanosine G, as products of heating of a dry-state mixture of sugar and base. Pyrimidines 71 and 73 are unreactive to these conditions.

The same problem exists for the pyrimidines (cytosine 71, uracil 73) with similar reasons; again, the sugar again prefers to sit as the pyranose form rather than the furanose form meaning less available reactant. Secondly, the N1 nitrogen’s electrons on the base are delocalized (Scheme 23) taking away crucial nucleophilic character and thirdly the hydrolysis to reagents is more favourable at neutral conditions. Due to both these kinetic and thermodynamic problems a different approach was required for a more widely accepted prebiotic synthesis to be established.

Scheme 23 Cytosine N1 lone pair delocalization precludes efficient glycosylation.
1.5.4.2 Glycosylations of 1′,2′-Cyclic Phosphates

Benner’s work on borate stabilization of preformed ribose 56 by complexing as a 1′,2′-boronate ester202,230 and Eschenmoser’s demonstration231 of 1′,2′-cyclic phosphate sugars lead Kim and Benner to the subsequent furanosyl constrained glycosylation of ribose 1′,2′-cyclic phosphate 74 (15 mM) with adenine 65 (7.5 mM) (Scheme 24)232. They regio-selectively assembled 2′-phosphorylated ribo-adenosine 75 in appreciable yield (15%). However, the concentrations of 74 and 65 have raised questions about this work’s prebiotic compatibility. Additionally, this route was expanded to be compatible with canonical pyrimidines (C and U), but was again low yielding (C ~7%, U, G < 1%), presumably due to the electronic reasons discussed (Scheme 23)233.

![Scheme 24](image)

Scheme 24 Kim and Benner’s glycosylation of ribose 1′,2′-cyclic phosphate 74 with adenine 65232.

1.5.4.3 FaPy Glycosylations

Further attempts to influence the regio- and stereoselectivity of prebiotic glycosylations have lead Carrell and co-workers to recently developing approaches to purine and pyrimidine nucleosides (Scheme 25) that leverage the greater reactivity of the activated formamidopyrimidine (FaPy) type bases234–236. These preferentially react with aldehydes at a masked N9/N1 position and after cyclization reveal canonical and non-canonical nucleosides with the correct regioselectivity (major product was the N9 and N1 for purines and pyrimidines respectively). However, the use of free ribose (invoking the formose reaction) lead to a non-stereo-, non-furanosyl selective mechanism and α-ribo and α/β-pyranosyl nucleosides were also formed.
Scheme 25 Carrell and co-workers synthesis of a) purine and b) pyrimidine nucleosides via FaPy compounds and ribose \(^{234-236}\).

Purine ribonucleosides were constructed from malonitrile and sodium nitrite to give (hydroxyimino)malononitrile 76 which reacts in the dry state (or in isoxazole 82 as a solvent) to give 5-nitroso-pyrimidines 78. These give the corresponding FaPy compounds 79 when reacted with formic acid and metals. When combined with wet-dry cycles in the presence of many equivalents of preformed ribose sugar 56 (or a mixture of glycolaldehyde 3 and glyceraldehyde 4) and strong base (borax/ammonia) gave the desired nucleosides (A, G) as well as access to a vast array of other non-canonical
products. Of particular interest here when tetrose sugars (e.g. 57) are added tetrose purines can be observed by HPLC but no yield was given, this will be discussed in more detail in Chapter 2.

Recently an improved method of pyrimidine synthesis allowed a unified approach that enabled both purines and pyrimidines to be synthesised (albeit some in very low yields, Table 2). The pyrimidines were constructed from cyanoacetylene 12 and subsequent reaction with hydroxylamine derivatives 81 gave isoxazole 82 (17-90%), reaction with urea 2 to give N-isoxazolyl-urea 83 (88%) required either Zn$^{2+}$ or Co$^{2+}$ to be efficient and dry down with boric acid and preformed ribose 56 gave α- and β-furanosyl as well as pyranosyl ribosylated species 84. Selective alkaline destruction in 125 mM borax at 95 °C allowed the furanosyl species to become the dominant product after 2-3 h and longer at lower temperatures. A strict requirement for Fe$^{2+}$ and thiols (DTT, cysteine etc) is enforced for the formation of canonical pyrimidines β-ribo-85 and β-ribo-86, via a reductive opening of the N-O bond of 84, followed by tautomerization, intramolecular cyclization and dehydration, when no Fe$^{2+}$ was used no pyrimidine formation was observed. Their previous purine synthesis was modified to work under this regime and the yields are summarised in Table 2.
<table>
<thead>
<tr>
<th>Entry</th>
<th>Anomer (α/β)</th>
<th>Ribose Conformer</th>
<th>Base</th>
<th>Yield (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>α</td>
<td>furanosyl</td>
<td>C</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>β</td>
<td>furanosyl</td>
<td>C</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>α</td>
<td>pyranosyl</td>
<td>C</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>β</td>
<td>pyranosyl</td>
<td>C</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>α</td>
<td>furanosyl</td>
<td>U</td>
<td>1.1</td>
</tr>
<tr>
<td>6</td>
<td>β</td>
<td>furanosyl</td>
<td>U</td>
<td>0.7</td>
</tr>
<tr>
<td>7</td>
<td>α/β</td>
<td>furanosyl and pyranosyl</td>
<td>A</td>
<td>35(^a)</td>
</tr>
<tr>
<td>8</td>
<td>β</td>
<td>furanosyl</td>
<td>A</td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td>α</td>
<td>pyranosyl</td>
<td>A</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>α</td>
<td>furanosyl</td>
<td>G</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>β</td>
<td>furanosyl</td>
<td>G</td>
<td>4</td>
</tr>
</tbody>
</table>

*Table 2 Optimised yields for the synthesis of ribonucleosides from the reaction of FaPy 79 and isoxazole 83 with ribose 56 as reported by Carell and co-workers\(^{36}\). \(^a\)combined yield of 35% for furanosyl-α-adenosine and pyranosyl-β-adenosine.*
1.5.5 Alternative Routes to Nucleic Acids

This section will describe routes towards nucleic acids that remove the need for glycosylation reactions by avoiding either free sugars, preformed bases or both. This has particular relevance to the work presented in later chapters.

Scheme 26 Orgel and Sanchez's synthesis of cytidine ribonucleosides.237

1.5.5.1 Alternate Routes to Pyrimidines

Orgel and Sanchez took an alternative approach to glycosylation by building the cytidine base onto the sugar (Scheme 26).238 They did however start with the free pentose sugars 56 and 59. Heating with ammonia and cyanamide 87 afforded the ribo-aminooxazoline RAO or AAO from 56 and 59. They then added cyanoacetylene 12,223,239, which formed cytidine ribonucleoside α-ribo-85 on hydrolysis (~10-20% for α-ribo-85 and β-ara-85). As discussed (vide supra, 1.5.2.1), 56 and 59 are unlikely to form in significant quantities as shown by Butlerow and subsequent experiments. A second problem was the remarkable but low yielding (at around 4%) photo-anomerization of the unnatural α-anomer α-ribo-85 to β-ribo-85. This reaction was investigated by Powner and Sutherland and the major
destructive pathway was identified as formation of the oxazolidinone 88 (27%) and concomitant nucleobase destruction. This process competes with the photoanomerisation at C1’ position\textsuperscript{240}. Both α-ribo-85 and β-ara-85 undergo an amount of photochemical anomerization to their corresponding α-anomers. Intriguingly, β-ara-85 also undergoes photochemical conversion to β-ribo-85 as well as interception of photoexcited intermediates to give ara-88, this epimerisation is quite surprising and will be discussed in later chapters. The clear involvement of the 2′-hydroxyl in this process revealed the potential for enhanced photoanomerization if the 2′-hydroxyl could be blocked, accordingly \textit{Powner} and \textit{Sutherland} also investigated the photochemical anomerization of 2′-phosphate-ribo-cytidine and found it to be more efficient with 36% total β-anomer formed and both cytidines and uridines being detected\textsuperscript{241}.

1.5.5.2 \textit{Powner} and \textit{Sutherland} Methodology
Similar to the motifs utilised by \textit{Orgel} and \textit{Sanchez}, \textit{Powner} and \textit{Sutherland} attempted to build an activated pyrimidine nucleotide by first condensing glycolaldehyde 3 and cyanamide 87 to give 2-aminooxazole 2AO in high yield (1 M P\textsubscript{i}, pH 7, >80%, 75% isolated)\textsuperscript{44,242}. Inorganic phosphate serendipitously provided general base catalysis and its inclusion at the latter stages indicated it could also be included from the start. Subsequent addition of glyceraldehyde 4 gave all the pentose aminooxazolines (Scheme 27) and in a one-pot sequential addition reaction the desired RAO and AAO were obtained in 25% and 15% respectively. The superb crystalline properties of RAO allowed separation from AAO\textsuperscript{201}. \textit{Powner} and \textit{Sutherland} identified AAO as a more suitable precursor en route to pyrimidine ribonucleotides as it would avoid the low yielding photoanomerisation step. Phosphate again was vital as when they cyanovinylated AAO to give arabino-anhydrocytidine ara-89 (92%) they observed chemoselective nucleobase construction and very low levels of off target O-cyanovinylation due to phosphates buffering. Urea/formamide phosphorylation of ara-89 gave activated pyrimidine 2′,3′-cyclic phosphate β-ribo-90 in good yield (46%), as well as the 5′-phosphate (β-ribo-91), via the intramolecular nucleophilic substitution at C2′ after C3′ phosphorylation. The advantage of phosphate inclusion at the start becomes key once more as this mixed chemistry makes the route more plausible. Here irradiation (254 nm, 3 d, pH 6.5) simply converts the stable cytidine β-ribo-90 to uridine β-ribo-92 (42%) with only minor base loss, although the α-anomer of β-ribo-90 was highly unstable to this process. This is one of the most prebiotically plausible pyrimidine synthesis and is a method that is directly applicable to the desired synthesis of TNA (Chapter 2). The simple modification being at the aminooxazoline stage instead of a molecule of glycolaldehyde 4 another molecule of glycolaldehyde 3 be used instead, this will give the tetrose-aminooxazoline TAO and should make it possible to proceed to threocytidine α-threo-85.
1.5.5.3 Problems with the Powner and Sutherland methodology

Despite being the most plausible route to pyrimidines there are concerns regarding spatio-temporal separations, prebiotic clutter and heterogeneous oligomerisation. The reaction of cyanamide 87 and a 1:1 mixture glycolaldehyde 3 and glyceraldehyde 4 gives an unselective mixture of products. Therefore, one of the main concerns with the work of Powner and Sutherland is the need for spatio-temporal separation of 3 and 4. This could be envisioned by the formation of 2-aminooxazole 2AO in one location and then its evaporation and condensation into a pool of 4. This is a restrictive condition to impose on a system that has to be fairly widespread to generate a large pool of nucleotides. A second concern is the unselective nature of the pentose aminooxazoline formation, RAO and AAO have been converted to β-ribo-85, but LAO and XAO have not despite subsequent work to interconvert between RAO and AAO.

A third concern is the problem of linkage isomerism. Due to the way activated pyrimidines are formed there may be issues with the selectivity of phosphorylation (indeed β-ribo-91 was also observed, Scheme 27) and a mixture of 5′, 3′ and 2′ phosphorylated nucleotides could be formed if hydrolysis of the 2′,3′-phosphate occurred, this would lead to linkage isomerism, 2′→5′ oligos being formed instead of 3′→5′, and a heterogenous RNA backbone. This could hinder (if not totally negate) accurate information transfer.

Finally, glyceraldehyde 4 (Scheme 28) exists in equilibrium with ketose dihydroxyacetone (DHA, 9) this sits vastly towards the side of 9 in the presence of phosphate buffer, a system...
fairly ubiquitous among prebiotic chemistry and the one employed by Powner and Sutherland.

Interestingly TNA (Chapter 2) has some innate advantages over RNA as it avoids the first and third of these problems as no glyceraldehyde 4 is required which means there are no chemoselectivity issues, additionally the step from glycolaldehyde 3 to TAO may be possible to achieve in one step, rather than isolating the 2AO intermediate. As there are only two hydroxyls to phosphorylate the issue of linkage isomerism is reduced, making TNA a seemingly more robust polymer from this perspective. Moreover, the two tetrose forms threo and erythro each have an L and D form and there is evidence from Sanchez et al. that the L-threo and D-erythro forms could both undergo photochemical processes (a C1′ anomerisation for the threo and a C2′ epimerisation for the erythro) to arrive at the desired stereochemistry for base pairing. This convergence effectively means that no material would be lost, this convergence is not possible in RNA due the number of stereocentres that would need to be inverted.

Scheme 28 Islam and Powner’s selective aminal crystallization of glycolaldehyde 3 and glyceraldehyde 4.

Recent work from the Islam and Powner may provide solutions to the above problems for the pentose series. This work (Scheme 28) describes a situation where aldehydes 3 and 4 can be separated from a complex mixture of sugars (C2, C3, C4, C5) after their
generation by the product of the reaction of beta-mercaptoacetaldehyde (BMA) and cyanamide 87: 2-aminothiazole 2AT. This sequesters 3 in approximately 2 hours as crystalline aminal 93 and prevents it from being consumed in aldol or degradative processes. Remarkably, 2AT can also pull 9 out into a crystalline aminal 94 in around 20 days and when a source of cyanamide 87 is introduced aminal 93 releases 2AO which can then react with aminal 94 in an exquisite piece of chemistry to give the ribo-aminooxazoline RAO. This can then be taken forward in the work previously discussed to afford the ribonucleotide. Additionally, they also observed the ability of 2AT to promote an aldehyde selective Strecker reaction and avoid the synthesis of α,α-disubstituted aminonitriles from mixtures of simple aldehydes and ketones.

This solves the problem of temporal separation of glycolaldehyde 3 and glyceraldehyde 4 for the synthesis of the ribo-nucleotides but also vastly improves the chances of 3 surviving in solution long enough to react with 87. Here 3 is effectively protected in crystalline form. A problem with this however is that there are no published prebiotic syntheses of BMA and therefore this may preclude the existence of 2AT; either way this means access to threose precursors are more likely to be able to be formed as 3 is effectively protected until 87 is introduced into the system.
1.5.5.4 Alternate Routes to Purines

The paucity of convincing prebiotic purine syntheses in the literature has been discussed (vide supra, 1.5.4.1) but lately the development of a totally furanosyl selective, divergent route to 8-oxo-purines and canonical pyrimidines from a common aminooxazoline scaffold has shed new light on the construction of purines\(^{192}\).

\[
\begin{align*}
\text{HCN} & \quad \xrightarrow{\text{80 °C, 2 h}} \quad \text{glycolaldehyde 3} \\
\text{glycolaldehyde 3} & \quad \xrightarrow{\text{HSCN, pH 7, 60 °C, 24 h}} \quad \text{thione 95 (85%)} \\
\text{thione 95 (85%)} & \quad \xrightarrow{\text{glyceraldehyde 4}} \quad \text{thione 96 (38%)}
\end{align*}
\]

Scheme 29 Stairs et al.’s prebiotic synthesis of oxazolidinone thione ribo-\(96\), ara-\(96\), xylo-\(96\), lyxo-\(96\) and lyxo-pyranosyl-\(96\). Efficient 2-thiooxazole 95 formation from glycolaldehyde 3 and thiocyanate (HSCN) followed by addition of glyceraldehyde 4.

8-oxo-purines are the non-canonical common oxidation products of purines, particularly guanosine. However, they are well tolerated by a number of enzymatic replicative processes\(^{245,246}\). They have also been implicated as potential flavin mimics in repair mechanisms for photolesions (CPDs, vide infra, 1.6)\(^{247}\). Here, Stairs et al. identified the oxazolidinone thiones as a potentially divergent scaffold en route to purines and pyrimidines. They report the high yielding reaction of glycolaldehyde 3 and thiocyanate (HSCN) to give 2-thiooxazole 95 (85%), followed by further reaction with glyceraldehyde 4 to give all four pentose oxazolidinonethiones 96 (Scheme 29). Desired arabino oxazolidinonethione ara-\(96\) was a major product and the reaction was highly arabino/ribo diastereoselective with xylo-, lyxo-\(96\) and lyxo-pyranosyl-\(96\) being minor products.

The divergent nature of this work became apparent when thione ara-\(96\) was converted to cyanovinyl adduct 97 with quantitative conversion. When 97 was subjected to multiple sequences of ammonolysis and re-activation with cyanoacetylene 12 it gave AAO in good yield (45%) which, as discussed (vide supra, 1.5.5.2), can be elaborated in high yields to give both canonical ribonucleotides (β-ribo-\(90\) and β-ribo-\(92\)). Using the activated thione 97 Stairs et al. investigated a displacement at pH 4 with 2-amino-2-cyanoacetamide 99 to give AICA 101 (33%, after cyclization at pH 9). This also revealed that facile thiol exchange (pH 6) with methanethiol (MeSH) could give methylthione 98 (60%). This allowed reactivity with HCN trimer 60 to give AICN 100 (15%). This was
required as thione 97 is incompatible with reaction at low pH and 60 is unstable above pH 3. This thiol substitution also increased the yield for the formation of AICA 101 to 59% when 99 was instead used.

Scheme 30 Stairs et al.’s divergent route to ribofuranosyl pyrimidines and 8-oxo-purines A β-ribo-105 and 1 β-ribo-106.

Incubation of AICN 100 and AICA 101 with formamidine 69 in formamide (HCON₂) gave 2′,8′-anhydro-arabino adenosine ara-102 (65%) and the 2′,8′-anhydro-arabino-inosine ara-103 (11%) respectively and phosphorylation in urea melt (140 °C, 20 mins) or in formamide (100 °C, 72 h) gave both corresponding 2′,3′-cyclic phosphates β-ribo-105 and β-ribo-106 (55-70%). Unfortunately, this methodology was unable to resolve a route to ribo-guanosine via ara-104. This chemistry is a key underpinning of the work explored in Chapter 3.
1.6 Prebiotic Photochemistry

Later chapters examine the role of UV-light photochemistry at the origins of life. UV-light was one of the most abundant sources of energy on the early earth; modelled emission spectra of the young sun and the early Earth’s predicted atmosphere (low in O₂/O₃, *vide supra*, 1.3.2) reveal a picture of a world bathed in UV-light and gives an available window for photochemistry on the early earth with all UV-flux attenuated below 200 nm. It is likely that the Origin of Life is thought to have occurred under these conditions.

Photochemistry has been shown to be highly useful for prebiotic transformations and it would leave any picture of the prebiotic landscape incomplete if ignored. Much of the work presented in later chapters explores the photochemistry of nucleic acids and their derivatives. Photochemistry has been used in origins of life research since the fields beginnings and continues to be an invaluable tool in the prebiotic chemist’s inventory, it is therefore useful to contextualise the work in later chapters if that has not already been discussed already, indeed many of the problems highlight above have had “new light” shed on them when UV-light has been used as a tool to investigate them.

Most of the research on UV-light has been on the negative impacts on human health. Nucleotides and by extension biopolymers containing them (RNA, DNA etc.) have been studied due to their reaction when exposed to UV-light. Photolesions caused by [2+2] photocycloadditions of thymines, uridines and cytidines to cyclobutane pyrimidine dimers (CPDs) have serious effects on the health (Scheme 31)²⁵². Recent work in this field has looked at how UV-light can also be involved in the repair mechanism for these photolesions²⁵³,²⁵⁴. This work has explored the ability of neighbouring charge transfer complexes to react with CPDs and cause reversion back to their two constituent bases. As the usual photodamage repair machinery (photolyases) is not needed, this has opened up the exciting possibility of a similar repair mechanism occurring at the origin of life on primitive oligonucleotide strands.

![Scheme 31 Cyclobutane Pyrimidine Dimer (CPD) formation for two adjacent thymines on a DNA strand. R= deoxyribose backbone.](image-url)

²⁵⁰,²⁵¹
With particular relevance to the work in Chapter 4, UV-light has also been used to assess the fitness of oligonucleotide strands, ribozymes, after irradiation. Saha and Chen have shown that some RNA aptamers (single stranded RNAs that bind a specific target molecule) are more affected by exposure to UV-light than others. Assays measuring the fluorescence after irradiation revealed that the MG Aptamer (Malachite Green Aptamer) was compromised in its ability to bind MG after irradiation indicated by a lowering of fluorescence. This indicates that irradiation affected the folded structure of enough to alter these properties. Whether the ability of RNA oligonucleotides to resist irradiation by UV-light is a strong selection pressure at the origin of life remains to be established.

The synthetic use of UV-light can further be applied to the generation of parts of modern-day metabolism that may have a part to play in the origins of life. Bonfio et al. showed that UV-light can drive the synthesis of iron-sulfur clusters. These clusters are found throughout biology and carry out various functions resulting in the belief by some that they are some of the most ancient cofactors. Ferric iron (Fe$^{3+}$) and thiolate species (glutathione in particular) can be used to synthesis polynuclear iron-sulfur clusters under 254 nm irradiation very rapidly. The process for this is synergistic with the formation of an alanine residue, from the glutathione’s cysteine, mid-peptide. Glutathione has a well-studied role in extant biology as an antioxidant and the relationship between antioxidant species and reactive photochemistry is clear.

**Scheme 32** Photochemical synthesis of Fe-S clusters from cysteine Cys by Bonfio et al. $^{256}$. $R =$ cysteine or glutathione.
1.6.1 Prebiotic Photochemical Reduction

The production of solvated (aqueous) electrons by irradiation of, first, cyanocuprates (CuCN$_3$)$^{59,257}$ and now hydrogen sulfide (H$_2$S)$^{47,208}$ and bisulfite (HSO$_3$)$^{58}$ has enabled coupling of UV-light to reductive processes and allowed a step change in the assembly of RNA and now DNA precursors. The photochemical reduction of prebiotically available HCN to simple aldehydic sugars has allowed the elimination of the formose reaction to generate RNA precursors$^{47,58,257,258}$. The solvated electrons reduce nitriles to imines which can then hydrolyse to aldehydes (Scheme 33a).

\[ \text{HCN} \xrightarrow{hv, 254 \text{ nm}} \text{HCN} \rightarrow \text{H} \xrightarrow{\text{via}} \text{O} \xrightarrow{\text{H}} \text{OH} \]

\[ \text{Scheme 33} \text{ Sutherland and co-workers, Killiani-Fischer type homologation of HCN to give two key prebiotic aldehydic sugars, C$_2$ glycolaldehyde 3 and C$_3$ glyceraldehyde 4; b) and the photochemical reduction of ribose 56 and ara 59 to deoxyribose sugars 107; c) conversion of 5-mercapto-uracil 109 to thymine 110 by photochemical reduction.}^{47,58,257,258} \]

As well as key RNA precursors this chemistry has now also allowed access to DNA precursors. Deoxyribose 107 can be accessed (Scheme 33b) by the photochemical reduction of both ribose 56 and arabinose 59 in the presence of sodium sulfide, potassium thiocyanate and copper cyanide$^{208}$. Yields of 52 and 43% from ribose 56 and arabinose 59 respectively mark a vast improvement over other syntheses of deoxyribose
through aldol-type chemistry (vide supra, 1.5.2.2). In the same work (Scheme 33c) they also highlight the photochemical synthesis of the DNA base thymine \(110\). Robertson and Miller had previously shown that uracil \(73\) can be formylated at C5 to 5-hydroxymethyluracil \(108\) which then undergoes near total conversion to 5-mercaptomethyluracil \(109\) when hydrogen sulfide is introduced\(^{259}\). Building on this, Ritson and Sutherland have shown that \(109\) can then undergo photochemical reduction (254 nm, 6 h, \(\text{H}_2\text{S}\)) to the DNA base thymine \(110\) in 36% yield. It remains unclear if this is relevant to the origins of life as similar reactivity has not been demonstrated to yield thymidine nor an efficient prebiotic glycosylation involving \(110\) and deoxyribose \(107\).

### 1.6.2 Prebiotic Photochemical Synthesis of Nucleic Acids

We have seen that photochemical anomerization can convert \(\alpha\)-ribo-85 to \(\beta\)-ribo-85 (vide supra, 1.5.5.1) but resistance to optimisation has meant it has remained a low-yielding process despite significant interest. This has meant that although many convergent and divergent syntheses towards RNA yield RAO and subsequent anhydro nucleoside ribo-89 they are oftentimes unutilised side products\(^{44}\).

\[ \text{Scheme 34 Xu et al.'s efficient photochemical anomerization of } \alpha\text{-ribo-111 to give } \beta\text{-ribo-111 is enabled by thiolysis in formamide of ribo-89} \]
Xu et al. have rectified this by demonstrating that by replacement of sulfur at C2 on the cytidine base they can achieve far higher yields of photochemical anomerization than previously reported\textsuperscript{260}. A high yielding thiolysis of anhydro ara-89 in formamide (up to 84\%) allowed exploration of the photochemistry of α-ribo-2-thio-cytosine α-ribo-111. When irradiated (254 nm, 2.5 d) in the presence of sulfide they observed that α-ribo-111 readily anomerized to β-ribo-111 in high yield of up to 76\%. Small amounts of starting material (11\%) and its hydrolysis product α-ribo-85 (9\%) were observed as well as base loss of 2-thiocytosine 113 (4\%). A slow hydrolysis (pH 7, 100 mM Pi, 84 d, 60 °C) of the C2 sulfur in water gave canonical pyrimidine β-ribo-85 and also thiouridine β-ribo-112 with a similarly slow acid hydrolysis (pH 3, 40 mM formic acid, 54 d, 60 °C).

1.6.2.1 Arabino-Nucleic Acids

The interplay between sulfur and UV-light (or hydrogen peroxide) has recently been used by Roberts et al. to deliver arabino sugars attached to canonical purine nucleobases by reduction of 8-thio-arabinosides\textsuperscript{159}. Arabinosides are the monomers for ANA (vide supra, 1.4.5.2) a non-canonical genetic polymer and their synthesis serves to inform whether these monomers would have been present under similar condition that give rise to ribonucleosides. Building upon work by Stairs et al. a further point of divergence was explored (Scheme 35) via the high yielding, aqueous thiolysis of arabino anhydronucleosides (e.g. ara-102) to regioselectively position a sulfur at C8. Subsequent irradiation (254 or 300 nm) or reaction with hydrogen peroxide (H\textsubscript{2}O\textsubscript{2}) resulted in effective reduction for ara-114 and ara-116, the two canonical purine nucleobases while ara-115 underwent extensive destruction under UV-light.

![Scheme 35](image)

\textbf{Scheme 35} Roberts et al.’s synthesis of arabino-purines by thiolysis and subsequent reduction with UV-light or hydrogen peroxide.
1.6.2.2 Photochemical Routes to DNA

Recently, further examples of sulfur’s potential as a photochemical substrate has been demonstrated for the construction of DNA nucleosides using chemistry that implicates a concomitant emergence of both RNA and DNA. DNA nucleosides, for the longest time, have remained difficult targets in prebiotic chemistry due to the incompatibility of the required deoxyribo sugar with the types of chemistry used to construct ribonucleotides that have been discussed previously (vide supra, 1.5.5).

**Scheme 36** Xu et al’s synthesis of deoxyribose nucleosides 121 and canonical deoxyribo adenosine 123. Inset: Proposed intermediate in the reduction of anhydro-122.

Xu et al. showed (Scheme 36) that, using the previously discussed thiolysis of ara-89 and subsequent (greatly more efficient) photochemical anomerization of α-ribo-111, could yield β-ribo-111 which then undergoes hydrolysis to β-ribo-112. They showed you could convert the 2-thio-uridine β-ribo-112 via a proposed 2',3'-cyclic phosphate intermediate to the 2,2'-thioanhydro 120, their substrate for photoreduction. The phosphorylation gave wide product distribution (50% total anhydro yield) as they also observed loss of their bases (15% 2-thiocytosine 113, 7% iso-cytosine 124) and variable 5'-phosphorylation. All the 2,2'-thioanhydouridines, phosphorylated or not, then undergo a photochemical reduction (254 nm) in the presence of sulfide (which is vital to product formation) yielding 2'-deoxy-2-thiouridine 121 in around 30% yield, they also observe the 2'-thio-2-thiouridine 122 (Scheme 36:Inset) (28% for 122) en route to nucleobase loss (113, 25%). The photochemical pathway implicates sulfide as an important source of solvated electrons by which this reduction is thought to occur. The solvated electrons generate an S-centered radical anion which then undergoes C-S bond homolysis and
subsequent capture of an H atom by the C-centred radical to give the noncanonical 2'-deoxy product 122. Interestingly, 2,2'-thioanhydrocytidine 125 did not exhibit any reduction but instead undergoes a more rapid destructive process to give 113 (29% yield) after only 3 h of irradiation, the propensity of 2-thio-nucleosides to destructively form episulfides has not been studied to a great extent and warrants further investigation. Subsequent transglycosylation with adenine 65 under dry conditions (100 °C, 31h) allowed access to canonical deoxyribo adenosine 123 in appreciable amounts (6% α, 4% β). 121 is unstable in both unbuffered and buffered solution but gives deoxyribose as a mixture of furanosyl and pyranosyl conformers in high yield (84-91%).

*Sutherland and co-workers* subsequently extended this photochemical reduction chemistry to afford 123 (Scheme 37). Transglycosylation of cytidine or uridine anhydros arabinose 89/126 by 8-thioadenine 127 gave a mixture of N⁹- and dead end N⁷-configured glycosylation products. The N⁹-configured products also underwent solvated-electron-mediated photoreduction in the presence of sulfide to 123 (39%) but when bisulfite was used the yield was improved (51%). The by-product of the reduction 8-thio-deoxy adenosine 129 underwent reduction (as expected based on *Roberts et al.*¹⁵⁹) to 123 in high yield (90%). The N⁷-configured products were destroyed far more rapidly by both photochemistry or hydrolysis than the N⁹-structures.

**Scheme 37** Xu *et al.*’s transglycosylation of anhydro and solvated electron reduction to give deoxyribo adenosine 123. A concomitant pathway to deoxyinosine 130 is also available from the reaction with sodium nitrite.
Prebiotic photochemistry has been shown to bring about powerful molecular change and with UV flux higher on the early Earth consideration for the photochemistry and photostability of prebiotic species is vital for a better understanding of both potential synthetic routes to desired biomolecules implicated in the origins of life and the potential selection pressures that could have acted on these biomolecules once formed.
2 Threose Nucleic Acid - TNA

2.1 Introduction

The structure of life’s first genetic polymer is a question of intense ongoing debate. Several theories rely on the non-enzymatic emergence of nucleic acids. The ‘RNA World’ theory (vide supra, 1.4.4) advocates for RNA as life’s first genetic polymer due to its dual genotypic and phenotypic functions as demonstrated by ribozymes and the core of the modern ribosome. However, other theories have been proposed. This chapter addresses one such theory; if a system can exchange information with RNA and is a simpler system is it likely to be a progenitor? For RNA to have evolved from another informational polymer, it follows that the system would have to be simpler, either structurally or generationally.

2.1.1 Etiology of Nucleic Acid Structure

The reason for Nature’s ‘choosing’ of RNA as its genetic polymer of choice is of decided interest. Eschenmoser et al. undertook extensive and influential work which tested these options by synthesising analogues of RNA with varied sugar backbones that could also be derived from a \((\text{CH}_2\text{O})_n\) aldosugar\(^{263,264}\). This was based on the belief that the number of atoms in the backbone was the main determinant of its ability to form Watson-Crick-Franklin duplexes (base pairs). They then evaluated each analogue system based on its ability to base pair.

![Scheme 38](image)

*Scheme 38* The non-natural nucleic acids synthesised by Eschenmoser et al. by varying the sugar backbone of the system (also included are RNA and TNA for completeness).\(^{264}\)
2.1.1.1 Hexose, Pento-pyranosyl and Pento-furanosyl Nucleic Acids

4′ → 6′ Homo-DNA, the hexose sugar derived by adding a CH₂ unit to DNA’s sugar was found to form a strong homo-duplex in the Watson-Crick-Franklin and reverse Hoogsteen modes. However it was unable to pair with RNA and DNA meaning no transfer of information could have been possible. It is also not thought to be prebiotically plausible to form, although the 4′→6′ allo and altro nucleotides are thought to be. Interestingly, the base pairing was much weaker in these other hexose analogues and the 4′→6′ gluco derivative exhibited very little pairing at all. These polymers did exhibit purine-purine pairing in the reverse-Hoogsteen mode, (except for 4′→6′ gluco). The lack of pairing has been attributed to additional strain of positioning the extra hydroxyl group when trying to achieve hydrogen-bonding conformations.

Next they investigated the 2′→4′ pyranose system, to probe why a ribo-furanose sugar was chosen over a ribo-pyranose one. p-RNA was found to show a Watson-Crick-Franklin base pair strength higher than that of DNA and RNA. Interestingly, this highlights that RNA was not solely chosen for its strong duplex formation ability but almost certainly a plethora of reasons include its catalytic capability (vide supra, 1.4.4.2). An exploration of the other 2′→4′ pyranose systems also revealed increased base pairing for all members with the 2′→4′ arabino being the strongest of the family. Unfortunately, the entire family do not pair with RNA, meaning it was unlikely to be a progenitor.

All the systems studied thus far had contained 6 bonded atoms in their backbone structures but did not cross-pair with DNA/RNA, the next logical step was seeing if this “rule” was concrete. The 3′→4′ (L)-lyxo pento-pyranosyl system proved to be an exception to this rule as this exhibited base-pairing both with strands of itself and, albeit weakly, with DNA and RNA. Lyxo has only 5 atoms in its back-bone and assumes an anti-periplanar geometry with respect to its two trans-vicinal phosphates, the (D)-ribo form does not adopt this geometry and therefore does not exhibit any cross pairing.

2.1.1.2 Threose Nucleic Acid (TNA)

Influenced by the 3′→4′ (L)-lyxo system, Eschenmoser et al. decided to explore systems with shorter than the conventional 6-atom backbone structure. This lead to the discovery of one such system: (3′→2′)-(L)-α-threofuranosyl/Threose Nucleic Acid (TNA). Remarkably, TNA can form stable duplexes not only with itself but also with RNA and DNA despite its reduced backbone. This allows TNA to exchange heritable genetic information with RNA and DNA and has generated considerable interest in TNA as a potential progenitor to RNA.
TNA’s *trans*-vicinal phosphodiester groups mean that it has just five atoms in its backbone and four carbon atoms making up the tetrose sugar. It also adopts the anti-periplanar structure akin to the 3′→4′-(L)-lyxo system (torsional angles between 158° and 169°)\(^{269}\). Structurally, it is thought to pair well due to the fact it can adjust conformationally more easily for base pairing due to having fewer carbons and hydroxyls and by stretching its tetrose phosphate moiety (3′-endo conformation) to match the repeat length of RNA\(^{276}\). It is also simpler in comparison to RNA in that it has one less chiral centre. Generationally, the tetrose sugar backbone is constitutionally simpler; conceptually it be formed from the condensation of two achiral C\(_2\) building blocks whereas RNA’s sugar moiety requires one C\(_2\) and one chiral C\(_3\) building block\(^{270}\). TNA has some other innate advantages over RNA. TNA oligomers were shown to have excellent hydrolytic stability compared to RNA, lasting up to months in comparison to similar RNA oligomers and are instead comparable to DNA\(^{157}\). This is due to the lack of free 2′-hydroxyl (Scheme 40) which cannot attack itself in TNA and causes cyclisation to 2′,3′-cyclic phosphates mid strand, although this has been postulated as a potential benefit of RNA allowing recombination.
Scheme 40 Mechanism of mid-strand 2',3'-cyclic phosphate formation only possible for RNA. TNA and DNA cannot undergo this process.

2.1.1.3 TNA as a Xeno-Nucleic Acid

TNA has been explored extensively as a xeno-nucleic acid and has been found to fold into tertiary structures that can bind targets with high affinity and specificity\textsuperscript{277–279}. Due to these remarkable properties, \textit{in-vitro} evolution of TNAs has been possible and TNAs have been synthesised by DNA polymerases\textsuperscript{273,280–283}, RNA polymerases\textsuperscript{284,285} and even TNA polymerases\textsuperscript{286}. TNAs can act as templates for DNA synthesis\textsuperscript{287} and be ligated on DNA templates\textsuperscript{288}, they can also be replicated and enriched \textit{in-vitro}\textsuperscript{289}. TNAs can also be incorporated into (hammerhead) ribozymes but reduce their catalytic activity\textsuperscript{290}, perhaps unsurprising due to their lack of 2'-hydroxyl. TNA has also been suggested as a viable candidate in antisense treatments due to its lack of recognition by nuclease enzymes and its hydrolytic stability\textsuperscript{291–293}. It is also potentially interesting to note that in many of the above mentioned studies the DNA pyrimidine thymine is used instead of uridine.

2.1.1.4 TNAs in Prebiotic Chemistry

This chapter focuses on the synthesis of TNA monomers under conditions that are also compatible with the synthesis of RNA monomers, this may then lead to the simultaneous emergence of TNA and RNA. Therefore, investigating TNA and RNA mixed systems is of clear interest. While not greatly studied, some work has been carried out on such systems. When mixed pools of TNA and RNA monomers are polymerized, TNA/RNA chimeras have been shown to promote the formation of homogenous backbones in template directed ligations (when amino-terminated TNA is used to speed up the ligation reactions)\textsuperscript{294}. The results of this study imply that even if a chimeric heterogenous system...
emerged due to unselective polymerisation of monomers a homogenous backbone can emerge.

Additionally, an important part of ascertaining the possibility of TNA as a progenitor or contemporary to RNA is its ability to undergo non-enzymatic replication (primer extension or ligation). This is well studied for many systems (vide supra, 1.4.6.1) but until recently TNA had not been examined in these systems despite Heuberger and Switzer having shown non-enzymatic synthesis of oligo-ribonucleotides on an all TNA template. This opened up the exciting possibility that TNA templates are amenable to non-enzymatic templated replication. Szostak and co-workers then showed the non-enzymatic template-directed primer extension using 2-methylimidazole activated-amino-terminated TNA on RNA, DNA and TNA templates. Thorough mechanistic studies and beautifully elucidated crystal structures for the mechanism of NERPE reveals that it proceeds first via the formation of a 2-aminoimidazole (2AI) bridged dinucleotide which then pairs with the template strand before attack of the primer’s (C3’)-OH on the bridged phosphate, extending the primer by one and releasing an activated 5’-monophosphate. Recently, canonical TNA has been thoroughly investigated. TNA is capable of being incorporated as the incoming dinucleotide, extended as the terminal primer nucleotide and copied as the template strand. Further crystallographic studies reveal that due to the conformation of the threose sugar the relatively slower rate (~100-fold) of primer extension observed for TNA is due in part to the distance between (C2’)-OH and incoming bridged dinucleotide.

Despite its clear appeal as potential progenitor to RNA and as a xeno-nucleic acid only limited attempts to understand the potentially prebiotic generation of TNA have been made.

### 2.1.2 Previous Prebiotic Syntheses of TNA Monomers

For TNA to be a plausible precursor to RNA it must have formed prebiotically. Despite the plethora of TNA’s merits as a genetic polymer discussed thus far and recent progress towards the chemical synthesis of RNA monomers, a thorough evaluation of the prebiotic synthesis of TNA monomers has not been carried out. Therefore, this chapter aims to find a synthetic route to TNA under prebiotically plausible conditions and answer the question: can TNA monomers be delivered alongside those of RNA?

Previously, the reasons behind inefficient and prebiotically-restricted glycosylations of sugars has been discussed (vide supra, 1.5.4.1). This remains a popular methodology
and represents the only attempts made towards prebiotic synthesis of TNA. While there have been some reports of threose sugar synthesis from the formose reaction and others (vide supra, 1.5.2.2), these reactions are not overly persuasive as a synthetic route due to low yields and unselective reactions. Additionally, threose 57 and its diastereomer erythrose 58 both readily undergo Lobry de Bruyn–Van Ekenstein reactions to erythrulose 131, possibly precluding their existence in aqueous prebiotic scenarios. Eschenmoser demonstrated (Scheme 41) that both phosphorylated tetrose sugar diastereomers are accessible in good yield (threo 132 48-80%, erythro 133 87% 9:1 with 134) by a regioselective phosphorylation with amidotriphosphate (AmTP) or (DAP) (diamidophosphate), but this still relies on pure tetrose sugars to begin with. Interestingly, for the reaction of erythrose 58 they also observed the 2′,3′-cyclic phosphate 134, this demonstrated the propensity of erythrose sugar’s cis-diol’s to undergo two condensation reactions even in aqueous solutions to give 2′,3′-cyclic phosphates (vide infra, 2.2.11).

Scheme 41  a) structures of potentially prebiotic phosphorylating agents diamidophosphate (DAP) and amidotriphosphate AmTP and (DAP); b) phosphorylation of 57 and 58 to their respective 1′,2′-cyclic phosphates 132/133 and 2′,3′-cyclic phosphate 134 for 58.
Building on this, some previous work by Kim and Benner (Scheme 42a) has shown that TNA 2'-phosphate adenosine 134 can be formed in low yields (<20%) from the pre-formed 1',2'-cyclic phosphate 132 in the presence of divalent metal ions (Ca²⁺), they were also able to glycosylate with non-canonical purine hypoxanthine 66 to yield inosine 134, also in low yields (<20%). This reaction does not appear to be compatible with pyrimidines as no cytidines are reported and TNA uridine 136 is only formed in very low yields (0.4%)³⁰¹. Using their FaPy chemistry (vide supra, 1.5.4.3) Becker et al. report the appearance of tetrose adenosines 137 when preformed tetrose is included in their reactions but do not report a yield²³⁴. These glycosylation reactions necessitate the use of preformed and activated nucleobases (FaPy-A 79) and sugars from formose-type processes.

**Scheme 42** Previous work towards TNA monomers a) Kim et al.'s synthesis of TNA A, I and U 134, 135 and 136 from preformed threose 1',2'-cyclic phosphate 132; b) reported tetrose purines from FaPy chemistry²³⁴.
Recent work has elucidated methods to avoid the need for these challenging glycosylation reactions and preformed sugars that invoke formose processes\textsuperscript{44,159,192,260}; this chapter explores the extension of this work to evaluate this method for the prebiotic synthesis of TNA.

A key element of the recent syntheses of RNA, ANA and DNA has been a common aminooxazoline structure. Interestingly, tetrose aminooxazolines (threose aminooxazoline TAO and erythrose aminooxazoline EAO) can also be efficiently synthesized (90\%) by the same reactions that yield pentose aminooxazolines RAO by the reaction of 2-aminooxazole 2AO with glycolaldehyde 3\textsuperscript{242}. Additionally, they are also the favoured products in reactions where both simple aldehydic sugars (gly 3 and glyceraldehyde 4) are incubated with 2AO, giving TAO (34\%) and EAO (38\%) after 6 d (phosphate buffer, RT) and only 5\% RAO and 3\% AAO\textsuperscript{244}. Therefore, the work in this chapter describes the development of a synthesis that could give TAO and would potentially therefore also yield TNA\textsuperscript{44,242,302}.

\begin{scheme}
\textbf{Scheme 43} Divergent synthesis of RNA and TNA cytidines through the reaction of cyanamide with prebiotically plausible aldehydes. The five-carbon RNA backbone requires sequential reaction of glycolaldehyde 3 (blue) and glyceraldehyde 4 (orange) generating four furanosyl pentose aminooxazolines (RAO, AAO, XAO and LAO) and one pyranosyl pentose aminooxazolines (p-LAO). The four-carbon TNA backbone requires only glycolaldehyde 3 (blue) and only generates two furanosyl isomers (TAO and EAO).
\end{scheme}
Irradiation by UV-light, as discussed (*vide supra*, 1.6.2), can be a powerful way to bring about molecular change in prebiotic chemistry, especially for nucleic acids. The pioneering work of Sanchez and Orgel and others (*vide supra*, 1.5.5) has eventually lead to the recent work by Sutherland and co-workers where the irradiation (254 nm) of α-ribo with hydrogen sulfide has achieved far higher anomerization yields than previously observed for cytidines even without blocking the 2′-hydroxyl to prevent oxazolidinone formation (*vide supra*, 1.5.5.2). Accordingly, a protecting-group-free, prebiotic synthesis from TAO to TNA involving photochemical anomerization was envisaged (Scheme 43) and will be described in this chapter.

### 2.1.3 Formation of 2-amino oxazole

2-Amino oxazole 2AO has been identified as a key building block of ribonucleotides and as one of three potentially privileged small molecules alongside 2AT (sequestration) and 2AI (NERPE rate enhancement) that are strongly implicated at the origins of life. Initial syntheses of these drew inspiration from early synthetic chemistry (NaOH in THF, pH 10-12) by Cockerill and it is now well established that a 1:1 mixture of cyanamide and glycolaldehyde can react efficiently to give 2AO at neutral pH under general base catalysis provided by inorganic phosphate (1 M, pH 7, >80% yield). The mechanism for this involves specific base (Cockerill) or general base promoted 5-exo-dig cyclisation and subsequent aromatisation to 2AO (Scheme 44).

![Scheme 44](image)

*Scheme 44* Mechanism for the formation of 2AO from glycolaldehyde 3 and cyanamide 87 the role inorganic phosphate represented as HA - general acid/base.

Additionally, recent work (*vide supra*, 1.5.5.3 and Scheme 45) has highlighted the ability of 2AT to effectively protect and sequester 3 by rapidly trapping (2 h) it as highly crystalline aminal. This averts any potential aldol processes destroying 3 after its production by photochemical HCN homologation (*vide supra*, 1.3.3.3). The glycolaldehyde 3 can then be released on reaction with 87 to give 2AO and addition of
another 3 can give both TAO and EAO\textsuperscript{242,302}. Taken together these realisations allows the confident use of 2AO as a key synthon in the assembly of TNA monomers as pentose aminooxazolines and TAO have been shown to be accessible from 2AO.

\textbf{Scheme 45} Islam and Powner's sequestration of glycolaldehyde 3 via crystalline aminal 93\textsuperscript{93}. Anastasi et al. have demonstrated that the efficient reaction 2AO and 3 to give both tetrose diastereomers TAO and EAO\textsuperscript{93}.
2.2 Results and Discussion

2.2.1 Tetrose-aminooxazoline

The elaboration of tetrose aminooxazolines to TNA is the subject of the investigation described in this chapter. Aminooxazolines can be formed by the reaction of 2AO with either glyceraldehyde or glycolaldehyde to generate pentose or tetrose aminooxazolines respectively\(^ {44,242,244}\). Accordingly, here 2AO was used for the synthetic preparation for tetrose aminooxazolines for further synthesis (Exp 5.4.1.2). When 2AO (9 mmol) and glycolaldehyde 3 (9 mmol) were incubated (RT, 16 h, pH 7) both diastereomers of the tetrose aminooxazolines (TAO/EAO) were afforded in good yield (78%) (Figure 7).

![Diagram of TAO and EAO](image)

**Figure 7** \(^ 1\)H NMR (600 MHz, D\(_2\)O, zg30, 3.0-6.0 ppm) spectrum of a mixture of threose aminooxazoline TAO and erythrose aminooxazoline EAO components showing the distinctive (C2')-H triplet and doublet for erythro and threo diastereomers respectively.

The reaction’s \(^ 1\)H NMR spectrum reveals the possibility to distinguish between the two diastereomers, TAO and EAO, and forms the basis for all analysis in this chapter. This is possible due to the splitting of the (C2')-H. As shown (Figure 7) the triplet at 4.85 ppm and a doublet at 4.80 are assignable and from \(^ 1\)H-\(^ 1\)H COSY data it is evident that these belong to two different species as they are not mutually coupled. On inspection of the structure of the two diastereomers it is evident that EAO must be the triplet as the (C3')-OH is equatorial giving resulting in the triplet environment. TAO has its (C3')-OH moiety
in the axial orientation giving the doublet environment observed. These environments are based on the dihedral angles and the couplings can therefore be deduced from the Karplus curve. **EAO** has the same dihedral angle between (C2′)-H and both (C1′)-H, (C3′)-H whereas **TAO** has an average dihedral angle of 90° between (C2′)-H and (C3′) -H giving a very low coupling constant and therefore no splitting by the (C3′)-H and the observed doublet environment. A characteristic triplet for the **EAO** (C4′)-H can also be observed.

**Scheme 46** Divergent routes to pentose (**RAO**, **AAO**, **XAO**, **LAO** and **p-LAO**) and tetrose aminooxazolines from cyanamide. The reaction of glyceraldehyde 4 (orange) can result in dead-end products 9 and 137. Putative one-pot reaction to yield **TAO/EAO** (green arrow) from addition of two molecules of 3 to 87.

The resulting mixture of two diastereomers was unsurprising as, for the pentose series, the reaction of **2AO** and the corresponding prebiotic aldehyde, glyceraldehyde 4, gives five diastereomers (Scheme 46). This is best explained mechanistically by the intermediacy of the open chain structure 138 which undergoes 5-exo-trig cyclisation with after the stereochemistry is set by the initial attack of **2AO** on 3 (Scheme 47).
These four furanosyl pentose aminooxazolines (RAO, AAO, XAO and LAO) and one pyranosyl pentose aminooxazoline (p-LAO) generate “prebiotic clutter” by diverting material from the desired aminooxazoline. Powner et al. utilised AAO in their original synthesis of activated ribonucleotides from pentose aminooxazolines but Xu et al.’s recent work has now also harnessed RAO and its superb crystalline properties in an efficient synthesis of ribonucleotides via a photochemical anomerization. Despite this the remaining pentose aminooxazolines have yet to be demonstrated as prebiotically useful and they would accordingly be converted by similar chemistry as for RAO/AAO to nucleotides and contaminate any later genetic polymers (vide supra, 1.4.6.1). Constructing aminooxazolines in the tetrose series therefore has several benefits over the pentose series. It simplifies the number of diastereomers to just two (TAO and EAO) versus five for the pentose series. Also, it revealed the possibility of a one pot synthesis from cyanamide with two molecules of glycolaldehyde whereas it would require one molecule of glycolaldehyde and one of glyceraldehyde to construct aminooxazolines in the pentose series.

The one pot synthesis of pentose aminooxazolines en route to RNA has been explored but either involves the spatio-temporal separation and sequential addition of glycolaldehyde and glyceraldehyde (as glycolaldehyde out competes glyceraldehyde, which is also unstable with respect to its conversion to DHA under the reaction conditions or sequestration of both aldehydes as aminals). The reaction between cyanamide and a 1:1 mixture of 3 and 4 revealed no selectivity with both 2AO and the dead-end product 137 of cyanamide were observed (Scheme 46).
Duly investigating this one-pot synthesis it was observed that when cyanamide 87 was incubated with varying amounts of glycolaldehyde 3 (>2 equiv., pH 7-11, 50-60 °C) TAO and EAO formation was observed (Figure 8 and Exp 5.3.1.1). As expected, at pH >8 the reaction was dominated by homo-aldol reactions of glycolaldehyde 3 forming complex mixtures that are unable to be resolved by 1H NMR spectroscopy.

It has been reported that the formation of 2AO can be promoted by basic conditions and when alkaline conditions (pH >9) were screened the formation of 2AO did occur, but the kinetics of the reaction favour the aldol products over rapid reaction with a second molecule of 3.

![Figure 8](image)

**Figure 8** 1H NMR (700 MHz, H2O/D2O 9:1, noesygpppr1d, 3.0-9.0 ppm) spectrum to show the reaction of cyanamide 87 (500 mM) and glycolaldehyde 3 (2.5 M) at pH 10, 60 °C after 1 d; 26% total yield 2:1 TAO/EAO.

Even when only two equivalents of glycolaldehyde 3 were used a small amount of TAO was observed (Exp 5.3.1.1). The yield of the reaction was low in most cases ~18% for TAO and ~10% EAO however this is similar to the yields achieved by Powner and Sutherland in their one-pot sequential addition scenario for RAO and AAO 25% and 15% respectively. However, 3 can all be added at once, whereas RAO/AAO synthesis require sequential addition of the separate two aldehydes 3 followed by 4.
Using multiple equivalents of 3 is consistent with recent work from the Sutherland group that indicates that 3 is easier to make than 4 and 87 as it is one of the first products of their HCN homologation pathway (*vide supra*, 1.3.3.3).

These results suggest that both the sequential addition of glycolaldehyde 3 to 2AO\textsuperscript{44,242,302} (even in mixtures with 4\textsuperscript{264}) and a one pot reaction of cyanamide 87 and multiple equivalents of 3 could give TAO/EAO in varying yields. Additionally, further investigation\textsuperscript{244} has found that addition of 2AO to mixtures of 4, DHA 9 and 3 favours reaction with 3 7:1; this is likely due to a combination of the lower steric impedance to reaction of 3 > 4, and the significantly increased reactivity of aldehydes over ketones combined with the loss of 3 to 9 at equilibrium (~20:1 9/4). With a prebiotic route to TAO and EAO established a subsequent elucidation of a route to TNA was investigated.

### 2.2.1.1 Tetrose Oxazolidinone thiones

Recently the divergent synthesis of 8-oxo-purine and pyrimidine ribonucleotides was demonstrated to proceed via a common *arabino*-oxazolidinone thione ara-96 (*vide supra*, 1.5.5.4). The reaction of glyceraldehyde 4 and 2-thiooxazole 95 (itself built from 3 and thiocyanate (HSCN)) gave all 5 pentose diastereomers of the oxazolidinone thiones. These could then be elaborated into purines by reaction with HCN trimer 60 (or 2-cyanoacetamide 99) after activation of the thione. Activation of the thione also allowed for conversion to AAO after ammonolysis. It was interesting to investigate the potential for this reaction in the tetrose series. Accordingly, the synthesis of 2-thiooxazole 95 was carried out following the procedure from *Stairs et al.*, glycolaldehyde 3 and thiocyanate react in a similar manner to 3 and cyanamide to make 2AO but instead here make 2-thiooxazole 95 (Scheme 48 and Exp 5.4.1.1). With the 2-thiooxazole 95 in hand a brief investigation of its reaction with 3 was undertaken.

When 2-thiooxazole 95 was incubated with glycolaldehyde 3 (2 equiv.) at pH 7, 60 °C for 2 d, oxazolidinone thione *threo*-96 (61%) and *erythro*-96 (23%) were observed. The characteristic resonances in the \textit{\textsuperscript{1}H} NMR for 2-thiooxazole 95 (15%) still remained indicating that there was not full conversion. Interestingly, the ratio of diastereomers was greatly in favour of the threose oxazolidinone thione *threo*-96 (3:1). This could provide a method of greater diasteroselectivity but further investigation of this pathway and its potential route towards TNA purines is beyond the scope of this thesis.
Scheme 48 Synthesis of tetrose oxazolidinone thiones from glycolaldehyde 3 and KSCN to give 2-thiooxazole 95 which then reacts slowly with 3 to give threo-96 and erythro-96.
2.2.2 Tetrose-anhydronucleosides

With the synthesis of TAO and AAO complete, attention was turned to constructing the remainder of the pyrimidine nucleobase. First Sanchez and Orgel and later Powner and Sutherland have demonstrated cyanoacetylene’s 12 remarkable ability to convert aminooxazoline structures into masked nucleobases/anhydronucleosides\(^{1,237}\). Cyanoacetylene 12 is a product of the spark discharge reactions of methane-nitrogen atmospheres\(^{223}\) both thought to present in the early Earth’s atmosphere (\textit{vide supra}, 1.3.2). It has also been observed in comets leading to being widely regarded as prebiotically plausible\(^{309}\). It has also been used historically and recently in the prebiotic construction of pyrimidine ribonucleosides\(^{225,236}\).

\begin{equation}
\text{Scheme 49} \quad \text{Powner and Sutherland’s synthesis of ara-89 from the phosphate-buffered cyanovinylation of AAO.}
\end{equation}

Sanchez and Orgel first revealed that RAO and AAO are converted directly to \(\alpha\)-ribo-85 and \(\beta\)-ara-85 when subjected to reaction with aqueous cyanoacetylene but despite a small excess of 12 they did not observe full conversion\(^{237}\). The pH of the reaction rises due to the consumption of protons to quench the incipient \(Z\)-cyanovinyl anion intermediate (Scheme 53). This in turn leads to in-situ hydrolysis of the anhydro (e.g. ara-89, Scheme 49) to the corresponding nucleoside in low yield (\(\sim 10\%\)).
The selectivity of this reaction is also low and Powner and Sutherland subsequently elucidated the full scope of the reaction products (Scheme 50)\textsuperscript{44}. Both RAO and AAO undergo 2', 3' and 5' hydroxyl cyanovinylolation in varying degrees as well as for the conversion of RAO its cis-hydroxyls also allow the formation of the 2',3'-cyclic structure.

Scheme 50 Cyanovinylolation products identified by Powner and Sutherland for the unbuffered reaction of RAO and AAO with cyanoacetylene 12\textsuperscript{44}. No anhydronucleoside was observed and instead a variety of O-cyanovinylated species were detected.
The reaction can be driven to completion by using at least 7 equiv. of cyanoacetylene 12 at pH 7 but when 1 equivalent of sodium phosphate was added, the same compound used to promote the formation of 2AO (Scheme 44), a remarkable shift in chemoselectivity was observed. Instead of the major product being a result of hydrolysis, almost total conversion (92%, pH 6.5, 7 equiv. 12, Scheme 49) to the corresponding anhydronucleoside ara-89 was observed and only very minor hydrolysis to β-ara-8544. This then allowed subsequent phosphorylation to give the desired activated ribonucleotides (vide supra, 1.5.5.2). This methodology should be compatible with the conversion of both TAO and EAO to threo-89 and erythro-89 respectively. The desire to explore simultaneous conversion of both diastereomers was influenced by the potential for convergent or divergent chemistry later in the synthesis was appealing.

Accordingly, this elaboration of the remaining pyrimidine nucleobase for the tetrose series was the next step. Commercially unavailable, the preparation of cyanoacetylene 12 involves the ammonolysis of methyl propiolate 139 and subsequent dehydration of the propiolamide 140 formed by P₂O₅ and dry distillation to give cyanoacetylene as a white solid that was immediately dissolved in a known amount of water to provide an aqueous solution of cyanoacetylene 12 of known concentration (Scheme 51 and Exp 5.4.1.4).

\[
\begin{align*}
\text{Me} & \quad \text{NH}_3 \\
\equiv & \quad -78 \degree C \\
\text{139} & \quad \equiv \\
\text{NH}_2 & \quad \Delta \\
\text{140} & \quad \equiv \\
& \quad \text{12}
\end{align*}
\]

*Scheme 51 Synthesis of cyanoacetylene 12 from methyl propiolate 139 via propiolamide 140.*

The reaction of the aminooxazolines TAO and EAO was then investigated. When a diastereomeric mixture of TAO and EAO were subjected to cyanovinylation (12 10 equiv., pH 6.5, RT, 16 h) appearance of characteristic resonances, in the crude ¹H NMR spectra, in the 8.05-8.10 ppm and 6.45-6.75 ppm regions were indicative of newly formed (C6)-H and (C5)-H protons along with downfield shifting of the (C1′)-H. Conversion was confirmed by 2D NMR and appeared to show full conversion as no signals indicative of TAO or EAO were observed. Both TAO and EAO were equally receptive to the cyanovinylation which presented no means by which to differentiate the synthesis. Attempted chromatographic separation proved unsuccessful and indeed even
separation of *threo*-89 and *erythro*-89 from CVP proved challenging. Instead a more accurate method of ascertaining the yield was sought.

Quantification of the reaction by carrying out the reaction in the presence of an internal standard was then developed. Methylsulfonylmethane (MSM) is inert under the reaction conditions and was used as an internal standard that gives has a characteristic singlet resonance in the $^1$H NMR, upfield of the area where TAO and EAO and their respective anhydronucleosides characteristic protons are observed (3.09 ppm). A control reaction indicated that, despite its reactivity, no reaction was observed between cyanoacetylene 12 (20 uL, 1M) and MSM (methylsulfonylmethane, 50 mM) when incubated for several days and the reaction progress followed by $^1$H NMR. This confirmed the applicability of MSM as an internal standard in this reaction.

When TAO (123 mmol) and EAO (123 mmol) were incubated with cyanoacetylene 12 (1 mol) at pH 6.5 with 1 equivalent of sodium phosphate (Na$_2$HPO$_4$) buffer clean cyanovinylation was observed with near quantitative conversion to *threo*-89 and *erythro*-89 (98\% by qNMR with MSM) (Scheme 52 and Exp 5.3.1.3). Only 1 equiv. of phosphate is needed which is advantageous as despite recent reports of carbonate lakes concentrating phosphate its presence in high amounts remains contentious due to its tendency to form apatite minerals in the presence of calcium\(^{45}\).

![Scheme 52](image_url)

_Scheme 52 High yielding, completely regioselective synthesis of *threo*-89 and *erythro*-89 from TAO and EAO (Cyanoacetylene 12 10 equiv., phosphate buffer). CVP was the only observable by-product._

The reaction was completely regioselective with no $O$-cyanovinylation observed due to the chemical and pH buffering of phosphate and accordingly characteristic resonances in the 7.20-7.30 ppm and 4.55-4.65 regions indicated concomitant formation of CVP. The reaction is believed to occur via attack of phosphate oxygen on the alkyne and protonation via another molecule of phosphate to give cis-$\beta$-cyanovinyl phosphate CVP as the sole by-product of the reaction in the tetrose series (Scheme 53b). CVP has been
suggested as potential phosphorylating agent in prebiotic chemistry and has been used to obtain pyrophosphate from orthophosphate by Ferris et al.\textsuperscript{310}. Phosphate also ensured negligible hydrolysis to β-\textit{threo}-85 and β-\textit{erythro}-85, observed in unbuffered reactions but not here, indicated by the absence of resonances 7.75-7.85 ppm in the $^1$H NMR.

\begin{center}
\textbf{Scheme 53} a) mechanism for the formation of both tetrose anhydronucleosides \textit{threo}-89 and \textit{erythro}-89; b) mechanism for the formation of cyanovinyl phosphate CVP
\end{center}

The mechanism (Scheme 53) is proposed to proceed via attack of the exocyclic imine atom of TAO/EAO on the alkyne of 12 with concomitant quenching of the resultant Z-cyanovinyl anion, presumably by phosphate, to give the N\textsuperscript{1} cyanovinyl adduct. This undergoes 6-\textit{endo}-dig ring closure by the exocyclic imine to give the anhydronucleosides \textit{threo}-89 and \textit{erythro}-89.

With a regioselective, high yielding synthesis of \textit{threo}-89 and \textit{erythro}-89 established a route to TNA could then be explored.
2.2.3 Thiolysis of tetrose anhydronucleosides in formamide

Sulfur-containing nucleobases have recently been investigated due to their ability to alter the native photochemistry of nucleosides. Xu et al. have demonstrated that (Scheme 54a), en route to β-ribo-85, an efficient (up to 84%) thiolysis in formamide of ribo-89 allowed access to the α-ribo-2-thiocytidine β-ribo-111 (vide supra, 1.6.2). Hydrogen sulfide and its conjugate base hydrosulfide (HS⁻, pKa = 7.2) had revealed itself to be a good source of solvated electrons when irradiated (254 nm) for use in Sutherland and co-worker’s cyanosulfidic protometabolism chemistry (vide supra, 1.3.3.3). Therefore, it appealed to them as an ideal reagent for thiolysis due to its highly plausible prebiotic origin. Hydrogen sulfide is widely produced in volcanism and can also be produced by the dissolution of a metal sulfide in cyanide solution, while formamide is the first hydrolysis product of hydrogen cyanide (vide supra, 1.3.3.2).

Additionally, Roberts et al. have shown (Scheme 54b) how thiolysis of several anhydropurines (vide supra, 1.6.2.1) can allow positioning of sulfur at C8 and with a subsequent peroxide reduction allow access to purine arabinosides.

Scheme 54 a) Xu et al.’s H₂S thiolysis of ribo-89 in formamide en route to β-ribo-85; b) Roberts et al.’s synthesis of purine arabinosides via aqueous H₂S thiolysis of anhydropurines.
These recent examples of anhydronucleoside thiolysis in formamide (or water) indicated that a similar reaction should be available in the tetrose series and would selectively position a sulfur at C2. This would result in the formation of β-tetose thionucleosides (β-threo-111 and β-erythro-111) with a cis-1′,2′ relationship that is incompatible with Watson-Crick-Franklin base pairing. Fortunately, as Xu et al. have shown, inclusion of sulfur at C2 can markedly increase the efficiency of photochemical anomerization for the ribo-cytidines and this should be accessible in the tetrose series. Accordingly, the thiolysis of the threo-89 and erythro-89 was investigated.

\[ \text{Figure 9} \quad ^1H \text{ NMR (700 MHz, D}_2\text{O, zg30, 5.0-8.5 ppm) spectra to show: a) threo-89 and erythro-89 (49 mM); b) the products of the reaction of a) with NaSH (635 mM) in formamide (1 mL) at room temperature overnight after then being lyophilized and dissolved in D}_2\text{O; c) spectra of authentic β-threo-111; d) spectra of authentic β-erythro-111. The products β-threo-2-thiocytidine β-threo-111 and β-erythro-2-thiocytidine X β-erythro-111 (78% yield) are labelled: (C6)-H, (C5)-H and (C1′)-H protons are labelled: β-threo-111 (▼) and β-erythro-111 (▲).} \]
While Xu et al. heat their reaction (50 °C, 7 h), formamide thiolysis with sodium sulfide (NaSH 635 mM) of a diastereomeric mixture of *threo*-89 and *erythro*-89 (49 mM) proceeds smoothly at ambient temperatures overnight (78% by qNMR). Confirmation by ¹H NMR was enabled by the shift of the (C1′)-H downfield (characteristic resonances at 6.85-6.88 ppm) for both diastereomers and the shift of the (C2′)-H resonance upfield of HOD resonance (4.75 ppm) as well as comparison with authentic samples of β-*threo*-111 and β-*erythro*-111 (Figure 9).

Due to the varying CVP and salt content of the reaction to form the tetrose anhydrocytidines (*threo*-89 and *erythro*-89) and the difficulty in removing formamide from the reaction mixture (b.p. 210 °C), a method was developed for accurately measuring the yield of this reaction (Exp 5.3.1.4 for full procedure). Here, the volatility of methylsulfonylmethane (MSM) was exploited as a noninteracting internal standard to calibrate the initial and final amounts of *threo*-89 and *erythro*-89.

With the thiolysis of *threo*-89 and *erythro*-89 giving high conversion to their respective thionucleosides (β-*threo*-111 and β-*erythro*-111) an investigation into the proposed photochemical anomerization could then be carried out. However, the chemistry of these key tetrose thionucleoside intermediates was also investigated. For this, a scalable synthetic method of preparing separable β-*threo*-111 and β-*erythro*-111 was sought.

### 2.2.4 Scalable thiolysis procedure

Due to the difficulty in purifying β-*threo*-111 and β-*erythro*-111 from formamide a two-step procedure was developed to produce the thiolysis products directly from TAO and EAO (Exp 5.4.1.6). Previous synthetic preparations of 2-thiocytosine nucleosides have used alternate solvents to water in order to enable regioselective thiolysis. DMF is an excellent solvent for polar nucleophilic reactions (such as this thiolysis) and has been employed in previous thiolyses and was used here¹⁵⁹,²⁶⁰.

TAO and EAO (2.1 mmol) were first incubated with cyanoacetylene (4 equiv., RT, 18 h, 1 equiv. phosphate) and then lyophilised. The lyophilizate was submitted to thiolysis (NaSH, 6.6 mmol) in DMF (1 equiv. NH₄HCO₃, RT, 18 h) to give β-*threo*-111 and β-*erythro*-111. As DMF is more facile to remove than formamide (b.p. 153 °C) this process could be scaled up if required, but the safety concerns of quenching large amounts of sodium sulfide preclude multigram scale.
Scheme 55 Two step synthesis of $\beta$-threo-111 and $\beta$-erythro-111 from TAO and EAO respectively.

With nucleosides $\beta$-threo-111 and $\beta$-erythro-111 both formed, attempts to separate the mixture of diastereomers was attempted in order to enable unambiguous studies of their photochemistry. Pleasingly, they readily separated when submitted to reverse phase flash column chromatography (RP-FCC). With access to diastereomerically pure $\beta$-tetrose thiocytidines an investigation into the photochemistry and oxidative behaviour of these compounds could be undertaken.
2.2.5 Hydrolysis of Anhydros

As discussed, it has previously been observed that hydrolysis of a transient anhydronucleoside to α-ribo-85 and β-ara-85 can occur when cyanovinylating RAO and AAO or after isolation of the anhydrocytidines. Indeed, hydrolysis to the oxygenic β-threocytidines (β-threo-85 and β-erythro-85) is one fate for anhydronucleosides threo-89 and erythro-89. This is accessible when threo-89 and erythro-89 are incubated at pH >7, even at room temperature (Scheme 56a). The higher the pH, the more extensive the amount of hydrolysis, as the reaction presumably relies on attack of hydroxide at the 2',2'-junction (Scheme 56b and Exp 5.3.1.6).

Scheme 56  a) alkaline hydrolysis of threo-89 and erythro-89; b) mechanism for the hydrolysis in a).

For the purpose of later photochemical studies an additional synthetic hydrolysis method for the preparation of was used involving DOWEX® (OH⁻ form) hydrolysis (RT, 4 h) and subsequent purification by RP-FCC (H₂O/MeOH) which yielded a pure diastereomeric mixture of β-threo-85 and β-erythro-85 (Exp 5.4.1.9).
2.2.6 Oxidative Recyclization of β-tetrose Thionucleosides

Previously, Roberts et al. had observed that β-ara-111 can undergo recyclization to ara-89 (Scheme 57a)\textsuperscript{159}. This implied that, for structures with cis-1′,2′ stereochemistry, intramolecular trapping of the proposed sulfone/sulfonate is far more rapid with the (C2′)-OH of the sugar than intermolecular reaction with bulk water which instead would lead to hydrolysis of the C=S bond\textsuperscript{159,311}. The exact nature of the sulfur species is as yet unclear (Scheme 57b). It may be eliminated as the sulfur dioxide and subsequently oxidized to sulfur trioxide (SO\textsubscript{3}) or be eliminated as sulfite after the sulfonate moiety is formed at C2. Regardless, this mode of reactivity was investigated for its ability to recyclize β-threo-111 and β-erythro-111 back to their corresponding anhydros (threo-89 and erythro-89).

\begin{center}
\begin{tabular}{c}
\textbf{Scheme 57} a) Roberts et al.’s peroxide-mediated recyclisation of β-ara-111 to ara-89\textsuperscript{159}; b) Proposed intermediate for the activation of the C2 sulfur by oxidation and subsequent re-cyclisation via attack of the (C2′)-OH\textsuperscript{311}.
\end{tabular}
\end{center}

Indeed, when β-threo-111 and β-erythro-111 (17 mM) were reacted with hydrogen peroxide (H\textsubscript{2}O\textsubscript{2}, 98 mM) in neutral water, the rapid and totally quantitative cyclization to the corresponding anhydro (threo-89 and erythro-89) was observed (Scheme 58 and Exp 5.3.1.7). Pleasingly, no hydrolysis was observed, therefore showing that this reactivity enables effective protection against potential slow C=S bond hydrolysis\textsuperscript{260} to β-threo-85 or β-erythro-85 during periods of dark chemistry (periods of time with very low UV surface actinic flux) (Scheme 58). This would prevent any photochemistry of β-threo-85 and β-erythro-85 taking place during subsequent periods of irradiation, which has been proposed to be less efficient and is subsequently investigated here (vide infra, 2.2.8.2). Additionally, it opens the potential for another thiolysis and subsequent photochemistry of β-threo-111 or β-erythro-111 when irradiation is available.
Stereochemistry dictates this reaction cannot occur for *trans*-1′,2′ systems meaning that if β-*threo*-111 and β-*erythro*-111 can anomerize efficiently they are then protected from recyclization and would likely instead undergo oxidase hydrolysis to the desired TNA cytidine.

**Scheme 58** Reaction of β-*threo*-111 and β-*erythro*-1111 (17 mM) with hydrogen peroxide (H₂O₂) (98 mM), pH 7 give quantitative conversion to *threo*-89 and *erythro*-89 and avoids hydrolysis to β-*threo*-85 and β-*erythro*-85. This enables another thiolysis and further reactivity towards TNA.

With an investigation into the thiolysis and hydrolysis of the tetrose anhydronucleosides completed and the oxidative chemistry of the thiolysis products investigated, the desired photochemical anomerization could then be investigated.
2.2.7 Synthesis of Authentic Tetrose Nucleic Acids

To provide unambiguous confirmation for the photochemical conversion of both tetrose thiocytidines (β-threo-111 to α-threo-111 and β-erythro-111 to α-erythro-111) authentic standards of these compounds were prepared. Unlike for the prebiotic synthesis of TNA, the contemporary chemical synthesis of TNA is comparatively widely studied. A modified version of the synthetic routes of Schöning et al. and Sau et al. was investigated here\(^{260,269,270,312}\). A synthesis was envisaged that would be modified to instead use global protection/deprotection strategies as there was no need to produce DMT functionalized nucleotides for solid phase oligonucleotide synthesis. Additionally, a 2-thiocytosine modified Vorbrüggen-Hilbert-Johnson glycosylation would also be a departure from Sau et al.’s use of benzoylated canonical nucleobases as glycosyl acceptors and instead recent work from the Sutherland lab was investigated as they had previously glycosylated ribose sugars with 113. The desired glycosyl donor precursor in this case, threose 57, is both commercially unavailable and unstable with respect to its conversion via Lobry de Bruyn–Van Ekenstein chemistry to erythrulose 131 necessitating its synthesis. Additionally, as there was a need for confirmation of the photochemical behaviour of β-erythro-111 a route towards the erythrose sugar glycosyl donor was also developed. Fortunately, erythronolactone 143 is commercially available hence the differing routes to the two diastereomers presented here.

![Scheme 59: Peroxide oxidative cleavage of L-ascorbic acid 141 to L-threonate 142.](image)

**Scheme 59** Peroxide oxidative cleavage of L-ascorbic acid 141 to L-threonate 142.

Accordingly, to synthesise α-threo-111 this synthesis started with the preparation of L-threonate 142 from commercially available L-ascorbic acid 141 (Scheme 59). This oxidative cleavage by \(\text{H}_2\text{O}_2\) and \(\text{CaCO}_3\) proceeded in good yield (60%) and was possible to carry out on a large scale (25 g).

Next the synthesis of L-threonolactone 143 was carried out according to the procedure reported by Sau et al. (Scheme 60) in which the calcium is chelated by oxalic acid and ring-closing lactonization is promoted by acid catalysis (p-Toluene sulfonic acid) at reflux (3 h). This afforded 143 in excellent yield (94%).
Scheme 60 High yielding, acid catalysed lactonization of 142 to 143 and simultaneous Ca\(^{2+}\) chelation by oxalic acid.

With the synthesis of threonolactone 143 complete and commercially available erythronolactone 144 in hand, the appeal of global protection of both 2’ and 3’ hydroxyls was apparent. The envisaged glycosylation reaction necessitates the use of a 2’-directing group, common directing groups for these glycosylations that are able to form the needed 1’,2’-cyclic structures with Lewis acids (SnCl\(_4\), TMOSTf) include both acetyl and benzoyl groups. Schöning \textit{et al.} have previously reported that 2’,3’-di-O-acetyl-threonolactone was too volatile to isolate. Instead the focus here was turned to benzoyl groups as they have proved more amenable to isolation as reported by Schöning \textit{et al.} This reaction was investigated here for its ability to yield bis-benzoylated tetrose lactones. Both \textit{threo}- and \textit{erythro}-lactones 143/144 were successfully converted to their corresponding 2’,3’-di-O-benzoyl-lactones 145 and 146 when reacted with benzoyl chloride (2-3 equivs) and catalytic DMAP in pyridine with no mono-benzoylation detected (Scheme 61).

Scheme 61 Bis-O-benzoylation of tetrose lactones with benzoyl chloride in pyridine.
A two-step procedure to convert the bis-benzoyl lactones to their corresponding nucleoside precursors was then developed (Scheme 62b). This involved the reduction of lactones 145 and 146 and subsequent 1'-acetylation of the incipient hydroxyl generated. This was followed by a Vorbrüggen-Hilbert-Johnson reaction with the now 1’-acetylated sugar as the glycosyl donor and transiently silylated 2-thiocytosine 113 as the glycosyl acceptor.

Scheme 62 a) Preparation of silylated 2-thiocytosine 113; b) two step procedure to α-threo-147/ α-erythro-147 from bis-benzoylated lactone 145/146; c) Overview of the proposed mechanism for the Silyl-Hilbert-Johnson glycosylation, shown here with 2',3'-di-O-benzoyl-tetrose sugars as glycosyl donors and silylated 2-thiocytosine 113 as the glycosyl acceptor. Note the intermediary of the 1',2'-cyclic sugar structure after activation by Lewis acid SnCl₄ which acts as a regioselective directing group to ensure total α-glycosylation.
When threonolactone 143 (2.15 mmol in THF) was subjected to reduction (DIBAL-H, 3 mmol, -78 °C) for 10 h and acetylated in-situ with acetic anhydride and DMAP in DCM (1.2 ml, 12.7 mmol/57 mg, 0.47 mmol/1.5 ml), analysis of the intermediate of this reduction by NMR spectroscopy revealed a mixture of α/β anomers were formed. However, when erythronolactone 144 (5 mmol in THF) was submitted to reductive conditions (DIBAL-H, 13.8 mmol, -78 °C, 10 h) and a similar acetylation, only one anomer was observed. 1H NMR analysis also revealed the tendency for this reaction to stall in its conversion, instead of observing potential over reduction of the 2′/3′ benzoyl groups at long reaction times (>12 h) instead remaining starting material was observed. This is consistent with previous work from Eschenmoser where a maximum yield of 53% was achieved\(^{270}\).

Xu et al. have previously synthesised authentic standards for their investigation of the photochemistry of β-ribo-111 (vide supra, 1.6.2) utilising transient silylation of 113 by Bis(trimethylsilyl)acetamide (BSA) and this seemed appropriate for these tetrose compounds as well\(^{260}\). When 113 was heated at reflux with BSA under anhydrous conditions it became rapidly homogenous, however under less strict anhydrous conditions this reaction was slow or simply did not occur. The precise intermediate is not known but presumably some combination of N\(^4\) and S\(^2\) silylation occurs (Scheme 62a), and when added to the corresponding glycosyl donor, after Lewis acid activation (SnCl\(_4\), DCE, 0 °C), the observed regio-selectivity of the glycosylation was high. This is presumably due to effective formation of the 1′,2′-cyclic sugar intermediate (Scheme 62c) which acts as a regioselective directing group to ensure total α-glycosylation. Indeed, the two major products of the reaction were the desired protected nucleosides (α-threo-147 and α-erythro-147) and their corresponding starting lactones (143/144) which had not undergone reduction. These proved facile to separate by FCC (DCM/MeOH) with the desired compounds eluting at 1-5% MeOH. As these reactions gave highly pure products in good yields (two step yields: 64% α-threo-147, 56% α-erythro-147) and were relatively simple to carry out they were not optimized further.

\[
\begin{align*}
&\text{BzO} & \text{OBz} & \text{NH}_2 \\
&\text{O} & \text{N} & \text{NH}_2 \\
&\alpha-\text{threo-147} & \alpha-\text{erythro-147} & \\
\text{NH}_3 & \rightarrow & \text{MeOH, 60 h} \\
&\text{HO} & \text{OH} & \text{S} \\
&\text{O} & \text{N} & \text{NH}_2 \\
&\alpha-\text{threo-111} & \alpha-\text{erythro-111} & \\
&94\% & 93\% \\
\end{align*}
\]

\textbf{Scheme 63 High yielding ammonolysis of tetrose 2′,3′-bis-O-benzoylthiocytidine}
Instead the subsequent deprotection to nucleoside was investigated (Scheme 63). Ammonolysis (7 M NH₃/MeOH, 72 h) in MeOH gave the corresponding deprotected nucleosides in near quantitative yields (α-threo-111 94%, α-erythro-111 93%); by-product benzamide is facile to remove by washing with DCM (not possible if acetyl protection was used), thereby requiring minimal purification. These were then used for sample spiking of subsequent NMR reactions.
2.2.8 Photochemical anomerization

Having synthesized the relevant authentic standards for product confirmation by sample spiking, the photochemical anomerization of β-threo-111 to α-threo-111 and β-erythro-111 to α-erythro-111 was investigated.

Despite the elegance of photochemical anomerizations, they have been historically low yielding for pyrimidines (less than 10% conversion) and have not been demonstrated for purines. The reasons for this have since been elucidated (vide supra, 1.5.5.2) and the discovery of oxazolidinone formation as the major destructive pathway lead to work investigating the blocking of the (C2′)-OH from participating in trapping excited nucleobases. This involved 2′-phosphorylation (among other strategies, acetylation, 2′-deoxygenation) and has modestly improved anomerization yields. Recently, 2-thiocytidines have been investigated for their ability to undergo photochemical anomerization. In stark contrast in both mode and efficiency to the photochemical reduction exhibited by 8-thiopurines (thiolysis products of anhydropurines, vide supra, 1.6.2.1) and the low yielding photochemistry of oxygenic cytidines (divergent hydrolysis product of anhdyrocytidines, vide infra, 2.2.8.2), 2-thiocytidines undergo remarkably efficient photochemical anomerization in neutral water in the presence of hydrogen sulfide (H₂S) and exhibit no propensity for photochemical reduction. Xu et al. showed conversion of α-ribo-111 to β-ribo-111 by anomerization (Scheme 64) in up to 76% yield. As they harnessed of the previously underutilised ribose aminooxazoline RAO and Powner and Sutherland has previously elucidated a route to RNA pyrimidine ribonucleotides from arabinose aminooxazoline AAO this opened opening up a convergent route for two of the 5 pentose aminooxazolines to become RNA pyrimidines260.

Mechanistic investigations for this anomerization (α-ribo-111 to β-ribo-111) have suggested (Scheme 64) an interrupted Norrish type 2 mechanism whereby heterolytic fission of the C=S bond allows an S-centered radical to abstract the anomeric (C1′)-H proton of the ribose sugar. This incipient C-centered radical causes planarization of the base with respect to the sugar and establishes an equilibrium where presumably the stereochemistry of the (C2′)-H proton enforces an equilibrium position such that a proton is picked up so that the base is now in the desired trans-1′,2′ β-ribo-stereochemistry.
With this understanding, it seemed likely that a similar mechanism for anomerization should be possible in the tetrose series and therefore warranted investigation as this would allow access, after a presumably facile oxidation or hydrolysis, to the canonical nucleobase TNA cytidine (or erythroxytidine) containing the required trans-1′,2′ stereochemistry for Watson-Crick-Franklin pairing.

The photochemistry of the isolated nucleobases (adenine, guanine, cytosine, uracil, thymine) has been studied in detail and is covered in detail in Chapter 4 but briefly, when studied by laser pump-probe experiments they exhibit ultrafast radiative decay to their ground states via readily accessible conical intersections. This alludes to a very low quantum yield for these species. UV-Vis absorption spectra of both β-threo-111 and β-erythro-111 (Exp 5.3.1.10) indicated strong absorbance close to 254 nm, this is in contrast to the effect sulfur has on purines where it causes a dramatic shift of the principle absorbance to ~300 nm. UV-light at these wavelengths (200-300 nm) is not believed to have been attenuated by atmospheric effects on the early Earth (vide supra, 1.3.2) allowing access to ideal wavelengths for photochemistry. Accordingly, the wavelength selected for irradiation here was 254 nm to maximize the potential absorption by the nucleobase and promote productive chemistry.
Figure 10 $^1$H NMR (700 MHz, $H_2O/D_2O$ (9:1), Robust5_ver_03b, 5.5-7.85 ppm) spectra to show the irradiation (254 nm) of β-threo-111 (25 mM) in $H_2O/D_2O$ (9:1) at pH 7 with NaSH (0.013 mmol, 26 mM): a) before irradiation, b) after 1 day of irradiation (46%), c) after 1.5 days of irradiation (51%), d) after 2 days of irradiation (51%), e) after 2.5 days of irradiation (49%), f) after 3 days of irradiation (49%), g) after 3.5 days of irradiation (43%) and h) after spiking g) with authentic α-threo-111. (C6)-H, (C5)-H and (C1′)-H protons are labelled: β-threo-111 (▼), α-threo-111 (●).
Indeed, after β-threo-111 (25 mM) was irradiated (254 nm) in neutral water in the presence of hydrogen sulfide (H₂S, 26 mM) photochemical anomerization was observed (Figure 10 and Exp 5.3.1.11). After 1 d good conversion to α-threo-111 was observed (46%) with maximum conversion after 1.5 d (51%). Anomerization was indicated by the appearance of a new characteristic singlet resonance in the 6.25 ppm region in the ¹H NMR spectrum and was confirmed by sample spiking with authentic sample of α-threo-111 (synthesised as discussed). Hydrogen sulfide was present to “minimize photooxidation by adventitious oxygen” and indeed no signals indicating recylization to threo-89 were observed, instead some destruction of the nucleosides occurred with the yield of α-threo-111 after 3.5 d being 43%. Additionally, this pathway does not seem to be susceptible to deleterious formation of oxazolidinones despite the cis-1′,2′ stereochemistry as resonances indicating their formation (doublet at ~5 ppm) were observed by ¹H NMR.

As thiolysis of the tetrose anhydronucleosides resulted in simultaneous production of both diastereomers, their concomitant irradiation was also investigated. When a mixture of β-threo-111 (18 mM) and β-erythro-111 (16 mM) was irradiated (254 nm, NaSH 113 mmol, pH 7), the rate of anomerization for threo and erythrothiocytidines were similar (Figure 11 and Exp 5.3.1.12) and after 1 d had furnished α-threo-111 and α-erythro-111 in 40% and 42% yield respectively. Similarly, after 2 d maximum conversion was observed of 52% and 51%. Sample spiking with authentic samples of α-threo-111 and α-erythro-111 confirmed that both threo and erythro can undergo simultaneous anomerization in and at similar rates.

The efficiency and selectivity of these experiments show that a route to TNA cytidine via photochemical anomerization of the products of thiolysed tetrose anydros may be viable. Accordingly, an investigation of the stability of the newly formed α-threo-111 and α-erythro-111 to UV-light was investigated, before an exploration of sulfur removal to give TNA cytidines.
Figure 11 $^1$H NMR (700 MHz, noesypppr1d, $H_2O/D_2O$ (9:1), 6.1-8.3 ppm) spectra showing the irradiation (254 nm) of $\beta$-threo-111 (18 mM) and $\beta$-erythro-111 (16 mM) at pH 7 with $H_2S$ (113 mM): a) before the irradiation; b) after 1 day of irradiation ($\beta$-threo-111 (40%), $\beta$-erythro-111 (42%)); c) after 2 days of irradiation ($\beta$-threo-111 (52%), $\beta$-erythro-111 (51%)); d) after 3 days of irradiation ($\beta$-threo-111 (40%), $\beta$-erythro-111 (36%)). Spectrum e) following addition of authentic $\alpha$-threo-111. Spectrum f) following addition of authentic $\alpha$-erythro-111. (C6)$H$ and (C1′)$H$ resonances are labelled: $\beta$-threo-111 (▼), $\beta$-erythro-111 (▲), $\alpha$-threo-111 (●) and $\alpha$-erythro-111 (●).
2.2.8.1 Irradiation of Tetrose α-Thiocytidines

Once formed, by anomerisation or otherwise, nucleic acids would continue to be irradiated by UV-light on the early Earth. The problem of nucleoside stability to UV-light is of paramount importance to understanding the key selection pressures acting at the origin of life and is the focus of Chapter 3. Here, the stability of α-threo-111 and α-erythro-111 was investigated by irradiation (254 nm) in neutral water with hydrogen sulfide. When α-threo-111 and α-erythro-111 (6.5 mM, 6.8 mM) were irradiated (254 nm) for multiple days appreciable destruction was observed as expected (Exp 5.3.1.21). The appearance of 2-thiocytosine 113 was expected and is indicative of a mechanism of destruction that proceeds via scission of the anomeric bond. However, anomerization to the undesired β-form was observed only in very low amounts (Exp 5.3.1.21). This is in contrast to Xu et al. where deuteration studies indicated a final ratio of β:α after 36 hours was 88:12 showing >10% anomerization to the undesired α-ribo-111.

![Scheme 65 Products of the irradiation of α-threo-111 and α-erythro-111 (6.5 mM, 6.8 mM), 2-thiocytosine 113 was the major product with only trace amounts of β-nucleosides observed and no observable sugar.]

These results indicate that while both α-threo-111 and α-erythro-111 are not indefinitely stable under UV-light they are stable enough to survive under the conditions of their own formation and are therefore suitable intermediates en route to TNA.

2.2.8.2 Attempted Anomerisation of Oxygenic Tetrose β-cytidines

The photochemical reactions of the oxygenic hydrolysis products of tetrose anhydronucleosides (threo-89 and erythro-89) was also of interest. If hydrolysis occurs during anhydronucleoside formation due to lack of buffering, e.g. if phosphate levels are low, then as previously described the anhydro can undergo hydrolysis to β-threo-85 and β-erythro-85. Additionally, the photochemical epimerisation of arabino-85 (vide supra, 1.5.5.1) to ribo-85 was considered to be potentially applicable here as potential convergent photochemistry could enable conversion of both threo and erythro diastereomers to the desired trans-1′,2′ relationship.
When both tetrose diastereomers (β-threo-85 and β-erythro-85) were irradiated (254 nm or 300 nm) in water (pH 6.5) with or without additives (sodium sulfide), no anomerization to α-threo-85 and α-erythro-85 was observed (Exp 5.3.1.15 and 5.3.1.16) and instead rapid formation of oxazolidinones threo-88 and erythro-88 and uridine β-threo-86 was observed, presumably by an identical mechanism to that proposed in previous work by Powner and Sutherland (vide supra, 1.5.5.2)\textsuperscript{240}.

To test this, bis-acetylation of the 2′ and 3′-hydroxyls was explored, as this was proposed to block formation of oxazolidinone formation\textsuperscript{240}. Acetyl imidazole 50 has been demonstrated to be a highly efficient prebiotically plausible regioselective acetylation agent and was used here (Exp 5.4.1.10)\textsuperscript{116}. 2′,3′-Bis-acetylated threose β-threo-148 exhibited no detectable anomerization or oxazolidinone formation when irradiated (254 nm) but do degrade over time (Exp 5.3.1.17).

Scheme 66 Hydrolysis of threo-89 and erythro-89 gives β-threo-85 and β-erythro-85 which instead of anomerization undergo destruction to oxazolidinones threo-88 and erythro-88 when irradiated. Blocking the (C2′)-OH by prebiotic acetylation reduces the extent of oxazolidinone formation but does not promote anomerization to α-threo-85 and α-erythro-85.

2.2.8.3 Irradiation of Threose Thiouridines
Thiouridines have been observed in tRNAs and have been investigated as a potential prebiotic solution to overcoming G:U wobble base pairing and its resulting lack of fidelity in NERPE (vide supra, 1.4.6.1). The greater thermodynamic stability of the s\textsuperscript{2}U:A pair helps to increase both the rate and fidelity of NERPE\textsuperscript{313}. The photochemical anomerization of ribo-2-thio-uridine α-ribo-112 has been investigated by Xu et al., they
found that no anomerization occurs. The reasons as to why this species not only does not anomerize but actually undergoes remarkable amounts of destructive process instead is of great interest.

Accordingly, the construction of the tetrose anhydrouridines was first investigated. Propiolamide 140, has been shown to enable the elaboration of aminooxazolines into anhydrouridines. When a diastereomeric mixture of tetrose aminooxazolines TAO and EAO were incubated with propiolamide 140 (RT, 18 h) formation of the corresponding tetrose-anhydrouridines (threo-126 and erythro-126) with exceptionally high conversion (97%) was observed by crude 

$^1$H NMR. Characteristic C6 and C5 resonances were observed in the 

$^1$H NMR and allowed confirmation. A subsequent thiolysis in DMF (H$_2$S, RT, 18 h) allowed isolation of β-threo-112 as a pure diastereomer after RP-FCC. Interestingly UV-Vis spectra of β-threo-112 showed a very different picture of absorbance compared to β-threo-111, with maximum absorbance closer to 275 nm (Exp 5.3.1.10). Subsequent irradiation (254 nm) in neutral water with hydrogen sulfide resulted in rapid destruction of β-threo-112 and signals indicated that free threose sugar 57 was the major product after just 16 h (Exp. 5.3.1.18). This indicates that despite anomerization of threo β-uridines not appearing an effective route towards TNA uridines that may allow for the recovery of TAO/EAO as the reaction of free sugars and cyanamide to give aminooxazolines has been shown by Orgel (vide supra, 1.5.5.1).

Scheme 67 Divergent routes from TAO towards TNA cytidine and TNA uridine.
2.2.9 Peroxide Oxidation

The success of the photochemical anomerization of β-threo-111 and β-erythro-111 enabled an apparent route to canonical TNA cytidine α-threo-85 (from α-threo-111). Xu et al. have shown (vide supra, 1.6.2) that a slow hydrolysis (pH 7, 100 mM Pi, 84 d, 60 °C) can allow conversion of ribose thiocytidine β-ribo-111 to its canonical oxygenic counterpart (β-ribo-85) but can also result in N^4 hydrolysis to 2-thiouridine β-ribo-112 (pH 3, 40 mM formic acid, 54 d, 60 °C)260. However, a more expedient method has recently been proposed, utilising hydrogen peroxide (H_2O_2), for the oxidation of 2-thio-cytosines and this was attempted here^159. A faster conversion of α-threo-111 to α-threo-85 would be advantageous to limit potential destructive chemistry (via UV-light or otherwise).

![Chemical structures](image)

**Figure 12** ^1^H NMR (700 MHz, H_2O/D_2O (9:1), Robust5_ver_03b, 5.05-8.20 ppm) spectra to show: a) β-threo-111 (12.6 mM) and α-threo-111 (18 mM); b) the reaction of a) with hydrogen peroxide (0.09 mmol, 196 mM) after 12 h. (C6)-H, (C5)-H and (C1′)-H protons (and (C2′)-H proton for threo-89) are labelled: β-threo-111 (▼), α-threo-111 (▲), threo-89 (●), α-threo-85 (●).

In contrast to the reactivity observed between peroxide and β-threo-111/β-erythro-111 (efficient re-cyclisation), peroxide enabled rapid, chemoselective, quantitative conversion of α-threo-111 and α-erythro-111 to the corresponding oxygenic cytidines α-threo-85 and α-erythro-85 (Exp 5.3.1.9). The trans-1',2' stereochemistry of α-tetroses

132
enabled by previous photochemical anomerization prohibits the rapid recyclization observed for the β-diastereomers. This is exemplified by the reaction (Figure 12) of β-threo-111 (12.6 mM) and α-threo-111 (18 mM) in phosphate buffer (0.1 M) with hydrogen peroxide (0.09 mmol, 196 mM) where after 12 h β-threo-111 had been quantitatively converted to anhydro threo-89 but α-threo-111 had been quantitatively converted to α-threo-85.

Scheme 68 This chapter’s prebiotic routes towards TNA cytidine α-threo-85 (and α-erythro-85). High yielding but oxidatively reversible thiolysis of anhydros threo-89 and erythro-89 to β-threo-111 and β-erythro-111 with sulfide then allows an effective anomerization of β-threo-111 and β-erythro-111 but anomerization of oxygenic β-threo-85 and β-erythro-85 does not occur, instead forming dead-end oxazolidinone products threo-88 and erythro-88. Irreversible peroxide oxidation of α-threo-111 and α-erythro-111 provides α-threo-85 and α-erythro-85.

This allows envisaging of a scenario where during or just after a period of photochemical anomerization, addition of peroxide could protect β-threo-111 against hydrolysis to β-threo-85 and its subsequent destructive photochemical pathway to dead end oxazolidinone threo-88 (vide supra, 2.2.8.2) while simultaneously providing the desired TNA cytidine α-threo-85 from α-threo-111.
Reliance on hydrogen peroxide was undesirable and other methodologies for sulfur removal were investigated. Accordingly, very slow (45 d) hydrolysis in phosphate buffer (0.1 M) at 40 °C can also convert α-threo-111 to α-threo-85 in low amounts (8%) (Scheme 69a and Exp 5.3.1.22). Cyanoacetylene can also accelerate hydrolysis by rapidly S-cyanovinylating α-threo-111 to give α-threo-149 (Scheme 69b and Exp 5.3.1.23) which activates the sulfur to hydrolysis similarly to oxidation. However, more work is required to ascertain whether this is a feasible method due to the lack of regioselectivity observed. The major products were α-threo-85 and α-threo-149, but many other unidentifiable species were also observed.

Scheme 69 Additional methods for sulfur removal of α-threo-111: a) slow hydrolysis (8% after 45 d) when heated at 40 °C in phosphate buffer (100 mM) b) rapid but unselective S-cyanovinylation also leads to hydrolysis.

The elucidated picture of tetrose nucleoside chemistry indicates very similar reactivity of threo and erythro diastereomers for the reactions investigated thus far (Scheme 68). Both aminooxazolines are quantitively converted to their anhydronucleosides. These are then both thiolysed efficiently or hydrolysed before undergoing efficient anomerization (thiolysis) or extensive conversion to oxazolidinones (hydrolysis). Peroxide can recyclize both β-thionucleosides to prevent hydrolysis to oxygenic β-ctydines and oxidizes both α-anomers to TNA cytidine and it’s erythro equivalent. Thus far none of these steps have revealed a method with which to enable differentiation of threo from erythro and in this way it is similar to the synthesis of the pentose precursors to RNA where the problem of mixed nucleosides also exists. However, the need to phosphorylate monomers to activate them for oligomerization presented an opportunity for this selection.
2.2.10 Prebiotic Phosphorylation

With a prebiotic route to TNA cytidine nucleoside α-threo-85 complete an examination of the role and timing of phosphorylation for the tetrose system was carried out. As phosphate and phosphorylation are vitally important but as yet unsolved issues in prebiotic chemistry a small digression follows to provide context. However, phosphate, as previously mentioned (vide supra, 1.3.3.1), is not without challenges in its geochemical provisioning making it one of the harder components to constrain in a prebiotic environment but due to its vital role is assumed to be available.

Phosphate is vital in extant biology and plays a major role in metabolism (ATP or primitive PEP, vide supra, 1.4.1.3), in membranes as phospholipids (phosphate provides the polar head group required, vide supra, 1.4.3) and in nucleic acids. Consequently, if the RNA world is as ancient as implied (vide supra, 1.4.4) then the special role of phosphorous must be considered. Phosphate is the third major component of nucleic acids. It exists as the phosphodiester linker between nucleic acids monomer units in all extant genetic polymers. RNA (including ribozymes) is highly dependent on the charged, flexible and relatively inert linker that phosphate provides.

Overall, there appears to be a requirement for genetic material to have a charged (positively or negatively) backbone. Westheimer came to this conclusion as to “why nature chose phosphates” and others have expanded upon this work 41,42. This negative charge has the dual purpose of repelling nucleophiles from hydrolysing genetic polymers (resulting in information loss without repair mechanisms) and in extant biology its use in signalling and regulation making use of charge-charge interactions.

The chemical structure of phosphate makes it uniquely suited for these roles. The negative charge (oxygen) and site of potential nucleophilic attack (phosphorous) are always close enough for nucleophiles to be affected by electrostatic repulsion, thereby helping to resist hydrolysis. Its first and second pKₐ (2.2, 7.2, 12.3) ensures that it remains charged in a majority of aqueous conditions (and vitally at physiological pH). Phosphates structure, with multiple acidic oxygens, allows it to maintain its negative charge while still bridging two sugar units as a phosphodiester. Additionally, the thermodynamics of organophosphate allow it to be simultaneously stable to hydrolysis but release large amounts of chemical energy when undergoing cleavage (hydrolysis) which can then be coupled to other chemical processes (e.g. ATP as a phosphorylating agent in extant biology with ADP as the by-product).
These properties make the incorporation of phosphate into prebiotic chemistry highly desirable and phosphorylation reactions are an area of ongoing research. This has recently been reviewed but several relevant examples for nucleic acid phosphorylation are provided here. The fundamental problem of nucleotides formation by phosphorylation in prebiotic chemistry is accessing the required condensation reactions, often with fairly sterically hindered secondary alcohols, in competition with hydrolysis by water (55 M). Overcoming these effects requires either very reactive phosphorylating agents that can result in unselective reactions or dry state conditions to negate the effect of water. A comprehensive overview of these thermodynamic issues has also recently been presented.

\[ H_2N\equiv N \quad N\equiv \equiv N \quad \equiv \equiv N \]
\[ \text{cyanogen} \quad \text{cyanogen} \quad \text{cyanate} \]

\[ \text{Scheme 70 a) common electrophilic phosphate activating agents; b) Lohrmann and Orgel's aqueous phosphorylation of uridine β-ribo-86 to give β-ribo-150 (1-4%).} \]

Aqueous phosphorylation remains a significant challenge. Early work towards this sought to make use of electrophilic anhydride-type phosphates generated from orthophosphate (PO₄³⁻) and activating agents. Cyanamide 87, cyanogen, cyanoacetylene 12 and cyanate have commonly been utilised (Scheme 70a). For example, when uridine β-ribo-86 was incubated (Scheme 70b) in 1 M phosphate with cyanoformamide 5'-monophosphate β-ribo-150 was observed, albeit in low yield (1-4%).

\[ \text{Scheme 71 Proposed mechanism of phosphorylation of alcohols mediated by urea under dry state conditions.} \]
To avoid the problems associated with water a great deal of work has been done on phosphorylations in the dry state and in formamide mixtures\cite{318,320,321}. Orgel and co-workers also showed that phosphorylation is improved by the addition of urea. In fact the urea can not only act as a pseudo-solvent for the reaction (m.p. 133-135 °C) but can play an active part in the mechanism (Scheme 71)\cite{317-319}. While the mechanism is not fully clear it is believed that the tautomeric form of orthophosphate (monoanionic) undergoes loss of water after nucleophilic attack by urea to give a ureidophosphate intermediate. This can then be intercepted by alcohols releasing the urea. The addition of urea has found good use in a number of prebiotic phosphorylation strategies. Powner and Sutherland made use of urea in formamide to achieve high yield phosphorylation of pyrimidine anhydronucleoside ara-89 to give 2′,3′-cyclic phosphate activated pyrimidine ribonucleotide β-ribo-90 (Scheme 72a)\cite{44}. Later work, in urea melts as well as with formamide, allowed access to 8-oxo-purine 2′,3′-cyclic phosphates from their corresponding anhydronucleosides (Scheme 72b)\cite{192}.

![Scheme 72](image)

Scheme 72 a) phosphorylation of arabino anhydro-pyrimidines leads to ribocytidine 2′,3′-cyclic phosphate\cite{44}; b) phosphorylation of arabino anhydro-purines leads to ribo-2′,3′-cyclic phosphates of several 8-oxo-purines\cite{192}.

Amidophosphates (DAP in particular) have also been demonstrated as powerful phosphorylating agents for a variety of prebiotic precursors (lipids, nucleosides, glycolysis products and peptides)\cite{118,322}. DAP has been shown to be prebiotically available\cite{323} and recently shown to promote formation of 2′,3′-cyclic phosphates from ribonucleosides (Scheme 73a)\cite{118}. Following from their work on sugar phosphorylation\cite{231}...
Krishnamurthy and co-workers demonstrated phosphorylation of all 4 canonical nucleosides to their 2',3'-cyclic phosphates in variable yields (27-89%) that were improved with the addition of imidazole (a prebiotic route for which is as yet uncertain) and metal ions (Mg$^{2+}$/Zn$^{2+}$). They have subsequently expanded this work to enable oligomerization of DNA with DAP and 2AI$^{324}$. Additionally, recent prebiotically plausible activation chemistry using methyl isonitrile can also be used to cyclise adenosine-3'-phosphate (56%, with AcOH)$^{133}$. Taken together these results hint that 2',3'-cyclic phosphate formation for cis-2',3' diols is inherently favoured.

Scheme 73 a) Gibard et al.’s ambient temperature DAP phosphorylation to ribo-2',3' cyclic phosphates. B = (U,C,A,G)$^{118}$; b) Xu et al.’s recent work towards DNA via 2,2'-thioanhydronucleotides; the phosphorylation of β-ribo-thionucleosides enables attack from C2-sulfur on a now activated C2’ and cyclization$^{261}$.

In the work from Xu et al. (vide supra, 1.6.2) phosphorylation had to occur after anomerization of α-ribo-111 en route to β-ribo-85. This was later shown to be enforced by the reaction of ribo-2',3'-cyclic phosphates which when in a trans-1',2' relationship with the 2-thiocytidine nucleobase undergo destructive cyclization to 2,2'-thioanhydronucleotides$^{260,261}$. This has allowed access to DNA (after irradiation and
transglycosylation, *vide supra*, 1.6.2.2) and also had implications for the proposed role and timing of phosphorylation en route to TNA.

The potential for selection of TNA over its erythro diastereomer was very attractive and phosphorylation appeared to provide such an opportunity. The reactivity between threo- and erythro-diastereomers when undergoing phosphorylation was believed to provide a handle for differentiation due to their trans- and cis-2′,3′ stereochemistry. It seemed likely that based on all previous work that threo-nucleosides could not form 2′,3′-cyclic phosphates whereas erythro nucleosides could. It was anticipated that then due to the previously observed efficient anomerization of β-*erythro*-111 to α-*erythro*-111 the C2-sulfur of α-*erythro*-1151 would be ideally positioned to attack the C2′ with the cyclic phosphate activating this reactivity. Conversely, threeo-nucleoside β-*threo*-111 should not form 2′,3′-cyclic phosphates and instead mono or bis phosphorylate but still undergo photochemical anomerization. Accordingly, these potential routes for selection of TNA were investigated.

### 2.2.11 Phosphorylation of Erythro-Nucleosides

Previous reports of efficient 2′,3′-cyclic phosphate formation encouraged investigation into similar reactivity for the β-*erythro* diastereomers here. It had previously been observed that 2′,3′-cyclic phosphate erythroses (and pentoses; ribose and its pyranosyl isomer) were observed products (1:9 with 1′,2′-cyclic phosphates) when phosphorylating sugars with DAP and AmTP (*vide supra*, 2.1.2) indicating that cyclic phosphate formation is inherently favoured for erythrose sugars or indeed any sugars with cis-2′,3′ stereochemistry. The proximity of the two hydroxyls, despite both being secondary alcohols, allows efficient reactivity.

*Scheme 74* Efficient 2′,3′-cyclic phosphate formation of β-*erythro*-151 allows its subsequent photochemical destruction.

Previous work has utilised phosphorylation, with ammonium phosphate as the phosphate source, in urea melts and whether this simple phosphorylation method could yield β-*erythro*-2′,3′-cyclic phosphates was explored here. Pleasently, this reactivity was
found to be highly favourable and indeed only 1 equivalent of ammonium phosphate is required to enable cyclic phosphate formation and when β-erythro-111 was subjected to Phosphorylation Method 1 (Exp 5.2.3), rapid and highly selective formation of 2′,3′-cyclic phosphate erythro-thiocytidine β-erythro-151 was observed as the major product, characterised by the appearance of downfield shifted (C2′)-H and (C3′)-H resonances and 31P resonances at 20 ppm. Very little mono-phosphorylation was observed even when more than 1 equiv. of phosphate was used.

The preparation of β-erythro-151 for investigation in photochemical anomerizations were required. While recent work by Yadav et al. has enabled the regio-specific cyclophosphorylation of cis-diols using a specialised cyclophosphate reagent325, the intrinsically favoured nature of β-erythro-151 allowed the synthetic preparation of β-erythro-151 in a simple urea-melt phosphorylation. When carried out with more than 1 equivalent of phosphate a facile RP-FCC (TEAB/MeOH) purification strategy and multiple lyophilisations to remove the volatile ammonium salts afforded the desired β-erythro-151 in good yield (50%).

The identified potential for photochemical destruction of β-erythro-151 was then able to be investigated. When irradiated (254 nm) in neutral water with hydrogen sulfide (184 mM), β-erythro-151 (7.1 mM) underwent rapid destruction to 2-thiocytosine 113 which, after 2 d, was the sole observed cytosine product with no remaining nucleotide observable (Exp 5.3.1.26). The proposed pathway for this (Scheme 75) highlights the key role of an efficient anomerization in positioning the C2-sulfur for attack on the C2′. Interestingly, unlike Xu et al., no 2,2′-thioanhydro erythro-120 was observed, presumably due to its photochemically unstable nature; it decays too rapidly via C-S bond homolysis and without good enough resonance stabilisation to 2-thiocytosine 113, similar to the when 2,2′-thioanhydrocytidine was irradiated to attempt to yield DNA cytidine (vide supra, 1.5.5.4). There has been some discussion as to whether the intermediacy of an 1′,2′-episulfide is involved but this has yet to be conclusively shown. This efficient 2′,3′-cyclic formation and rapid photochemical destruction seemingly provided a way to selectively form TNA cytidines from mixtures of tetrose thionucleosides.
**Scheme 75** Proposed mechanism for the photochemical destruction of β-erythro-151. Anomerization followed by attack on the C2’ position yields the 2,2′-thioanhydrole structure which can then fragment to give the base and the sugar phosphate.
**Figure 13** $^1$H NMR (700 MHz, noesyppr1d, H$_2$O/D$_2$O (9:1), 5.85-8.3 ppm) spectra showing the irradiation (254 nm) of β-threo-111 (12 mM) and β-erythro-151 (12 mM) at pH 7 with H$_2$S (109 mM): a) before the irradiation; and after b) after 1 d; c) 2 d; d) 3 d; e) 4 d irradiation. Spectrum f) following addition of authentic α-threo-111. Spectrum g) following addition of authentic 113.
2.2.12 Selection of TNA cytidines

Accordingly, the concomitant photochemical anomerization of \( \beta\text{-threo-111} \) and photochemical destruction of \( \beta\text{-erythro-151} \) was investigated (Figure 13 and Exp 5.3.1.25). When \( \beta\text{-threo-111} \) (12 mM) and \( \beta\text{-erythro-151} \) (12 mM) were irradiated (254 nm) in neutral water with sodium sulfide (109 mM) rapid destruction of \( \beta\text{-erythro-151} \) was observed and once again resonances indicating the formation of 2-thiocytosine 113 were detected. The \(^{31}\text{P} \) NMR gave perhaps the most striking example as the characteristic resonance at 20 ppm rapidly diminished (Exp 5.3.1.25) clearly demonstrating the destruction of the 2',3'-cyclic phosphate. However, concomitant efficient photochemical anomerization of \( \beta\text{-threo-111} \) to \( \alpha\text{-threo-111} \) was also observed. Spiking with authentic samples of \( \alpha\text{-threo-111} \) and 2-thiocytosine 113 allowed confirmation of effective selection of TNA cytidine. This highly desirable selection has the potential to totally remove the erythro nucleotides leaving a diastereomerically pure pool of TNA available for incorporation into genetic polymers.

The \( \text{trans-2',3'} \) stereochemistry of threo-diastereomers ensures that TNA nucleosides cannot form 2',3'-cyclic phosphates and can therefore be protected from these destructive pathways available for \( \beta\text{-erythro-151} \). However, the phosphorylation of TNA is presumably required before its incorporation as a nucleotide into a genetic polymer. The proposed photoanomerization that allowed selection of TNA was intriguing in that it could presumably also occur efficiently for threo-thionucleotides. Therefore, the phosphorylation of \( \beta\text{-threo-111} \) and subsequent photochemical anomerization to \( \alpha\text{-threonucleotides} \) was investigated.

2.2.13 Phosphorylation of \( \beta\text{-Threo-thiocytidine} \)

In similar urea-melt phosphorylations of \( \beta\text{-threo-thiocytidine} \) \( \beta\text{-threo-111} \) was confirmed to be unable to form \( \beta\text{-threo-2',3'-cyclic phosphates} \). Instead, extensive investigation (oxic/anoxic atmosphere, multiple equiv. of phosphate, varying reaction length) revealed that it furnishes a mixture of nucleoside phosphates (Scheme 76 and Exp 5.3.1.28) including mono-, bis- and pyrophosphates. Wanting to explore the potentially efficient photochemical anomerization of \( \beta\text{-threo-thionucleotides} \) this mixture was subjected to irradiation (254 nm) in neutral water with hydrogen sulfide (Exp 5.3.1.27). The enzyme alkaline bovine phosphatase has been used extensively to disentangle complex phosphorylations. Interestingly, the enzyme can hydrolyse poly/pyrophosphates but has no effect on 2',3'-cyclic phosphates. After the irradiation was complete it was used to simplify the distribution of products and in conjunction with sample spiking of authentic standards allowed confirmation of effective photochemical anomerization of these mixed phosphorylated species (Exp 5.3.1.27). This demonstrates that for TNA phosphorylation
is timing agnostic as it does not prevent the required photochemical anomerization but does allow for concomitant destruction of erythro-2,3-cyclic phosphates that are necessarily formed when mixtures of thiolysed tetrose diastereomers are phosphorylated.

![Scheme 76 Phosphorylation of β-threo-111 in urea-melts yields a mixture of β-threo-thiocytidine phosphates that undergo anomerize to their α-anomers. Thermal recyclization of β-threo-111 to threo-89 and subsequent phosphorylation gives photostable α-erythro-cytidine α-erythro-152. R = H or (PO3-)n.](image)

The elevated urea-melt temperatures (130-140 °C) also appear to promote the recyclization (similar to hydrogen peroxide) of β-threo-111 to threo-89 which then undergoes rapid phosphorylation on the sterically unhindered (C3')-OH followed by similar reactivity to Powner and Sutherland and Stairs et al. (Scheme 72) in which rapid attack by the phosphate forms, in this case, α-erythro-cytidine-2',3'-cyclic phosphate α-erythro-152. This was confirmed by characteristic downfield shifting of (C2')-H and (C3')-H proton resonances in the ¹H NMR and appearance of a signal at 20 ppm in the ³¹P NMR (Exp 5.3.1.27). This reactivity may be assisted by oxygen as carrying out these phosphorylations under N₂ atmospheres (highly plausible for the early Earth) reduced the amount of α-erythro-152. During irradiations α-erythro-152 exhibits photochemical stability, unsurprising for these (trans-1',2') 2',3'-cyclic phosphate structures. Interestingly, this mode of reactivity should be equally accessible during the phosphorylation of β-erythro-111 but clearly the rapid and highly favoured formation of 2',3'-cyclic phosphates outcompetes any other reactivity.
To improve phosphorylation as a selection method for TNA a potentially milder (lower temperature) phosphorylation may allow erythro-specific phosphorylation due to the favoured nature of the 2',3'-cyclic phosphates it forms, while the more easily hydrolysed monophosphorylation of TNAs may be reversible and will not promote recyclization.

2.2.14 Phosphorylation of α-tetrose Nucleic Acids

If anomerization of both tetrose thionucleosides occurs before phosphorylation the products of any subsequent phosphorylation are also of interest, especially as the photochemical destruction of β-erythro-151 presumably proceeds through its α-anomer and that anomerization of β-threo-111 and β-erythro-111 are both equally efficient. This then could also provide a method of selecting for threo over erythro nucleotides. When α-threo-111, α-erythro-111 and α-threo-85 were phosphorylated (ammonium phosphate, >1 equiv.) under the same dry-state conditions as the β-anomers (urea, 140 °C, 15-30 mins) differing reactivity was observed (Scheme 77). As anticipated an unselective complex mixture of mono/bis/pyro phosphates were observed by $^1$H/$^31$P NMR spectroscopy for α-threo-111 (Exp 5.3.1.29) and also for its oxygenic hydrolysis product α-threo-85 (Exp 5.3.1.30). Both appear to be robust to these reactions as they are unable recyclize due to their trans-1',2'-configuartion.

![Scheme 77](image)

Scheme 77 Phosphorylation in urea of both α-threo-111 and α-threo-85 results in mixtures of mono/bis/pyrophosphates.

Interestingly, when α-erythro-111 was subjected to phosphorylation 2-thiycytosine 113 was observed even on very short timescales (Scheme 78). The complex product mixture indicates that α-erythro-111 is unstable to urea-melt phosphorylation and that this indeed may be a viable method of selecting for TNA (Exp 5.3.1.31).
Scheme 78 Phosphorylation in urea of α-erythro-111. 113 was observed after only 10 mins but a complex mixture of products presumably including 2,2′-thioanhydros make unambiguous assignment impossible.

Taken together, these phosphorylations of tetrose α-anomers have twofold implications. Firstly, the role of pyrophosphate TNA is an exciting prospect. It seems likely from the phosphorylations demonstrated here that, in urea, bis-phosphorylation of both β-threo-111 and α-threo-85 is common. Previously, undiscovered chemistry would have to revealed for a subsequent activation not to simultaneously activate both phosphates which would lead to pyrophosphate oligomers. Interestingly, pyrophosphates can patch/repair strands of RNA in NERPE (vide supra, 1.4.6.1)\(^{194}\). Secondly, that the role of phosphorylation in TNA could not only be to activate its monomers but to select for it as either the β- or α-anomer, this sequentially agnostic approach to nucleotide construction has not been demonstrated for any other nucleic acid’s monomers.

2.2.15 Conclusion

This chapter has demonstrated that TNA cytidines can be constructed in high yields from aminooxazoline TAO via a prebiotically plausible, protecting-group-free synthesis involving a quantitative cyanovinylation, high-yielding thiolysis and an efficient photoanomerization of the 2-thio-nucleosides in a manner that also yields RNA monomers. A concomitant route to nucleosides is possible for erythro-diastereomers but phosphorylation, instead of activation to nucleotides, enables selection of TNA due to the innately favoured formation of 2′,3′-cyclic phosphates for erythro-nucleosides. The sequentially agnostic approach to nucleotide formation for TNA points towards a scenario where erythro-thionucleotides necessarily are destroyed whenever they are formed. While formation of some oxygenic 2′,3′-cyclic phosphate (α-erythro-152) is observed for TNA the sluggish reactivity of 2′,3′-cyclic phosphates\(^{326-330}\) may therefore provide the final means of selecting TNAs as threose can be activated as an
imidazolide\textsuperscript{906,331} and cannot form these structures meaning that even if oxygenic \(\alpha\)-\textit{erythro}-152 can form it is unlikely to be incorporated into genetic polymers. These results warrant further investigation into the mild phosphorylation of threose nucleosides, the selective ligation of the resultant nucleotides as well as an investigation of concomitant purine synthesis.
Towards a prebiotic synthesis of DNA

3.1 Introduction

DNA is ubiquitous across all domains of extant biology; it is the stable genetic polymer that has allowed life to thrive on our planet. One clear advantage it has over other genetic polymer systems is its hydrolytic stability due to its lack of (C2')-hydroxyl, this has however resulted in a lack of prebiotic synthesis for DNA and a focus on RNA instead\textsuperscript{158}. As discussed (\textit{vide supra}, 1.4.4), RNA is widely thought to have been the first genetic polymer to emerge during the evolution of life due to its dual functions of information storage and catalysis in extant biology\textsuperscript{151,152,273,332,333}. While endogenous DNA-zymes have not (yet) been found in nature, DNA has been evolved \textit{in vitro} to yield deoxyribozymes that can catalyse chemical reactions\textsuperscript{155,156,334}. DNA's stability is thought to have driven its biological selection during the early stages of evolution and despite the presence of transcriptional genes in LUCA (\textit{vide supra}, 1.2.1), it is widely assumed that an evolutionary mechanism of "genetic takeover" occurred in early evolution to allow life to transition from an RNA-based information storage to the extant DNA system now observed\textsuperscript{158}. One intermediate on this evolutionary pathway could have been chimeric polymers with mixed RNA-DNA backbones. Such systems have recently been investigated by several groups and were found to support replicable catalytic function\textsuperscript{335}. They also exhibit a tendency to produce homogenous polymers from heterogenous chimeras when involved in template directed ligation\textsuperscript{294}. Additionally, of all the current studies on alternate sugar backbones (e.g. TNA, ANA), DNA suffers from the least rate reduction in NERPE (\textit{vide supra}, 1.4.6.1) implicating its incorporation into primitive polymers if a convergent synthesis can be demonstrated\textsuperscript{191}. Therefore, the simultaneous emergence of RNA and DNA monomers must be addressed to provide the evolutionary capability for the incorporation of DNA into early evolution. There have been some attempts to access DNA monomers by a Hilbert-Johnson-type glycosidation of nucleobases but most do little to persuade that it would be a reasonable prebiotic method\textsuperscript{336}. The difficulties of forming a deoxyribose sugar from an aldol formose-type reaction process have not been overcome\textsuperscript{210,337}, no formose process has produced 2'-deoxyribose in appreciable yield, however Ritson and Sutherland have demonstrated that 2'-deoxyribose synthesis is achieved by the convergent reduction of ribose and arabinose by cyanosulfidic UV α-deoxygenation (\textit{vide supra}, 1.6.1)\textsuperscript{208}. Unfortunately, the synthesis of ribose/arabinose is (at best) very low yielding under prebiotic constraints and the subsequent glycosylation of preformed nucleobases is presumed to be impossible for pyrimidines and disfavoured for purines leaving glycosylation a currently unviable routes to DNA monomers\textsuperscript{227,333,338,339}. 
Recently several new routes towards DNA have emerged. As discussed (vide supra, 1.6.2.2), recent work from Sutherland and co-workers have demonstrated two routes towards DNA purines\(^{261,262}\). Firstly, via a low yielding (α/β, 6/4%) and unregioselective transglycosylation of non-canonical deoxy-thiouridine \(121\) with adenosine (Scheme 36) and secondly via a transglycosylation of the pyrimidine ribonucleotide precursor anhydronucleoside \(\text{ribo-89}\) (and \(\text{ribo-126}\)) and subsequent photochemical reduction of the mixture of thio-anhydropurines formed (Scheme 37). Trapp and co-workers have also demonstrated that all four canonical deoxynucleosides can be synthesized, albeit in very low yields (Scheme 79), but their route requires the use of preformed and acetaldehyde activated nucleobases\(^{340}\). These activated vinyl nucleobases are able to undergo slow but regioselective nucleophilic attack on glyceraldehyde \(4\) in water (50 °C, 4 d) where a 5-exo-trig cyclization gives the desired furanosyl deoxyribose conformer. Glyceraldehyde \(4\) is presumably mostly converted to DHA \(9\) under these conditions and may contribute to the low yields and is the reason for detection of a wide spectrum of products including the deoxyapurionucleosides.

\[ \text{Scheme 79 Recent work by Trapp and co-workers towards DNA by reaction of vinyl nucleobases with glyceraldehyde 4}^{340}. \]
While good advances have been made, these routes suffer from low yields and a lack of selectivity or an inability to make all DNA pyrimidines. Additionally, the potential for divergent prebiotic chemistry to form both RNA and DNA has yet to be fully investigated. This chapter seeks to address all three of these issues.

Scheme 80 Progress towards purine and pyrimidine nucleic acids: arabino-anhydronucleosides ara-89 and ara-102 (and ara-126) can be formed from a divergent pathway from ara-96 (ara-89 and ara-126 can also be synthesised from AAO) whose reactivity form the focal point of this chapter. DNA monomers, 155 and 156, are the desired products of this chapter and are proposed to be formed by thiolysis and reduction of thiols positioned at C2′.

Towards this, it is of relevance that it has now been demonstrated that aminoazoline AAO and oxazolidinone thione ara-96 (Scheme 80) can be synthesised and elaborated to give access to anhydronucleosides ara-89 and ara-102, respectively, and subsequently both pyrimidine and 8-oxo-purine ribonucleotides by divergent prebiotic chemistry (vide supra, 1.5.5.4)\textsuperscript{14,192}. Thiolysis of anhydronucleosides (including ara-102) has also recently been shown to enable subsequent reduction by peroxide or UV-light\textsuperscript{159}. It was of particular relevance to this chapter that the C8 regioselectivity of hydrogen sulfide addition (H\textsubscript{2}S) was high. The reversible addition of H\textsubscript{2}S at C2’ presumably precludes formation of a C2’ thiol. However, as others have found (vide infra) alkyl thiols can be added at this position. Therefore, the thiolysis of anhydronucleosides with alkyl
thiols and reduction of the resultant compounds to deoxy sugars was investigated here (Scheme 80).

In the 1970s *Ikehara and co-workers* observed, during an investigation of synthetic nucleotide manipulations, that upon incubation of **ara-102** with ethanethiolate (EtS⁻), nucleophilic thiol addition to the (C2)⁻-position, to give **2'-thio-nucleosides** (Scheme 81)\(^{341}\). This was achieved with ethanethiol (EtSH) solution that had been partially deprotonated with sodium hydride (NaH, 1 equiv). *Ikehara et al.* obtained the **2'-thio-2'-deoxyribo-8-oxoadenine 157** derivative in good yield (60%) from the **ara-102** alongside 8-thio derivative **158** and an unknown compound, for which they do not provide yields but infer them to be minor species\(^{342}\). **2'-thio-adenosine 157** was subjected to a Raney-Nickel reduction to give the nucleoside **2'-deoxyribo-8-oxo-adenosine 159**. This is of note because **159** exhibits the correct sugar for a DNA nucleoside. Although **159** is furnished with an 8-oxo-adenine nucleobase, these have been proposed to have possibly played a key role in primordial nucleic acid evolution (*vide supra*, 1.5.5.4)\(^{192}\).

*Scheme 81* *Ikehara et al.*’s method for the synthesis of 2',8-cycloadenosine **ara-102**. (orange arrow) and *Stair’s et al.*’s alternative method for the synthesis of 2',8-cycloadenosine **ara-102** that will be used in this chapter (blue arrow). *Ikehara’s thiolysis of ara-102 in DMF with ethanethiol (EtSH) (pink arrow) leads to 8-oxo-deoxyribo-adenine 159 after Raney-Nickel deuslfurisation.*

It is of interest then that *Stairs et al.* have developed a prebiotic pathway to **ara-102**, and also that the same synthesis could divergently lead to **AAO**, therefore it is of note that if the same chemical strategy for thiolysis and C2'-selective reduction could be applied to
Anhydrocytidine ara-89 as has been applied to ara-102 by Ikehara and could give both RNA and DNA precursors from the same key intermediates.

The route to make 2'-thio pyrimidines was influenced by previous work on similar substrates. Reese and co-workers carried out extensive studies on 2,2'-anhydouridine ara-126 and discovered that in aprotic solvents such as DMF, as Ikehara used, they could furnish either the 2' or 3'-thio derivative by altering the order of addition343,344. If they pre-mixed the base (NaH as Ikehara) and ara-126, this would afford the isomeric epoxide 161 and then when EtSH was added this would be converted to the 3'-thio adduct 162. If they pre-mixed the base and the thiol and then added ara-126 they achieved the desired 2'-thio compound 163. Todd et al., in 1958, also found that if the ethanthiolate salt was used in DMF at 100 °C they would proceed directly to the 3'-thio 162 in 55% yield345. This implies that ara-126 can rapidly deprotonate and form 161 implying that to form 2'-thio 163 there needs to be a balance of nucleophile anionic character to add to the preferred position.

\[ \text{Scheme 82} \quad \text{Addition order of base and thiol influences product selectivity; a) base and anhydouridine react to give epoxide 161 which undergoes regioselective thiol opening at C3'; b) deprotonating EtSH and then adding ara-126 selectively gives the 2'-thioethyl compound 163}^{343,344}. \]

The chemistry of uridine has been explored by both Reese et al. and Todd et al. but a novel synthesis of 2'-sulfur cytidine derivates was also explored here; as if these compounds can be desulfurised successfully, it will yield the canonical DNA nucleoside 155, and as ara-89 has been shown to lead to RNA a route to both RNA and DNA could be possible.

Although Ikehara’s work points the way to a selective C2' reduction strategy and the potential to unite pyrimidine and purine reduction, exploiting the inherent activation of anhydronucleoside, it is clear that Ikehara et al.’s methods are not prebiotically plausible.
Therefore, a translation of this synthetic strategy to a methodology that is prebiotically plausible was also explored.

As mentioned, it is widely thought that RNA was the first system to emerge\textsuperscript{141} but the potential divergence of this method means both RNA and DNA monomers, for both purines and pyrimidines, may be able to be synthesised from the same precursors, the anhydronucleosides \textit{ara-89} and \textit{ara-102}. Additionally, as \textit{ara-104} has not yet been shown to be formed prebiotically it was not investigated here.
3.2 Results and Discussion

3.2.1 Thiol addition

3.2.1.1 Synthetic Reactions in DMF

The systematic investigation of the reactivity of ara-102 and ethanethiol was undertaken. While DMF is certainly not prebiotic in nature, it was thought to provide a reasonable model for formamide (HCONH₂) which is widely exploited in prebiotic chemistry, and has been identified on comets and other interstellar media. It has been proposed to have been present in prebiotic environments as it is formally the hydration product of HCN. Using water as a solvent for these reactions certainly would make the reactions more prebiotic, but previous work has demonstrated that anhydropurines (such as ara-102) have a tendency to be attacked by sulfur nucleophiles, particularly H₂S, at the nucleobase moiety (C8 for purines and C4 or C2 in pyrimidines). Moreover, ethanethiol (EtSH) also has poor water solubility; in order to counteract this the remit of this work includes investigations on other sulfur nucleophiles (enabling the use of water as a solvent) that will also be discussed.

The reaction of 2′,8-cylooadensine ara-102 with NaH and ethanethiol was attempted in DMF and after 1 h at 70 °C with ethanethiol (8 equiv.) the reaction had gone to completion as the characteristic ABX doublets at 5.80 and 6.61 ppm (J = 5.5 Hz for both) had disappeared in the ¹H NMR spectrum of the crude reaction mixture. Confirmation that the correct product had formed was also obtained from ¹H-¹³C HMBC NMR data which demonstrated a characteristic cross-coupling between proton resonance (2.44 ppm) and carbon resonance (49 ppm, C2') (Exp 5.3.2.1). The cross-coupling between the SCH₂CH₃ proton resonances and the C2' clearly identify the ethylmercapto moiety at the 2' position. Significantly, under these conditions, there is no cross peak correlating the ethanethiol moiety to the base showing no thiol addition at C8 has occurred. It is of note that Ikehara et al. observed addition at C8, but it is only anticipated for reactions of ara-102 in water with sulfur nucleophiles. However, it is possible that the unidentified species in their reaction is either the 2',3'-epoxide 164 or its subsequent 3'-thio addition product 165 as the epoxide has been shown to be an isomer of ara-102 at high pH (pH 13) and subsequent addition of thiol could give 165. While the large ³JHH coupling for the (C1')-H (9.7 Hz) is indicative of dihedral angles of 40° or 140° (for trans or north pucker configuration), NOESY NMR confirmed that 2'-addition to give 157 here is directly from ara-102 and does not go via epoxide 164.
Attempts to translate this to prebiotic chemistry and conversion to canonical DNA nucleoside 155 were then investigated. The NaH promoted reaction of 2,2′-anhydrocytidine ara-89 with EtSH (8 equivalents, 70 °C, 24 h) was therefore investigated. The 1H NMR of the reaction (Figure 14) shows that when this reaction is carried out at the same temperature as for the purines (70 °C), two products can be observed. 2D NMR data reveals that the minor product (76%: 24%) is indeed the 3′-thio addition product 167 and not simply the hydrolysis to cytidine β-ara-85.

![Chemical structure and NMR spectrum](image)

*Figure 14* 1H NMR (600 MHz, D2O, noesygppr1d, 5.05-8.95 ppm) spectrum to show: the reaction of ara-89 with EtSH, NaH in DMF, 70 °C for 24 h. (C6)-H, (C5)-H and (C1′)-H protons are labelled: 166 (●), 167 (▼).

A plausible reason for this distribution of products may be the formation of 2′,3′-epoxides like 161 in agreement with the work of Reese et al. in that when the base and ara-126 are mixed together there is the possibility of accessing epoxide 161 (Scheme 83)343. Interestingly, the desired place of attack for cytidines the C2′, is still the most favourable showing that direct attack on the ara-89 is still possible for anhydropyrimidines.

![Mechanism diagram](image)

*Scheme 83* Potential mechanism of 3’-thioethyl-uridine 162 formation via epoxide 161
Carrying out this reaction at room temperature resulted in total regioselective addition at C2′ and accordingly only one product, the desired 2′-thio adduct 166, was observed (Exp 5.3.2.2). These results suggest that clearly both temperature and the anion character of the sulfur nucleophile an anhydronucleoside is exposed to dictates its reactivity.

2,2′-Anhydouridine ara-126 is also product of prebiotic nucleotide syntheses via aminooxazolines (Scheme 80) but would not yield canonical DNA nucleotides via this process (as DNA uses thymidine), however, it is readily commercially available reagent and has been studied in depth as previously mentioned344,345. It was assessed as a model substrate also under the reduced temperature reactivity for ara-89. Pleasingly, under similar conditions it also underwent selective thiolysis to give 163 as the major product (Exp 5.3.2.3).

3.2.1.2 Thiol conjugate bases as a more prebiotic model
Anticipating that formamide can be incompatible with strong bases such as NaH a different method was sought. Ethanethiolate (EtSNa), the conjugate base of ethanethiol, was used by Todd et al. and has been investigated as a replacement for NaH and was explored here345.

A comprehensive exploration of conditions for the reaction of 8,2′-cycloadenosine ara-102 (1mM) and EtSH/EtSNa (10:0-0:10, 10 mM) in DMF (25-70 °C, 1-16 d) were undertaken (Exp 5.3.2.5) and the product distribution is indicated (Scheme 84). It was of note that the reaction required at least 1% of EtSNa to occur with desired efficiency as after 1 day of incubation with 0% conjugate base (EtSNa) no reaction was observed. However, if too much EtSNa was present then ara-102 gave predominantly epoxide 164 (77%) with a small amount of desired 2′-thioethyl 157 (23%). This is thought to be due to the increased basicity promoting deprotonation of the 3′-hydroxyl moiety followed by intramolecular inversion192,343,344. 10 equivalents of EtSNa gave the desired product, albeit in low yield (23%) but the 3′ product 165 was not observed perhaps due to the low temperature making thiol addition to the epoxide more challenging. 3′-thioethyl adduct 165 and 8-thio adduct 158 were not observed in quantities large enough to be fully characterised, however they can be assigned by NMR data using characteristic HMBC couplings to the base in the case of 8-thio 158 and to the observable (C3′)-H for 165. This optimisation allowed for the development of a thiolysis that was totally regioselective (no 165 observed) and exceptionally high yielding. With a ratio of EtSH/EtSNa of 8:2 equivalents, 157 was obtained in 99% yield by NMR and 92% by isolation. This is an improvement on Ikehara’s method which affords 157 in 60% yield. Some other trends
suggest that heating the reaction not only expedites it, as expected, but also increases regioselectivity for the desired 157.

Scheme 84 Product distribution of the reaction of ara-102 (1mM) with EtSH/EtSNa (10:0-0:10, 10 mM) in DMF (25-70 °C, 1-16 d).

When investigating the pyrimidines its was observed that ara-89 and U ara-126 also totally selectively, convert to their respective 2’-thio adducts 166 (94%) and 163 (96%) when incubated with EtSH/EtSNa (8:2, 10 equiv.) at room temperature for 24 h (Exp 5.4.2.3 and Exp 5.4.2.4). As before, ambient temperatures increased product selectivity for the pyrimidines. The 1H NMR revealed both the high $^{3}J_{HH}$ coupling for the (C1′)-H (8.7/8.9 Hz for 166 and 163) a characteristic upfield shifted (C2′)-H resonance (3.58/3.59 ppm for 166 and 163) which give provide evidence that sulfur has been added at the correct C2′ position. 166 is the potential precursor to DNA nucleoside 155 and so a prebiotic route to it is of considerable interest as a synthesis of this has yet to be shown prebiotically.

It appears important to the reactivity of the sulfur as a nucleophile species to have sufficient basicity such that (partial) thiol deprotonation may occur to promote $S_N_2$ type nucleophilic displacement at the C2′ position of the anhydronucleosides.

However, a balance of thiol deprotonation seems to be important particularly for the pyrimidines$^{344}$. The $pK_{a}$ of ethanethiol is 10.61 (at 25 °C)$^{349}$, making it not only a plausible reagent but one that can also be utilised in aprotic environments and in exact quantifiable amounts. As the initial reactions were carried out in DMF a prebiotic approach cannot be claimed. However, it was judged to be important to satisfactorily explore the chemistry of this system before moving toward prebiotically constrained reactivity.
3.2.1.3 Potentially prebiotic reactions in formamide

DMF is clearly a non-prebiotic solvent but with an optimised synthesis of the 2’-thioethyl nucleosides developed, the solvent could now be exchanged for formamide (HCONH$_2$) in an attempt to achieve correct reactivity in a more prebiotic solvent. Formamide has been discussed in the context of prebiotic chemistry (vide supra, 1.3.3.2) and has been observed in appreciable quantities on comets and in interstellar media$^{346,347}$. NaH was observed to react with formamide as it commonly has some percentage of water as an impurity, as it surely would on the early Earth, hence the need to use EtSNa instead. However, the advantage of formamide aside, from its supreme solvation of the compounds being investigated, was the higher boiling point of 210 °C, this means higher temperatures can be used to drive reactions toward kinetic products where water is unable to, this has been shown to be beneficial in for the formation of 157. While the high boiling point enables formamide to be used in high reaction temperatures, the same thing hampers its use, as its lack of volatility makes it difficult to remove to analyse reactions.

Accordingly, the temperature of the reaction was first investigated here, as it was believed that the optimised 8:2 stoichiometric ratio of EtSH:EtSNa would give the best regioselectivity as established previous in DMF.

Formamide facilitates reaction of ara-102 at the C8-position seemingly with a higher selectivity at low temperatures than in DMF (158 70%, 157 7%, at RT, 2 d, Exp 5.3.2.6 Entry C). This regioselectivity is not unexpected as previous work has shown that, in water, ara-102 can react with H$_2$S selectively at the C8-position$^{159}$. To decrease the formation of the undesired C8-adduct 158 the temperature of the reaction was increased. Surprisingly, at 70 °C, the temperature needed for totally selective reaction in DMF, these conditions proved to be ineffective at producing desired 157 in anything above yields of 23%, with 158 again the major product (65%).

One advantage of formamide here is the reaction is facile to carry out at or above 100 °C, but further increasing the temperature to 120 °C and varying the ratio of EtSH/EtSNa (both a 5:5 and 9:1 yielded lower amounts of 157 over a one-day timescale) did little to increase the yield or selectivity (Exp 5.3.2.6). However when the temperature was raised further to 125 °C the opposite was observed; the reaction showed greater selectivity with the highest amount of 157 observed for any temperature investigated (37%) and was the major product observed (22% 158, 27% 164).

For the temperature to affect the distribution of products, the reaction of ara-102 must be influenced by kinetic vs thermodynamic control. Here epoxide 164 seems to be a
reversible product that can be converted back to ara-102 and then to 157. In contrast, the addition at the 8-position to give 158 seems to be irreversible regardless of the temperature. However, as this clearly shows that 157 can be selected for in a prebiotically plausible solvent, and its subsequent reduction to deoxyribo-8-oxo-nucleoside should be investigated.

The reactivity of the pyrimidines in formamide was then investigated. Anhydro-uridine ara-126 does not react to form the desired 2′-thioethyl 163 in formamide with an 8:2 ratio of EtSH/EtSNa at room temperature or 70 °C. However, selective formation of 2′-thioethyl 163 (77%) was observed by 1H NMR (compared to authentic sample) after 4 d when the temperature was raised to 120 °C using a ratio of 5:5 EtSH/EtSNa (Exp 5.3.2.7).

Scheme 85 The major products for the reaction of a) ara-126 and b) ara-89 with EtSH/EtSNa in formamide (120 °C, 4d).

In contrast, the reaction under the same conditions (120 °C, 5:5 EtSH/EtSNa, 4 d) of ara-89 does not selectively afford potential DNA precursor 166 but gave a mixture of products: the desired 166 (35%), 3′-thio 167 (39%) and hydrolysis to arabinoside β-ara-85 (19%) (Exp 5.3.2.8). Under the same conditions but with EtSH/EtSNa 8:2 the reaction was far more selective and 166 (42%), β-ara-85 (42%) were observed. While not as selective as for ara-126, this work does show that 166 can be formed under prebiotic conditions with high enough conversion for its reduction to DNA nucleoside to warrant investigation.

Once the 2′-thio compounds had been shown to be formed in prebiotically plausible conditions (HCONH₂), an initial investigation into their prebiotic desulfurisation could be attempted as this would presumably yield the relevant 2′-deoxy-nucleosides.
3.2.2 Other Thiols
The use of sulfur nucleophiles that had a higher aqueous solubility than EtSH were also of interest and a variety were trialled (Exp 5.3.2.9). Disappointingly, many of the sulfur nucleophiles exhibited sluggish or no reaction with either ara\textsuperscript{-}102 or ara\textsuperscript{-}126/89. It is of note however, that despite several being alkyl thiols (e.g. cysteamine 168), (Exp 5.3.2.10) gave selective C8 addition with ara\textsuperscript{-}102 in water with high conversion and the C2′-adduct was not observed. This is consistent with the work by Roberts et al. that show that in water, ara\textsuperscript{-}102 reacts with H\textsubscript{2}S selectively at C8\textsuperscript{-159}.

3.2.3 Desulphurisation
3.2.3.1 Raney-Nickel
Desulphurisation of nucleosides in particular with sulfur at the C2′, C3′ and C8 positions is known and was reported by Ikehara as the desulphurisation of 157 to 159\textsuperscript{341}. Raney-Nickel was employed but is not a prebiotically plausible reagent\textsuperscript{350}. Raney-Nickel can give relatively low yields for these reactions and it has been proposed that a significant amount of the substrate may become adsorbed onto the catalyst hence the low yield\textsuperscript{351}. Ikehara et al. do not report their yield but Reese et al. achieved 73% of 3′-deoxyuridine 156 in water after 1 hour.

The desulphurisation of 157, when investigated in collaboration with the Sutherland lab (LMB-MRC), with either Raney-Nickel or hypophosphite nickel (sponge nickel) revealed no \textsuperscript{1}H NMR resonances indicative of 159 (Exp 5.3.2.11). As these are not prebiotically plausible other methods to prepare 159 were sought. However, this did indicate the potential stability of the thioether bond.

3.2.3.2 UV-light
Work by Danishefsky and co-workers revealed that cysteine could be desulphurised via a radical process to give alanine, also a canonical amino acid\textsuperscript{352}. Their proposed mechanism is enabled by TCEP which is not a prebiotic reagent. Recently work has sought to bring about similar desulphurisations in prebiotic chemistry using UV-light and several other methods (hydrogen peroxide mediated hydrolysis, thiol acetylation and subsequent elimination) have reported cleavage of thioethers similar to those of interest here\textsuperscript{93,159,256,261,262,353–355}. This prompted the investigation into the UV chemistry of these 2′-thioethyl compounds. Noting the recent observations that thiols at C2′ are usually precursors to degradation it was hoped the thioether moiety here would allow for photochemical cleavage. It was envisaged that the potential homolytic cleavage of the C-S bond would lead to a thyl radical and the desired secondary alkyl radical, which could abstract a proton to give the desired deoxy-ribose sugar\textsuperscript{262}. 

160
Surprisingly, when 157 (2.5 mM) was subjected to irradiation (254 nm) no desulfurisation was observed (no new (C1')-H triplet resonance ~7 ppm or new signals in the region 2.1-2.5 ppm) and instead photochemical destruction dominated after 4 days (Exp 5.3.2.12). Similarly, when 166 (15 mM) was irradiated (254 nm) total destruction was observed after just 18 h (Exp 5.3.2.13). Presumably these reactions are not accessible due to the stability of the thioether bond to homolytic fission by UV-light and also due to the photochemical instability of the nucleobases. This indicates that preferential absorption by the nucleobase may prevent any absorption of UV-light by the thioether and therefore the nucleobase offers some level of protection from irradiation, this is discussed in more detail in Chapter 4. While unsuccessful these reactions do inform us that while C2' thiols may be unstable to UV-light that the stability of thioethers is too great to be used as precursor to DNA when UV-light is used to attempt desulfurisation.

3.3 Conclusion and future work

This chapter demonstrates that anhydronucleosides can be thiolysed to give DNA precursor C2'-thioethers under prebiotic conditions. Alkyl thiols can indeed enable irreversible addition at C2' that hydrogen sulfide cannot. The difference in reactivity, particularly the balance between thiol and its conjugate base in the explored model system DMF and prebiotic solvent formamide is also interesting and warrants further exploration. There is a clear need for plausibly prebiotic sulfur nucleophiles to be investigated under a broader range of conditions such as pH (to increase available anion as with EtSH), temperature and solvent. Photochemical shielding of the thioether by the nucleobase prevented access to a reduced C2' and thus DNA. Alternative ways of generating thiol radicals on the early Earth with thioethers at C2' would be very interesting and perhaps allow access to this chemistry. Wavelength dependence has been explored for some prebiotic photochemistry and this system would also benefit from this. If a wavelength could be identified where the nucleobase absorbs weakly but the thioether does so strongly this could also provide access to thiol radicals and thus potentially to DNA.
4 UV Stability of Nucleic Acids

4.1 Introduction

Prebiotic chemistry is primarily focused on the design and optimisation of plausible synthetic routes to molecules that could have significance at the origin of life. However, recent work has also shown that it is important to consider the influence of the prebiotic environment in which molecules of relevance are formed\textsuperscript{37,307,356,357}. Environments on the early Earth would have crossed a diverse conditions, ranging from hot\textsuperscript{177,179} to cold\textsuperscript{358,359}, with extremes of pH and salinities\textsuperscript{184} and likely exposure to UV-light\textsuperscript{255,307,357}. Considering the environments of where prebiotically important molecules would have formed and accumulated is therefore vital to a further understanding potential scenarios for the origin of life. These environmental conditions could have been decisive selection pressures for the biomolecules present in extant biology. For example, many have speculated that a lack of stability or longevity of nucleobases would preclude their inclusion into primitive biology. Indeed, Miller and Levy argued that due to a lack of hydrolytic stability cytidine would degrade relatively rapidly compared to the other canonical nucleosides (\(t_{1/2} = 19\) days at 100 °C)\textsuperscript{360}.

As early as the 1970s, Sagan argued that the UV stability of nucleosides could have been an important selection pressure in establishing the nucleosides we see in extant biology\textsuperscript{361}. Yet today, for many possible conditions (temperature, pH, UV-light), the effect of these potential selection pressures is still not fully understood, despite many recent advancements in the prebiotic synthesis of nucleosides. Motivated by this and work in previous chapters, where the stability of nucleosides was important to prebiotic synthesis or selection, this chapter presents a preliminary investigation of the stability of nucleic acid monomers under irradiation by UV-light.

As discussed (\textit{vide supra}, 1.3.2), not only do models of the young sun suggest that solar flux was higher in the short-wavelength higher-energy region (200-300 nm) in the Archean era but that the amount of ozone in the Earth’s atmosphere was also too low to attenuate this higher solar flux\textsuperscript{16,36,37,248,362,363}. In tandem, this means that a greater proportion of high energy UV-light was hitting the surface of the early-Earth compared to after the Great Oxidation Event (GOE) when \(O_2/O_3\) could then attenuate these high energy wavelengths. This means that chemistry occurring in surficial environments was likely to have been exposed to UV-light as a potential selection pressure.
UV-light can negatively affect molecules (vide supra, 2.2.8.1 and 3.2.3.2), indeed much work has focused on the negative impacts of UV-light on human health as skin cancer is one of the most common forms of cancer. Nucleotides and by extension biopolymers containing them (i.e. RNA, DNA) have been studied due to their reaction with UV-light. Photolesions have serious effects on health and are caused by [2+2] photocycloadditions of thymines, uridines and cytidines to cyclobutane pyrimidine dimers (CpDs); with the reaction of adjacent thymines being shown to be particularly rapid (Scheme 87). Recent work has examined how UV-light can also be involved in the repair mechanism for these photolesions, potentially opening up avenues for self-repair of photolesions during the origins of life before complex photolyase enzymes were evolved. Pyrimidine nucleosides can also form photohydrates when irradiated and is an intermediate in the UV-light driven deamination of cytidine to uridine.

Scheme 87 Cyclobutane pyrimidine dimer (CPD) formation from two adjacent thymidine residues by the action of UV-light.

Despite this UV-light has also been shown to be a powerful tool in prebiotic chemistry (vide supra, 1.6). Therefore, compounds vital to the origin of life may have required some element of resistance to the greater flux of UV-light hitting the surface of the earth to be incorporated into life. Indeed, there are compelling arguments that the canonical nucleobases in extant biology were selected in large part for their photochemical stability. This UV stability could have arisen via some mechanism of environmental protection or inherent molecular stability to excitation by UV-light, both of which was explored in this chapter.

Previous work has investigated some aspects of molecular UV stability but only for nucleobases (i.e. not attached to sugars as they would be in informational polymers) and mostly in the gas phase, most commonly explored by ultrafast spectroscopy or computational simulation. These findings have revealed that the nucleobases are supremely photostable (Figure 15) due to internal conversion and vibrational cooling.
that allow them to return to their ground state. They dissipate energy efficiently to their surroundings as heat as shown by their very low fluorescence quantum yield which suggests excited-state lifetimes shorter than 1 ps ($10^{-12}$ s). These short lifetimes are due to readily accessible conical intersections connecting the first excited ππ* state with the ground state$^{374-376}$. This stability of the isolated nucleobases has led to work proposing it as the reason for the central presence of the canonical bases in extant biology$^{369,373}$. However recent work has demonstrated that an investigation of other intermediates en route to nucleic acids may be valuable and indeed the UV stability of the nucleosides themselves is important to consider, this is the focus of the work presented in this chapter.

**Figure 15** Lowest singlet excited-state ($S_1$) lifetimes of canonical bases (blue squares) and non-canonical bases (red circles) in aqueous solution measured by femtosecond transient absorption. Reproduced from Beckstead et al. with permission from the PCCP Owner Societies$^{369}$.

Janicki et al. showed that key RNA (and 8-oxo-purine) precursors AAO arabinoaminooxazoline and arabino-oxazolidinone thione ara-96 were highly stable when irradiated (254/300 nm) for several hours$^{377}$. They proposed two reasons for this, either they did not absorb in the range required (major absorbance was <200 nm for AAO) or they do absorb UV-light in the 200-300 nm range but they can undergo a process of non-
radiative deactivation and thus avoids destructive bimolecular radical reactions. As these structures were intermediates on route to RNA other prebiotic building blocks, that are used in non-glycosylating routes to RNA, have also been investigated.

Indeed the UV stability of three azoles, that are all implicated in the origins of life, 2-aminooxazole \textit{2AO}, 2-aminothiazole \textit{2AT} and 2-aminoimidazole \textit{2AI}, has also been investigated (Figure 16)\textsuperscript{307}. They have distinct uses with \textit{2AO} being a precursor to RNA pyrimidines, \textit{2AT} as a powerful sequestration agent and \textit{2AI} as a activating agent to generate nucleotide phosphorimidazolides\textsuperscript{34,189,305}. \textit{Todd et al.} investigated their stabilities at a range of wavelengths (210-290 nm) and report that \textit{2AO} is the least photostable, while 2-aminoimidazole is the most photostable.

\textit{Figure 16 The structures of the three prebiotically important oxazoles (2AO, 2AT and 2AI) investigated for their UV stability\textsuperscript{307,357}.}

Interestingly, simply increasing the aqueous concentration of \textit{2AO} is sufficient to offer a significant enhancement in photochemical half-life\textsuperscript{357}. \textit{2AO} (0.1 mM) has a wavelength-dependent half-life of 7 h but when the concentration is increased to 1 mM or 10 mM the half-life is extrapolated to be 12 and 180 h respectively when irradiated (254 nm). Key to the work presented here, a similar half-life enhancement was also observed when nucleosides (A, G, I, C, U) were included. As discussed, the ability of nucleobases to rapidly return to their ground state after photoexcitation makes them candidates as sun-screening molecules\textsuperscript{378}. A (0.1 mM) when co-irradiated with equimolar \textit{2AO} (0.1 mM) gives a lifetime enhancement of 300% but the identity of the nucleoside affects the extent of lifetime enhancement with the purines (A and I) offering on average 200% lifetime enhancement but G and both pyrimidines (C and U) offering far less. Presumably, for the pyrimidines this is due to photohydrate formation (or other destructive photochemistry) but the explanation of G is less easily understood and clearly warrants further investigation.

However, there is no current literature addressing the UV stability of nucleosides themselves in aqueous solution. A considerable gap in the literature given that the RNA world theory is one of the most prevalent theories in the field of abiogenesis and that
almost all origins of life scenarios invoke an aqueous environment as the main solvent for RNA monomer forming reactions to occur in. This chapter seeks to address this gap in the literature by investigating the UV stability of the canonical nucleoside purines by irradiating them in water and observing their stabilities by quantifying their destruction over time. The pyrimidines are known to be far less photostable and accordingly were not investigated here.

![Figure 17 Canonical nucleosides adenosine (A) and guanosine (G) and non-canonical nucleosides Inosine (I), Diaminopurine (D), Xanthosine (X) for which the UV stability was investigated here.]

The destruction of the products from complex prebiotic chemistry would seemingly preclude them from involvement in a major role in the origin of life if no other method of protection can be invoked. Therefore, the photochemical screening of nucleosides with other nucleosides was also investigated. The assertion in the present literature that a compound’s UV stability may have been the key selection pressure by which extant canonical nucleoside structures were selected was particularly intriguing and to this end an investigation of the stability of some common non-canonical nucleosides: Inosine I, Diaminopurine D and Xanthosine X was also carried out.
4.2 Results and Discussion

4.2.1 General trends

To investigate their UV stability, solutions of the nucleosides being investigated (in degassed water, at various concentrations (20 mM, 10 mM, 2 mM)) were irradiated at 254 nm in a Rayonet photoreactor in quartz cuvettes (Exp 5.3.3.1, for full procedure). Aliquots were removed at the desired timepoints and analysed by $^1$H NMR with the addition of an internal standard, potassium hydrogen phthalate (KHP), for quantification. In some experiments, the pH was buffered by the addition of sodium phosphate (monobasic). The purines investigated were adenosine A and guanosine G, the two canonical purines, as well as non-canonical nucleosides diaminopurine D, inosine I, xanthosine X.

Adenosine A was chosen as the standard with which to compare all other nucleobases' stability to. A is one of the most stable nucleobases based on all data from previous studies$^{369,379}$. It is also the most abundant nucleobase found in modern cells (ATP, ~3 mM vs GTP ~0.5 mM) as it is used in genetic polymers but also in energy transfer, as the ubiquitous ATP, as well as other cofactors and may even be a hydrotrope, essential for peptide function$^{380,381}$.

A was observed, across all experiments investigated, to be uniquely stable. Over the course of several days of continuous irradiation A often remains largely intact even at low concentrations (Exp 5.3.3.8). Perhaps the stability of A to UV-light, explains biology's selection of it as its favoured purine and is a legacy of its UV stability at the origins of life.

This is presumably due to adenosine's (A) ability to, while photoexcited, access its conical intersection and decay back to its ground (dark) state very rapidly, giving it a very short excited state lifetime compared to other nucleosides which can have longer excited state lifetime giving it more opportunities to do destructive photochemistry.

One surprising result was the distinct lack of observable destruction products of the purine nucleosides. In previous work it has been shown that a common destructive pathway for the pyrimidine nucleobases (such as α-ribocytidine) is via the observable and isolatable oxazolidinone (e.g. ribo-88, vide supra, 1.5.5.2), but no such products were observed here. Indeed, the mechanism of destruction remains an open question as no precipitates were observed over the course of the reactions and no characteristic signals indicated free sugar and base were observed that would indicate depurination at the anomeric position. The only characteristic species that was observed as these
reactions proceed is formate (HCO$_2^-$), presumably formed by various oxidatively destructive processes and identified by a singlet resonance at ~8.25 ppm.

It was also observed, similarly to Todd et al., that concentration has a large impact on the amount of destruction\textsuperscript{957}. Whenever the concentration was increased the amount of destruction decreased. This will be explored in more detail here.

To ensure that differing amounts of destruction was due to the UV stabilities of the purine nucleobases, and not factors such as the sugar, control reactions with a different sugar were examined. A (2 mM) was irradiated (254 nm, pH 7, Exp 5.3.3.2) with 117 (2 mM) and no difference in the rate of destruction was observed over 3 d (remaining: A (25%), 117 (27%)), indicating that differing reactivity was indeed due to the structure of the purine.

To investigate the effect of changing purine two non-canonical purines of prebiotic interest were investigated. This was to determine which was more stable by directly comparing the amount of destruction.

Inosine is currently being investigated for a number of reasons in a prebiotic context. It’s deoxyribo analogue 130 is able to be formed alongside deoxyribo-adenosine 123 in recent work by Sutherland and co-workers (\textit{vide supra}, 1.5.5.4) and it also shows excellent properties when used in NERPE (\textit{vide supra}, 1.4.6.1)\textsuperscript{191,193,262}. There are also reports in the literature which have led to the assumption that inosine I is a the most photostable nucleoside, potentially more stable than A (Figure 15)\textsuperscript{369}. This has been inferred from pump-probe laser and simulation studies with hypoxanthine (inosine’s nucleobase; many of these studies have been carried out on the nucleobase without the ribose sugar attached) having one of the shortest excitation wavelengths implying it can, once excited, quickly decay back to its ground state and avoid destructive photochemistry. However, when A (2.1 mM) and I (2.1 mM) were irradiated (254 nm, pH 6.5, Exp 5.3.3.3) A was observed to be greatly more stable than I over 3 d of irradiation (remaining: A (41%), I (0%)) with very little I remaining even after 2 d (4%). This result appears to indicate that, despite the enviable photostability properties of hypoxanthine (the nucleobase of I), A is more stable than I and this appears to be independent of concentration (\textit{vide infra}). This also provides a warning that while calculations can be useful, over-inferring from these calculations can lead to incorrect assumptions without adequate experimental data.
Diaminopurine has also been investigated in NERPE\textsuperscript{382} but has a long lived excited state and is therefore believed to be unstable over long periods of irradiation\textsuperscript{372}. Surprisingly, when A (2.2 mM) and D (2.2 mM) were irradiated (254 nm, pH 6.5, Exp 5.3.3.4) A was indeed more stable than D over 3 d of irradiation (remaining: A (44%), D (5%)). However, D appears to be more stable than I (2 d: D (22%), I (4%), 3 d: D (5%), I (0%)) as it survives even to a low degree at these extended irradiation times. Despite the effect of concentration, even at such a difference as 2.1 vs 2.2 mM, I appears to be conclusively less stable than D which in turn is less stable than A.

\textit{Figure 18 Concentrations over time for the irradiation (254 nm) of A vs a) I b) D.}
As discussed, it has been proposed that during prebiotic chemistry taking place there could have been species that shielded more delicate molecules from UV-light\textsuperscript{371,378}. Additionally, Todd et al. have recently shown that A provided the best protection to 2AO when co-irradiated, therefore it was of great interest to probe whether A can be used in similar as a similar sunscreen for other nucleosides. To investigate this A was co-irradiated with nucleosides at a higher concentration.

Indeed, when A (8.7 mM) and D (1.8 mM) were irradiated (254 nm, pH 6.5, Exp 5.3.3.5) A helps D survive over 3 d of irradiation, but only to a small degree (remaining: D (10%) vs D (5%) with equimolar A). As expected, A exhibits higher stability at these higher concentrations (3 d: A 2.2 mM (44%); A 8.7 mM (65%)).

Shielding by A has a more pronounced effect on the survival of I. When A (8.3 mM) and I (1.7 mM) were irradiated (254 nm, pH 6.5, Exp 5.3.3.6) A shields I over 3 d of irradiation, allowing it to survive (remaining: I (25%) vs I (0%) with equimolar A). This is interesting as this was observed despite a lower starting concentration of A (8.3 mM vs 8.7 mM with D) which should cause greater destruction as well as the behaviour observed at lower equimolar concentrations of A. However, this may be explained by the rate of destruction of A as when it is destroyed less it may help other species survive more, here after 1 d 97% of A remained but with D only 87% did. This may indicate some co-operative behaviour available to I and A that D cannot undergo or that if A is destroyed quickly the knock-on effect, on other co-irradiated nucleosides, is large.

A also shields fellow canonical ribonucleoside G. However, contrary to the assertions discussed earlier that the canonical nucleosides were chosen for their UV stability, G is not as stable as A. Co-irradiation (254 nm, pH 6.5, Exp 5.3.3.7) of A (9.1 mM) and G (1.0 mM) revealed that G is relatively stable over the first day of irradiation (86%), but decays rapidly at longer times and is totally destroyed after 3d. The lower initial concentration of G (1.0 mM), based on concerns about its solubility, could contribute to this behaviour because, as discussed, the lower the concentration the higher the rate of destruction. Compared to D and I under similar conditions G is less stable than both (after 3 d: G (0%), D (10%), I (25%)).

To examine this intriguing behaviour of G an investigation into its place compared to other purines was carried out by the concomitant irradiation of five purines (A, G, I, D, X). Solubility was a factor that set the upper limit of concentrations especially for guanosine whose solubility is such that only concentrations of 1-2 mM or lower were able to be used. Additionally, the (C1')-H proton resonances of G and D are coincident. This
instead prompted the separate irradiation of four of the five purines (A, I, G, X and A, I, D, X) to provide unambiguous data.

Figure 19 $^1$H NMR (700 MHz, $H_2O/D_2O$ (9:1), noesygppr1d, 5.7-6.3 ppm) spectra to show the irradiation (254 nm) of adenosine A (2.0 mM), guanosine G (1.8 mM), inosine I (2.0 mM) and xanthosine X (1.1 mM) with KHP internal at pH 6.5: a) before irradiation; b) after 1 day of irradiation, A (95%), G (87%), I (80%), X (50%); c) after 2 d, A (94%), G (25%), I (65%), X (0%); d) after 3 d, A (58%), G (0%), I (26%), X (0%). (C1′)–H protons are labelled: A (●), G (▼), I (▲), X (■).

First investigating the mixture of both canonical nucleosides, A and G, their concomitant irradiation (254 nm, pH 6.5, Figure 19 and Exp 5.3.3.8) revealed that indeed that G (2 d: (25%), 3 d: (0%)) is less stable than A (2 d: (94%), 3 d: (58%)), as expected when at similar concentrations (A 2.0 vs G 1.8 mM) and also that X is the least stable, being rapidly destroyed in 2 d. I again benefits from increased overall concentration (6.9 mM total conc.) and survives well after 3d (26% vs 25% when total conc. 10 mM with A, Exp 5.3.3.6). Although only by a modest margin, the confirmation that G is less stable than I is particularly enticing given a dearth of prebiotic routes to G and recent work on I as an alternative to G$^{193,262}$. It is also of note that not only is it increasingly plausible that 2-thiocytosines may have played a role in early RNA and DNA$^{260,261}$ (and now TNA with the work demonstrated in Chapter 2) but it also forms stronger base pairs with inosine.
than β-ribo-85, leading to some proposing that together they could have encompassed an early surrogate system without G\(^{191}\). However, whether differing UV stability alone is enough to influence this is beyond the scope of this thesis.

To complete the picture an irradiation (254 nm, pH 6.5) of A, I, D and X was carried out (Exp 5.3.3.9). This added evidence to X having the lowest UV stability of the nucleosides tested, as here despite being in higher initial concentration (1.6 mM vs 1.1 mM (Exp 5.3.3.8) it also underwent extensive destruction (1 d: X (40%), A (93%)) with total destruction after 2 d. I exhibited greater UV stability than D under direct comparison (3 d: I (23%), D (0%)) but D appears more stable than G (2 d D (39%), G (25%)) and along with the irradiations previously discussed some level confidence can be ascribed to D being the more stable nucleoside.

### 4.3 Conclusion and future work

The final picture that emerges from this work is that A is uniquely photostable amongst purines in aqueous irradiations. Whether this is a clue as to the special role of A in extant biology remains to be further investigated. I is interesting in that its stability can be increased greatly alongside A allowing it to be more stable than D. The other canonical purine G is less stable than both I and D, but more stable than the very unstable X. Critically important is also the concentration of these irradiations with a strong trend in that the higher the concentration the lower the destruction and the more photostable the nucleobases are. While the solubility of several purines (G, X, D) preclude investigations of mixtures of higher than ~1-2 mM, some work should seek to fully probe the full effect and the thresholds behind this.

No rates were derived here but this chapter opens the door for further interesting and vital into understanding the fundamental properties of the ribonucleosides and the role they played on the early Earth under a more intense young sun.

Further work should examine a broader irradiation profile and wavelength dependence. As shown here the purines do not produce observable degradation products, these may tell us a great deal about potential degradation pathways and may hold the key to understanding the differing stabilities. It also allows for, as mentioned by Todd et al., a more precise quantification of these species and may allow accurate rates to be found for each nucleobase and collaborative work is ongoing towards this.
5 Experimental

5.1 General Experimental

All compounds were obtained from Sigma Aldrich, Alfa Aesar, Hopkins and Williams, Fisher Scientific, Carbosynth, BDH, Lancaster and VWR and used without further purification unless specified. Water (H$_2$O) refers to deionised water produced by an Elga Option 3 purification system. Flash Column Chromatography (FCC) and Reverse Phase Flash Column Chromatography (RP-FCC) were carried out using either Biotage SNAP or Kinesis TELOS cartridges in a Biotage Isolera One purification system. Sodium hydrogen sulfide (NaSH) was purchased from Sigma Aldrich as a hydrate, NaSH.$x$H$_2$O, and used without further purification. After reactions were complete, solutions were sparged of H$_2$S by bubbling argon or nitrogen gas for 1 h+ through the solution, whilst maintaining the pH of the solution at acidic pH to ensure efficient degassing and H$_2$S was quenched with sodium hypochlorite solution in 2 or 3 glass bubblers. Reactions using with anhydrous solvents were carried out in flame dried glassware that was cooled under N$_2$ with a Schlenk adapter. $^1$H, $^{13}$C and $^{31}$P NMR spectra were recorded on Bruker NMR spectrometers AVANCE Neo 700, AVANCE III 600, AVANCE III 400 and AVANCE 300, equipped with a Bruker room temperature 5 mm multinuclear gradient probe (700 MHz), 5 mm DCH cryoprobe (600 MHz) and a gradient probe (400 and 300 MHz). All reported chemical shifts (δ) are given in parts per million (ppm) relative to residual solvent peaks, and $^1$H and $^{13}$C spectra calibrated using the residual solvent peaks relative shift to TMS. Water supressed $^1$H NMR spectra were obtained using a 1D nuclear Overhauser enhancement spectroscopy (NOESY) pulse sequence (noesygpppr1d, Bruker) and all spectra were recorded at 298 K. Rapidly exchanging proton (O–H, N–H) resonances are not detected due to signal broadening and coalescence with the HOD signal in 9:1 H$_2$O/D$_2$O. Coupling constants (J) are given in Hertz (Hz). The following abbreviations refer to spin multiplicities: s (singlet); d (doublet); t (triplet); q (quartet); m (multiplet); br (broad signal); obs. (obscured/coincidental signals), app. (apparent) or any combination of these. Diastereotopic geminal (AB) spin systems coupled to an additional nucleus are reported as ABX. NMR data are stated as follows: chemical shift (multiplicity, coupling constants (J), number of protons, nuclear assignment). An electrothermal standard digital apparatus was used to record the melting points for all crystalline solids. Melting points are uncorrected and are quoted to the nearest C. Infrared spectra (IR) were recorded on a Shimadzu IR Tracer 100 FT-IR spectrometer. Absorption maxima are reported in wavenumber (cm$^{-1}$). Ultraviolet (UV) spectra were recorded on a Shimadzu UV-1800 UV spectrophotometer. Absorption maxima are reported in wavelength (nm). Mettler Toledo Seven Compact pH meter with a Mettler Toledo InLab semi-micro pH probe and a Corning pH meter 430 with a Fischerbrand (FB68801) semi-micro pH probe.
were used to measure the pH of solutions. UV irradiations were performed using a Rayonet RPR-200 with RPR-3000A lamps or with RPR-2547A lamps or using a Ultra-Violet Products Ltd (UVP, LLC) Pen-Ray 254 nm mercury lamp in a water-cooled Ace Glassware quartz microreactor. Mass spectra and accurate mass measurements were recorded by Thermo Finnigan MAT 900XP, VG70-SE, Waters LCT Premier XE Q-TOF or Thermo Orbitrap Q Exactive Plus instruments at the Department of Chemistry, University College London.

5.2 General Methods

5.2.1 Irradiation Method 1
Nucleosides/nucleotides (0.022 mmol) and NaSH (3 mg, 0.03 mmol) were dissolved in degassed water (0.5/1 mL, containing 10% D₂O). The solution was adjusted to pH 6.5 or 7 with HCl, transferred to a quartz NMR tube and irradiated with a mercury lamp (principle emission at 254 nm) for the desired duration. The reaction was periodically monitored by ¹H NMR spectroscopy.

5.2.2 Irradiation Method 2
Nucleosides were dissolved in degassed water (0.5/1 mL, containing 10% D₂O) to give the desired concentration (5 mM). The solution was adjusted to pH 6.5 or 7 with HCl, transferred to a quartz cuvette, re-degassed by N₂ ebullition (30 mins) and irradiated with a mercury lamp (principle emission at 254 nm) for the desired duration. The mixture was allowed to relax for 4 h, heated at 90 °C for 12 h to remove photohydrates and lyophilised. The lyophilizate was dissolved in D₂O and analysed by ¹H spectroscopy.

5.2.3 Phosphorylation Method 1
Adapted from Stairs et al.¹⁹²

Nucleosides (0.01 mmol), ammonium phosphate (1, 2, 4 equivalents) and urea (1.6 mmol) were heated to 140 °C for the specified time (5, 10 or 30 minutes) and then allowed to cool to room temperature. The resultant solids were dissolved in D₂O (2 mL) and lyophilised. The lyophilisate was dissolved in D₂O and analysed by ¹H and ³¹P NMR spectroscopy.

5.2.4 Phosphorylation Method 2
Nucleosides (0.01 mmol), ammonium phosphate (1, 2 or 4 equivalents) and urea (1.6 mmol) placed under an inert gas atmosphere (N₂ or Ar) and heated to 140 °C for the specified time (5, 10 or 30 minutes) and then allowed to cool to room temperature. The
resultant solids were dissolved in D$_2$O (2 mL) and lyophilised. The lyophilisate was redissolved in D$_2$O and analysed by $^1$H and $^{31}$P NMR spectroscopy.

### 5.2.5 Phosphatase Method 1

Nucleotides were dissolved in H$_2$O/D$_2$O (0.5/1 mL), then NaCl (5.9 mg) and MgCl$_2$ (2.0 mg) were added and the solution was adjusted to pH 8.0. Bovine intestinal phosphatase (10 µl, Sigma Aldrich) was added, and the solution maintained at 37 °C for 24 hours. The reaction was then lyophilized. The lyophilisate was dissolved in D$_2$O (0.5 mL) and analysed by NMR spectroscopy and sample spiking with authentic standards.

### 5.2.6 Thionucleoside Oxidation prep

The desired thionucleoside/tide (0.01 mmol) was dissolved in degassed phosphate buffer (0.5 mL, 0.1 M, 10% D$_2$O) in an eppendorf and the pH adjusted to 7 with NaOH/HCl. The reaction was transferred to an NMR tube and H$_2$O$_2$ (30% w/w, 3/5/10 µL, 0.03 mmol) was added and the tube shaken vigorously for 20 s then monitored by $^1$H NMR until the reaction was observed to be complete.

### 5.2.7 Dowex Preparation Method

DOWEX® ion exchange resin in the OH$^-$ form was prepared as follows:

- The required amount of DOWEX® in the chloride (Cl$^-$) form was weighed
- The DOWEX® was placed in a sintered funnel under vacuum pressure
- Aliquots of appropriate volume of 4 M NaOH were used to wash the DOWEX®
- The process repeated 4-5 times
- Residual NaOH washed away with water until filtrate was neutral (pH 7).
5.3 Supplementary Figures

5.3.1 TNA

5.3.1.1 One pot formation of TAO/EAO

Experimental Figure 1 $^1$H NMR (700 MHz, H$_2$O/D$_2$O 9:1, noesygppr1d, 3.0-8.0 ppm) spectra for the reaction of glycolaldehyde 3 (500 mM) with cyanamide 87 (100 mM) at pH 10 after 1 day at 50 °C. Internal standard of potassium hydrogen phthalate (KHP, 50 mM) allows for quantification of TAO (18%) and EAO (10%). (C1')-H and (C2')-H protons are labelled for: TAO (*). (C1')-H, (C2')-H and (C4'')-H protons are labelled for: EAO (▼). (C4)-H and (C5)-H protons are labelled for 2AO (●).
Experimental Figure 2 $^1$H NMR (600 MHz, H$_2$O/D$_2$O 9:1, zg30, 3.0-9.0 ppm) spectrum of the reaction of cyanamide 87 (333 mM) with glycolaldehyde 3 (666 mM) after 1 day at 60 °C at: a) pH 9; b) pH 10; c) pH 11. (C1')-H protons are labelled for: TAO (★), EAO (▼). (C4)-H and (C5)-H protons are labelled for 2AO (●).
Experimental Figure 3 $^1$H NMR (700 MHz, D$_2$O, zg30, 3.0-7.5 ppm) spectrum to show the crude reaction of 2-thioxazole 95 (420 mM) and glycolaldehyde 3 (360 mM) at 60 °C after 2.5 d. (C1')-H protons are labelled for: threo-96 (●), erythro-96 (▼). (C4)-H and (C5)-H protons are labelled for 95 (●).
5.3.1.3 Reaction of TAO and EAO with cyanoacetylene

Experimental Figure 4 $^1$H NMR (700 MHz, $\text{H}_2\text{O}/\text{D}_2\text{O}$ (9:1), noesygppr1d, 3.0-9.0 ppm) spectra to show: a) threose aminooxazoline TAO and erythrose aminooxazoline EAO (123 mM) with MSM (50 mM) internal standard; b) the reaction of a) with cyanacetylene (1 M, 1 mL) and sodium phosphate (67 mM) at pH 6.5, room temperature after 1 day. The reaction proceeds with near quantitative conversion (98%) by NMR spectroscopy.
5.3.1.4 Formamide thiolysis procedure

A mixture of diastereomeric 2,2'-anhydrocytidne (5.28 mg, 0.027 mmol) and methylsulfonylmethane (MSM, 10.55 mg, 0.11 mmol) were dissolved in D$_2$O (0.5 mL) and analysed by $^1$H NMR to determine initial concentrations. MSM and D$_2$O were removed by lyophilisation and the lyophilisate dissolved in formamide (1 mL). Sodium sulfide (59.3 mg, 0.63 mmol) was added and the resulting mixture was stirred for 18 h at room temperature. Water (5 ml) was added to the resulting solution and HCl under air ebullition until the solution reached pH 5, to degas and sparge H$_2$S from the solution. The sparged solution was adjusted to pH 7 and lyophilised. To the lyophilisate was added MSM (10.57 mg, 0.11 mmol) and redissolved in D$_2$O (1 mL) for $^1$H NMR analysis.
5.3.1.5 Thiolysis of *threo*-89 and *erythro*-89

![Chemical structures and NMR spectra](image)

**Experimental Figure 5** $^1$H NMR (700 MHz, D$_2$O, zg30, 3.0-9.0 ppm) spectra to show: a) *threo*-anhydrocytidine 89 and *erythro*-anhydrocytidine 89 (49 mM); b) the products of the reaction of a) with NaSH (635 mM) in formamide (1 mL) at room temperature overnight after then being lyophilized and dissolved in D$_2$O. The products β-*threo*-2-thiocytidine 111 and β-*erythro*-2-thiocytidine 111 (78% yield) are labelled, as well as the internal standard methylsulfonylmethane (MSM, 3.09 ppm) which was used to quantify the reaction yield. (C6)-H, (C5)-H and (C1′)-H protons are labelled: β-*threo*-2-thiocytidine 111 (▼) and β-*erythro*-2-thiocytidine 111 (▲).
5.3.1.6 Hydrolysis of threo-89 and erythro-89

Experimental Figure 6 $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), (a)Robust5_ver_03b ; b) noesygpprd) 3.5-8.5 ppm) spectra to show: a) authentic sample of threo-89 and erythro-89; incubation of threo-89 and erythro-89 (50 mM) at RT for 5 d at: b) pH 8 ($\beta$-threo-85 (33%), $\beta$-erythro-85 (27%), threo-89 (17%), erythro-89 (23%)); c) pH 9 ($\beta$-threo-85 (38%), $\beta$-erythro-85 (32%), threo-89 (12%), erythro-89 (18%)); d) pH 10 ($\beta$-threo-85 (45%), $\beta$-erythro-85 (41%), threo-89 (5%), erythro-89 (9%)). (C6)-H and (C1′)-H protons are labelled: $\beta$-threo-85 (▼) and $\beta$-erythro-85 (▲). (C6)-H and (C2′)-H protons are labelled: threo-89 (●), erythro-89 (∗).
5.3.1.7 Reaction of $\beta$-threo-111 and $\beta$-erythro-111 with hydrogen peroxide

**Experimental Figure 7** $^1$H NMR (700 MHz, $D_2O$, zg30, 3.0-8.5 ppm) spectra to show:

a) $\beta$-threo-111 and $\beta$-erythro-111 (17 mM); b) the reaction of a) with hydrogen peroxide (98 mM) at room temperature after 10 mins. (C6)-H, (C5)-H and (C1')-H protons for are labelled: $\beta$-threo-111 (▼), $\beta$-erythro-111 (▲), (C6)-H, (C5)-H, (C1')-H and (C2')-H protons for are labelled: threo-89 (●) and erythro-89 (★).
5.3.1.8 Hydrogen peroxide oxidation of α-threo-111

 Experimental Figure 8 $^1$H NMR (700 MHz, H$_2$O/D$_2$O 9:1, (a) zg30, b) noesypppr1d), 3.7-8.5 ppm) spectra to show: a) α-threo-111 (20 mM); b) the reaction of a) with hydrogen peroxide (59 mM) after 10 mins. (C5)-H, (C5′)-H and (C1′)-H protons are labelled: α-threo-111 (▼), α-threo-85 (●).
5.3.1.9 Reaction of β-threo-111 and α-threo-111 with hydrogen peroxide

\[
\text{\(\text{H}_2\text{O}_2, \text{P}_1\)} \rightarrow \text{\(\text{H}_2\text{O}_2\)}
\]

\[
\beta\text{-threo-111} \quad \alpha\text{-threo-111} \quad \text{threo-89} \quad \alpha\text{-threo-85}
\]

**Experimental Figure 9** \(^1\text{H} \text{NMR (700 MHz, H}_2\text{O/D}_2\text{O (9:1), Robust5_ver_03b, 3.0-9.0 ppm)} \) spectra to show: a) \(\beta\text{-threo-111} \) (12.6 mM) and \(\alpha\text{-threo-111} \) (18 mM); b) the reaction of a) with hydrogen peroxide (0.09 mmol, 196 mM) after 12 h. (C6)-H, (C5)-H and (C1′)-H protons (and (C2′)-H proton for threo-89) are labelled: \(\beta\text{-threo-111} \) (▼), \(\alpha\text{-threo-111} \) (▲), threo-89 (●), \(\alpha\text{-threo-85} \) (★).
5.3.1.10 UV-Vis Spectra

**Experimental Figure 10** Normalised UV-Vis absorption spectra for β-threo-111.

**Experimental Figure 11** Normalised UV-Vis absorption spectra for β-erythro-111.
Experimental Figure 12 Normalised UV-Vis absorption spectra for β-threo-112.
5.3.1.11 Photoanomerization of β-threo-111

\[
\begin{align*}
\text{β-threo-111} \quad \text{▼} & \quad \text{hv, 254 nm} \quad \text{H}_2\text{S, pH 7.0} \\
& \quad \text{→} \\
\text{α-threo-111} \quad \text{★}
\end{align*}
\]

**Experimental Figure 13** \(^1\)H NMR (700 MHz, \(\text{H}_2\text{O}/\text{D}_2\text{O} \ (9:1), \ 	ext{Robust5} \ _\text{ver}_03b, \ 3.8-8.0 \ \text{ppm})\) spectra to show the irradiation (254 nm) of β-threo-thiocytidine β-threo-111 (25 mM) in \(\text{H}_2\text{O}/\text{D}_2\text{O} \ (9:1)\) at pH 7 with NaSH (0.013 mmol, 26 mM), MSM internal standard (73 mM, resonance at 3.09 ppm omitted): a) before irradiation; b) after 16 h of irradiation (43%); c) after 1 day of irradiation (46%); d) after 1.5 days of irradiation (51%); e) after 2 days of irradiation (51%). \((\text{C6})-\text{H, (C5)}-\text{H and (C1')}-\text{H protons are labelled: β-threo-111 (▼), α-threo-111 (★)).\)
Experimental Figure 14 $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), Robust5_ver_03b, 3.0-8.5 ppm) spectra to show the irradiation (254 nm) of β-threo-thiocytidine 111 (25 mM) in H$_2$O/D$_2$O (9:1) at pH 7 with NaSH (0.013 mmol, 26 mM), MSM internal standard (73 mM, resonance at 3.09 ppm): a) before irradiation, b) after 1 day of irradiation (46%), c) after 1.5 days of irradiation (51%), d) after 2 days of irradiation (51%), e) after 2.5 days of irradiation (49%), f) after 3 days of irradiation (49%), g) after 3.5 days of irradiation (43%) and h) after spiking g) with authentic α-threo-111.
5.3.1.12 Photoanomerization of β-threo-111 and β-erythro-111

Experimental Figure 15 $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), noesygpr1d, 3.0-8.5 ppm) spectra to show the irradiation (254 nm) of 18 mM β-threo-111 and 16 mM β-erythro-111 in H$_2$O/D$_2$O (9:1) at pH 7 with NaSH (0.1 mmol, 113 mM), MSM internal standard (18.6 mM, resonance at 3.09 ppm): a) before the irradiation; b) after 1 day of irradiation (α-threo-111 (40%), α-erythro-111 (42%)); c) after 2 days of irradiation (α-threo-111 (52%), α-erythro-111 (51%)); d) after 3 days of irradiation (α-threo-111 (40%), α-erythro-111 (36%)). (C6)-H and (C1′)-H protons are labelled: β-threo-111 (▼), β-erythro-111 (▲), α-threo-111 (●), α-erythro-111 (★).
5.3.1.13 Photoanomerization of β-threo-111 and β-erythro-111 spiked with authentic α-threo-111 and α-erythro-111

Experimental Figure 16 ¹H NMR (700 MHz, H₂O/D₂O (9:1), noesygpppd, 3.0-8.5 ppm) spectra to show the irradiation (254 nm) of 18 mM threo-9 and 16 mM erythro-10 in H₂O/D₂O (9:1) at pH 7 with NaSH (0.1 mmol, 113 mM), MSM internal standard (18.6 mM, resonance at 3.09 ppm): a) before the irradiation; b) after 1 day of irradiation; c) after 2 days of irradiation; d) after 3 days of irradiation; e) after 3 days of irradiation and spiked with authentic α-threo-111 and f) spiked with authentic α-erythro-111. (C6)-H and (C1′)-H protons are labelled: β-threo-111 (▼), β-erythro-111 (▲), α-threo-111 (●), α-erythro-111 (★).
Experimental Figure 17. $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), noesygppr1d, 5.9-8.5 ppm) spectra to show the irradiation (254 nm) of 18 mM β-threo-111 and 16 mM β-erythro-111 in H$_2$O/D$_2$O (9:1) at pH 7 with NaSH (0.1 mmol, 113 mM), MSM internal standard (18.6 mM, resonance at 3.09 ppm omitted): a) before the irradiation; b) after 1 day of irradiation; c) after 2 days of irradiation; d) after 3 days of irradiation; e) after 3 days of irradiation and spiked with authentic α-threo-111 and f) spiked with authentic α-erythro-111. (C6)-H and (C1′)-H protons are labelled: β-threo-111 (▼), β-erythro-111 (▲), α-threo-111 (●), α-erythro-111 (★).
5.3.1.15 Irradiation of β-threo-85

\[
\begin{align*}
\text{β-threo-85} & \xrightarrow{hv, 254 \text{ nm}} \text{β-threo-86} \\
\text{pH 6.5} & \xrightarrow{254 \text{ nm, Irradiation Method 2}} \text{β-threo-85} (30\%), \text{threo-88} (37\%), \text{β-threo-86} (33\%). \\
(C5) & \text{-H protons are labelled: β-threo-85 (●), β-threo-86 (●). (C1')-H and (C2')-H protons are labelled: threo-88 (▼).}
\end{align*}
\]

Experimental Figure 18 \(^1\)H NMR (600 MHz, H\(_2\)O/D\(_2\)O (9:1), (a) noesygpppr1d, b) zg30), 3.5-8.5 ppm) spectra to show: authentic standard of β-threo-85; b) the irradiation (254 nm, Irradiation Method 2) of β-threo-85 (2.5 mM) at pH 6.5 for 2 d to give β-threo-85 (30%), threo-88 (37%), β-threo-86 (33%). (C5)-H protons are labelled: β-threo-85 (●), β-threo-86 (●). (C1')-H and (C2')-H protons are labelled: threo-88 (▼).
Experimental Figure 19: 1H NMR (600 MHz, H2O/D2O (9:1), noesygppr1d, 3.0-9.0 ppm) spectra to show: authentic standard of β-threo-85; b) the irradiation (254 nm, Irradiation Method 1) of β-threo-85 (5 mM) at pH 6.5 for 1 d. (C1′)-H protons are labelled: β-threo-85 (û). (C1′)-H and (C2′)-H protons are labelled: threo-88 (▼).
5.3.1.16 Irradiation of β-erythro-85

Experimental Figure 20 $^1$H NMR (600 MHz, H$_2$O/D$_2$O (9:1), (a) zg30, b) noesygpr1d, c) noesygpr1d), 3.35-7.85 ppm) spectra to show: a) authentic β-erythro-85; b) the irradiation (254 nm) of a) (38 mM) after 1 d; c) after 2 d. (C6)-H, (C5)-H and (C1′)-H protons are labelled for β-erythro-85 (★). (C1′)-H, (C2′)-H and (C4′)-H protons are labelled for erythro-88 (▼).
5.3.1.17 Irradiation of β-threo-148

Experimental Figure 21 $^1$H NMR (600 MHz, H$_2$O/D$_2$O, (a) zg30, b) noesypppr1d), 1.75-8.0 ppm) spectra to show: a) authentic β-threo-148; b) the irradiation (254 nm, Irradiation Method 2) of β-threo-148 (10 mM) at pH 6.5 for 3 d. (C1')-H and (C2')-H protons are labelled: β-threo-148 (★). No characteristic signals for threo-88 were observed.
5.3.1.18 Irradiation of β-threo-112

Experimental Figure 22 $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), (a) zg30, b) Robust5_ver_03b, c) Robust5_ver_03b), 3.5-8.5 ppm) spectra to show: the irradiation (254 nm) of β-threo-thiouridine 112 (0.02 mmol, 44 mM) and NaSH.xH$_2$O (0.06 mmol, 120 mM) at pH 7 a) before irradiation, b) after 16 h c) after 1 d.
5.3.1.19 Irradiation of α-threo-111

Experimental Figure 23 ¹H NMR (700 MHz, H₂O/D₂O (9:1), noesygpr1d, 3.0-8.5 ppm) spectra to show the irradiation (254 nm) of α-threo-111 (41 mM) in H₂O/D₂O (9:1) at pH 7 with NaSH (0.11 mmol, 227 mM), MSM internal standard (35 mM, resonance at 3.09 ppm): a) before the irradiation; b) after 1 day of irradiation; c) after 2 days of irradiation and d) after 3 days of irradiation. (C6)-H, (C5)-H and (C1′)-H protons (where applicable) are labelled: α-threo-111 (▼), 2-thiocytosine 113 (∗).
Experimental Figure 24 $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), noesygppr1d, 3.0-8.5 ppm) spectra to show the irradiation (254 nm) of $\alpha$-threo-111 (41 mM) in H$_2$O/D$_2$O (9:1) at pH 7 with NaSH (0.11 mmol, 227 mM), MSM internal standard (35 mM, resonance at 3.09 ppm): a) after 1 day of irradiation; b) after 2 days of irradiation; c) after 3 days of irradiation; d) after 4 days of irradiation; e) after 5 days of irradiation and f) after 6 days of irradiation.
5.3.1.20 Irradiation of α-erythro-111

Experimental Figure 25 $^1$H NMR (700 MHz, $\text{H}_2\text{O}/\text{D}_2\text{O}$ (9:1), noesygppr1d, 3.0-8.5 ppm) spectra to show the irradiation (254 nm) of $\alpha$-erythro-111 (41 mM) in $\text{H}_2\text{O}/\text{D}_2\text{O}$ (9:1) at pH 7 with NaSH (0.12 mmol, 248 mM), MSM internal standard (43 mM, resonance at 3.09 ppm): a) before the irradiation; b) after 1 day of irradiation; c) after 2 days of irradiation; d) after 3 days of irradiation; e) after 4 days of irradiation and f) after 5 days of irradiation. (C6)-H, (C5)-H and (C1')-H protons (where applicable) are labelled: $\alpha$-erythro-111 (▲), 2-thiocytosine 113 (★)
5.3.1.21 Irradiation of α-threo-111 and α-erythro-111

Experimental Figure 26 $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), noesygpp1d, 3.0-8.5 ppm) spectra to show the irradiation at 254 nm of α-threo-111 and α-erythro-111 (6.5 mM threo: 6.8 mM erythro) in H$_2$O/D$_2$O (9:1) at pH 7 with NaSH (0.14 mmol, 139 mM), MSM internal standard (13.3 mM, resonance at 3.09 ppm): a) before the irradiation; b) after 1 day of irradiation; c) after 2 days of irradiation and d) after 3 days of irradiation. (C6)-H, (C5)-H and (C1′)-H protons (where applicable) are labelled: α-threo-111 (▼), α-erythro-111 (▲), 2-thiocytosine 113 (◆).
5.3.1.22 Hydrolysis of α-threo-111

Experimental Figure 27 $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), Robust5_ver_03b, 4.0-8.0 ppm) spectra to show: conversion of α-threo-111 (92%) to α-threo-85 (8%) on incubation in phosphate buffer (0.1 mM) for 45 d at 40 °C. (C6)-H, (C5)-H and (C1’)-H protons are labelled: α-threo-111 (★), α-threo-85 (▼).
5.3.1.23 Cyanoacetylene accelerated hydrolysis

Experimental Figure 28 $^1$H NMR (700 MHz, $\text{H}_2\text{O}/\text{D}_2\text{O}$ (9:1), (a) zg30, b)noesygpppr1d), 4.05-8.25 ppm) spectra to show: a) authentic sample of $\alpha$-threo-111; b) the reaction of $\alpha$-threo-111 (33 mM) with cyanoacetylene (100 μL, 1 M) after 18 h. The major identifiable products of the reaction are $\alpha$-threo-85 and S-cyanovinylated intermediate $\alpha$-threo-149. (C6)-H, (C5)-H and (C1')-H protons are labelled: $\alpha$-threo-111 (●). (C6)-H, (C5)-H and (C1')-H and cyanovinyl protons are labelled: $\alpha$-threo-149 (●).
5.3.1.24 Phosphorylation of β-erythro-111

Experimental Figure 29 $^1$H NMR (700 MHz, D$_2$O, (a) zg30, b) lc1pngpf2) 3.0-8.52 ppm) spectra to show the phosphorylation of β-erythro-111 (0.06 mmol) with ammonium phosphate (0.06 mmol): a) β-erythro-111; b) crude products of the reaction of β-erythro-111 under Phosphorylation Method 1.
5.3.1.25 Irradiation of $\beta$-threo-111 and $\beta$-erythro-151

**Experimental Figure 30** $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), noesyppr1d, 3.0-9.0 ppm) spectra to show the irradiation at 254 nm of 12 mM $\beta$-threo-111 and 12 mM $\beta$-2',3'-cyclic-phosphate $\beta$-erythro-151 in H$_2$O/D$_2$O (9:1) at pH 7 with NaSH (0.11 mmol, 109 mM), MSM internal standard (18.1 mM, resonance at 3.09 ppm): a) before the irradiation; b) after 1 day of irradiation ($\beta$-threo-111 (33%), 113 (36%)); c) after 2 days of irradiation ($\beta$-threo-111 (35%), 113 (38%)); d) after 3 days of irradiation ($\beta$-threo-111 (47%), 113 (52%)); e) after 4 days of irradiation ($\beta$-threo-111 (29%), 113 (30%)); f) after 4 days of irradiation and spiked with authentic $\alpha$-threo-111 and g) spiked with authentic 2-thiocytosine 113. (C6)-H, (C1')-H and (C5)-H (for 12) protons are labelled: $\beta$-threo-111 (▼), $\beta$-2',3'-cyclic-phosphate-erythro-151 (▲), $\alpha$-threo-111 (●), 2-thiocytosine 113 (★).
Experimental Figure 31 $^{31}$P NMR (284 MHz, H$_2$O/D$_2$O (9:1), zgpg30, -200-150 ppm) spectra to show the irradiation (254 nm) of 12 mM threo-111 and 12 mM β-2',3'-cyclic-phosphate-erythro-151 in H$_2$O/D$_2$O (9:1) at pH 7 with NaSH (0.11 mmol, 109 mM), MSM internal standard (18.1 mM, resonance at 3.09 ppm): a) before the irradiation; b) after 1 day of irradiation; c) after 2 days of irradiation and d) after 3 days of irradiation.
5.3.1.26 Irradiation of β-erythro-151

![Chemical Structure](image)

Experimental Figure 32 $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), Robust5_ver_03b, 3.8-8.5 ppm) spectra to show the irradiation at 254 nm of 7.1 mM β-2′,3′-cyclic-phosphate-β-erythro-151 in H$_2$O/D$_2$O (9:1) at pH 7 with NaSH (0.09 mmol, 184 mM), MSM internal standard (18.1 mM, resonance at 3.09 ppm omitted): a) before the irradiation; b) after 1 day of irradiation (113 (31%)); c) after 2 days of irradiation (113 (18%)); d) after 3 days of irradiation (113 (18%)). (C6)-H, (C5)-H and (C1′)-H protons are labelled for β-2′,3′-cyclic-phosphate-erythro-151 (▲) and (C6)-H and (C5)-H protons for 2-thiocytosine 113 (*).
5.3.1.27 Phosphorylation and irradiation of β-threo-111 and subsequent treatment by phosphatase

Experimental Figure 33 to show the phosphorylation and subsequent irradiation of β-threo-111 and treatment by phosphatase.
Experimental Figure 34 $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), Robust5_ver_03b, 4.9-8.5 ppm) spectra to show: a) phosphorylation of β-threo-111 by Phosphorylation Method 2; b) irradiation (254 nm) of a) with NaSH (0.1 mmol, 199 mM) after 1 day; c) after 2 days of irradiation; d) after 3 days of irradiation; e) spectra d) after undergoing Phosphatase Method 1; f) spectra e) spiked with authentic standard of α-threo-111. (C6)-H protons are labelled: β-threo-111 ($\blacklozenge$), β-threo-polyphosphates ($\blacktriangle$), α-2',3'-cyclic-phosphate-erythro-152 ($\blacklozenge$), α-threo-polyphosphates ($\blacktriangledown$), α-threo-111 ($\blacktriangledown$), 2-thiocytosine 113 ($\blacklozenge$).
5.3.1.28 Phosphorylation of β-threo-111

**Experimental Figure 35** $^1$H NMR (700 MHz, D$_2$O, zg30, 3.75-8.5 ppm) spectra to show: phosphorylation of β-threo-111 by Phosphorylation Method 2 for varying times and equivalents of phosphate: a) 25 mins, P$_i$ 2 equiv., β-threo-111 (30%), β-threo-polyphosphates (36%), α-erythro-111 (34%); b) 10 mins, P$_i$ 6 equiv., β-threo-111 (36%), β-threo-polyphosphates (40%), α-erythro-152 (24%); c) 5 mins, P$_i$ 10 equiv., β-threo-111 (39%), β-threo-polyphosphates (29%), α-erythro-152 (32%), (C6)-H, (C5)-H and (C1')-H protons are labelled: β-threo-111 (◆), β-threo-polyphosphates (▲), (C6)-H, (C5)-H, (C1')-H, (C2')-H and (C3')-H protons are labelled: α-erythro-152 (★).
Experimental Figure 36 $^{31}$P NMR (284 MHz, D$_2$O, zgpg30, -50-50 ppm) spectra to show the phosphorylation of $\beta$-threo-111 by Phosphorylation Method 2 (5 mins, P, 10 equiv.).
5.3.1.29 Phosphorylation of α-threo-111

**Experimental Figure**

$^1$H NMR (700 MHz, D$_2$O, zg30, 4.0-8.0 ppm) spectra to show: a) authentic sample of α-threo-111; b) phosphorylation of α-threo-111 (0.03 mmol) by Phosphorylation Method 1 giving α-threo-111 (35%) and α-threo-polyphosphates (65%); c) phosphorylation of α-threo-111 (0.02 mmol) by Phosphorylation Method 2 giving α-threo-111 (36%) and α-threo-polyphosphates (64%). (C1′)-H protons are labelled: α-threo-polyphosphates (●), α-threo-111 (▼).
Experimental Figure 38: $^{31}P$ NMR (284 MHz, D$_2$O, zgpg30, -23-30 ppm) spectra to show the phosphorylation of α-threo-111 by Phosphorylation Method 1.
5.3.1.30 Phosphorylation of α-threo-111

Experimental Figure 39 ¹H NMR (700 MHz, D₂O, (a)nosesygpr1d, b) Robust5_ver_03b, 4.0-8.0 ppm) spectra to show: a) authentic sample of α-threo-111; b) phosphorylation of α-threo-111 (0.03 mmol) by Phosphorylation Method 1 giving >5 distinct phosphorylation products but not α-erythro-152 (confirmed by ³¹P HMBC). (C1')-H protons are labelled: α-threo-polyphosphates (●), α-threo-111 (▼).
Experimental Figure 40 $^{31}$P NMR (284 MHz, D$_2$O, zgpg30, -25-30 ppm) spectra to show the phosphorylation of α-threo-111 by Phosphorylation Method 1.
5.3.1.31 Phosphorylation of α-erythro-111

Experimental Figure 41 $^1$H NMR (700 MHz, D$_2$O, (a) noesyppr1d, b) zg30), 3.5-8.5 ppm) spectra to show: a) authentic sample of α-erythro-111; b) phosphorylation of α-erythro-111 (0.02 mmol) by Phosphorylation Method 2 (10 mins, P$_2$i 2.5 equiv.) giving 2,2'-thioanhydro (20%), α-erythro-polyphosphates (23%), α-erythro-111 (31%), 2-thiocytosine 113 (13%), cytosine 71 (13%); c) phosphorylation of α-erythro-111 (0.03 mmol) by Phosphorylation Method 1 (10 mins, P$_2$i 1 equiv.) giving 2,2'-thioanhydro (13%), α-erythro-polyphosphates and α-erythro-111 (67%), 2-thiocytosine 113 (11%), cytosine 71 (9%). (C5)-H protons are labelled: α-erythro-111 (▼), α-erythro-polyphosphates (▲), 2-thiocytosine 113 (▲), cytosine 71 (■). (C6)-H protons are labelled: 2,2'-thioanhydro (●).
**Experimental Figure 42** $^{31}$P NMR (284 MHz, D$_2$O, zgpg30, -30-30 ppm) spectra to show the phosphorylation of α-erythro-111 (0.02 mmol) by Phosphorylation Method 2 (10 mins, P, 2.5 equiv.).
5.3.2 DNA

5.3.2.1 Reaction of *ara*-102 with ethanethiol in DMF

*Experimental Figure 43* $^1$H-$^{13}$C HMBC ($^1$H: (600 MHz, 0-8.5 ppm), $^{13}$C: (151 MHz, -10-220 ppm), $D_2$O) spectrum of the reaction of 2',8-cycloadensine *ara*-102 with ethanethiol (8 equiv.) in DMF at 70 °C after 1 d. Characteristic cross-coupling between $SCH_2CH_3$ proton resonance (2.44 ppm) and C2' carbon resonance (49 ppm).
5.3.2.2 Reaction of *ara*-89 with EtSH and NaH

**Experimental Figure 44** $^1$H NMR (600 MHz, D$_2$O, zg30, 1.0-8.4 ppm) spectrum to show: a) authentic standard of *ara*-89; b) the crude reaction of *ara*-89 (71 mM) with EtSH (58 mM) and NaH (268 mM), in DMF for 24 h at room temperature. (C1')-H and (C2')-H protons are labelled: *ara*-89 (■), 166 (●).
5.3.2.3 Reaction of *ara*-126 with EtSH and NaH

**Experimental Figure 45** $^1$H NMR (600 MHz, D$_2$O, noesygpp1d, 1.15-7.95 ppm) spectra to show: the crude reaction of *ara*-89 (71 mM) with EtSH (58 mM) and NaH (268 mM), in DMF for 24 h at room temperature. (C1')-H and (C2')-H protons are labelled: 163 (●).
5.3.2.4 Thiolysis Screening Method 1

Premixed thiol(s) (0.1-1.0 mmol) in DMF/formamide/water/acetonitrile/dioxane (1 mL) were added to the pyrimidine/purine anhydronucleoside (0.1 mmol) and heated for the desired amount of time (1-16 d). The mixture was degassed by N₂ ebullition (30 mins) to remove excess thiol. The solvent was removed in vacuo and the residue was redissolved in D₂O (1 mL) and a sample of this (0.5 mL) removed for analysis by ¹H NMR spectroscopy.
5.3.2.5 Reactions of *ara*-102 in DMF with EtSH/EtSNa

Reactions of *ara*-102 under Thiolysis Screening Method 1 in DMF.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Temp (°C)</th>
<th>Time (days)</th>
<th>Equivalents</th>
<th>(^1)H NMR integration relative to other species</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>EtSH</td>
<td>EtSNa</td>
</tr>
<tr>
<td>A</td>
<td>RT</td>
<td>1</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>B</td>
<td>70</td>
<td>1</td>
<td>2.5</td>
<td>7.5</td>
</tr>
<tr>
<td>C</td>
<td>70</td>
<td>1</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>D</td>
<td>70</td>
<td>1</td>
<td>7.5</td>
<td>2.5</td>
</tr>
<tr>
<td>E</td>
<td>70</td>
<td>1</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>F</td>
<td>RT</td>
<td>1</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>G</td>
<td>RT</td>
<td>4</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>H</td>
<td>RT</td>
<td>16</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>I</td>
<td>70</td>
<td>1</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>J</td>
<td>70</td>
<td>1</td>
<td>9.9</td>
<td>0.1</td>
</tr>
<tr>
<td>K</td>
<td>RT</td>
<td>1</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

\(^1\) MUS – minor unidentifiable species
### 5.3.2.6 Reactions of *ara-102* in formamide with EtSH/EtSNa

Reactions of *ara-102* under Thiolysis Screening Method 1 in formamide.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Temp. °C</th>
<th>Time (days)</th>
<th>Equivalents</th>
<th>EtSH</th>
<th>EtSNa</th>
<th><strong>H NMR integration relative to other species</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>157 158 164 MUS² <em>ara-102</em></td>
</tr>
<tr>
<td>A</td>
<td>70</td>
<td>1</td>
<td>8 2</td>
<td>23</td>
<td>65</td>
<td>4 6 2</td>
</tr>
<tr>
<td>B</td>
<td>RT</td>
<td>1</td>
<td>8 2</td>
<td>5</td>
<td>50</td>
<td>10 0 35</td>
</tr>
<tr>
<td>C</td>
<td>RT</td>
<td>2</td>
<td>8 2</td>
<td>7</td>
<td>70</td>
<td>9 0 14</td>
</tr>
<tr>
<td>D</td>
<td>RT</td>
<td>4</td>
<td>8 2</td>
<td>8</td>
<td>80</td>
<td>5 2 5</td>
</tr>
<tr>
<td>E</td>
<td>125</td>
<td>1</td>
<td>8 2</td>
<td>37</td>
<td>22</td>
<td>27 9 5</td>
</tr>
<tr>
<td>F</td>
<td>100</td>
<td>1</td>
<td>8 2</td>
<td>13</td>
<td>37</td>
<td>12 21 17</td>
</tr>
<tr>
<td>G</td>
<td>120</td>
<td>1</td>
<td>9 1</td>
<td>5</td>
<td>20</td>
<td>5 36 34</td>
</tr>
<tr>
<td>H</td>
<td>120</td>
<td>1</td>
<td>5 5</td>
<td>27</td>
<td>41</td>
<td>11 12 9</td>
</tr>
</tbody>
</table>

² MUS – minor unidentifiable species
5.3.2.7 Reaction of ara-126 in formamide

Experimental Figure 46 $^1$H NMR (600 MHz, HCONH$_2$, lc1pngpps, 1.0-9.0 ppm) spectrum to show the crude reaction of ara-126 (97 mM) with EtSH/EtSNa (5:5, 10 equiv.), in HCONH$_2$, 120 °C a) after 0 h; b) after 4 d (163, 77%), minor unidentifiable species (33%). (C1′)-H and (C2′)-H protons are labelled: ara-126 (●), 163 (▼).
5.3.2.8 Reaction of ara-89 in formamide

Experimental Figure 47 $^1$H NMR (600 MHz, HCONH$_2$, 1c1pngpps, 1.0-9.0 ppm) spectrum to show the crude reaction of ara-89 (97 mM) in HCONH$_2$ at 120 °C after 4 days with: a) EtSH/EtSNa (8:2, 10 equiv.) giving 166 (42%), β-ara-85 (42%) and minor unidentifiable species (16%); b) EtSH/EtSNa (5:5, 10 equiv.) giving 166 (35%), 167 (39%), β-ara-85 (19%) and minor unidentifiable species (16%). (C1′)-H, (C5)-H and (C2′)-H protons are labelled for 166 (●), 167 (▼). (C1′)-H and (C5)-H protons are labelled for β-ara-85 (▲).
### 5.3.2.9 Reaction of *ara-102/ara-126* with prebiotic sulfur nucleophiles

Reactions of *ara-102*, or *ara-126* in water under Thiolysis Screening Method 1.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Anhydro (100 mM)</th>
<th>Solvent</th>
<th>pH</th>
<th>Temp. (°C)</th>
<th>Time (days)</th>
<th>Thiol</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>H₂O</td>
<td>8</td>
<td>60</td>
<td>1</td>
<td>(\text{HS}^\text{-NH}_2)</td>
<td>8 addition</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
<td>H₂O</td>
<td>8</td>
<td>60</td>
<td>1</td>
<td>(\text{HS}^\text{-O-NH}_2)</td>
<td>No reaction</td>
</tr>
<tr>
<td>C</td>
<td>A</td>
<td>H₂O</td>
<td>8</td>
<td>60</td>
<td>1</td>
<td>(\text{HS}^\text{-OH-SH})</td>
<td>No reaction</td>
</tr>
<tr>
<td>D</td>
<td>A</td>
<td>H₂O</td>
<td>Contact, 9</td>
<td>50</td>
<td>10</td>
<td>(\text{HS}^\text{-N}^\text{-})</td>
<td>8 addition</td>
</tr>
<tr>
<td>E</td>
<td>A</td>
<td>H₂O</td>
<td>Contact</td>
<td>70</td>
<td>1</td>
<td>(\text{O}^\text{-S}^-)</td>
<td>No reaction</td>
</tr>
<tr>
<td>F</td>
<td>U</td>
<td>H₂O</td>
<td>Contact</td>
<td>70</td>
<td>1</td>
<td>(\text{O}^\text{-S}^-)</td>
<td>No reaction</td>
</tr>
<tr>
<td>G</td>
<td>A</td>
<td>H₂O:dioxane, DMF</td>
<td>Contact</td>
<td>RT, 70, reflux</td>
<td>2</td>
<td>(\text{O}^\text{-SH})</td>
<td>No reaction</td>
</tr>
<tr>
<td>H</td>
<td>U</td>
<td>H₂O:dioxane, DMF</td>
<td>Contact</td>
<td>RT, 70, reflux</td>
<td>2</td>
<td>(\text{O}^\text{-SH})</td>
<td>No reaction</td>
</tr>
<tr>
<td>I</td>
<td>A</td>
<td>H₂O</td>
<td>Contact</td>
<td>60</td>
<td>4</td>
<td>KSCN</td>
<td>No reaction</td>
</tr>
<tr>
<td>J</td>
<td>U</td>
<td>H₂O</td>
<td>7</td>
<td>RT</td>
<td>1</td>
<td>(\text{HS}^\text{-SO}_2^-)</td>
<td>No reaction</td>
</tr>
<tr>
<td>K</td>
<td>A</td>
<td>H₂O</td>
<td>Contact</td>
<td>70</td>
<td>1</td>
<td>Na₂S₂O₅</td>
<td>No reaction</td>
</tr>
<tr>
<td>L</td>
<td>U</td>
<td>H₂O</td>
<td>Contact</td>
<td>70</td>
<td>1</td>
<td>Na₂S₂O₅</td>
<td>Hydrolysis</td>
</tr>
</tbody>
</table>
5.3.2.10 Reaction of cysteamine with *ara*-102 in water

Reaction of *ara*-102 (100 mM) with cysteamine 168 (10 equiv.) at pH 8, 60°C.

*Experimental Figure 48* $^1$H-$^{13}$C HMBC ($^1$H: (600 MHz, 2.15-9.0 ppm), $^{13}$C: (151 MHz, -10-220 ppm), D$_2$O) spectrum of the reaction of 2′,8-cycloadenosine *ara*-102 (100 mM) with cysteamine 168 (1 M). Characteristic cross-coupling between diastereotopic proton resonances 3.76 (m, 1H, $\text{SCH}_2\text{CH}_2\text{NH}_2$), 3.50 (m, 1H, $\text{SCH}_2\text{CH}_2\text{NH}_2$) and C8 carbon resonance (152.5 ppm).
5.3.2.11 Attempted desulfurisation reaction of 157 with Raney-Nickel

Experimental Figure 49 $^1$H NMR (600 MHz, $D_2$O, zg30, 0.5-8.2 ppm) spectrum to show: a) authentic sample of 157; b) crude reaction of 157 (0.08 mmol) with an excess of Raney-Nickel after 24 h; c) separate reaction after 4 h (carried out with Raney-Nickel at LMB-MRC)
5.3.2.12 Irradiation of 157

Experimental Figure 50 $^1$H NMR (600 MHz, (a)D$_2$O b) H$_2$O/D$_2$O (9:1)), zg30, 0.5-9.0 ppm) spectra to show: a) authentic sample of 157 ; b) the irradiation (254 nm) of 157 (2.5 mM) after 1 d; c) after 4 d. (C1')-H proton is labelled for 157 (●).
5.3.2.13 Irradiation of 166

![Chemical structures](image)

Experimental Figure 51 $^1$H NMR (400 MHz, H$_2$O/D$_2$O (9:1), noesygppr1d, 1.0-9.0 ppm) spectra to show: a) 166 (15 mM); b) the irradiation (254 nm) of a) with NaSH (120 mM) after 1 d. (C6)-H, (C1')-H and (C5)-H protons are labelled for 166 (●).
5.3.3 UV stability

5.3.3.1 General Method for Investigating UV Stability

Nucleosides/Nucleotides with or without buffer (sodium phosphate monobasic) were dissolved in H$_2$O (10% D$_2$O, degassed) to give the desired concentration (2, 5, 10, 20 mM) and the pH adjusted to 7. The desired amount of the solution (3.5, 4 mL) was transferred to a quartz cuvette and re-degassed by N$_2$ ebullition (30 mins). The cuvette was sealed and transferred to a Rayonet Photoreactor and irradiated at 254 nm with Rayonet mercury lamps (principle emission 254 nm). After the desired time an aliquot was removed, the remaining solution was re-degassed N$_2$ ebullition (30 mins) and subjected to further irradiation. Removed aliquots were analysed by NMR spectroscopy after an internal standard (potassium hydrogen pthalate) had been added.
5.3.3.2 Irradiation of A 2mM vs 117 2 mM

Experimental Figure 52 $^1$H NMR (600 MHz, H$_2$O/D$_2$O (9:1), noesygppr1d, 3.0-9.0 ppm) spectra to show the irradiation (254 nm) of adenosine A (2 mM) and arabino-adenosine 117 (2 mM) with KHP internal standard: a) before irradiation; b) after 1 day of irradiation; c) after 2 d; d) after 3 d. (C1′)-H protons are labelled: A (●), 117 (◆).
Experimental Figure 53 Concentration over time for the irradiation (254 nm) of adenosine A (2 mM) and arabino-adenosine 117 (2 mM).
5.3.3.3 Irradiation of A 2 mM vs I 2 mM

Experimental Figure 54 \(^1\)H NMR (600 MHz, \(\text{H}_2\text{O}/\text{D}_2\text{O} \ (9:1), \text{noseygppr1d, 3.0-9.0 ppm})\) spectra to show the irradiation (254 nm) of adenosine A (2.1 mM) and inosine I (2.1 mM) with KHP internal standard at pH 6.5: a) before irradiation; b) after 1 day of irradiation, A (88%), I (27%); c) after 2 d, A (81%), I (4%); d) after 3 d, A (41%), I (0%). (C1')-H protons are labelled: A (●), I (▲).
**Experimental Figure 55** Concentration over time for the irradiation (254 nm) of adenosine A (2 mM) and inosine I (2 mM).
5.3.3.4 Irradiation of A 2 mM vs D 2 mM

Experimental Figure 56 $^1$H NMR (600 MHz, H$_2$O/D$_2$O (9:1), noesygppr1d, 3.0-9.0 ppm) spectra to show the irradiation (254 nm) of adenosine A (2.2 mM) and diaminopurine D (2.2 mM) with KHP internal standard at pH 6.5: a) before irradiation; b) after 1 day of irradiation, A (100%), D (73%); c) after 2 d, A (74%), D (23%); d) after 3 d, A (44%), D (5%). (C1')-H protons are labelled: A (●), D (●).
Experimental Figure 57 Concentration over time for the irradiation (254 nm) of adenosine A (2.2 mM) and diaminopurine D (2.2 mM).
5.3.3.5 Irradiation of A 10 mM vs D 2 mM

Experimental Figure 58 $^1$H NMR (600 MHz, H$_2$O/D$_2$O (9:1), noesygpprd1, 3.0-9.0 ppm) spectra to show the irradiation (254 nm) of adenosine A (8.7 mM) and diaminopurine D (1.8 mM) with KHP internal standard at pH 6.5: a) before irradiation; b) after 1 day of irradiation, A (87%), D (62%); c) after 2 d, A (81%), D (33%); d) after 3 d, A (65%), D (10%). (C1’)-H protons are labelled: A (●), D (●).

---

238
Experimental Figure 59 Concentration over time for the irradiation (254 nm) of adenosine A (8.7 mM) and diaminopurine D (1.8 mM).
5.3.3.6 Irradiation of A 10 mM vs I 2 mM

Experimental Figure 60 \(^1\)H NMR (600 MHz, \(\text{H}_2\text{O} / \text{D}_2\text{O} \, 9:1\), noesygpr1d, 3.0-9.0 ppm) spectra to show the irradiation (254 nm) of adenosine A (8.3 mM) and inosine I (1.7 mM) with KHP internal standard at pH 6.5: a) before irradiation; b) after 1 day of irradiation, A (97%), I (70%); c) after 2 d, A (78%), I (45%); d) after 3 d, A (66%), I (25%). (C1\textsuperscript{-})-H protons are labelled: A (●), I (▲).
Experimental Figure 61 Concentration over time for the irradiation (254 nm) of adenosine A (8.3 mM) and inosine I (1.7 mM).
5.3.3.7 Irradiation of A 10 mM G 2 mM

Experimental Figure 62 $^1$H NMR (600 MHz, H$_2$O/D$_2$O (9:1), noesypppr1d, 3.0-9.0 ppm) spectra to show the irradiation (254 nm) of adenosine A (9.1 mM) and guanosine G (1.0 mM) with KHP internal standard at pH 6.5: a) before irradiation; b) after 1 day of irradiation, A (99%), G (86%); c) after 2 d, A (83%), G (9%); d) after 3 d, A (68%), G (0%). (C1')-H protons are labelled: A (●), G (▼).
Experimental Figure 63 Concentration over time for the irradiation (254 nm) of adenosine $A$ (9.1 mM) and guanosine $G$ (1.0 mM).
5.3.3.8 Irradiation of A, I, G and X 2 mM

Experimental Figure 64 $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), noesygppr1d, 3.0-9.0 ppm) spectra to show the irradiation (254 nm) of adenosine A (2.0 mM), guanosine G (1.8 mM), inosine I (2.0 mM) and xanthosine X (1.1 mM) with KHP internal at pH 6.5:

a) before irradiation; b) after 1 day of irradiation, A (95%), G (87%), I (80%), X (50%); c) after 2 d, A (94%), G (25%), I (65%), X (0%); d) after 3 d, A (58%), G (0%), I (26%), X (0%). (C1′)-H protons are labelled: A (●), G (▼), I (▲), X (■).
Experimental Figure 65 Concentration over time for the irradiation (254 nm) of adenosine A (2.0 mM), guanosine G (1.8 mM), inosine I (2.0 mM) and xanthosine X (1.1 mM).
5.3.3.9 Irradiation of A, I, D and X 2 mM

Experimental Figure 66 $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), noesypppr1d, 3.0-9.0 ppm) spectra to show the irradiation (254 nm) of adenosine A (1.8 mM), inosine I (2.0 mM), diaminopurine D (1.9 mM) and xanthosine X (1.6 mM) with KHP internal standard at pH 6.5: a) before irradiation; b) after 1 day of irradiation, A (93%), D (77%), I (67%), X (40%); c) after 2 d, A (92%), D (39%), I (35%), X (0%); d) after 3 d, A (65%), D (0%), I (23%), X (0%). (C1')-H protons are labelled: A (●), D (●), I (▲), X (■).
Experimental Figure 67 Concentration over time for the irradiation (254 nm) of adenosine A (1.8 mM), inosine I (2.0 mM), diaminopurine D (1.9 mM) and xanthosine X (1.6 mM).
5.4 Synthesised Compounds

5.4.1 TNA

5.4.1.1 2-thioxazole

![2-thioxazole structure]

Procedure adapted from Stairs et al.\textsuperscript{192}

Glycolaldehyde (1 g, 16.7 mmol) and potassium thiocyanate (KSCN) (3.24 g, 33.3 mmol) were dissolved in water (3 mL) and the mixture cooled to -5 °C. HCl (37%, 2.10 mmol) was added dropwise over 30 min and allowed to come to RT for 2 h before heating at 80 °C for 24 h. The mixture was allowed to cool to room temperature before extraction with EtOAc (3 x 50 mL). The combined organics were washed with brine (3 x 50 mL), dried over MgSO\textsubscript{4}, filtered and the solvent removed in vacuo to give 2-thioxazole (0.83 g, 49%) which was used without further purification.

<table>
<thead>
<tr>
<th>( ^1H ) NMR ( (700 \text{ MHz, DMSO-D}_6) )</th>
<th>( \delta = 13.07 \text{ (s (br), 1H, S-H), 7.75 (d, } J = 1.59 \text{ Hz, (C4)-H)}, ) 7.37 (d, ( J = 1.59 \text{ Hz, (C5)-H)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( ^{13}C ) NMR ( (176 \text{ MHz, DMSO-D}_6) )</td>
<td>( \delta = 178.4 \text{ (C2), 137.8 (C4), 117.4 (C5)} )</td>
</tr>
<tr>
<td>HRMS</td>
<td>ESI-HRMS (pos. m/z): [M+H]+ calculated for C\textsubscript{3}H\textsubscript{4}NOS+: 102.0008; found 102.0013.</td>
</tr>
<tr>
<td>IR (solid, cm\textsuperscript{-1})</td>
<td>3113 (broad, weak, N-H), 1586 (C=C), 1476 (strong, C=S)</td>
</tr>
</tbody>
</table>
Experimental Figure 68 $^1$H NMR (700 MHz, DMSO-D$_6$, zg30, 2.25-3.5 ppm) spectrum to show 2-thiooxazole 95.

Experimental Figure 69 $^{13}$C NMR (176 MHz, DMSO-D$_6$, 80-200 ppm) spectrum to show 2-thiooxazole 95.
2-Aminooxazole (756 mg, 9 mmol) and glycolaldehyde (540 mg, 9 mmol) were dissolved in water (10 mL) and the solution adjusted to pH 7. The solution was stirred at room temperature for 16 h and then lyophilised to afford tetrose aminoxazoline \textbf{TAO/EO} (1.02 g, 78\%) as an orange crystalline solid.

\begin{align*}
\text{\textsuperscript{1}H NMR (600 MHz, D}_2\text{O)} & \delta = 5.85 (d, J = 5.1 \text{ Hz}, 1 \text{ H}, (C1')-H-threo), 5.72 (d, J = 5.2 \text{ Hz}, 1 \text{ H}, (C2')-H-erythro), 4.85 (t, J = 5.2 \text{ Hz}, 1 \text{ H}, (C2')-H-erythro), 4.80 (d, J = 5.1 \text{ Hz}, 1 \text{ H}, (C2')-H-threo), 4.33 (d, J = 2.6 \text{ Hz}, 1 \text{ H}, (C3')-H-threo), 4.28 (ddd, J = 5.1, 6.7, 10.0 \text{ Hz}, 1 \text{ H}, (C3')-H-erythro), 3.91 (dd, J = 6.7, 9.5 \text{ Hz}, 1 \text{ H}, (C4')-H-erythro), 3.80 (dd, J = 0.8, 11.0 \text{ Hz}, 1 \text{ H}, (C4')-H-threo), 3.57 (dd, J = 2.6, 11.0 \text{ Hz}, 1 \text{ H}, (C4'')-H-threo), 3.21 (t, J = 9.5 \text{ Hz}, 1 \text{ H}, (C4'')-H-erythro).
\end{align*}

\begin{align*}
\text{\textsuperscript{13}C NMR (151 MHz, D}_2\text{O)} & \delta = 167.0 \text{ (C2-erythro), } 165.8 \text{ (C2-threo), } 98.8 \text{ (C1'-erythro), } 98.3 \text{ (C1'-threo), } 87.4 \text{ (C2'-threo), } 82.0 \text{ (C2'-erythro), } 74.2 \text{ (C3'-threo), } 71.1 \text{ (C3'-erythro), } 70.5 \text{ (C4'-threo), } 66.1 \text{ (C4'-erythro).}
\end{align*}

\textbf{HRMS} \text{ ESI-HRMS (pos. m/z): [M+H]+ calculated for C}_5\text{H}_9\text{N}_2\text{O}_3^-: 145.0613; found: 145.0615.}

\textbf{IR (solid, cm}^{-1}) \text{ 3328 (broad, weak, O–H), 1661 (strong, sharp C=\text{N}).}

\textbf{Melting Point (°C)} \text{ 82–90\textsuperscript{3}}

\textsuperscript{3} \text{NB: Melting point recorded for diastereomeric mixture of products.}
**Experimental Figure 70** $^1$H NMR (600 MHz, D$_2$O, zg30, 3.0-6.0 ppm) spectrum to show a mixture of threo-aminooxazoline TAO and erythro-aminooxazoline EAO.
Experimental Figure 71 $^{13}$C NMR (151 MHz, D$_2$O, 65-170 ppm) spectrum to show a mixture of threeo-aminooxazoline $\text{TAO}$ and erythro-aminooxazoline $\text{EAO}$. 
5.4.1.3 Synthesis of propiolamide 140

\[
\begin{array}{c}
\text{O} \\
\text{CH}_3 - \text{C} = \text{N} \\
\end{array}
\]

To condensed NH\(_3\) (~350 mL) at -78 °C was added methyl propiolate (25 ml, 23.63 g, 281 mmol) over 10 minutes. This was reacted for 30 minutes before warming the solution to room temperature to allow the excess NH\(_3\) to evaporate. The residue was dissolved in Et\(_2\)O (200 mL) dried over MgSO\(_4\), filtered and evaporated. The crude product was recrystallised from DCM to yield yellow shard-like crystals of propiolamide (19.32 g, 99.6%).

5.4.1.4 Synthesis of cyanoacetylene 12

\[
\begin{array}{c}
\text{C} = \text{C} = \text{N} \\
\end{array}
\]

Methyl propiolamide (15 g, 211.03 mmol), dried sea sand (75 g) and P\(_2\)O\(_5\) (45 g, 317.03 mmol) were mixed together with a pestle and mortar. The mixture was dry distilled at 135 °C under low pressure to collect cyanoacetylene 12 as a white solid in a flask maintained at 78 °C. The white solid (8.05 g, 76.2%) was dissolved in water as a 1 M solution and used directly or stored at -80 °C.
5.4.1.5 Tetrofuranosyl-2,2'-anhydrocytidine hydrochloride

![Chemical Structure](image)

Three-aminooxazoline **TAO** and erythro-aminooxazoline **EAO** (0.359 mg, 2.5 mmol) and sodium dihydrogen phosphate (0.357 mg, 3 mmol) were dissolved in water (6 mL) and the solution was adjusted to pH 6.5. Cyanoacetylene (1.276 g, 25 mmol) was added and the reaction was stirred for 18 h at room temperature. The resulting yellow solution was lyophilised to give a yellow powder (906 mg) that contained tetrofuranosyl-2,2'-anhydrocydines **threo-89** and **erythro-89**. The lyophilisate was dissolved in D$_2$O (0.5 mL) and analysed by NMR spectroscopy. Cyanovinyl phosphate, but no residual aminooxazoline (**TAO** or **EAO**) was observed. This material was carried forward to the next step without purification.
**H NMR (600 MHz, D$_2$O)** \( \delta = 8.03 \) (d, \( J = 7.3 \) Hz, 1 H, (C6)-H-threo), 8.01 (d, \( J = 7.3 \) Hz, 1 H, (C6)-H-erythro), 6.55-6.52 (m, 3 H, (C5)-H-threo, (C5)-H-erythro, (C1')-H-threo), 6.43 (d, \( J = 5.2 \) Hz, 1 H, (C1')-H-erythro), 5.47 (t, \( J = 5.2 \) Hz, 1 H, (C2')-H-erythro), 5.43 (d, \( J = 5.2 \) Hz, 1 H, (C2')-H-threo), 4.64 (d, \( J = 2.7 \) Hz, 1 H, (C3')-H-threo), 4.59 (ddd, \( J = 5.2, 6.6, 9.5 \) Hz, 1 H, (C3')-H-erythro), 4.18 (dd, \( J = 6.6, 9.5 \) Hz, 1 H, (C4')-H-erythro), 4.10 (dd, \( J = 1.0, 10.2 \) Hz, 1 H, (C4')-H-threo), 3.86 (dd, \( J = 2.7, 10.2 \) Hz, 1 H, (C4'')-H-threo), 3.60 (t, \( J = 9.5 \) Hz, 1 H, (C4'')-H-erythro).

**C NMR (151 MHz, D$_2$O)** \( \delta = 168.2 \) (C2-threo), 167.6 (C2-erythro), 160.1 (C4-threo), 160.1 (C4-erythro) 141.0 (C6-threo) 140.8 (C6-erythro), 103.9 (C5-threo), 103.8 (C5-erythro), 91.8 (C1'-erythro), 91.7 (C1'-threo), 88.7 (C2'-threo), 83.6 (C2'-erythro), 73.8 (C4'-threo), 72.1 (C3'-threo), 70.1 (C3'-erythro), 69.1 (C4'-erythro).

**HRMS** ESI-HRMS (pos. m/z): [M+H]$^+$ calculated for C$_8$H$_{11}$N$_3$O$_3$+: 196.0722; found: 196.0720.

**IR (solid, cm$^{-1}$)** 3097 (broad, O-H), 1637 (sharp, C=N), 1499 (medium, C=C).

**Melting Point (°C)** 150-165 °C (decomp.)
Experimental Figure 72 $^1$H NMR (600 MHz, D$_2$O, zg30, 3.0-8.5 ppm) spectrum to show a mixture of threo-anhydronucleoside threo-89 and erythro-anhydronucleoside erythro-89 and cyanovinyl phosphate CVP (*).
Experimental Figure 73 $^{13}$C NMR (176 MHz, D$_2$O, 65.5-170 ppm) spectrum to show a mixture of threo-anhydronucleoside threo-89 and erythro-anhydronucleoside erythro-89.
5.4.1.6 Two step synthesis of β-*tetro*-thiocytidines

![Chemical structure]

*Threeo*-aminooxazoline TAO and *erythro*-aminooxazoline EAO (303 mg, 2.1 mmol) and sodium dihydrogen phosphate (252 mg, 2.1 mmol) were dissolved in water (5 mL) and the solution was adjusted to pH 6.5. Cyanoacetylene (9.35 mL, 8.41 mmol) was added and the reaction was stirred for 18 h at room temperature, the resulting yellow solution was lyophilised to give tetofuranosyl-2,2′-anhydrocytidine *threo*-89 and *erythro*-89 as a yellow powder. The crude mixture of *threo*-89 and *erythro*-89 and ammonium bicarbonate (498 mg, 6.30 mmol) was dissolved in anhydrous DMF (20 mL) and NaSH (612 mg, 6.6 mmol) was added. The resulting blue solution was stirred for 18 h at room temperature. Water (40 mL) was added to the resulting solution and HCl under air ebullition until the solution reached pH 5, to degas and sparge H2S from the solution. The sparged solution was adjusted to pH 7 and the solvent removed in vacuo. The crude product was purified by reverse phase RP-FCC eluting with water/MeOH to give β-*threo*-111 and β-*erythro*-111 (178 mg, 37%).
5.4.1.7 Characterisation data for β-threo-111

\[ \text{H NMR (700 MHz, D}_2\text{O)} \delta = 7.78 (d, J = 7.6 \text{ Hz, 1H, (C6)-H}, 6.88 (d, J = 3.6 \text{ Hz, 1H, (C1')-H}, 6.29 (d, J = 7.6 \text{ Hz, 1H, (C5)-H}), 4.70 (d, J = 3.6 \text{ Hz, 1H, (C2')-H}), 4.39-4.36 (m, 2, (C3')-H, (C4')-H), 4.03 (d, J = 9.7 \text{ Hz, 1H, (C4'')-H})} \]

\[ \text{C NMR (176 MHz, D}_2\text{O)} \delta = 179.0 \text{ (C2), 161.8 (C4), 143.7 (C6), 99.3 (C5), 93.0 (C1'), 76.2 (C3'), 74.7 (C4'), 74.2 (C2')} \]

HRMS ESI-HRMS (pos. m/z): [M+H]^+ calculated for C_{8}H_{12}N_{3}O_{3}S^+: 230.0555; found 230.0599.

IR (solid, cm\(^{-1}\)) 3171 (broad, weak, O-H), 1659 (strong, sharp, C-N), 1081 (strong, sharp, C-N)

Melting Point (°C) 168-172 (decomp.)
Experimental Figure 74 $^1$H NMR (700 MHz, D$_2$O, zg30, 3.0-9.0 ppm) spectrum to show β-threo-111.
Experimental Figure 75 $^{13}$C NMR (176 MHz, D$_2$O, 72-182 ppm) spectrum to show β-threo-111.
5.4.1.8 Characterisation data for β-erythro-111

\[
\begin{align*}
\text{HO} & \quad \text{O} & \quad \text{NH}_2 \\
\text{H} & \quad \text{S} & \quad \text{N}
\end{align*}
\]

$^1$H NMR (700 MHz, D$_2$O) \( \delta = 7.88 \) (d, \( J = 7.6 \) Hz, 1H, (C6)-H), 6.86 (d, \( J = 5.0 \) Hz, 1H, (C1′)-H), 6.30 (d, \( J = 7.6 \) Hz, 1H, (C5)-H), 4.69 (t, \( J = 5.0 \) Hz, 1H, (C2′)-H), 4.48 (app. q, \( J = 5.5 \) Hz, 1H, (C3′)-H), 4.12 (dd, \( J = 5.5, 9.3 \) Hz, 1H, (C4′)-H), 4.06 (dd, \( J = 5.5, 9.3 \) Hz, 1H, (C4″)-H)

$^{13}$C NMR (151 MHz, D$_2$O) \( \delta = 179.3 \) (C2), 161.8 (C4) 144.0 (C6), 99.4 (C5), 91.5 (C1′), 72.1 (C4′), 70.6 (C2′), 70.1 (C3′)

HRMS ESI-HRMS (pos. m/z): [M+H]$^+$ calculated for C$_8$H$_{12}$N$_3$O$_3$S$: 230.0555; found 230.0599.

IR (solid, cm$^{-1}$) 3474-3197 (broad, weak, O-H), 1636 (strong, sharp, C-N), 1076 (strong, sharp, C-N)

Melting Point (°C) 210 (decomp.)
Experimental Figure 76 $^1$H NMR (700 MHz, D$_2$O, zg30, 4.0-8.0 ppm) spectrum to show $\beta$-erythro-111.
Experimental Figure 77 $^{13}$C NMR (176 MHz, D$_2$O, 64-184 ppm) spectrum to show β-erythro-111.
5.4.1.9 Synthesis of tetrose β-cytidines β-\textit{threo}-85 and β-\textit{erythro}-85

DOWEX® (in the OH\textsuperscript{-} form) was prepared by Dowex Preparation Method 1 (Exp.5.2.7)

The desired amount of DOWEX® (~10 g) was stirred in water (10 mL) for 10 mins. A mixture of \textit{threo}-89 and \textit{erythro}-89 (392 mg, 2 mmol) were added and the solution was stirred for 4 h at RT. The resulting suspension was filtered under vacuum and washed with water (3 x 10 mL). The filtrate was lyophilised to give a racemic mixture of tetrose β-cytidines (β-\textit{threo}-85 and β-\textit{erythro}-85) (237 mg). The solid was redissolved in D\textsubscript{2}O for NMR analysis. The mixture of diastereomers was separated by RP-FCC eluting with a gradient of water/MeOH to give 27 mg of β-\textit{threo}-85 and 25 mg of β-\textit{erythro}-85 as analytical samples.
$^1$H NMR (600 MHz, D$_2$O)  \( \delta = 7.69 \) (d, \( J = 7.5 \) Hz, 1 H, (C6)-H-erythro), 7.59 (d, \( J = 7.5 \) Hz, 1 H, (C6)-H-threo), 6.07 (d, \( J = 3.5 \) Hz, 1 H, (C1')-H-erythro), 6.03 (d, \( J = 5.0 \) Hz, 1 H, (C1')-H-threo), 5.94 (d, \( J = 7.5 \) Hz, 1 H, (C5)-H-erythro), 5.94 (d, \( J = 7.5 \) Hz, 1 H, (C5)-H-threo), 6.07 (d, \( J = 3.5 \) Hz, 1 H, (C1')-H-erythro), 4.42 (t, \( J = 5.0 \) Hz, 1 H, (C2')-H-erythro), 4.41 (dd, \( J = 5.0, 9.9 \) Hz, 1H, (C3')-H-erythro), 4.26 (dd, \( J = 3.5 10.1 \) Hz, (C4')-H-erythro), 4.02-3.96 (m, 2 H, (C4')-H-erythro, (C4'')-H-erythro), 3.92 (d, \( J = 10.1 \) Hz, 1 H (C4'')-H-threo).

$^{13}$C NMR (151 MHz, D$_2$O)  \( \delta = 164.1 \) (C2-threo) 158.1 (C4-threo) 143.2 (C6-threo), 95.6 (C5-threo), 88.7 (C1'-threo), 76.2 (C2'-threo), 75.1 (C3'-threo), 74.2 (C4'-threo). 167.0 (C2-erythro) 158.4 (C4-erythro), 143.6 (C6-erythro), 95.7 (C5-erythro), 86.9 (C1'-erythro), 72.1 (C2'-erythro), 71.3 (C30-erythro), 70.3 (C4'-erythro).

HRMS  ESI-HRMS (pos. m/z): [M+H]$^+$ calculated for C$_8$H$_{12}$N$_3$O$_4$$^+$ 214.0828; found 214.0830.

IR (solid, cm$^{-1}$)  3312-3194 (broad, weak, O–H), 1749 (weak), 1641 (strong, sharp), 1601 (medium), C=C, C=N, C=O, 1489 (medium), 1091 (strong, broad, C–N).

Melting Point (°C)  184 °C (decomp.)
5.4.1.10 Synthesis of tetrose 2',3'-di-O-acetyl-β-cytidines β-threo-148 and β-erythro-148

1-Acetylimidazole (551 mg, 5 mmol) and a mixture of β-threo-85 and β-erythro-85 (92.5 mg, 0.5 mmol) were dissolved in water and the pH adjusted to 8 until all the crystals of 1-acetylimidazole had dissolved (~30 mins). The solution was stirred for 2 h at RT and then lyophilised to give a white powder of a racemic mixture of 2',3'-di-O-acetyl-tetrocytidines (β-threo-148 and β-erythro-148). The racemic mixture was separated by RP-FCC eluting with a gradient of water/MeOH. Fractions containing the UV chromophore were combined and concentrated in vacuo, then redissolved in water and lyophilised to yield 31 mg of β-erythro-148, and 33 mg of β-threo-148 as white powders.
5.4.1.11 Characterisation data for β-threo-148

![Chemical structure](image)

$^1$H NMR (600 MHz, D$_2$O)  
δ = 7.55 (d, J = 7.6 Hz, 1 H, (C6)-H), 6.23 (d, J = 4.0 Hz, 1 H, (C1')-H), 5.96 (d, J = 7.6 Hz, 1 H, (C5)-H), 5.48 (d, J = 4.0 Hz, 1 H, (C2')-H), 5.35 (m, 1 H, (C3')-H), 4.51 (dd, J = 4.2, 10.7 Hz, (C4')-H), 4.12 (app. d, J = 10.7 Hz, 1 H, (C4'')-H), 2.14 (s, 3 H, OC(CH$_3$)), 1.96 (s, 3 H, OC(CH$_3$)).

$^{13}$C NMR (151 MHz, D$_2$O)  
δ = 164.2 (C2), 158.4 (C4), 143.9 (C6), 96.0 (C5), 87.0 (C1'), 76.9 (C2'), 73.0 (C3'), 69.8 (C4')

**HRMS**  
ESI-HRMS (pos. m/z): [M+H]$^+$ calculated for C$_{12}$H$_{16}$N$_3$O$_6$$: 298.1039; found 298.1038.

**IR (solid, cm$^{-1}$)**  
3344-3192 (weak, broad, N−H), 1744 (medium, C=O), 1634 (strong, broad, C=O), 1481 (medium, C−C), 1209 (strong, C−O, C−N Aromatic amine), 1043 (medium, sharp, C−N).

**Melting Point (°C)**  
175-187 °C.
5.4.1.12 Characterisation data for β-erythro-148

**1H NMR (600 MHz, D₂O)**  
δ = 7.80 (d, J = 7.6 Hz, 1 H, (C6)-H), 6.20 (d, J = 6.0 Hz, 1 H, (C1')-H), 6.00 (d, J = 7.6 Hz, 1 H, (C5)-H), 5.65 (t, J = 6.0 Hz, 1 H, (C2')-H), 5.48 (m, 1 H, (C3')-H), 4.39 (dd, J = 1.9, 11.0 Hz, 1 H, (C4')-H), 4.19 (dd, J = 4.2, 11.0, 1 H, (C4'')-H), 2.01 (s, 3 H, OC(CH₃)), 1.89 (s, 3 H, OC(CH₃)).

**13C NMR (151 MHz, D₂O)**  
δ =167.0 (C2), 153.9 (C4), 142.8 (C6), 95.3 (C5), 85.7 (C1'), 73.7 (C2'), 71.8 (C3'), 71.3 (C4').

**HRMS**  
ESI-HRMS (pos. m/z): [M+H]⁺ calculated for C₁₂H₁₆N₃O₆⁺: 298.1039; found 298.1038.

**IR (solid, cm⁻¹)**  
3344-3192 (weak, broad, N-H), 1744 (medium, C=O), 1634 (strong, broad, C=O), 1481 (medium, C-C), 1209 (strong, C-O, C-N Aromatic amine), 1043 (medium, sharp, C-N).

**Melting Point (°C)**  
175-187 °C.
5.4.1.13 Synthesis of rac-Tetrofuranosyl-2,2'-anhydouridine

rac-Tetrose-aminooxazoline TAO/ EAO (144.13, 1 mmol) and was dissolved in water (2 mL). Methyl propiolate (0.2 mL, 2.2 mmol) was added and the reaction was stirred for 18 h at room temperature. The resulting solution was lyophilised to give an off-white powder (189.5 mg, 97%) that contained rac-tetrofuranosyl-2,2'-anhydouridine (threo-126 and erythro-126) which was used without further purification.

1H NMR (700 MHz, D2O) \( \delta = 7.90 \) (d, \( J = 7.4 \) Hz, 1H, (C6)-H-threo), 7.88 (d, \( J = 7.4 \) Hz, 1H, (C6)-H-erythro), 6.51 (d, \( J = 5.3 \) Hz, 1H, (C1')-H-threo), 6.39 (d, \( J = 5.3 \) Hz, 1H, (C1')-H-erythro), 6.20 (app. t, \( J = 7.4 \) Hz, 2H, (C5)-H-threo, (C5)-H-erythro), 5.43 (t, \( J = 5.4 \) Hz, 1H, (C2')-H-erythro), 5.39 (d, \( J = 5.4 \) Hz, 1H, (C2')-H-threo), 4.66 (d, \( J = 2.8 \) Hz, 1H, (C3')-H-erythro), 4.60 (m, (C3')-H-erythro), 4.21 (dd, \( J = 6.6, 9.5 \) Hz, (C4')-H-erythro), 4.13 (dd, \( J = 1.2, 11.2 \) Hz, 1H, (C4')-H-erythro), 3.90 (ABX, \( J = 2.8, 11.2 \) Hz, 1H, (C4'')-H-threo), 3.54 (t, \( J = 9.5 \) Hz, 1H, (C4'')-H-erythro)

13C NMR (176 MHz, D2O) \( \delta = 176.3 \) (C4), 176.3 (C4), 162.3 (C2), 161.8 (C2), 139.2 (C6), 139.0 (C6), 109.6 (C5), 109.6 (C5), 91.1 (C1'), 91.1 (C1'), 88.2 (C2'), 82.6 (C2'), 73.3 (C3/4), 73.3 (C3/4), 71.1 (C3/4), 70.2 (C3/4).

HRMS ESI-HRMS (neg. m/z): [M+H]+ calculated for C8H9N2O4+: 197.0557; found 197.0562.

IR (solid, cm⁻¹) 3235 (broad, O−H), 1649 (sharp, C=N), 1476 (medium, C=C).

Melting Point (°C) 80.5-95
Experimental Figure 78 $^1$H NMR (700 MHz, D$_2$O, zg30, 3.5-8.5 ppm) spectrum to show threo-126 and erythro-126.

Supplementary Figure 1 $^{13}$C NMR (176 MHz, D$_2$O, -20-260 ppm) spectrum to show threo-126 and erythro-126.
5.4.1.14 Synthesis of β-threo-112

rac-tetrofuranosyl-2,2'-anhydrouridine (98.08 mg, 0.5 mmol), ammonium bicarbonate (158.1 mg, 2 mmol) and NaSH.xH₂O (224 mg, 2.4 mmol) were dissolved in anhydrous DMF (5 mL) and the reaction stirred at room temperature for 18 h. Water (5 ml) was added to the resulting solution and HCl (4 M) under air ebullition until the solution reached pH = 5. The solution was readjusted to pH 7 and lyophilised. The crude product was purified by RP-FCC eluting with water/MeOH to give β-threo-thiouridine β-threo-112 (35 mg, 30%) as an off-white powder.

| ¹H NMR (700 MHz, D₂O) | δ = 7.75 (d, J = 8.1 Hz, 1H, (C6)-H), 6.78 (d, J = 3.5 Hz, 1H, (C1')-H), 6.09 (d, J = 8.1 Hz, 1H, (C5)-H), 4.65 (app. d, J = 3.5 Hz, 1H, (C2')-H), 4.37 (m, 2H, (C3')-H, (C4')-H), 4.03 (app. d, J = 9.5 Hz, 1H, (C4'')-H), |
| ¹³C NMR (176 MHz, D₂O) | δ = 176.1 (C2), 164.8 (C4), 143.8 (C6), 105.8 (C5), 92.6 (C1'), 76.0 (C3'), 74.9 (C4'), 74.2 (C2') |
| HRMS | ESI-HRMS (pos. m/z): [M+H]⁺ calculated for C₈H₁₁N₂O₄S⁺: 231.0434; found 231.0435. |
| IR (solid, cm⁻¹) | 3202 (broad, weak, O-H), 1648 (strong, sharp, C-N), 1081 (strong, sharp, C-N) |
| Melting Point (°C) | 135 decomp. |
Experimental Figure 79 $^1$H NMR (700 MHz, D$_2$O, zg30, 3.5-8.5 ppm) spectra to show β-threo-thiouridine β-threo-112.
Experimental Figure 80: $^{13}$C NMR (176 MHz, D$_2$O, 74-184 ppm) spectrum to show β-threo-thiouridine β-threo-112.
5.4.1.15 Synthetic route for authentic standard of α-threo-111

Experimental Figure 81 Synthetic route for the synthesis of α-threo-111 as an authentic standard.
5.4.1.16 Synthesis of Calcium L-threonate 142

![Calcium L-threonate](image)

Procedure adapted from Sau et al.\textsuperscript{312}.

To a solution of L-ascorbic acid 141 (25.1 g, 0.14 mol) in water (200 mL) was slowly added of CaCO\textsubscript{3} (25.2 g, 0.25 mol) over 30 min, whilst the solution temperature was maintained between 0–5 °C. H\textsubscript{2}O\textsubscript{2} (30%, 50 mL) was added dropwise to the resultant slurry over 1 h whilst stirring at 0–5 °C. The reaction mixture was allowed to warm to room temperature and stirred overnight. The mixture was filtered, and the solids washed with water (2 × 20 mL). To the filtrate, activated carbon (5 g) was added and stirred at 50 °C, until peroxide was no longer detected by peroxide test strips (Quantofix\textsuperscript{®} Peroxide 100). The hot suspension was filtered, and the solids washed with water (2 × 10 mL). The volume of the filtrate was then reduced in vacuo to 20 mL. The resulting solution was precipitated in ice cold MeOH and filtered to give calcium L-threonate 142 as a white powder (26.7 g, 60%).

\textsuperscript{1}H NMR (600 MHz, D\textsubscript{2}O) \[ \delta = 3.96 (d, J = 2.4 Hz, 1H, (C2')-H), 3.89 (m, 1H, (C3')-H), 3.52 (ABX, J = 5.2, 11.5 Hz, 1H, (C4')-H) 3.62 (ABX, J = 7.7, 11.5 Hz, 1H, (C4'')-H). \]

\textsuperscript{13}C NMR (151 MHz, D\textsubscript{2}O) \[ \delta = 179.5 (C1'), 73.4 (C3'), 73.1(C2'), 63.5 (C4'). \]

HRMS ESI-HRMS (pos. m/z): [M+H]\textsuperscript{+} calculated for C\textsubscript{4}H\textsubscript{9}O\textsubscript{5}: 137.0405; found: 137.0444

IR (solid, cm\textsuperscript{-1}) \[ 3218.59 (broad, weak, O-H (acid)), 1581.59 (strong, C=O) \]

Melting Point (°C) \[ > 380 \]
Experimental Figure 82 $^1$H NMR (700 MHz, D$_2$O, zg30, 3.0-5.1 ppm) spectrum to show calcium L-threonate 142.
Experimental Figure 83 $^{13}$C NMR (176 MHz, D$_2$O, 60-192 ppm) spectrum to show calcium L-threonate 142.
5.4.1.17 Synthesis of L-threonolactone 143

Procedure adapted from Sau et al.\textsuperscript{312}

To a suspension of calcium L-threonate 142 (25 g, 0.081 mol) in dry acetonitrile (125 mL) were added anhydrous oxalic acid (7.26 g, 0.081 mol) and para-toluenesulfonic acid monohydrate (0.258 g). The heterogeneous mixture was stirred at reflux for 3 h. The hot mixture was allowed to cool to room temperature and filtered. The solids were washed with 20 mL of acetonitrile, and the combined filtrate was evaporated under reduced pressure to produce a colourless syrup. The residue was dissolved in EtOAc (25 mL) and evaporated to dryness to give L-threonolactone 143 as a white solid (12 g, 94%) which was used without further purification.

\begin{align*}
\text{\textsuperscript{1}H NMR (600 MHz, MeOD)} & \quad \delta = 4.41 \text{ (dd, } J = 6.8, 8.9 \text{ Hz, } 1\text{H, (C4')-H)}, \quad 4.29 \text{ (q, } J = 7.2, \text{1H, (C3')-H)}, \quad 4.19 \text{ (d, } J = 7.2 \text{ Hz, } 1\text{H, (C2')-H)} \quad 3.94 \text{ (dd, } J = 7.2, 8.9 \text{ Hz, } 1\text{H, (C4'')-H}) \\
\text{\textsuperscript{13}C NMR (100 MHz, MeOD)} & \quad \delta = 177.4 \text{ (C1)}, \quad 74.8 \text{ (C3)}, \quad 74.1 \text{ (C2)}, \quad 71.2 \text{ (C4)}. \\
\text{HRMS} & \quad \text{ESI-HRMS (pos. m/z): [M+H]\textsuperscript{+} calculated for C}_{4}H_{7}O_{4}\textsuperscript{+}: 119.0339; \text{found 119.0687.} \\
\text{IR (solid, cm}\textsuperscript{-1}) & \quad 3344 \text{ (strong, broad, O-H), 1732 (strong, C=O)} \\
\text{Melting Point (°C)} & \quad 66-70 (61-76\textsuperscript{383})
\end{align*}
Experimental Figure 84  $^1$H NMR (600 MHz, MeOD, zg30, 3.0-5.1 ppm) spectrum to show L-threonolactone 143.
Experimental Figure 85 $^{13}$C NMR (151 MHz, MeOD, 45-180 ppm) spectrum to show L-threonolactone 143.
5.4.1.18 Synthesis of 2′,3′-di-O-benzoyl-threolactone 145

![Diagram of 2',3'-di-O-benzoyl-threolactone]

Procedure adapted from Schöning et al.\textsuperscript{384}

L-Threonolactone 143 (4.17 g, 35 mmol) and 4-dimethylaminopyridine (DMAP) (30 mg, 0.25 mmol) were dissolved in anhydrous pyridine (30 mL) and anhydrous dichloromethane (DCM) (10 mL) at 0 °C. Benzoyl chloride (10 mL, 86 mmol) was added dropwise under an N\textsubscript{2} atmosphere over 30 min. The homogenous mixture was allowed to warm to room temperature and then stirred for 2 d. The heterogenous mixture was then filtered and the filtrate concentrated in vacuo. The residue was diluted with DCM (25 mL) and washed with 1 M HCl (3 x 25 mL), water (25 mL), NaHCO\textsubscript{3} sat. (25 mL), and brine (25 mL). The organics were then dried over MgSO\textsubscript{4}, filtered and concentrated in vacuo. The crude product was purified by normal phase FCC eluting with hexane/EtOAc (80:20) to give 2′,3′-di-O-benzoyl-threolactone 145 (3.01 g, 26 %).

\[^{1}H\text{ NMR (700 MHz, D}_{2}\text{O)}\]

\[
\delta = 8.10 (d, J = 7.3 Hz, 2H, (C2′)-Ar-H), 8.05 (d, J = 7.3 Hz, 2H, (C2′)-Ar-H), 7.62 (t, J = 7.5 Hz, 2H, Ar-H), 7.48 (td, J = 1.6, 8.1 Hz, 4H, Ar-H), 5.81 (m, 2H, (C2′)-H, (C3′)-H)
\]

\[^{13}C\text{ NMR (176 MHz, D}_{2}\text{O)}\]

\[
\delta = 169.7 (C1′), 165.8 (COCOAr), 165.5 (COCOAr), 134.2 (Ar-para), 134.2 (Ar-para), 130.4 (2C, Ar-ortho), 130.1 (2C, Ar-ortho), 128.8 (2C, Ar-meta), 128.8 (2C, Ar-meta), 128.4 (Ar-ipso), 128.2 (Ar-ipso), 73.4, 72.3, 69.6 (C4′)
\]

HRMS

ESI-HRMS (pos. m/z): [M+H]\textsuperscript{+} calculated for C\textsubscript{18}H\textsubscript{15}O\textsubscript{6}+: 327.0863; found 327.0865.

IR (solid, cm\textsuperscript{-1})

1678 (strong, C=O), 1287 (strong), 703 (strong)

Melting Point (°C)

114-119 (114\textsuperscript{385})
Experimental Figure 86 $^1$H NMR (700 MHz, CDCl$_3$, zg30, 4.0-8.5 ppm) spectrum to show 2',3'-di-O-benzoyl-threonolactone 145.
Experimental Figure 87 $^{13}$C NMR (176 MHz, CDCl$_3$, 68-180 ppm) spectrum to show 2',3'-di-O-benzoyl-threonolactone 145.
5.4.1.19 Silylation of 2-thiocytosine 113

2-Thiocytosine 113 (402 mg, 3.16 mmol) was added to a flame dried flask under N$_2$ atmosphere and dissolved in Bis(trimethylsilyl)acetamide (BSA) (1.58 mL, 6.46 mmol) and anhydrous acetonitrile (22.5 mL). The heterogenous mixture was refluxed for 2 h or until homogenous. The mixture was evaporated to an orange residue, redissolved in anhydrous dichloroethane (DCE) (5 mL) under N$_2$ atmosphere and used immediately without purification.
5.4.1.20 Synthesis of 2′,3′-di-O-benzoyl-2-thiothreocytidine α-threo-147

2′,3′-Di-O-benzoyl-threonolactone 145 (700 mg, 2.15 mmol) was added under a N₂ atmosphere and dissolved in anhydrous THF (10 mL). The mixture was cooled to -78 °C and DIBAL-H (1.2 M in toluene, 2.5 mL, 3 mmol) added dropwise over 30 minutes. The reaction was maintained at -78 °C for 10 h and monitored by TLC. When the starting material was consumed Ac₂O/DCM/DMAP (1.2 ml, 12.7 mmol/1.5 ml/ 57 mg, 0.47 mmol) was added at -78 °C, over 20 mins, under a N₂ atmosphere. The reaction was allowed to warm to room temperature over 16 h. Hexane (30 mL) was added and the mixture was added to rapidly stirring 1 M HCl (50 mL) and stirred vigorously for 1 h. The aqueous layer was separated and washed with EtOAc (3 x 50 mL). The combined organic layers were separated and washed with water (50 mL), NaHCO₃ sat. (50 mL) and brine (50 mL), dried over MgSO₄ and evaporated to give a crude yellow oil. To this, DCM (15 mL) was added and co-evaporated 3 times before anhydrous dichloroethane (DCE) (3 mL) was added under N₂ atmosphere. To this mixture, freshly prepared, silylated 2-thiocytosine 113 (402 mg, 3.16 mmol, dissolved in 5 mL anhydrous DCE) was added and the mixture cooled to 0 °C before SnCl₄ (1.2 mL, 10.3 mmol) was added over 20 minutes. The now orange homogenous mixture was allowed to come to room temperature over 16 h and monitored by TLC. When the reaction was judged complete the reaction was quenched with NaHCO₃ sat. (15 mL) and stirred vigorously for 2 h. The mixture was filtered through Celite® 545, washed with DCM (30 mL) and NaHCO₃ sat. (15 mL) and the organics separated. The aqueous phase was washed with DCM (3 x 10 mL). The combined organics were washed with brine (30 mL) and dried over MgSO₄ before the solvent was removed in vacuo. The crude product was purified with FCC eluting with a mixture of DCM/MeOH to give 2′,3′-di-O-benzoyl-2-thiothreocytidine α-threo-147 (581.2 mg, 62%) as a brown solid.
<table>
<thead>
<tr>
<th>NMR</th>
<th>δ (MHz, solvent)</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1$H</td>
<td>700 MHz, CDCl$_3$</td>
<td>$\delta = 8.14$ (d, $J = 7.1$ Hz, 2H, (C2')-Ar-H-ortho), 7.85 (d, $J = 7.1$ Hz, 2H, (C3')-Ar-H-ortho), 7.81 (d, $J = 7.5$ Hz, 1H, (C6)-H), 7.63 (dd, $J = 7.4$, 7.5 Hz, 1H, (C2')-Ar-H-para), 7.60 (dd, $J = 7.4$, 7.5 Hz, 1H, (C3')-Ar-H-para), 7.50 (dd, $J = 7.6$, 7.9 Hz, 2H, (C2')-Ar-H-meta), 7.43 (dd, $J = 7.6$, 7.9 Hz, 2H, (C3')-Ar-H-meta), 6.91 (s, 1H, (C1')-H), 6.03 (s, 1H, (C2')-H), 5.95 (d, $J = 7.5$ Hz, 1H, (C5)-H), 5.51 (d, $J = 3.3$ Hz, 1H, (C3')-H), 4.62 (ABX, $J = 3.3$, 11.4 Hz, 1H, (C4')-H), 4.59 (ABX, $J = 11.4$ Hz, 1H, (C4'')-H)</td>
</tr>
<tr>
<td>$^{13}$C</td>
<td>176 MHz, CDCl$_3$</td>
<td>$\delta = 180.5$ (C2), 165.0 (C3'-O$\text{C}$OAr), 164.9 (C2'-O$\text{C}$OAr), 160.2 (C4), 141.6 (C6), 134.0 (C3'-Ar-para), 133.8 (C2'-Ar-para), 130.2 (2C, C2'-Ar-ortho), 130.0 (2C, C3'-Ar-ortho), 129.1 (C2'-Ar-ipso), 128.7 (2C, C3'-Ar-meta), 128.7 (2C, C2'-Ar-meta), 128.6 (C3'-Ar-ipso), 97.2 (C5), 94.8 (C1') 80.3 (C2'), 75.9 (C3'), 75.6 (C4')</td>
</tr>
</tbody>
</table>

HRMS: ESI-HRMS (pos. m/z): [M+H]$^+$ calculated for C$_{22}$H$_{20}$N$_3$O$_5$S$: 438.1124; found 438.1153.

IR (solid, cm$^{-1}$): 1720 (strong, C=O), 1644 (strong, C=C), 1095 (medium, broad C-N)

Melting Point (°C): 185-190 (decomp.)
Experimental Figure 88 $^1H$ NMR (600 MHz, CDCl$_3$, zg30, 4.5-8.25 ppm) spectrum to show 2',3'-di-O-benzoyl-2-thio-threocytidine α-threo-147.
Experimental Figure 89 $^{13}$C NMR (176 MHz, CDCl$_3$, 72-183 ppm) spectrum to show 2',3'-di-O-benzoyl-2-thio-threocytidine α-threo-147.
5.4.1.21 Synthesis of authentic standard of \(\alpha\)-threo-thiocytidine \(\alpha\)-threo-111

\[
\begin{align*}
\text{HO} & \quad \text{O} \\
\text{N} & \quad \text{N} \\
\text{S} & \quad \text{NH}_2 \\
\end{align*}
\]

2',3'-Di-O-benzoyl-2-thio-threocytidine 146 (200 mg, 0.46 mmol) was dissolved in 7 M \(\text{NH}_3/\text{MeOH}\) (3.15 mL) and MeOH (12.6 mL) and the heterogenous mixture stirred at room temperature for 2.5 days. The solvent was evaporated and the residue redissolved in \(\text{D}_2\text{O}\) (7 mL) and washed DCM (2 x 7 mL). The aqueous layer was separated and lyophilised giving \(\alpha\)-threo-thiocytidine \(\alpha\)-threo-111 (98.2 mg, 94%) as a white powder. A sample was removed for NMR analysis.

\(^1\text{H NMR}\) (400 MHz, \(\text{D}_2\text{O}\)) \(\delta = 7.89\) (d, \(J = 7.6\) Hz, 1H, (C5)-H), 6.32 (s, 1H, (C1\(^1\)))-H), 6.24 (d, \(J = 7.6\) Hz, 1H, (C6)-H), 4.45 (app. s, 1H, (C2\(^2\))-H), 4.42 (app. td, \(J = 2.7\), 10.3 Hz, 2H, (C4')-H, (C4")-H), 4.28 (app. dt, \(J = 1.1\), 2.7 Hz, 1H, (C3\(^3\))-H)

\(^{13}\text{C NMR}\) (100.61 MHz, \(\text{D}_2\text{O}\)) \(\delta = 178.8\) (C2), 162.0 (C4), 143.0 (C5), 99.1 (C1\(^1\)), 97.4 (C6), 80.9 (C2\(^2\)), 77.7 (C4'), 75.1 (C3')

HRMS ESI-HRMS (pos. \(\text{m/z}\)) \([\text{M}+\text{H}]^+\) calculated for \(\text{C}_8\text{H}_{12}\text{N}_3\text{O}_3\text{S}^+\): 230.0592; found 230.0599.

IR (solid, cm\(^{-1}\)) 3396 (broad), 3103 (broad), 1634 (strong), 1476 (medium, C=C)

Melting Point (°C) 158-167 decomp.
Experimental Figure 90 $^1$H NMR (400 MHz, D$_2$O, zg30, 4.0-8.5 ppm) spectrum to show α-threo-thiocytidine α-threo-111.
Experimental Figure 91. $^{13}$C NMR (100 MHz, D$_2$O, 70-200 ppm) spectrum to show α-threo-thiocytidine α-threo-111.
5.4.1.22 Synthetic route for authentic standard of α-erythro-111

\[ \text{[Chemical Structures]} \]

1) THF, DIBAL, -78 °C, 5h
2) Ac₂O/DMAP/DCM, -78 °C, 16h
3) 2-thiocytosine, BSA, MeCN, reflux, 5h
4) DCE, SnCl₄, 0 - 5 °C

α-erythro-111
5.4.1.23 Synthesis of 2',3'-di-O-benzoyl-erythronolactone 146

L-Erythronolactone 144 (5 g, 42.3 mmol) and DMAP (100 mg, 0.82 mmol) were dissolved in anhydrous pyridine (80 mL) at 0 °C. Benzoyl chloride (12 mL, 103 mmol) was added dropwise under a N₂ atmosphere over 30 min. The mixture was allowed to come to room temperature and stirred for 2 days. The now heterogenous mixture was filtered and solvent removed in vacuo. The residue was co-evaporated with DCM (4 x 50 mL) and washed with 1 M HCl (3 x 50 mL), water (50 mL), NaHCO₃ sat. (50 mL), and brine (50 mL) before drying over MgSO₄, filtration and removing the solvent in vacuo to give 2,3-di-O-benzoyl-erythronolactone 146 (12.2 g, 88.3 %) as an amorphous off-white solid.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Mass (m/z)</th>
<th>Type</th>
<th>Charge</th>
<th>Number of Ions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+H]⁺</td>
<td>327.0863</td>
<td>ESI-HRMS</td>
<td>pos.</td>
<td>1</td>
<td>calculated for C₁₈H₁₅O₆⁺:</td>
</tr>
</tbody>
</table>

**1H NMR (700 MHz, D₂O)**

δ = 7.98 (app. d, J = 7.5 Hz, 4H, Ar-H), 7.59 (tt, J = 1.2, 7.4 Hz, 1H, (C3′)-Ar-H), 7.55 (tt, J = 1.2, 7.4 Hz, 1H, (C2′)-Ar-H), 7.43 (app. t, J = 7.5 Hz, 2H, (C3′)-Ar-H), 7.37 (app. t, J = 7.5 Hz, 2H, (C2′)-Ar-H), 5.99 (dd, J = 3.4, 5.4 Hz, 1H, (C3′)-H), 5.95 (d, J = 5.4 Hz, 1H, (C2′)-H), 4.71 (dd, J = 3.4, 11.4 Hz, 1H, (C4′)-H), 4.66 (d, J = 11.4 Hz, 1H, (C4′′)-H)

**13C NMR (176 MHz, D₂O)**

δ = 170.4 (C1), 165.4 (C3'OOCOAr), 165.1 (C2'OOCOAr), 134.0 (C2'/3'-Ar-para), 134.0 (C2'/3'-Ar-ortho), 130.2 (2C, C2'/3'-Ar-meta), 128.7 (2C, C3'-Ar-meta), 128.6 (2C, C2'-Ar-meta), 128.2 (C2'-Ar-ipso), 70.1 (C4), 69.8 (C3), 67.9 (C2)

**IR (solid, cm⁻¹)**

1794 (strong, C=O), 1723 (strong, C=O)

**Melting Point (°C)**

100-108 (110 - 111°)

---

294
Experimental Figure 92 $^1$H NMR (700 MHz, CDCl$_3$, zg30, 4.0-8.5 ppm) spectrum to show $2',3'$-di-O-benzoyl-erythronolactone 146.
Experimental Figure 93 $^{13}$C NMR (176 MHz, CDCl$_3$, 64-172 ppm) spectrum to show 2',3'-di-O-benzoyl-erythonolactone 146.
### 5.4.1.24 Synthesis of 2′,3′-di-O-benzyol-erythrothiocytidine α-erythro-147

2′,3′-Di-O-benzyol-erythronolactone 146 (1.63 g, 5.0 mmol) was added under a N₂ atmosphere and dissolved in anhydrous THF (50 mL). The mixture was cooled to -78 °C and DIBAL-H (1.2 M in toluene, 11.5 mL, 13.8 mmol) added dropwise over 2 h. The reaction was maintained at -78 °C for 10 h and monitored by TLC. When the starting material was consumed Ac₂O/DCM/DMAP (4.7 ml, 49.7 mmol/3.5 ml/ 100 mg, 0.82 mmol) was added at -78 °C, over 20 mins, under a N₂ atmosphere. The reaction was allowed to warm to room temperature over 16 h. Hexane (100 mL) was added and the mixture was added to rapidly stirring 1 M HCl (100 mL) and stirred vigorously for 1 h. The aqueous layer was separated and washed with EtOAc (3 x 50 mL). The combined organic layers were separated and washed with water (50 mL), NaHCO₃ sat. (50 mL) and brine (50 mL), dried over MgSO₄ and evaporated to give a crude yellow oil (1.66 g, 90%). A portion (1 g, 2.7 mmol) of this was dissolved in DCM (15 mL) and co-evaporated 3 times before anhydrous dichloroethane (DCE) (15 mL) was added under N₂ atmosphere. To this mixture, freshly prepared silylated 2-thiocytosine 113 (700 mg, 5.5 mmol, dissolved in 15 mL anhydrous DCE) was added and the mixture cooled to 0 °C before SnCl₄ (2 mL, 17.1 mmol) was added over 30 minutes. The now orange homogenous mixture was allowed to come to room temperature for 16 h and monitored by TLC. When the reaction was judged complete the reaction was quenched with NaHCO₃ sat. (25 mL) and stirred vigorously for 2 h. The mixture was twice filtered through a Celite® 545, washed with DCM (50 mL) and NaHCO₃ sat. (25 mL) and the organics separated. The aqueous phase was washed with DCM (3 x 20 mL). The combined organics were washed with brine (50 mL) and dried over MgSO₄ before the solvent was removed in vacuo. The crude product was purified with FCC eluting with a mixture of DCM/MeOH to give 2′,3′-di-O-benzyol-2-thiothreoctydine α-erythro-147 (547.9 mg, 56% (2 steps)) as a dark brown solid.
| **1H** NMR (700 MHz, CDCl₃) | δ = 8.08 (d, J = 1.3, 8.4 Hz, 2H, (C2′)-Ar-H-ortho), 7.90 (d, J = 1.3, 8.4 Hz, 2H, (C3′)-Ar-H-ortho), 7.66 (d, J = 7.5 Hz, 1H, (C6)-H), 7.57 (tt, J = 1.3, 7.4 Hz, 1H, (C2′)-Ar-H-para), 7.53 (tt, 1H, J = 1.3, 7.4 Hz, (C3′)-Ar-H-para), 7.50 (app tt, J = 1.7, 7.5 Hz, 2H, (C2′)-Ar-H-meta), 7.35 (app tt, J = 1.76, 7.5 Hz, 2H, (C3′)-Ar-H-meta), 7.32 (d, J = 3.6 Hz 1H, (C1′)-H), 6.08 (d, 1H, J = 7.5 Hz, (C5)-H), 5.90 (dd, J = 3.6, 5.2, 1H, (C2′)-H), 5.64 (q, J = 6.1 Hz, 1H, (C3′)-H), 4.62 (ABX, J = 6.3, 9.9 Hz, 1H, (C4′)-H) 4.28 (ABX, J = 6.3, 9.9 Hz, 1H, (C4″)-H) |
| **13C** NMR (151 MHz, D₂O) | δ = 180.2 (C2), 165.7 (C’3-OOCAr), 165.4 (C’2-OOCAr), 160.0 (C4), 140.2 (C6), 133.7 (C3′-Ar-para), 133.7 (C2′-Ar-para), 130.2 (2C, C3′-Ar-ortho), 130.0 (2C, C2′-Ar-ortho), 129.1 (ipso), 128.9 (ipso), 128.6 (2C, C3′-Ar-meta), 128.6 (2C, C2′-Ar-meta), 99.6 (C5), 92.3 (C1′), 75.7 (C2′), 71.6 (C4′), 70.6 (C3′) |
| **HRMS** | ESI-HRMS (pos. m/z): [M+H]+ calculated for C₂₂H₂₀N₃O₅S+: 438.1124; found 438.1153. |
| **IR (solid, cm⁻¹)** | 1720 (strong, C=O), 1648 (strong, C=C), 1247 (medium, broad C-N) |
| **Melting Point (°C)** | 128 (decomp.) |
Experimental Figure 94 $^1$H NMR (700 MHz, CDCl$_3$, zg30, 4.0-8.25 ppm) spectra to show 2',3'-di-O-benzoyl-erythrothiocytidine α-erythro-147.
Experimental Figure 95 $^{13}$C NMR (176 MHz, CDCl$_3$, 70-182 ppm) spectrum to 2',3'-di-O-benzoyl-erythrothiocytidine $\alpha$-erythro-147.
5.4.1.25 Synthesis of α-erythro-thiocytidine α-erythro-111

2,3-O-benzoyl-erythrothiocytidine 147 (109.1 mg, 0.25 mmol) was dissolved in 7 M NH₃/MeOH (2 mL) and MeOH (8 mL) and the heterogeneous mixture stirred at room temperature for 2.5 days. The solvent was evaporated and the residue redissolved in D₂O (7 mL) and washed DCM (3 x 7 mL). The aqueous layer was separated and lyophilised giving α-erythro-thiocytidine α-erythro-111 (53.4 mg, 93 %) as a white powder. A sample was removed for NMR analysis.

<table>
<thead>
<tr>
<th>1H NMR (700 MHz, D₂O)</th>
<th>δ = 7.76 (d, J = 7.6 Hz, 1H, (C5)-H), 6.70 (d, J = 3.4, 1H, (C1′)-H), 6.33 (d, J = 7.6 Hz, 1H, (C6)-H), 4.44 (dd, J = 5.2, 9.8 Hz, 1H, (C4′)-H), 4.38 (m, 2H, (C2′)-H, (C3′)-H), 4.00 (dd, J = 4.7, 9.8 Hz, 1H, (C4″)-H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13C NMR (176 MHz, D₂O)</td>
<td>δ = 179.7 (C2), 161.5 (C4), 141.7 (C6), 100.5 (C1′), 95.2 (C5), 76.4 (C2′), 73.4 (C4′), 70.0 (C3′)</td>
</tr>
<tr>
<td>HRMS</td>
<td>ESI-HRMS (pos. m/z): [M+H]+ calculated for C₈H₁₂N₂O₃S+: 230.0594; found 230.0597.</td>
</tr>
<tr>
<td>IR (solid, cm⁻¹)</td>
<td>3201 (broad, weak O-H), 1640 (strong, sharp, C-N), 1045 (strong, sharp, C-N)</td>
</tr>
<tr>
<td>Melting Point (°C)</td>
<td>149 (decomp.)</td>
</tr>
</tbody>
</table>
**Experimental Figure 96** $^1$H NMR (700 MHz, D$_2$O, zg30, 3.75-8.0 ppm) spectrum to show α-erythrothiocytidine α-erythro-111.
Experimental Figure 97 $^{13}$C NMR (176 MHz, D$_2$O, 65-185 ppm) spectrum to show α-erythrothiocytidine α-erythro-111.
**5.4.1.26 Peroxide oxidation of α-threo-111**

α-threothiocytidine (2.29 mg, 0.01 mmol) was dissolved in phosphate buffer (0.5 mL, 0.1 M, 10% D₂O) in an eppendorf and the pH adjusted to 7 with 1 M NaOH/HCl. The reaction was transferred into an NMR tube and H₂O₂ (30% w/w, 3.06 μL, 0.03 mmol) was added and the tube shaken vigorously for 20 s then monitored by NMR over 10 h after which the reaction was observed to be complete.

<table>
<thead>
<tr>
<th>¹H NMR (700 MHz, D₂O)</th>
<th>δ  =  7.71 (d, J = 7.41 Hz, 1H, C6), 5.96 (d, J = 7.63 Hz, 1H, C5), 5.72 (d, J = 0.90 Hz, 1H, C1’), 4.29-4.27 (m, 2H, C2’ C3’), 4.25-4.23 (m, 2H, C4’,C4”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>¹³C NMR (176 MHz, D₂O)</td>
<td>δ  =  166.92 (C4), 157.91 (C2), 142.59 (C6), 95.64 (C5) 93.58 (C1) 80.54 (s, 3 C4) 76.56 (s, 3 C2) 75.34 (s, 3 C3)</td>
</tr>
<tr>
<td>HRMS</td>
<td>ESI-HRMS (neg. m/z): [M-H]⁻ calculated for C₈H₁₀N₃O₄⁻ 212.0666; found 212.0676.</td>
</tr>
<tr>
<td>Melting Point (°C)</td>
<td>208-214 (decomp.)</td>
</tr>
</tbody>
</table>
Experimental Figure 98. $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), noesyggr1d, 4.0-8.25 ppm) spectrum to show α-threocytidine α-threo-85.

Experimental Figure 99. $^{13}$C NMR (176 MHz, D$_2$O, 74-176 ppm) spectrum to show α-threocytidine α-threo-85.
5.4.1.27 Peroxide oxidation of α-erythro-111

α-erythro-thiocytidine (2.29 mg, 0.01 mmol) was dissolved in phosphate buffer (0.5 mL, 0.1 M, 10% D₂O) in an eppendorf and the pH adjusted to 7 with 1 M NaOH/HCl. The reaction was transferred into an NMR tube and H₂O₂ (30% w/w, 3 μL, 0.03 mmol) was added and the tube shaken vigorously for 20 s then monitored by NMR over 10 h after which the reaction was observed to be complete.

\[ \begin{align*}
\text{1H NMR (700 MHz, D}_2\text{O)} & \quad \delta = 7.59 (d, J = 7.55 \text{ Hz, 1H, C6}), 6.00 (d, J = 7.53 \text{ Hz, 1H, C5}), 5.76 (d, J = 5.45 \text{ Hz, 1H, C1'}), 4.44 (m, 1H, C2'), 4.36 (m, 2H, C3', C4'), 3.94 (m, 1H, C4'') \\
\text{13C NMR (176 MHz, D}_2\text{O)} & \quad \delta = 166.75 (C4), 158.13 (C2), 142.66 (C6), 96.95 (C5), 91.82 (C1'), 75.27 (C2'), 74.00 (C4'), 70.87 (C3')
\end{align*} \]

Melting Point (°C) HRMS  ESI-HRMS (pos. m/z): [M+H]⁺ calculated for C₈H₁₂N₃O₄⁺ 214.0822; found 214.0818.
Experimental Figure 100 $^1$H NMR (700 MHz, H$_2$O/D$_2$O (9:1), noesygpr1d, 3.5-8.0 ppm) spectrum to show $\alpha$-erythrocytidine $\alpha$-erythro-85.

Experimental Figure 101 $^{13}$C NMR (176 MHz, D$_2$O, 65-170 ppm) spectrum to show $\alpha$-erythrocytidine $\alpha$-erythro-85.
5.4.1.28 Synthesis of β-erythro-thiocytidine-2',3'-cyclic phosphate β-erythro-151

β-erythro-thiocytidine (4.45 mg, 0.02 mmol), urea (96 mg, 1.6 mmol) and ammonium phosphate (12 mg, 0.1 mmol) were heated at 140 °C for 30 mins, allowed to cool and dissolved in water (3 mL) and lyophilised. The resultant brown residue was dissolve in triethyl ammonium carbonate (TEAB, 1 mL 10 mM, pH 9) and purified by FCC (C18) eluting with TEAB/MeOH. Multiple lyophilisations gave β-erythro-thiocytidine 2',3'-cyclic phosphate β-erythro-151 as white powder (2.8 mg, 50 %).

$^1$H NMR (700 MHz, D$_2$O) δ = 7.92 (d, J = 7.7 Hz, 1H, (C6)-H), 6.61 (app. dd, J = 1.9, 4.0 Hz, 1H, (C1')-H), 6.32 (d, J = 7.7 Hz, 1H, (C5)-H), 5.37 (m, 1H, (C2')-H), 5.12 (td, J = 4.1, 6.5 Hz, 1H, (C3')-H), 4.36 (d, J = 11.6 Hz, 1H, (C4')-H), 3.98 (dt, J = 3.6, 11.6 Hz, 1H, (C4'')-H)

$^{13}$C NMR (176 MHz, D$_2$O) δ = 178.8 (C2), 161.9 (C4), 144.0 (C5), 99.5 (C1'), 90.6 (C5), 78.2(C2'), 76.4(C3'), 72.0 (C4')

$^{31}$P NMR (283 MHz, D$_2$O, $^1$H-decoupled) δ = 20.33 (s)

$^{31}$P NMR (283 MHz, D$_2$O) δ = 20.33 (t, J = 7.3 Hz)

HRMS ESI-HRMS (neg. m/z): [M+H]$^+$ calculated for C$_8$H$_{11}$N$_3$O$_5$PS$^+$: 292.0151; found 292.0150.
Experimental Figure 102 $^1$H NMR (700 MHz, D$_2$O, zg30, 3.3-8.3 ppm) spectrum to show β-erythro-thiocytidine-2',3'-cyclic phosphate β-erythro-151.
**Experimental Figure 103** $^{13}$C NMR (176 MHz, D$_2$O, 50-200 ppm) spectrum to show β-erythro-thiocytidine-2',3'-cyclic phosphate β-erythro-151.

**Experimental Figure 104** $^{31}$P NMR (283 MHz, D$_2$O, $^1$H-decoupled, -50-150 ppm) spectrum to show β-erythro-thiocytidine-2',3'-cyclic phosphate β-erythro-151.
5.4.2 DNA Chapter

5.4.2.1 2',8-cyclo-adenosine

Adapted from Stairs et al.\textsuperscript{192}

8-bromo-\textit{ribo}-adenosine (100 mg, 0.29 mmol) and ammonium formate (73 mg, 1.16 mmol) were suspended in water (50 mL) and the pH adjusted to 9 with aqueous ammonia solution (30% w/w). The mixture was stirred for 18 h, neutralised with formic acid (conc.) concentrated under a stream of air and filtered. The filtrate was washed with cold water (2 x 25 mL) and acetone (2 x 25 mL) and dried under vacuum to give 2',8-cyclo-adenosine (63.1 mg, 82.4%) as an off-white powder.

| Table 5.4.2.1 2',8-cyclo-adenosine
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>\textsuperscript{1}H NMR (700 MHz, D\textsubscript{2}O)</strong></td>
</tr>
<tr>
<td><strong>\textsuperscript{13}C NMR (700 MHz, D\textsubscript{2}O)</strong></td>
</tr>
<tr>
<td><strong>HRMS</strong></td>
</tr>
<tr>
<td><strong>IR (solid, cm\textsuperscript{-1})</strong></td>
</tr>
<tr>
<td><strong>Melting Point (°C)</strong></td>
</tr>
</tbody>
</table>
5.4.2.2 Synthesis of 2′-deoxy-2′-ethylmercapto-8-oxy-adenosine

Adapted from Ikehara et al.\textsuperscript{387}

NaH (60% suspension in mineral oil) (44 mg), was dissolved in 5 mL of DMF and ethanethiol (0.3 mL, 4.05 mmol, 8.14 eq) was added slowly and stirred under N\textsubscript{2} until the mixture was homogenous. 8,2′-anhydro-arabino-adenosine (79) (132 mg, 0.5 mmol) was added in 2 mL of DMF and the reaction heated at 70 °C for 1 hour. The reaction was neutralised with 1 M HCl and excess thiol blown off with N\textsubscript{2}. The solvent was removed under vacuo, redissolved in water (10 mL) and lyophilised and analysed by \textsuperscript{1}H NMR spectroscopy by dissolving the lyophilisate in 1 mL D\textsubscript{2}O and removing 0.5 mL for analysis.

Improved method of 2′-deoxy-2′-ethylmercapto-8-oxy-adenosine

Ethanethiolate (170 mg, 2 mmol) was dissolved in DMF (10 mL) and 8,2′-anhydro-arabino-adenosine (270 mg, 1 mmol) and ethanethiol (590 µL, 8 mmol) were added and the reaction heated at 70 °C for 24 hours. The excess thiol was blown off with N\textsubscript{2} ebullition. The solvent was removed under vacuo and redissolved in water (10 mL) and lyophilised. 2′-deoxy-2′-ethylmercapto-8-oxy-adenosine was obtained as an orange crystalline solid 309.3 mg (92% yield).
**$^1$H NMR (600 MHz, D$_2$O)**  
$\delta = 7.95$ (s, 1H, (C2)-H), 5.90 (d, $J = 9.7$ Hz, 1H, (C1')-H), 4.41 (d, $J = 5.1$ Hz, 1H, (C3')-H), 4.3 (m, 2H, (C2')-H, (C4')-H), 3.87 (app. qd, $J = 3.0, 12.7$, 2H, (C5')-H), (C5'')-H), 2.44 (m, 2H, SCH$_2$CH$_3$) 1.02 (t, $J = 7.4$ Hz, 3H, SCH$_2$CH$_3$).

**$^{13}$C NMR (151 MHz, D$_2$O)**  
$\delta = 155.1$ (C8), 151.3 (C2), 149.4 (C5), 149.0 (C6), 148.2 (C4), 88.0 (C1'), 49.6 (C2'), 73.7 (C3'), 87.6 (C4'), 63.3 (C5'), 54.3 (SCH$_2$CH$_3$), 6.00 (SCH$_2$CH$_3$).

**MS**  
ESI-MS (pos. m/z): [M+H]$^+$ calculated for C$_{12}$H$_{18}$N$_5$O$_4$S$: 328.1; found 328.1.

**UV-Vis**  
$\lambda_{\text{max}} = 270, 207$ nm.

**IR (solid, cm$^{-1}$)**  
3188-3324 (weak, broad, N−H, O−H), 1614 (strong, C=O), 1450 (medium, C-H), 1337 (strong, C-N)

**Melting Point (°C)**  
144
5.4.2.3 Synthesis of 2′-deoxy-2′-ethylmercapto-ribocytidine

![Structure of 2′-deoxy-2′-ethylmercapto-ribocytidine]

Ethanethiolate (101 mg, 1.2 mmol) was dissolved in DMF (6 mL) and ancitabine (135 mg, 0.6 mmol) and ethanethiol (250 µL, 4.8 mmol) were added and the reaction stirred for 24 hours. The excess thiol was blown off with N₂. The solvent was removed under vacuo and redissolved in water (10 mL) and lyophilised. 2′-deoxy-2′-ethylmercapto-ribocytidine was recovered as an off white oily solid 135.1 mg (94%).

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ (600 MHz, D₂O)</td>
<td>δ = 7.92 (d, J = 7.6 Hz, 1H, (C6)-H), 6.12 (d J = 8.7 Hz, 1H, (C1')-H), 6.10 (d, J = 7.6 Hz, 1H, (C5)-H), 4.38 (dd, J = 2.5, 5.6 Hz, 1H, (C3')-H), 4.14 (m, 1H, (C4')-H), 3.81 (ABX, J = 4.6, 12.5 Hz, 2H, (C5')-H, (C5'')-H), 3.58 (dd, J = 5.6, 8.7 Hz, 1H, (C2')-H), 2.53 (ABX, J = 1.5, 7.4 Hz, 2H, S-CH₂CH₃), 1.13 (t, J = 7.4, Hz, 3H, S-CH₂CH₃).</td>
</tr>
<tr>
<td>δ (151 MHz, D₂O)</td>
<td>δ = 166.7 (C4), 158.0 (C2), 142.2 (C6), 97.8 (C5), 90.2 (C1'), 86.5 (C4'), 72.2 (C3'), 62.1(C5'), 52.6 (C2'), 26.1 (S-CH₂CH₃), 15.1 (S-CH₂CH₃)</td>
</tr>
<tr>
<td>UV-Vis</td>
<td>λmax = 270, 253, 195</td>
</tr>
<tr>
<td>MS</td>
<td>ESI-MS (pos. m/z): [M+H]+ calculated for C₅₁H₁₈N₃O₄S+: 288.1; found 288.1</td>
</tr>
<tr>
<td>IR (solid, cm⁻¹)</td>
<td>3198-3330 (weak, broad, N-H, O-H), 1643 (strong, C=C), 1489 (medium, C-H)</td>
</tr>
<tr>
<td>Melting Point (°C)</td>
<td>186-203</td>
</tr>
</tbody>
</table>
5.4.2.4 Synthesis of 2′-deoxy-2′-ethylmercapto-ribouridine

Ethanethiolate (170 mg, 2 mmol) was dissolved in DMF (10 mL) and 2,2'-anhydouridine (226 mg, 1 mmol) and ethanethiol (590 µL, 8 mmol) were added and the reaction stirred for 24 h. The excess thiol was blown off with N₂. The solvent was removed in vacuo and redissolved in water (10 mL) and lyophilised. 2′-deoxy-2′-ethylmercapto-ribouridine was recovered as a yellow oily solid 275.3 mg (96%).

\[ ^1H \text{ NMR (600 MHz, } D_2O) \delta = 7.73 \text{ (d, } J = 7.7 \text{ Hz, 1H, (C6)-H}), 6.11 \text{ (d, } J = 8.9 \text{ Hz, 1H, C1′), 5.87 \text{ (d, } J = 7.7 \text{ Hz, 1H, (C5)-H}), 4.38 \text{ (dd, } J = 2.2, 5.7 \text{ Hz, 1H, (C3′)-H}), 4.13 \text{ (m, 1H, (C4′)-H}), 3.80 \text{ (ABX, } J = 4.68, 12.47 \text{ Hz, 2H, (C5′)-H, (C5′′)-H)), 3.59 \text{ (dd, } J = 5.7, 8.9 \text{ Hz, 1H, (C2′)-H}), 2.53 \text{ (ABX, } J = 1.9, 7.4 \text{ Hz, 2H, SCH}_2CH_3), 1.13 \text{ (t, } J = 7.4 \text{ Hz, 3H, SCH}_2CH_3). \]

\[ ^13C \text{ NMR (151 MHz, } D_2O) \delta = 174.4 \text{ (C4), 158.3 \text{ (C2), 141.2 \text{ (C6), 104.2 \text{ (C5), 89.9 \text{ (C1′), 86.5 \text{ (C4′), 72.6 \text{ (C3′), 62.2 \text{ (C5′), 52.1 \text{ (C2′), 25.9 \text{ (SCH}_2CH_3), 15.1 \text{ (SCH}_2CH_3).}}}}\]

UV-Vis \[ \lambda_{\text{max}} = 261, 190 \text{ nm} \]

MS ESI-MS (pos. m/z): [M+H]^+ calculated for C_{11}H_{17}N_2O_5S^+: 289.1; found 289.1

IR (solid, cm⁻¹) \[ 3302-3366 \text{ (weak, broad, N–H, O–H), 1633 \text{ (strong, C=C), 1507 \text{ (medium, C–H)}} \]

315
6 References


70. Milsheteyn, D., Cooper, G. & Deamer, D. Chemiosmotic energy for primitive cellular life: Proton gradients are generated across lipid membranes by redox reactions coupled to meteoritic quinones. 1–12 (2019). doi:10.1038/s41598-019-48328-5


87. Stubbs, R. T., Yadav, M., Krishnamurthy, R. & Springsteen, G. A plausible metal-


89. Kurland, C. G. The RNA dreamtime: Modern cells feature proteins that might have supported a prebiotic polypeptide world but nothing indicates that RNA world ever was. *BioEssays* **32**, 866–871 (2010).


185. Inoue, T. & Orgel, L. E. Substituent Control of the Poly(C)-Directed Oligomerization of Guanosine 5’-Phosphoroimidazolide. Journal of the American


238. Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis. V. Synthesis and


276. Wilds, C. J., Wawrzak, Z., Krishnamurthy, R., Eschenmoser, A. & Egli, M. Crystal structure of a B-form DNA duplex containing (L)-α-threofuranosyl (3'→2') nucleosides: A four-carbon sugar is easily accommodated into the backbone of


342

5847 (2013).


