Chemoselective Syntheses of Pantetheine with Prebiotic Nitrile Chemistry and an Investigation of the Photochemical Reactivity of Thioacids in Water

Jasper Henry Fairchild

UCL

Thesis submitted to University College London for the degree of

Doctor of Philosophy

2024

'I, Jasper Henry Fairchild 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.'

Signed: Jasper Fairchild

Date: 28/02/2024

Abstract

Co-factors are vital nodes in metabolism and perform a wide range of functions at the heart of cellular machinery. Despite the prevalence and conservation of co-factors in extant biology it is unclear why and how these specific molecules came into being. For example, coenzyme A (CoA) is found in all domains in life and is essential in autotrophic carbon fixation pathways, fatty acid, polyketide, and non-ribosomal peptide syntheses. CoA has a complex structure which can be broken down into an adenosine nucleotide and pantetheine residue. It is the thiol bearing pantetheine fragment that is catalytically active, underpinning its metabolic value. This work focuses on prebiotic syntheses of pantetheine which is comprised of pantoate, β-alanine and cysteamine sub-fragments. Previous attempts to synthesise pantetheine from pantolactone, β-alanine and cysteamine gave near negligible yields. However, by utilising the innate reactivity of nitriles this work demonstrates several high yielding routes towards pantetheine in water. The reactivity of nitriles provided a chemical rationale for each element of the canonical structure of pantetheine over non-biological analogues and is concurrent with peptide synthesis. As the functional unit within CoA, pantetheine is involved in many enzymatic reactions (~ 4%) and is implicated in various origins of life scenarios.

α-Peptides are an essential component of life, enabling catalysis, replication, and structure in cells. Previous work in the Powner group exploited α-aminonitriles in a high-yielding (biomimetic) peptide ligation that's precedes *via* thioacid intermediates. The final part of this thesis investigated the UV irradiation of thioacids. The *N*-acetylation of amines with thioacetate under UV irradiation was demonstrated. However, attempts to ligate α-amidothioacids with α-aminonitriles upon UV irradiation were unsuccessful due to radical dethiocarboxylation. Interestingly, β-amidothioacids did not undergo dethiocarboxylation and were activated towards peptide coupling under UV irradiation. During these studies it was observed that acidic conditions (in the dark) resulted in ligation of α-amidothioacids with glycine nitrile. Proteins are the catalytic workhorse of biology and often require co factors to function efficiently. This work sheds light on the key chemical reactions required to build life and these synergistic qualities present a possible route to a protometabolic network present in the beginning.

Impact Statement

Work on the Origins of Life captures the imagination of scientist and the public alike. The deeply existential grand questions of 'Where do we come from?' or 'Is there life beyond Earth?' triggers deeply profound questions of science, religion, philosophy for many people and the possible answers this field provides may impact our views of our place within the Universe. As such, abiogenesis has the means to spark curiosity in the most unlikely of places and is a worthy topic to invest researching.

Proteins form a core part of cell metabolism and are essential to life as we know it. Proteins provide the means for biology to enact complex chemical transformations catalytically through enzymes. The group has been working on peptide synthesis in water and this work led to the discovery of catalytic peptide ligations that use thiols as the catalysis. CoA is a thiol that is ubiquitous to all life on Earth and is present in approximately 4% of enzymes found in extant biology. CoA sits at the interface between many different biological processes and is quite a complex molecule consisting of many fragments, therefore an understanding how this molecule came into being is important.

Society wants to get close to the answer to the question of how life began on earth and this work offers a small step towards that. This work will help to inform the understanding of how small molecules can be highly functional at producing useful species such as cofactors, which are known by scientifically aware members of the public to be important in extant biological systems.

The work on a prebiotically plausible chemoselective pantetheine synthesis in water has been published in *Science*. This work has been presented at international symposia (Simmons Collaboration on the Origins of Life 2023, CRC Emergence of Life 2023). The collaborative nature of the field has led to the work on the photochemistry of thioacids being investigated by Prof. Dimitar Sasselov and his group at the Harvard University.

Published Works

Parts of this thesis were published during the submission of this thesis:

J. Fairchild, S. Islam, J. Singh, D-K Bučar and M. W. Powner, Prebiotically plausible chemoselective pantetheine synthesis in water. *Science* **383**, 911-918 (2024). Published.

Manuscript doi: https://www.science.org/doi/10.1126/science.adk4432

UCL Research Paper Declaration Form
Referencing the doctoral candidate's own published work(s)

1. For a research manuscript that has already been published

   a) **What is the title of the manuscript?**

   Prebiotically plausible chemoselective pantetheine synthesis in water

   b) **Please include a link to or doi for the work**

   https://www.science.org/doi/10.1126/science.adk4432

   c) **Where was the work published?**

   Science

   d) **Who published the work?** (e.g. OUP)

   AAAS

   e) **When was the work published?**

   22nd Feb 2024

   f) **List the manuscript's authors in the order they appear on the publication**

   Jasper Fairchild, Saidul Islam, Jyoti Singh, Dejan-Krešimir Bučar and Matt Powner

   g) **Was the work peer reviewed?**

   Yes

   h) **Have you retained the copyright?**

   Copyright © 2024 the authors, some rights reserved; exclusive licensee AAAS

i) **Was an earlier form of the manuscript uploaded to a preprint server?**
(e.g. medRxiv). If 'Yes', please give a link or doi)

No

If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

☒

*I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.*

2. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)

The author Matthew W. Powner. conceived and coordinated the research.

Jasper Fairchild, Saidul Islam, Jyoti Singh, and Matthew W. Powner designed and analysed the experiments.

Dejan-Krešimir Bučar carried out the x-ray crystallography.

Jasper Fairchild and Saidul Islam contributed equally to the experiments.

1. Attempted syntheses of pantothenic acid and pantetheine from pantolactone, β-alanine, and Cysteamine. Jasper Fairchild and Saidul Islam

2. Syntheses of pantetheine from pantolactone using nitrile activation. Jasper Fairchild

3. Chemoselective aldol synthesis of hydroxypivaldehyde and pantolactone at neutral pH. Jasper Fairchild, Saidul Islam and Jyoti Singh

4. Reactions of hydroxypivaldehyde with amines Jasper Fairchild, Saidul Islam and Jyoti Singh

5. Preparative syntheses of authentic standards. Jasper Fairchild and Saidul Islam

6. Single crystal X-ray diffraction data. Dejan-Krešimir Bučar

Saidul Islam and Matthew W. Powner wrote the paper and compiled the supplementary information

Jasper Fairchild, Saidul Islam, Jyoti Singh and Matthew W. Powner wrote the SI

All authors approved the final submission.

3. In which chapter(s) of your thesis can this material be found?

Chapter 2. Coenzyme A - Results & Discussion

Sub-sections

2.1 Introduction

2.2 Syntheses of pantothenic acid & pantetheine in water

2.3 Syntheses of pantetheine using nitrile activation

2.4 Reactions of pantothenic acid nitrile with cysteamine

2.5 Chemoselective aldol synthesis of hydroxypivaldehyde and pantolactone at neutral pH

2.6 Reactions of hydroxypivaldehyde with amines

2.7 Conclusions

4. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

Candidate

Jasper Fairchild

Date:

23/02/2024

Supervisor/ Senior Author (where appropriate)

Matthew W. Powner

Date

23/02/2024

Acknowledgments

Although a PhD may feel like a solo endeavour, you are never truly alone. When I was young, I never thought in a I would complete a PhD and become a scientist. I am very fortunate to be able to do something that I love and have only reached this point through the special people that have come with me on this journey.

Firstly, I would like to thank Matt for giving me the opportunity to attempt a PhD. I count myself extremely lucky that you were my supervisor and I'm amazed by the enthusiasm you show daily. It reflects on the group too and I finish these studies in greater awe and wonder of the beauty of nature. I finish my PhD with a great appreciation of the skill and dedication it takes to be a great scientist.

I would like to thank Lello for his support and pushing me to reach my potential. I also appreciate red wine now thanks to you! When I started this PhD, I told myself if I knew half as much as Daniel at the end of my PhD I would consider it a successful PhD – I'm still learning. I would like to thank Daniel for inspiring me to become a better scientist and supporting me along the way. To Jyoti for encouraging me along the way and always providing a shoulder to lean on when the times got tough.

To my cohort, Alvaro and Benji. Special thanks to Alvaro as a work mate and friend, I consider myself very lucky to have the pleasure of living with you and seeing us both grow during our PhDs. To Benji for your thoughtfulness and witty humor, I'll miss sitting next to you in the office. I'm very excited to see both your journeys post PhD.

To the other members of Matt's group, Abid, Andrew, Arran, Ben, Cal, Christian, Leonardo and Shane, thanks for the equally good banter and support. To all of members of the group, you have all made this such a rich period of my life and there are no words I can express for all your support, humor, the deep discussions, and the group outings that make even the daily grind of research bearable.

Of course, none of this happened without the analysis, so thanks also to Abil for help with NMR and his letting me work in the NMR room to fund my reef tank addiction. To Tony in stores for the quick chats whilst helping me and the group function upstairs.

To my parents, Barry and Josie, you have been so supportive of my studies from the time I started my A Levels all the way to finishing this PhD, I want to thank you for enabling me to pursue my dreams and enable me to reach this point. I want to thank my brother, Tom for being there for me all the time and supporting me in the good and bad times. To my grandparents Nan (Shirley), Tom, Grandad (Ken) and Silvia who were so encouraging when I was younger.

I'd like to thank all my friends near or far, all of you have helped me through this journey. A shout out to Michelle, Maria and My; my Notts friends Alastair, Alex, Andy, Matt, Rob, the Ramzey's, Tom; to members of the Anderson group, if it wasn't for Chia Hao, I wouldn't have started this PhD! To my friends from Colindale Ed, Anca, baby Saul and Vishal. To my friends from home who were with me at the beginning. UCL is amazing place to study and work and there so many friends and colleagues I see every day. The quick chats and support, it really makes a difference, so thanks to all you guys at UCL!

Last but not least, I would like to thank Noemi who has been my biggest fan and kept me going even when I did not believe in myself. I love you and I can't wait to spend my future with you.

The Author

Jasper Fairchild was born in Freedom Fields Hospital, Plymouth on February 1989. He attended Tavistock Primary School, then Keyham Barton Catholic Primary School. He studied for GCSEs at St. Boniface's Catholic College and stayed for his A-levels studies, where he won the Brother Rafferty prize for chemistry. He received his MSci University of Nottingham in 2015, and then worked for Oxford Analytical Ltd for 18 months. The author decided to resume his studies at UCL under Professor James C Anderson synthesising Luciferin Analogues, earning a MRes in 2018. He decided to resume his studies at UCL under Professor James Anderson synthesising Luciferin Analogues, earning a MRes in 2018. He remained at UCL to study prebiotically plausible chemoselective pantetheine syntheses and to investigate the photochemistry of thioacids under Professor Matthew Powner in 2019. He completed his PhD in 2024 and has stayed at UCL as an experimental officer in organic chemistry to support teaching at the university.

"What's the matter Danny, you never taken a shortcut before"

Sgt Nicholas Angel – Hot Fuzz

""I wish it need not happen in my time," said Frodo. "So do I," said Gandalf, "and so do all who live to see such times. But that is not for them to decide. All we have to decide is what to do with the time that is given us"""

J.R.R Tolkien's The Lord of the Rings – The Fellowship of the Ring

"Try not to become a man of success, rather become a man of value."

Albert Einstein

"20 years from now you will be more disappointed by the things you didn't do, than by the ones you did do. So, throw off the towlines, sail away from safe harbour catch the trade winds in your sails. Explore, Dream, Discover"

Mark Twain

"Roads go ever on, over rock and under tree, By caves where never sun has shone, By streams that never find the sea; Over snow by winter sown, And through the merry flowers of June, Over grass and over stone, And under mountains in the moon. Roads go ever on, Under cloud and under star, yet feet that wandering have gone, turn at last to home afar. Eyes that fire and sword have seen, And horror in the halls of stone, Look at last on meadows green, And trees and hills they long have known""

J.R.R Tolkien's The Lord of the Rings – The Fellowship of the Ring

"Sooo Jasper, you got those NMRs yet?"

Professor Matthew Powner

Abbreviations

Å	Angstrom
AA-CN	α-aminonitrile
Ac	Acetyl
AICN	4-amino-imidazole-5-carbonnitrile
AMP	adenosine monophosphate
Ar	aryl
ATP	adenosine triphosphate
Boc	*tert*-butyloxycarbonyl
Bu	butyl
CAA	*N*-carbamoyl amino acids
CDI	*N*, *N′*-carbonyldiimidazole
CoA	coenzyme A
d	doublet
D	dextrorotary
DAP	diamidophosphate
DCC	*N,N′*-dicyclohexylcarbodiimide
DMAP	4-dimethylaminopyridine
DMF	dimethyl formamide
DMSO	dimethylsulfoxide
δ	chemical shift in parts per million
EDC	*N*-(3-Dimethylaminopropyl)-*N′*-ethylcarbodiimide
EG-F-GTP	prokaryote elongation factor G/translocase complex
Et	ethyl
Ga	gigaannum (1 billion/1 x 10^9 years)
h	hour(s)
HD	hydantoin
HDA	hydantoic acids
HATU	hexafluorophosphate azabenzotriazole tetramethyl uronium

HOAt	1-hydroxy-7-aza-benzotriazole
HOBt	hydroxybenzotriazole
HRMS	high resolution mass spectroscopy
Hz	hertz
^{i}Pr	isopropyl
IR	infra-red
J	coupling constant in NMR spectroscopy
LRMS	Low resolution mass spectrometry
LUCA	last universal common ancestor
m	multiplet
Me	methyl
min	minutes
MSM	methylsulfonylmethane
NADH	nicotinamide adenine dinucleotide
NCA	*N*-carboxy amino acid anhydrides
NCL	native chemical ligation
NTA	*N*-thiocarboxy anhydride
NMR	nuclear magnetic resonance
PCP	peptidyl carrier protein
Pfp	pentafluorophenyl
Ph	phenyl
pK_a	acid dissociation constant
pK_{aH}	pK_a of conjugate acid
PPAN	4'-phospopanterheinyl cofactor
PLP	pyridoxal-5'-phosphate
ppm	parts per million
S_N2	substitution nucleophilic bi-molecular
spt.	septet
TIPS	triisopropylsilane
TFA	triflouroacetic acid
TLC	thin layer chromatography

| Trt | trityl (triphenylmethyl) |
| UV | ultra-violet radiation |

Nomenclature and Numbering

Pentose sugar numbering

α-ribose   β-ribose
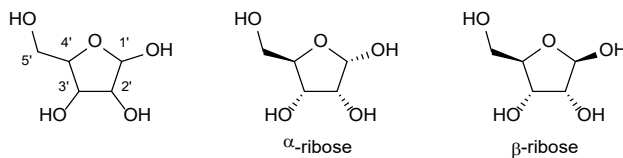
Nucleoside numbering
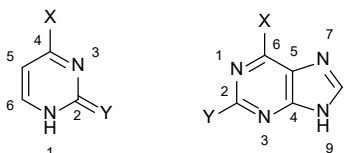
Heterocycle numbering
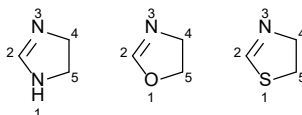
Amino acid nomenclature

Ac-**AA**-CN   **AA**-SNH$_2$ **AA**-X

Ac-**AA**-CN   Ac-**AA**-SNH$_2$ Ac-**AA**-X

Ac-**(AA)**$_n$-CN   Ac-**(AA)**$_n$-SNH$_2$ Ac-**(AA)**$_n$-X

Ac-**Ser^{Ac}**-CN   Ac-**Cys^{Ac}**-CN   Ac-**Cys^{STrt}**-STrt

Ac-**Gly-Gly**-X   Ac-β-**Ala-Cys$_{Ac}$**-CN
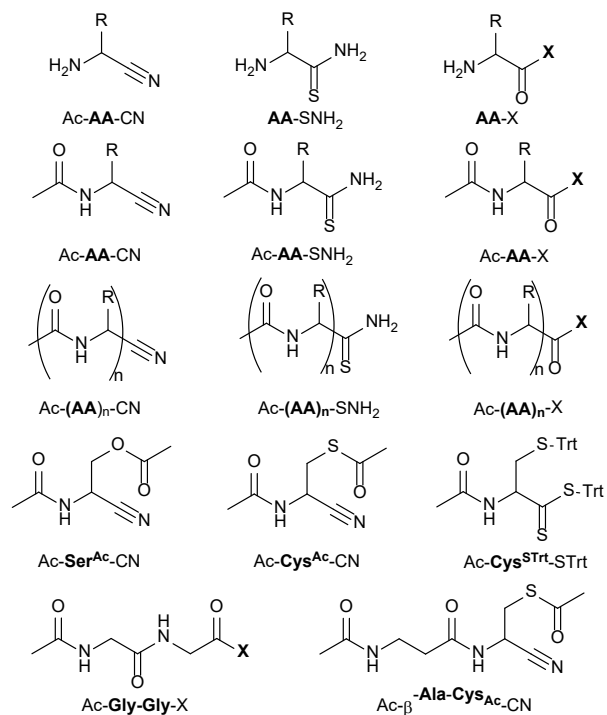
AA = Any Amino Acid eg Glycine (Gly), Alanine (Ala), Cysteine (Cys) etc, **X** = SH, NH$_2$, OH

Contents

1. Introduction

Ever since ancient times humans have asked the question of "where do we come from?" and throughout history humans have tried to explain and rationalize this problem through religion and science. How can it be that the Earth orbiting the Sun, which is considered an average star amongst the billions of stars in our galaxy, demonstrate such abundant and diverse lifeforms? Indeed, the fact that there are beings on Earth that can contemplate their place within the grand tapestry of the universe and attempt to even ask this existential question is remarkable.

How life came to be is one of the great unanswered questions and it may turn out to be impossible to precisely determine the events and processes which led to abiogenisis.[1] However, despite the enormity of the task many great minds through the ages have tried to explain and understand this. In ancient Greece the philosopher Anaximander of Mellites (610-546 BC) proposed from the apeiron (chaos) that all terrestrial beings are derived from the interactions between the non-living elements of earth, air, fire and water.[2,3] This interplay between hot and cold, wet and dry, light and dark has been thematic for thinkers over the millennia. One of Anaximander's proposals was that the first animals originated in the water and that humans could be descended from these aquatic organisms. Although this hypothesis is not Darwinist in the modern sense it could be considered a proto evolutionary theory. In another school of thought, Plato's student Aristotle (384-322 BC) in his book 'On the Generation of Animals' suggested that different forms of life emerged from non-living material.[2,4] This theory was born from his observations that some organisms such as maggots could form from rotting matter or the apparent generation of insects, clams, snails, and other simple organisms from mud.

Much of the evolutionary ideas were lost after the fall of the Roman Empire with the Aristotelian theory of 'spontaneous' generation and the belief in essentialism becoming the dominant force during medieval times. These doctrines greatly influenced Western Christian thought. However, scholars such as Al Jahiz (776 – 868 AD) or Ibn Kaldun (1332 – 1406 AD) had also explored ideas relating to the

transition of non-living matter to living beings and perhaps a biological theory of evolution by asserting that humans developed from 'the world of monkeys' during the Islamic Golden Age.[5]

By the 17th Century scientists had begun to propose that humans had descended from primitive creatures originating in the world's oceans, however many were still convinced that even this was the result of some divine plan. In 1861 Louis Pasteur demonstrated that bacteria and fungi did not spontaneously appear in sterile vessels filled with nutrient rich beef broth until this mixture was exposed to the environment.[6] At the turn of the empiricism and the movement towards rational scientific enquiry culminated in Darwin's theory of evolution and indicated that humans were the children of the forces of natural selection and the race of the survival of the fittest. Even in the 19th century Charles Darwin had contemplated a chemical origin of life, for when writing to his good friend Hooker in 1871 he said:

"But if (and oh what a big if!) we could conceive in some warm little pond with all sorts of ammonia and phosphoric salts, light, heat, electricity etcetera present, that a proteine compound was chemically formed, ready to undergo still more complex changes (..)"[7]

It is a curious phenomenon that despite the ample quantities of organic material on Earth, life does not form *de novo* from these mixtures in a 'warm little pond'. The tools available to scientists have exponentially improved since the first synthesis of urea by Wohler in 1828 through to subsequent discoveries of DNA and the genome in the 20th century.[8–10] Currently, the field of origin of life research encompasses a wide spectrum of scientific disciplines such as physics, astronomy, engineering, computational sciences, geology, biology, and chemistry. With the improvement and development of new technology, distant exoplanets can be observed and analysed with new telescopes, spacecraft can reach further and analyse Martian soil or bodies beyond in greater detail, dating rock samples based on isotope ratios are more accurate, or computational modelling can simulate more variables and components due to exponentially increasing computational power.

Abiogenesis can be considered fundamentally a chemical problem, since life functions through a complex array of systems, which have increased in complexity over generations from the reactions of structurally simple molecules. Modern organic chemistry is a powerful tool, and prebiotic chemists can investigate the chemical reactions that could have plausibly occurred prior to the earliest life forms establishing itself on Earth ca. 3.5 Ga.[11] By performing prebiotic chemical reactions, organic chemists seek to understand why life is built from specific sets of molecules over others and to explore the relationships between these essential molecules by elucidating their fundamental common chemical roots.

With the development of new synthetic and analytical techniques such as nuclear magnetic resonance (NMR) and high-performance liquid chromatography (HPLC), studies not feasible even a decade ago become trivial. The collegiate nature of this field is extremely important and enables scientists to continue learning and propose plausible scenarios which can be tested *via* numerous complementary lines of investigation. So that a convincing consensus of how life may have emerged.

Major strides have been achieved in the field and now possible to show how simple feedstocks are able form nucleic acids, amino acids, fatty acids and the other core components of extant biology. From these relatively simple building blocks, chemists seek to find plausible routes to more complicated biopolymers such as lipids, proteins or RNA oligomers which give form and functionality in cellular life.

1.1. What is Life?

To begin the process of understanding how life is formed and why nature 'chose' the system presently found in life on Earth, a precise definition of what life is may help towards this end goal. This is surprisingly difficult, as life is the quintessential complex system and there may or not be a discrete point where these systems are considered alive.[12–15] However, a useful working definition from NASA stipulates that:

"A self-sustaining chemical system capable of Darwinian evolution."[16]

However, this definition is quite broad and vague, the precise nature of life is up for philosophical, religious and of course scientific debate that is unique for each person and scientist interested in this conundrum. Some origins of life scientists such as Sutherland assert that there is a gradual transition from 'inanimate' chemical systems to one that maybe considered 'alive'.[17] This process envisages taking inherently favoured prebiotic chemistry such as the synthesis of simple building blocks, which over many incremental cycles can increase in complexity to macromolecules which could undergo chemical evolution.[18] However, this point in evolutionary innovation is a far cry to an organism like the last common ancestor (LUCA).[18] Nevertheless, to begin the process of replicating primeval chemistry, the timeframe, the conditions of the early Earth and the potential chemical environments at this time need to be considered.

1.2. The Early Earth

1.2.1. The Sun

Today, the light emitted from the Sun is the undisputed energy source for many processes on the Earth. The Sun influences the weather, climate, ocean currents, the seasons, photosynthesis and in turn our food sources. It can also affect us as a species through climate change or individually through cancer caused by the photo destruction of DNA found in our skin cells.[19,20]

The constellations of stars in the night sky have fascinated humans over the eons, but it was only in the early 20th century that stars including our Sun were empirically categorised in order to understand them. Scientists began to plot stars based on their colour (Luminosity, L_x) and their surface temperature (T_{eff}) which was imaged as the Hertzsprung-Russel diagram.[21] The discovery of binary stars enabled the mass (M_x) of both stars to be measured relative to the surface temperature (T_{eff}). Since these star systems were effective calibration points, subsequent theoretical models were able to plot the stars according to their mass (M_x) and their age (t_x). Based on the observations of similar classes of star, computational simulations and radiometric studies of primitive meteorites known as carbonaceous chondrites; the sun is believed to have formed from the remnants of a supernova ca. 4.57 Ga and that the resulting dust and gas accreted to begin the formation of a proto planetary disk.[22,23] Our Sun evolved from a T Tauri star to a Main Sequence star (Type G) over the course of several millions of years and was able to contract and reach hot enough temperatures to enable fusion of hydrogen to helium. From theoretical studies and observations of other type G stars in the galaxy (the same class as the Sun), it is thought that the light emitted by the young Sun (ca. 4.0 Ga) was 30% less luminous than the present day and strongly emitted light in the mid to near UV wavelengths (< 300 nm).[24,25]

Considering the prevalence of UV irradiation on the early Earth,[26–29] photochemistry has the potential to unlock alternative paths towards prebiotically plausible molecules (Chapter 3 will explore the photochemistry of thioacids and thiols). UV light also has the capacity to destroy molecules and exert a selection pressure on molecules to perhaps provide a rational why nature 'selected' specific molecules found in extant biology.

1.2.2. Geological Environment

The early Earth or Gaia was formed approximately 4.56 billion years ago in the protoplanetary disk and coalesced from the impacts of proto planets and

accumulation of mass through gravity.[30,31] The Moon is thought to be created during the early Hadean period ca 4.54 Ga from the collision between a body approximately the size of Mars (Theia) and the early Earth.[32] The resulting debris from the impact accreted to form the Moon. The oldest known minerals discovered to date are the Jackson Hill zircons, which through oxygen isotope analysis gave an age of 4.4 ± 0.08 Ga.[33] During the Hadean period the high frequency of meteorite impacts, and a molten surface suggest conditions were not yet ripe for life. Water vapour is considered a major component of the atmosphere from the outset and geological ^{18}O isotopic evidence from the Jackson Hill zircons indicated that water condensed to form the first oceans on Earth ca. 4.35 Ga.[34] The early Earth would require conditions where a period of abiotic syntheses and then subsequent accumulation of essential organic molecules could occur. The general consensus is that towards the end of the late heavy bombardment ca. 4.0 - 3.6 Ga, as impacts from meteorites abated a plausible window emerged where the early Earth was stable enough to possibly support life (Figure 1).[35-36]
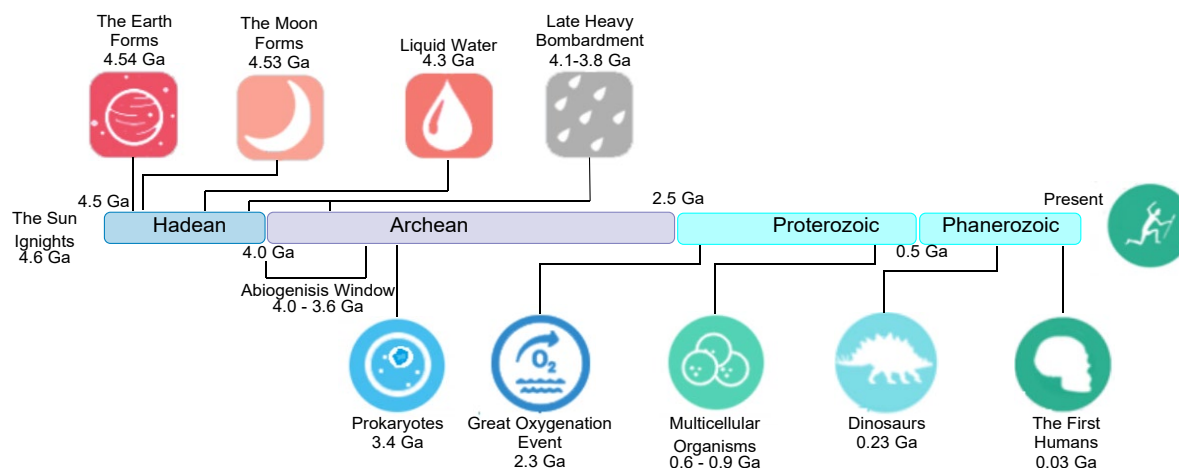
Figure 1. A chronological timeline of the geological and biological events on Earth.

1.2.3. Purple Haze - Chemical Feedstocks

Even though the atmosphere and the hydrosphere of Earth only represents 0.000088% and 0.022% of the total mass of the Earth (~ 6 x 10^{24} Kg), the emergence and evolution of life played out on this stage.[21]   The consensus is that in the beginning the atmosphere of the Earth was comprised mostly hydrogen and helium

accreted from the protoplanetary nebula. However, studies measuring the isotopic abundances of helium or xeon (^{4}He and ^{129}Xe gas are the product or by-products of fusion) found the relative abundances significantly lower than the Sun's.[21,37–39] This suggested that the primordial atmosphere was lost and replaced with a secondary atmosphere.

Initially Oparin postulated that the composition of the early atmosphere of the Earth was reducing and comprised mostly gases such as ammonia (NH_3), carbon dioxide (CO_2), hydrogen (H_2), methane (CH_4), nitrogen (N_2), and water (H_2O).[40,41] This primordial degassing is assumed to come from mostly volcanic activity which also produces reduced sulphur gases such as H_2S, COS and CS_2.[42,43] However, in recent years the reduced atmosphere model has come under scrutiny; theoretical modelling of the solar system, geological, moon or meteorite data indicate the Earth may have had a neutral atmosphere comprising mostly of nitrogen and carbon dioxide in a 80:20 ratio and traces of ammonia and methane.[29,44,45] A weakly reducing atmosphere comprising of mostly N_2, CO_2 with low levels of CO, NH_3, H_2 and CH_4 would attenuate incoming UV light (< 200 nm) from the Sun and create a greenhouse effect with mild temperatures (0-50 °C). For models with CO_2 dominated atmospheres, the average surface temperatures could potentially increase all the way up to 250 °C.[46,47] Today, oxygen (O_2) and therefore ozone (O_3) shields the Earth's surface from damaging UV radiation (O_3 absorbs 100-280 nm). However, Kasting predicted that substantial levels of O_2 were not present before the Great Oxygenation Event (2.4 - 2.0 Ga) and so an ozone layer could not have formed.[48] Hydrogen cyanide (HCN) is thought to be formed in the weakly reducing atmosphere from high energy processes such as lightning, UV irradiation, meteor impacts or volcanism. Sequestration of HCN in the atmosphere with ample ferrous (Fe^{2+}) found on Earth yields water soluble ferricyanide ($Fe(CN)_6^{-4}$) anion which upon wet-dry cycling can form concentrated cyanide salts ($^-$CN).[49–52] The innate energy stored within nitrile functional groups has been utilised by many scientists in the field and will be a core foundation for the work presented in this thesis.

It is generally accepted that even after attenuation in the atmosphere, the early Earth would have still been subjected to intense UV radiation. As the Earth began to cool in ~ 4.3 Ga, water condensed forming rivers, oceans and lakes within a range of pH 5 to pH 9. Water is essential for extant life and abundant on the Earth. Therefore, water can be considered an excellent solvent to conduct prebiotic chemical reactions. Some scientists consider water as a 'poison' for abiotic reactions, and whilst water as a solvent may have issues such as causing unwanted reactivity (e.g., hydrolysis) reactions in the absence of water e.g., dry, wet-dry cycles, formamide also have significant challenges. Water can also affect photochemical reactions by attenuating incoming UV light.[34] Ranjan *et al* measured the UV absorption of water at the surface and at a depth of 1 m (Figure 2) and calculated that the early Earth could have been subjected to a broad range of 224 - 350 nm of UV light.[53,54] It can be deduced that during the abiogenesis window, the chemical environment could be considered anoxic, aqueous, pH 5-9 and subjected to UV radiation > 224 nm up to a depth of 1 m in water.
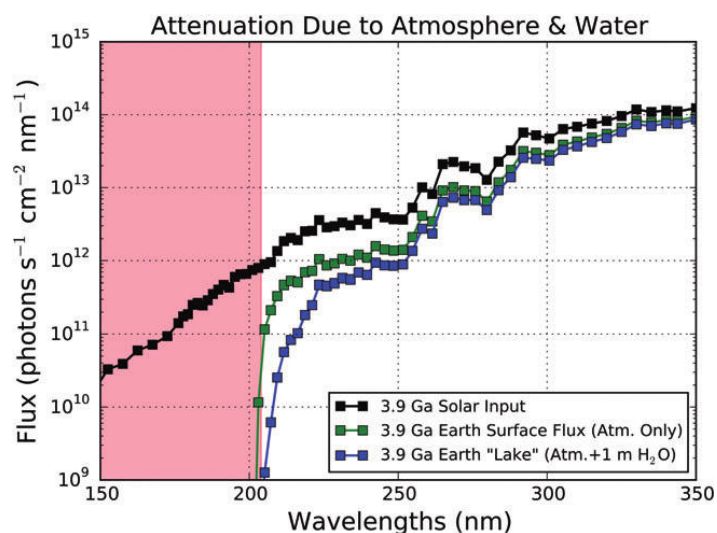
Figure 2. A comparison of the UV flux attenuation of the Sun ca 3.9 Ga in vacuum, 1 μm and 1 m of H_2O.

The heterotrophic theory of the origin of life, proposed by Oparin and Haldane envisaged a 'prebiotic soup' where the essential molecules of life are believed to have originated.[55] The environmental conditions and accumulation of organic molecules that could have precluded 'life' could have occurred in many places on

the early Earth. For example, reactions could have occurred in rivers, shallow lagoons, lakes, oceans, volcanos, hydrothermal vents, glaciers or in the atmosphere. Although the exact composition of the atmosphere of the early Earth is difficult to accurately define and open to interpretation, the identification and availability of possible feedstock molecules can be inferred from many varied sources such as geochemistry, interstellar dust clouds,[56] celestial bodies,[56] carbonaceous chondrites (Murchison meteorite)[57] and lab experiment.[58,59] Many planets and moons such as Titan or Enceladus may also give clues to what chemical feedstocks were possibly present on the early Earth.[60] A comprehensive list of 'plausible' species has been extensively discussed and some are depicted in figure 3.[61]
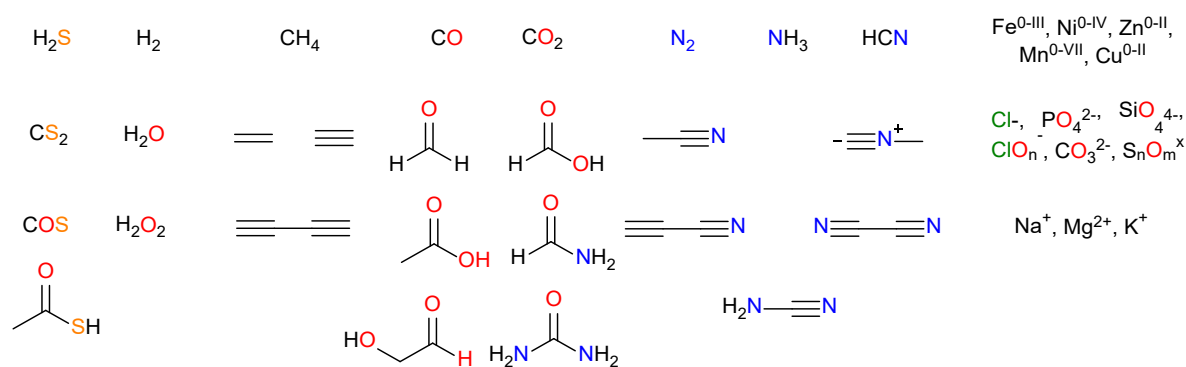
Figure 3. Commonly accepted prebiotic chemical feedstocks. The list is not exhaustive.

Although there is much dispute about what is prebiotically 'plausible', this pool of chemical species is a good representation of simple and acceptable chemical feedstocks in the field. If this is the case, what are the requirements for a successful prebiotic synthesis? Orgel stipulated a set of loose guidelines that are generally accepted in the field.[62] The first is that the synthesis must be plausible, at least to the proposers of a prebiotic synthesis; secondly that the reagents should be present in sufficient quantities at the location of the synthesis; the reactions must occur in water; and finally, the yield must be 'significant' at least to the proposers of the synthesis.

1.3. Theoretical Approaches Towards the Origins of Life

Currently, scientists in the field take two approaches when tackling this complex and difficult problem. The first is a top-down approach which uses current biology to work backwards down to the molecular level and can infer the plausible systems at play during abiogenesis. The second is a chemistry up process, this strategy utilises plausible feedstocks and mild chemistry to synthesise the basic building blocks of life and works upwards (in complexity) to arrive at a chemical system that has the hallmarks of life such as replication, storage of information, metabolism of nutrients, etc. In the field a combination of these approaches has led to significant progress through conducive and thorough investigation.

1.3.1. Biological Approach

Through diligent observation Darwin was able to propose a theory of evolution by natural selection and this presented an elegant and logical system of explaining the development of complex life throughout natural history. Using life as we know it today, biologists investigating the origins of life can use phylogeny to trace backwards in time through the evolutionary pathways designated by fossil and genetic information to what is believed to be the last universal common ancestor (LUCA).[63] However, the oldest known fossils trace back to approximately 3.4 billion years and was a group of prokaryote organisms which thought to be significantly more complicated than LUCA (Figure 4).[11,64] Despite the gap in the fossil record, LUCA is important because it represents the point where there ceases to be variations between species. Past this point there is an information event horizon for comparative biology and so LUCA's biological systems are the closest point where we can infer plausible chemical origins to life.

Unfortunately, LUCA is still very intricate organism and is only a representation of the biological commonality found in all species today, indeed LUCA could have comprised of different populations contesting evolutionary space. LUCA can be considered the simplest reduced form a cell can take. Despite this, LUCA would have a fully formed translational system, including the world most complex molecular

machine (the ribosome; see below section 1.3.3).[9,65–67] Despite this issue, studies on the various species indicate that they all bear close biochemical similarities to each other.[68,69] Such signature properties include the universal reliance on nucleic acids (RNA & DNA) for propagation of genetic information, a single near-universal genetic code for protein synthesis and the presence of the same ~20 amino acids in proteins.[69] From this, one can infer a set of functions that are essential to life and that would be present in LUCA.
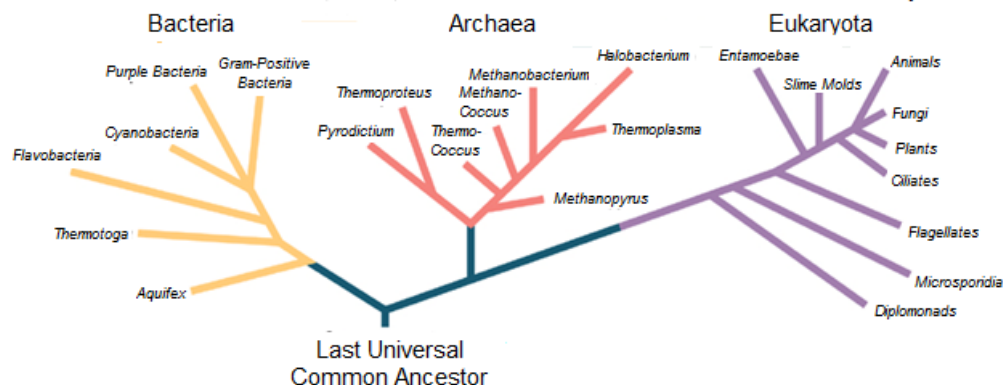
**Figure 4.** Phylogenetic tree of life adapted from Woese.[38]

These functions are the capacity to store, edit and transmit information. To release, store and transfer energy through metabolic processes. The ability to catalyse useful chemical reactions. And finally, to compartmentalise all these processes with a membrane so that the inside of the cell is separated from the environment. This is a significant factor, since compartmentalisation enables a cell to accrue the advantage of evolutionary change.[70] The complexity of the network of systems found in a proto cell resulted in many scientists focusing on one aspect of these, such as the genome or catalysis as being the sole progeny to the origin of life. Instead, a holistic approach is to consider all the components that comprise a cell (Figure 5).[71] This is important to consider, since feedstocks are precursors to a myriad of prebiotic molecules which are intrinsically interlinked in the synthesis of biological compounds such as sugars, lipids, nucleobases, and amino acids. The functionality of many these biological compounds is co-dependent, for example RNA codes for the sequence of amino acids found in peptides.
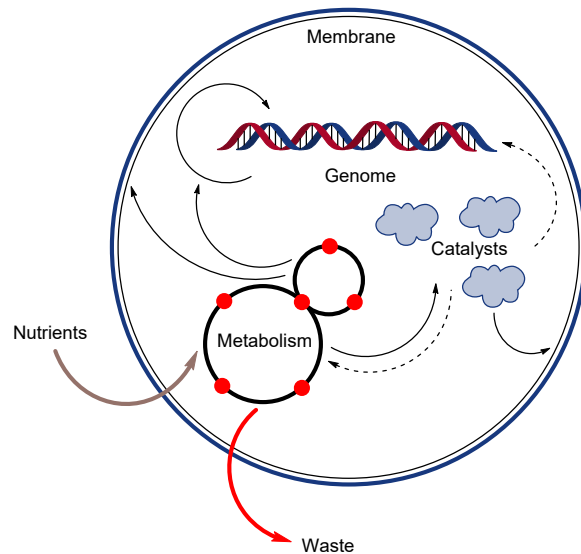
Figure 5. Schematic of a basic protocell depicting a simplified network of systems that are essential to life.

1.3.2. Central Dogma of Molecular Biology

Although the theory of evolution was well established at the turn of the last century, there was a lack of knowledge at the molecular level on how mutations could then give rise to natural selection. The discovery of the structure of DNA by Crick, Franklin and Watson showed that genetic information is stored in the genome sequences of DNA and is preserved during replication of the genome.[10,72] In 1958, Crick proposed the 'Central Dogma of Molecular Biology' which described the relationship between DNA, RNA and proteins (Figure 6).[73]
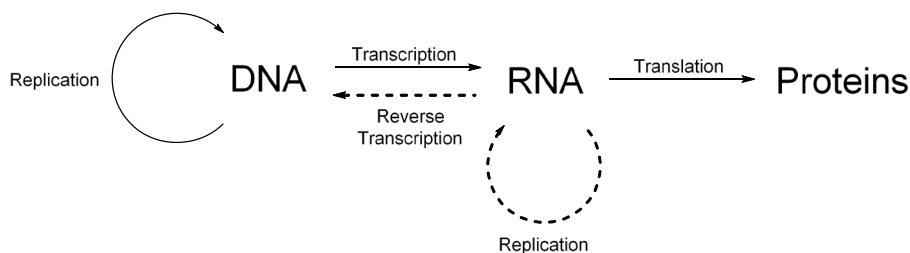
Figure 6. Central Dogma of Molecular Biology. The dashed arrows represent the recently discovered RNA replication from RNA ribozymes and subsequently added to the central dogma.

DNA acting a genetic store of information is transcribed into mRNA by RNA polymerase from which information on the mRNA can be translated by the ribosome into proteins which carry out functional tasks in all extant organisms. This doctrine was reinforced by Crick in 1970 with the discovery of the genetic code which showed how RNA acts as the 'middleman' between a genotype (DNA) which is then expressed as a phenotype (proteins).[74] The genetic information flows downhill from DNA, however there is currently no known transfers uphill back towards DNA.

One of the key features of the central dogma of molecular biology is the RNA codon and the triplet system which codes for proteinogenic amino acids. Barnett, Brenner, Crick, Watts-Tobin found that the proteinogenic amino acids were coded for by a set of three base pairs which was defined as a codon.[75] By accounting for the four different base pairs in RNA (A, C, G and U) and use of three bases to code for amino acids, it was calculated by Gamow that there would be 4^3 = 64 combinations of different base codes for amino acids. This prompted a race to decipher the code between the scientists Nirenberg and Ochoa who were able to decode 54 of 64 codons.[76,77] Khorana was able to validate the results and completed identification of the rest of the genetic code in 1968.[78] Remarkably, the degeneracy of the code implies that only the 20 proteinogenic amino acids are coded for by a triplet system of RNA base pairs in all biological systems (Table 1).

The complexity of the biological system has led some to suggest that it is unlikely that all these the systems simultaneously formed. This led scientists to question what could have preceded the central dogma and whether DNA or proteins may have formed first. A chicken and egg paradox ensued since DNA and RNA are needed for proteins to be formed and proteins are required for RNA translation, DNA replication and for DNA to be transcribed.

Table 1. Triplet base codons of RNA that corresponding to the 23 encoded proteinogenic amino acids. Many amino acids have degenerate codes, whereas methionine and tryptophan (blue) only have one triplet codon each and there are only three stop codons (red). Lower degeneracy is assumed be a result of amino acids that are considered late additions to the genetic code.

Second Letter

First Letter		U	C	A	G	Third Letter
U		Phe	Ser	Tyr	Cys	U
		Phe	Ser	Tyr	Cys	C
		Leu	Ser	**Stop**	**Stop**	A
		Leu	Ser	**Stop**	**Trp**	G
C		Leu	Pro	His	Arg	U
		Leu	Pro	His	Arg	C
		Leu	Pro	Gln	Arg	A
		Leu	Pro	Gln	Arg	G
A		Ile	Thr	Asn	Ser	U
		Ile	Thr	Asn	Ser	C
		Ile	Thr	Lys	Arg	A
		Met	Thr	Lys	Arg	G
G		Val	Ala	Asp	Gly	U
		Val	Ala	Asp	Gly	C
		Val	Ala	Glu	Gly	A
		Val	Ala	Glu	Gly	G

1.3.3. The 'RNA World' Hypothesis

To overcome the problem of whether information or catalysis came first, Crick, Orgel and Woese both proposed that there might be a biopolymer that is able to have both genotypic and phenotypic properties.[79,80] Palade observed the ribosome in 1955 and not long after, Cech and Altman discovered ribozymes which are catalytic strands of RNA. Further investigation of ribosomes by Ramakrishan, Steiz and Yonath resulted in concrete determination of the structure of the prokaryote ribosome with X-ray crystallography (Figure 7).[45,81,82]
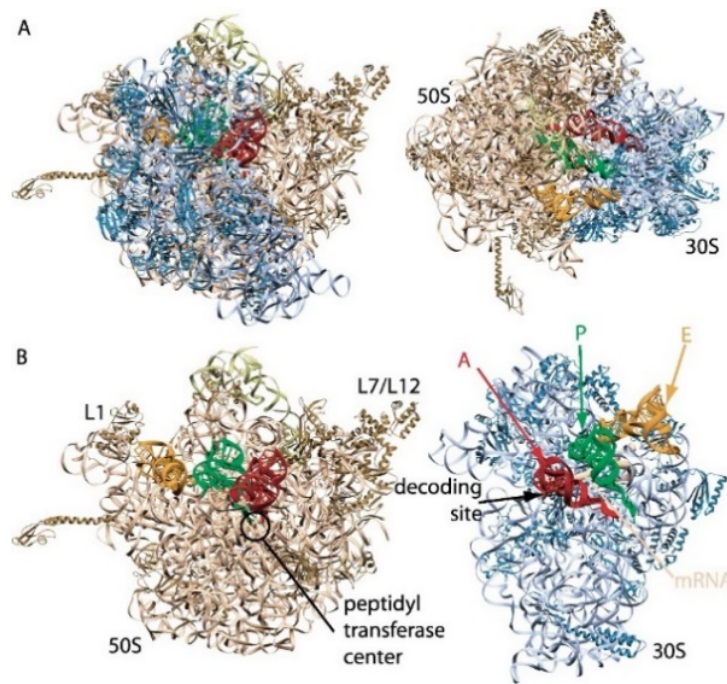
Figure 7. Crystal structure of prokaryotic ribosome.[65–67,82] (**A.**) Two views of the 70S ribosome complexed with mRNA and tRNA (top left image); view from the 30S side (top right image); (**B.**) Expanded view of the 50S (bottom left image) and the 30S (bottom right image) subunits in the 70S ribosome, showing the location of the A-, P-, and E- site RNAs.

The bacterial (70S) ribosome consists of a small (30S) and a large (50S) subunit. The 30S subunit consists of about 20 different proteins and a sequence of ribosomal RNA (rRNA) containing about 1600 nucleotides. The 50S subunit consists of about 33 different proteins and a rRNA sequence with about 3000 nucleotides. The catalytic site for peptide formation was comprised solely of RNA residues, with the nearest amino acids almost 20 Å away. Ribosomes from eukaryotes are larger and more complex than those from prokaryotes, and as far as we know ribosomes from all three kingdoms of life function according to the same principles.[82,83,84]

The idea that RNA could be a progeny of modern biology gained traction because RNA displays both genotypic (information storage) and phenotypic (catalytic) qualities. The 'RNA world' hypothesis, which was first coined by Gilbert in 1986 envisages the formation of RNA oligomers from the prebiotic milieu which would evolve over time to incorporate catalytic functions with the aid of co-factors. Many co-factors have non-functional nucleobases imbedded within their structure and these nucleotides are considered molecular relics from an earlier period that was not

evolved out of biology. Co-factors such as CoA, NADH and ATP contain RNA adenosine nucleotide in their molecular structure (Figure 8).[85] Eventually peptides/proteins synthesised from ribozymes would eventually supersede ribozymes due to the superior catalytic properties of enzymes.
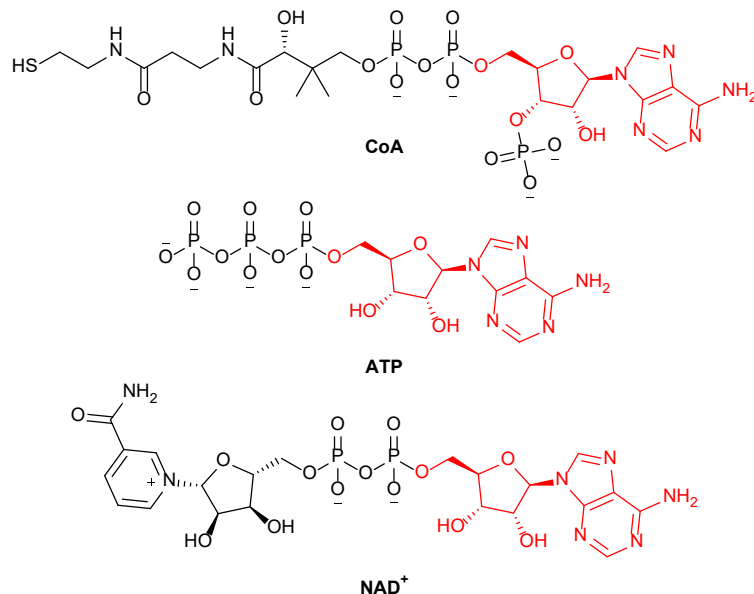
Figure 8. Cofactors containing nucleotides in red, coenzyme A (CoA), adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide (NAD$^+$).

The 'RNA world' hypothesis is an elegant and attractive theory; however, proponents of this theory often neglect the roles other biopolymers have on life. Some challenging hurdles such as the non-enzymatic synthesis and polymerisation of nucleotides; the replication of these polymers and their incorporation into a self-sustaining system need to be resolved.

1.4. Geology Up Approach

At the dawn of organic chemistry scientists such as Jöns Jacob Berzelius started to characterise chemical compounds based on whether they required living organisms to produce (organic) or not (inorganic).[86] Over the course of the 19th century this came to be known as the vitalism hypothesis where it was believed that only living organisms could only make the chemical components of living systems. Vitalism implied that living organisms contain some vital force that is absent from inanimate matter, therefore organic substances could only be made by living things. It seemed that prebiotic chemistry lay far beyond the means of these early pioneers in chemistry. In 1828 Berzelius's student Wöhler accidently discovered the synthesis of urea by heating ammonium chloride and silver cyanate in water.[8,86] Wöhler is considered the father of organic chemistry, and his discovery dispelled the myth that organic molecules could not be derived from inanimate matter. This spelled the beginning of chemists looking to seek the nature of matter and indirectly finding the chemical components of life.[55,87]

1.5. Amino Acids

1.5.1. Amino Acid Synthesis

Not long after Wöhler's synthesis of urea, a synthesis of amino acids was discovered in 1850 by Adolph Strecker.[88] Strecker found that α-aminonitriles could be formed from the condensation of either an aldehyde or a ketone with ammonia to form an imine. Cyanide attacks the imine to give the desired α-aminonitrile **AA**-CN in alkaline conditions. Subsequent acid hydrolysis of **AA**-CN yields racemic α-aminoamide **AA**-NH$_2$, then further hydrolysis to α-amino acid **AA**-OH (Scheme 1). The reaction of cyanide with aldehyde is faster than the addition of ammonia and formation of imine. The formation of cyanohydrin **1** is reversible back to the aldehyde or ketone and the equilibrium favours the formation of α-aminonitriles especially above pH 8.[89,88] The rate of hydrolysis of cyanohydrin **1** is significantly lower than the hydrolysis of the α-aminonitrile ($k_1 \gg k_2$) enabling the liberation and accumulation of amino acids.
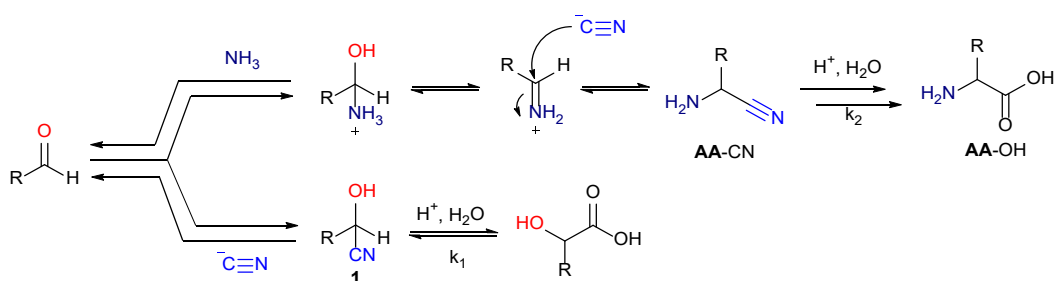
Scheme 1. The mechanism of the Strecker reaction.

1.5.2. The Miller-Urey Experiment

One of the first and most famous examples of prebiotic chemistry was performed by Miller in 1953.[90] The Miller-Urey Experiment demonstrated that some of the basic molecules of life could be formed from simple feedstocks. Miller was able to show that when an electric discharge was initiated in a reaction vessel containing reducing gases which were thought to be present in the early atmosphere, amino acids formed by Strecker type reactions. In the resulting soup which formed after several days; basic amino acids such as glycine, alanine, β-alanine, aspartic acid, and α-aminobutyric acid were found by chromatography (Figure 9).[90,91,92] Throughout his career Miller experimented with different combinations of gases to simulate different early Earth atmospheres. For example, he used H_2S and/or COS in addition to the original gaseous mixture to simulate the atmosphere close to active geology (volcanos, geysers, etc) where he detected many more amino acids as well as sulphur derived amino acids.[58] Johnson *et al* revisited the Miller-Urey experiment and re-analysed old samples from the original electric-discharge experiments performed by Miller. With modern analytical techniques the group was able to find significantly more amino acids, unfortunately not all these amino acids were the proteinogenic amino acids found in extant biology today.[59] The extra 'clutter' of non-proteinogenic acids found in these reactions, begs the question of why only 20 proteinogenic acids are found in biology today.
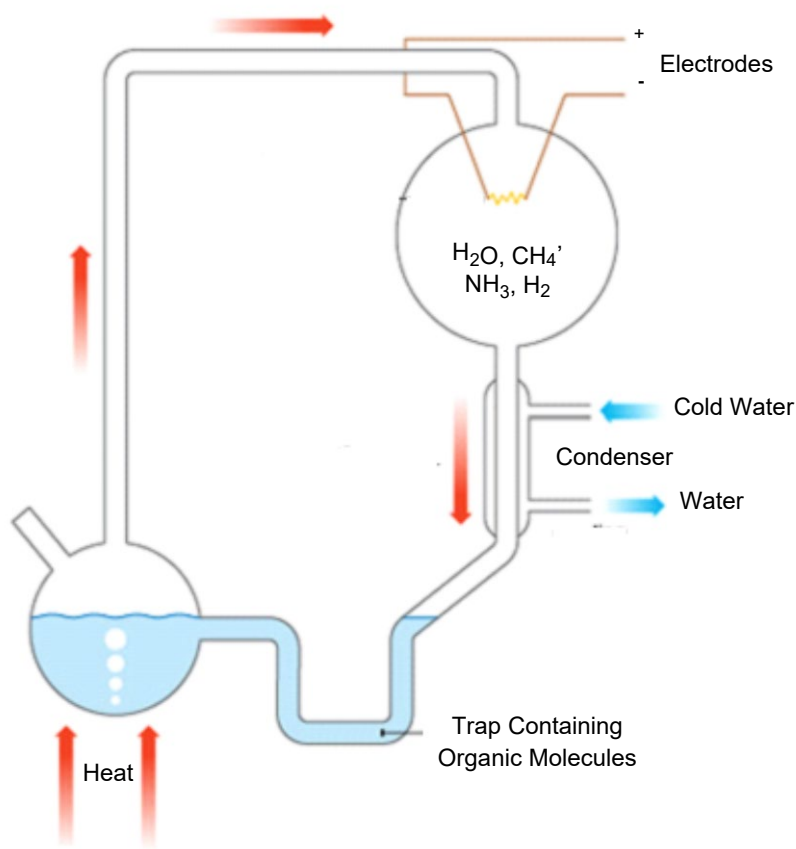
Figure 9. Diagram depicting the synthesis of the plethora of amino acids that form from a reducing atmosphere under electrical spark discharges in the famous Miller-Urey experiment.

1.5.3. Amino Nitrile Hydrolysis

There are other routes to amino acids such as the Bücherer-Bergs reaction or the Commeyras' ketone assisted hydrolysis. These procedures result in the hydrolysis of α-aminonitriles **AA**-CN to amino acids **AA**-OH. For example, the Bücherer-Bergs utilises the reaction of amino acid **AA**-CN with CO_2 to form carbamic acid derivative **2** which can undergo intramolecular cyclisation to hydantoin **HD**. The hydantoin **HD**, then hydrolyses to the *N*-carbamoyl amino acid (**CAA**-OH).[93] However, there are significant problems since the Bücherer-Bergs only works for free α-aminonitriles and hydrolysis to the amino acid from both hydantoin **HD** and **CAA**-OH is very slow.[94] An alternative route to amino acids from aldehydes involves the phosphoro-Strecker. In an analogous fashion to the Strecker reaction diamidophosphate **DAP** condenses with an aldehyde to form a phosphoroimine, which is attacked by cyanide

to form *N*-phosphoroaminonitrile **3**.[95,96] The advantage of this method is that the reaction takes place at pH 7 and is specific for aldehydes.   The reaction with ketones is sluggish thus avoiding non-proteinogenic α,α-disubstituted amino acids.   It is possible that thioamides formed from the addition of H_2S to the phosphoaminonitriles may have been involved in prebiotic peptide formation.[95]
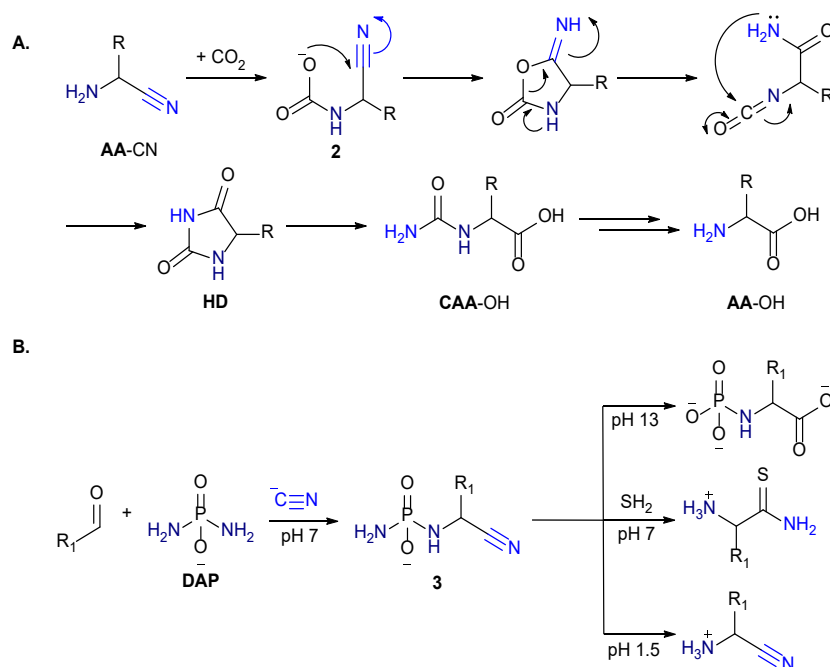
Scheme 2. (**A.**) The Bücherer-Bergs reaction; (**B.**) Phosphoro-Strecker synthesis.

1.5.4. Synthesis of Amino Acids with α-Ketoacids *via* Transamination

In metabolism transamination is the process where the amino group of amino acids is transferred to a α-ketoacid with the support of a pyridoxal 5'-phosphate (PLP) co-factor in the transaminase enzyme. A Schiff base is formed by the linkage of a ε-amino lysine residue to PLP. The amino acid displaces the ε-amino lysine residue and the PLP acts as an electron sink as the amino acid is reduced to a α-ketoacid. The amino acid acts like a reservoir of ammonia which can be transferred to α-ketoacids formed in the r-TCA cycle. The PLP now acts as an amine carrier to another α-ketoacid which is subsequently aminated into an amino acid after hydrolysis (Scheme 3). All amino acids except lysine, threonine and proline can undergo transamination.

46
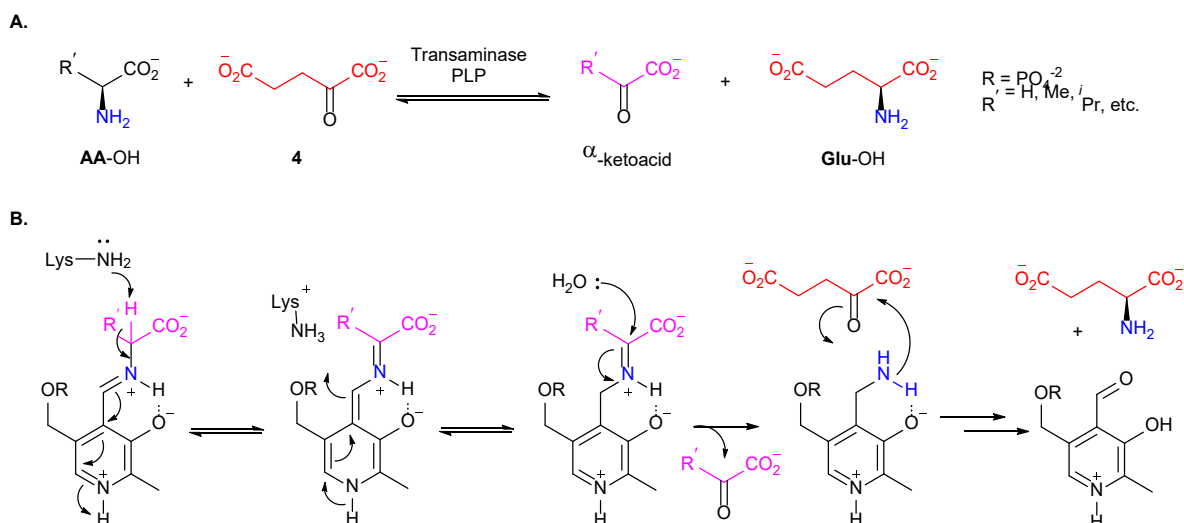
Scheme 3. (**A.**) General reaction for transamidase catalysed transamination between an amino acid and α-ketoglutamate **4** to form glutamic acid **Glu**-OH and α-ketoacid; (**B.**) The mechanism for pyridoxal 5'-phosphate (PLP) mediated transamination.

In the 1950s Snell, discovered vitamin B_6 (PLP) and set about understand its role in biochemistry. In a set of non-enzymatic reactions,[97–99] Snell was able to replicate transamination in water. In the first set of reactions Snell was able to deaminate glycine, threonine or allothreonine with PLP and Al^{+3} in water.[99] Again, under similar conditions glutamate was transamidated to α-ketoglutarate with a variety of metal salts (Cu^{+2}, Fe^{+2}, Ni^{+2}, Al^{+3} etc).[98] Inspired by the work by Snell and enzymatic transaminations with PLP cofactors, Moran, Springsteen and Krishnamurthy recently re-investigated the same processes in order to improve the yields of these reactions and uncover mechanistic insights with modern techniques and instrumentation.

Moran *et al*. were able to confirm the formation of glycine **Gly**-OH (63%), aspartate **Asp**-OH (27%), alanine **Ala**-OH (2%) and glutamate **Glu**-OH (9%) with an iron (Fe^0) catalyst and ammonium hydroxylamine as an amine source at 70 °C from the r-TCR products glyoxylate **5**, oxaloacetic acid **6**, pyruvate **7** and α-ketoglutamic acid **4** respectively (Scheme 4A).[100] Most recently, Mayer *et al*. was able to observe the effect of different metal cations on the transamination reaction of glyoxylate **5** with glutamine **Gln**-OH as an amine source to form glycine **Gly**-OH.[101] The same reactions were tried with pyruvate **7** to yield alanine **Ala**-OH, however, only Cu^{2+} gave good yields (up to 67%) which was also observed by Snell. Most of the other

metals tested only yielded < 10% alanine **Ala**-OH (Scheme 4B).[101] Springsteen and Krishnamurthy were also able to form alanine **Ala**-OH in 19% yield over one day and glutamic acid **Glu**-OH in 44% yield after 4 hours under mild conditions at pH 5 using the mineral alunite (AlK(SO$_4$)$_2$ (Scheme 4C).[102]

Although the formation of amino acids from transamination of α-ketoacids formed in a presumably protometabolic r-TCA cycle gives good yields and is an attractive proposition, from the 'metabolism first' camp. The process relies on an external amine source, the yields do appear to flow any trend or indicate a satisfactory rational for amino acid selectivity and some of the metals used may not be considered 'plausibly' prebiotic.
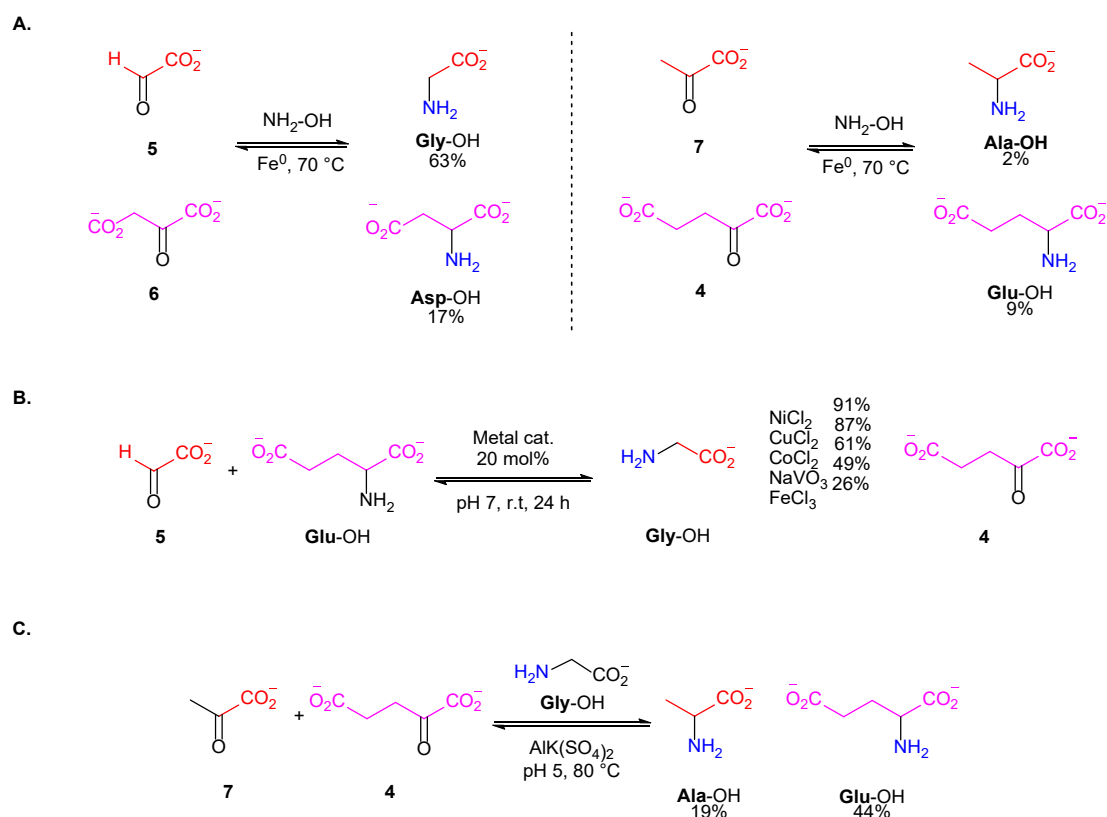
Scheme 4. (**A.**) Muchowaska *et al*. on the conversion of α-ketoacids to amino acids with reduced iron (Fe^0) and hydroxylamine (NH$_2$OH);[100] (**B.**) Moran's metal ion catalysed transamination of glutamate **Glu**-OH to glycine **Gly**-OH with glyoxylate **5**;[101] (**C.**) Springsteen and Krishnamurthy's metal free transamination reaction of pyruvate **7** and α-ketoglutamic acid **4** with glycine **Gly**-OH as the ammonia source.[102]

1.6. Nucleotides

1.6.1. Prebiotic RNA Synthesis

The RNA world (Section 1.3.3) attracted much interest from chemists and biochemists who then endeavoured to synthesise ribonucleotides to demonstrate that the 'RNA world' hypothesis was feasible. The classical retrosynthetic analysis of nucleotides split the RNA monomer into the nucleobase, sugar and inorganic phosphate components.[103] However, the disconnection approach highlights the difficulty of nucleotide synthesis. Firstly, the sugar must be ribose of the correct stereochemistry. Secondly, the four canonical nucleobases must attach in the correct position on the nucleobase with the correct anomeric configuration. On top of this the RNA monomer would also have to specifically oligomerise on the 3'-5' positions of the ribose sugar.
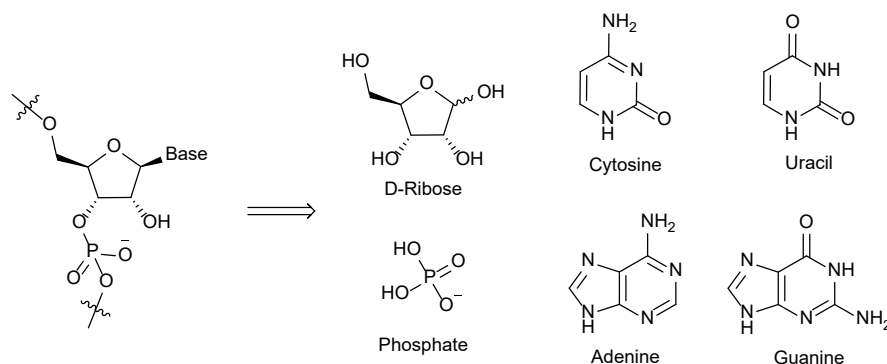
Figure 10. Classical retrosynthetic analysis of RNA into its component parts.

1.6.2. Synthesis of Sugars

The Formose reaction was discovered by Butlerow in 1861[104] and highlighted the importance of formaldehyde as the possible prebiotic source for the ribose sugar in RNA. The formation of sugars proceeds through uncontrolled successive aldol type reactions from formaldehyde **8** heated at 60-80 °C with a metal salt such as calcium hydroxide at pH 10-11.[104,105] The full mechanism for the Formose reaction is not clear, however Breslow showed that the first step proceeded from a metal assisted umpolung reaction with formaldehyde **8** to form glycolaldehyde **9**. Glycolaldehyde **9**

49

is a useful intermediate for nucleotide and serine synthesis, however **9** enolises at a fast rate under basic conditions and quickly forms C^3 and higher order sugars. It was also found that impurities in formaldehyde, such as glycolaldehyde **9**, are required to instigate the dimerization of formaldehyde and the successive uncontrolled homologation.
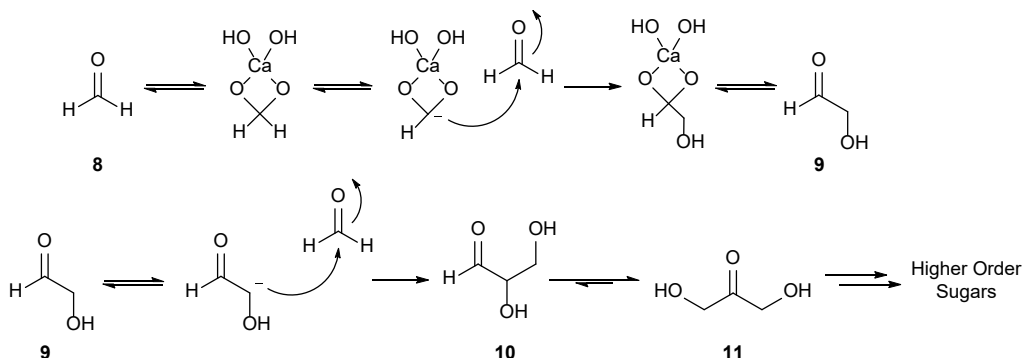

Scheme 5. A simplified depiction of the Formose reaction, with the proposed mechanism which proceeds *via* the aldol cross-coupling of formaldehyde **8**. This reactivity enabled by metal assisted reversing of the polarity 'umpolung' of the carbonyl functional group of formaldehyde **8**.

Unfortunately, the Formose reaction is not selective and uncontrolled homologation results in a brown tar with a plethora of different sugars. Decker measured the ribose content of the Formose reaction and found in that ribose only comprised less than 1% of the total mixture.[106] Another problem is that in aqueous medium, ribose can interchange rapidly between its pyranose, furanose and open chain configurations. This lack of selectivity for α-*D*-ribose in the Formose reaction and its relative instability has posed a major conundrum in its synthesis. Benner was able to use borate minerals to preferentially synthesise pentoses from glycolaldehyde and glyceraldehyde, through the stabilisation of *cis*-diols.[107] However ribose was not selectively preferred over the other sugars, indeed it was the minor observed isomer.[108] An alternative route proposed by Sutherland, utilises UV light, cyanide[28,109,110] or sulfide (hydrogen sulfide H_2S or bisulfite HSO_3^-) reductants, and Cu (I) or Cu (II) as a catalyst in a photoredox system to synthesise the basic building blocks of the sugars glycolaldehyde **9**, glyceraldehyde **10** and dihydroxyacetone **11** from HCN.[109] The cyanosulfidic homologation process from cyanohydrins after H_2S reduction could in principle unlock higher order sugars through sequential

homologations.[109] Powner, Sutherland and co-workers developed a new synthesis of ribonucleotides through the reaction of glycolaldehyde **9** and cyanamide **22** to form 2-aminooxazole **2AO**. Then the reaction of **2AO** with glyceraldehyde **10** to form ribose and arabinoside (Section 1.6.4).[111] This was significant development since this procedure formed the correct ribose sugar and incorporated the glycosidic bond in good yields. Despite much progress, there is still not a high yielding and selective synthesis of stereochemically pure ribose sugars.

1.6.3. HCN Oligomerisation and Nucleobase Syntheses

In 1960 Oró was able show that the canonical nucleobase adenine **12** could be formed in poor yield (0.5%) by heating HCN at 70°C over a few days in the presence of ammonia.[112]   Orgel showed that diaminomaleonitrile **DAMN** was formed in concentrated HCN solutions to give a steady state yield of 15%.[113] When a solution of tetramer **DAMN** is exposed to UV light at 350 nm, photo-isomerisation formed 4-amino-imidazole-5-carbonnitrile **AICN** in an excellent 80% yield. The reaction of **AICN** with HCN results in a very poor yield of adenine **12** (11%, < 0.5% directly from HCN).[114]   Guanine **13** was also formed from **AICN** in a good yield (60%) from exposing tetramer **AICA** to cyanate **14** (Scheme 6).[115]   Despite this achievement, many other HCN homologation species were found.
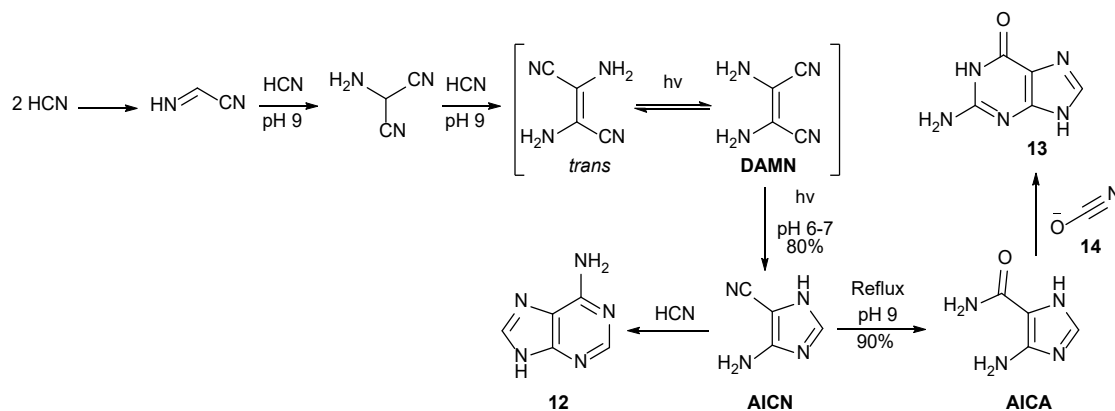
Scheme 6.   Tetramerisation of HCN to purines adenine **12** or guanine **13** from the photo-isomerization of **DAMN** to **AICN**.

The syntheses of pyrimidines have proven a challenging synthetic target and the attempts discussed here have resulted low selectivity and poor yields. In 1966 Orgel was able to synthesise the canonical pyrimidine cytosine **18** by heating cyanoacetylene **15** with potassium cyanate **14** at 100 °C over 24 h. Unfortunately, the reaction conditions are harsh, and the total yield was low (7.6%). The low yield could be attributed to the hydrolysis of cyanate **14** faster than it could react with **15**.[116] The yield of cytosine **18** can be improved up to 50% with the treatment of cyanoacetaldehyde **16** and urea **17** at 100 °C. The hydrolysis of cytosine **18** is also led to the formation of uracil **19**.[117]
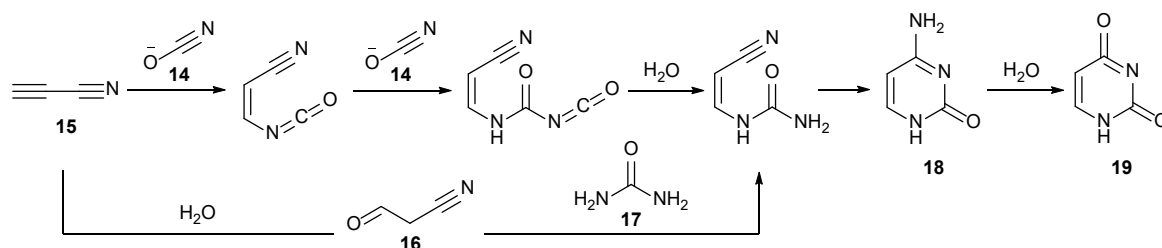
Scheme 7. Orgel's syntheses of pyrimidines cytosine **18** and uracil **19** from cyanoacetylene **15**.

Attempts at the glycosidation of the ribose sugar with nucleobases is another significant hurdle when following the classical retrosynthetic strategy towards the synthesis of RNA (Section 1.6.1). Originally, Fuller, Sanchez and Orgel have tried to form nucleosides directly by trying to heat adenine **12**, guanine **13**, cytosine **18** or inosine with pure *D*-ribose **20** at high temperatures.[118]
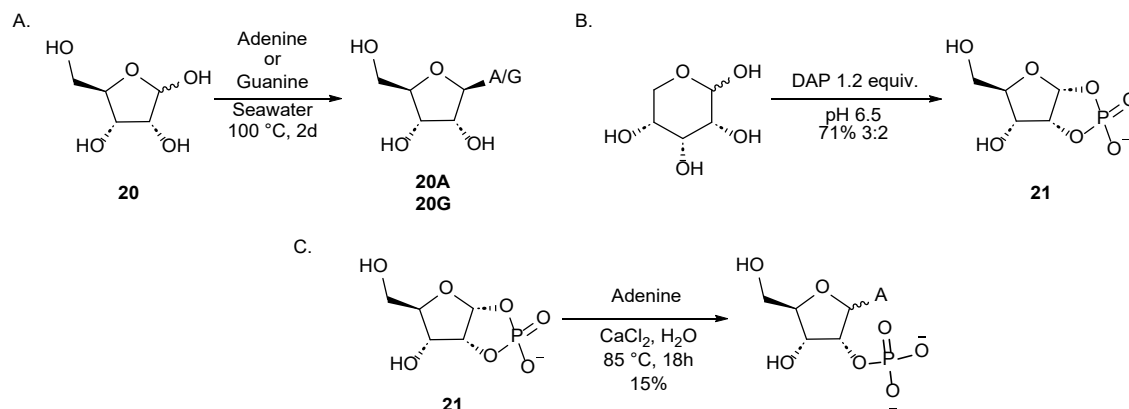
Scheme 8. (**A.**) Fuller's attempts of the glycosidation of ribose **20** with either adenine **12** or guanine **13** in dry-down experiments;[118] (**B.**) Eschenmoser's 1',2'-cyclicphosphate synthesis;[119] (**C.**) Kim and Benner's glycosylation of 1',2'-cyclicphosphate **21** with adenine **12**.

52

Various conditions were attempted, such as drying down the reaction at 100 °C or by using MgCl$_2$ or by trying prebiotic activating agents such as cyanamide **22**. Unfortunately, the yields were poor (β-adenosine **20/A** 4%, β-guanosine **20/G** =9%)) or in the case of the pyrimidines, cytosine, and uracil no reaction occurred at all (Scheme 8A).[118] Kim and Benner also tried direct glycosylation with Eschenmoser's 1,2-cyclic phosphate sugar **21** and adenine **12** to limited success (Scheme 8C).[119,120]

1.6.4. Synthesis of Pyrimidine Nucleotides

The failure of direct glycosidation of ribose and nucleobase led Orgel and Sanchez to look at mixed nitrogenous and oxygenous chemistry to overcome the low yields and poor selectivity (Scheme 8). [121] They found the reaction of cyanamide **22** with ribose **20** and ammonia, yielded ribofuranosyl aminooxazoline **23** 87% after recrystallisation. Heating **23** with cyanoacetylene **15** at 60 °C in basic conditions (pH 8.5) gave anhydrocytidine intermediate **24**. The subsequent hydrolysis of **24** to α-ribofuranosyl cytosine α-**20C** was achieved in a good yield of 64% over two steps. This procedure avoided the troublesome glycosidation step by building the cytidine stepwise onto the ribose skeleton. However, the non-canonical α anomer is formed instead and photo-anomerisation only resulted in 4% of the correct β-ribofuranosyl cytosine β-**20C** being formed (Scheme 9).[121] Although this route provided an elegant solution to glycosylation, the reaction still required enantiomerically pure ribose starting material and this procedure assumed that this synthetically challenging sugar which is produced in low yields (< 4%) in the Formose reaction (Section 1.6.2) could come from another source.
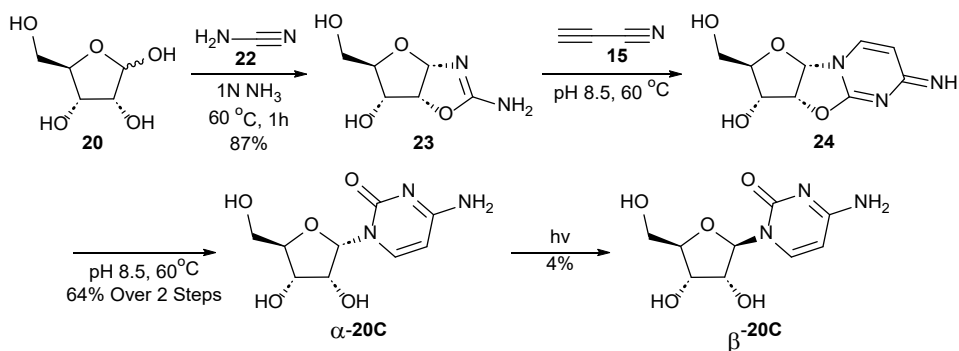
Scheme 9. Orgel's synthesis of β-ribofuranosyl cytosine β-**20C** *via* aminooxazoline **23**.

Nagyvary was able to find a way to the stereochemically correct β-ribofuranosyl from arabinose anhydrocytidine-3'-phosphate **25** (Scheme 10). Between pH 6 and 8 the C3' phosphate of **25** undergoes intramolecular cyclisation and the subsequent S_N2 attack inverts the stereochemistry to give β-cytidine-2',3'-cyclicphosphate **26C** as the desired biological anomer in high yields.[122] However, at higher pH's hydrolysis competes with cyclisation to give the arabinose-3'-phosphate sideproduct.
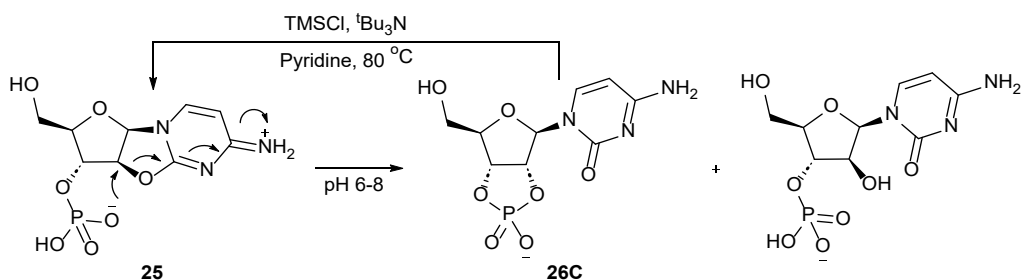

Scheme 10. Nagyvary's ring opening of anydrocytidine riboside **25** to the canonical β-riboside anomer **26C**.

Powner and co-workers were able to build on the advances of Orgel and Sanchez to give the most convincing prebiotic pyrimidine synthesis to date (Scheme 11). The difficulty of *D*-ribose synthesis (Section 1.6.2) was overcome by utilising phosphate as an acid/base catalyst to condense cyanamide **22** with glycolaldehyde **9** at pH 7 to form 2-aminooxazole **2AO** (> 80%, 75% isolated). A clever solution to pyrimidine glycosylation involved installing the pentofuranosyl backbone through nucleophilic attack of **2AO** with glyceraldehyde **10** to give preferential yields of ribo-**23** (25%) and ara-**23** (15%). Despite the mixture of stereoisomers, serial crystallisation resulted in the isolation of pure ribo-**23** and the resulting supernatant was enriched with ara-**23**. The ara-**23** remaining in solution was reacted with cyanoacetylene **15** to yield arabino-anhydrocytidine ara-**24**. Then the solution of ara-**24** and urea **16** (or formamide) diluted in phosphate buffer was heated until dry to give phosphorylated cytidine **26C** (63%). UV irradiation converted the cytidine **26C** to uridine **26U** (42%) and prolonged irradiation destroyed any unwanted by-products.[123,124,125,126,111,127]

Although this work is one of the most convincing prebiotic syntheses of pyrimidine nucleotides to date, it is important to recognise the limitations of this method. This

set of reactions requires the sequential delivery of reagents and contain mutually reactive feedstocks. For example, both glycolaldehyde **9** and glyceraldehyde **10** can react with cyanamide **22** to give unwanted aminooxazoles, all these products can react further with aldehydes to give an unselective mixture of products. The formation of 2-aminooxazole **2AO** helps to alleviate these issues but this process requires **2AO**'s transport to another site where the other reagents are available to form the furanose backbone **23**. Another issue is that the 2',3'-cyclic-phosphate could hydrolyse in either the 2', 3' or 5' positions. As a result of this, the canonical 3'→5', non-canonical 2'→5' or mixed oligomers could all be formed. The synthesis of pyrimidine nucleotides is a significant achievement that was realized over several generations of scientists. Despite this progress, the prebiotic synthesis of RNA is not complete since a robust purine synthesis remains elusive so far.
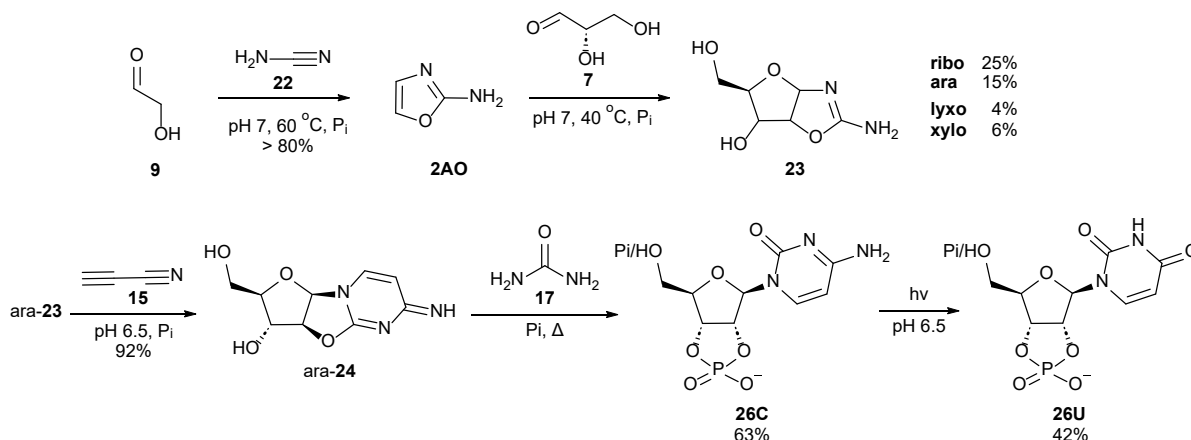
Scheme 11. Powner and co-workers' synthesis of pyrimidine nucleotides cytidine **26C** and uridine **26U** from glycoaldehyde **9**, cyanamide **22**, glyceraldehyde **7** and cyanoacetylene **15**.

1.6.5. Synthesis of Purine Nucleotides

Current research in the field has turned towards purine synthesis, and this has proved a harder challenge than prebiotic pyrimidine synthesis so far. Direct glycosylation of ribose with adenine or guanine had been attempted previously and resulted in only low yields (Section 1.6.3). Different groups have either utilised the synthetic routes, techniques and chemical motifs that were used in the prebiotic synthesis of prebiotic pyrimidine synthesis or devised new strategies to solve purine

nucleotide synthesis. There are currently several incomplete routes to purine nucleotides which are briefly discussed in this review.

Sutherland had previously demonstrated an incomplete synthesis of a potential purine anhydronucleoside **27** from a multicomponent Mannich type reaction from **AICN**, **2AO** and glyceraldehyde **10**.[103,128] Becker and Carrel achieved adenosine by reacting aminopyridine **28** with open chain ribose **20** to form directly the *N*-glycosylic bond *via* imine formation, the intermediate **29** then collapses to adenosine or guanosine **20A/G** under basic conditions.[129–131] Stairs was able to recognise that 2-oxazolidinone thiones **2TO** could be used as a divergent scaffold towards purines and pyrimidine and was able to form 8-oxopurines **20OA/OI** from intermediate **31** in good yields.[132] Finally, Roberts *et al*. was able to form arabino purine nucleosides ara-**20A/G** by the reduction of 8-thio-arabanosides **33** which were formed from **32**.[133] Although much has been achieved, the prebiotic purine synthesis remains elusive.

Although RNA is an attractive candidate for the first information polymer (Section 1.3.3), it is entirely plausible that different and perhaps constitutionally simpler nucleotides may have preceded it. From the plethora of molecules in the prebiotic soup it is also possible that the first genetic material was chimeric in nature, composed of several different classes of nucleotide which were eventually superseded by RNA after a period of chemical evolution. Although this is beyond the scope of this review, current research is experimenting with other informational monomers which could of be present in the 'prebiotic soup' and to deduce whether there was a selection pressure that yielded solely the RNA and DNA genetic material we observe today.
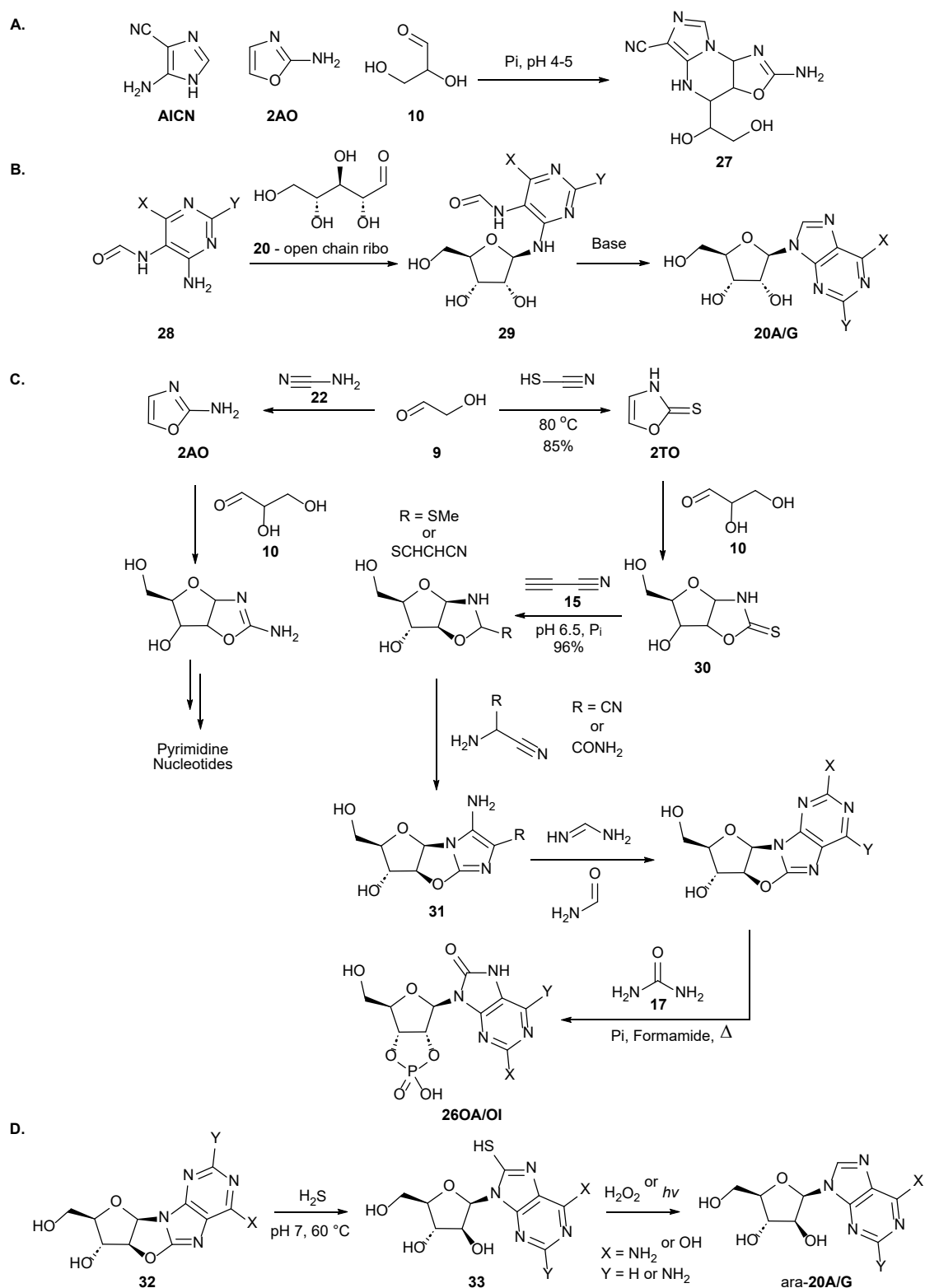
Scheme 12. (**A.**) Sutherland's multicomponent Mannich reaction to potential purine precursor **27**.[128] (**B.**) Becker and Carrel's prebiotic adenosine synthesis[129–131] (**C.**) Stairs *et al*. divergent synthesis to ribofuranosyl pyrimidines and 8-oxopurine nucleotides **26OA/OI**.[132] (**D.**) Roberts *et al*. synthesis of arabino-purines by thiolysis of **32** and subsequent reduction of **33** with UV light or H₂O₂ to ara-**20A/G**.[133]

1.7. Systems Chemistry - Cyanosulfidic Protometabolic Network

Systems chemistry explores chemical reactions that can arise from a common pool of simple molecules, which react and increase in complexity to form a hierarchical network of interacting molecules. The role or origins of the molecules of life cannot be understood in isolation, rather that the molecules must be considered within the context of the interlocked systems they find themselves in. These molecules can be linked in a broad web and systems chemistry seeks to find generational links between these chemical sub-systems. Therefore, more information can be obtained by cross referencing these molecules rather than analysing these molecules in isolation. For example, by considering the relationship between amino acid (formed by Strecker reaction) and sugars the Sutherland group applied systems chemistry thinking to overcome the nonselective umpolung chemistry required to initiate the Formose reaction. They utilised hydrogen cyanide in Killani-Fischer homologation with stoichiometric H_2S reducing the nitrile group under UV irradiation.
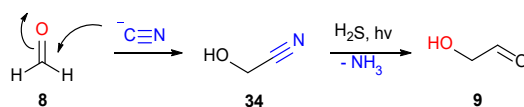
Scheme 13. Kiliani-Fischer type homologation of formaldehyde **8** to glycoaldehyde **9**.

The route begins with the Kiliani-Fischer homologation of formaldehyde **8** with cyanide to form glycolonitrile **34** (Scheme 13). Glycolonitrile **34** reacts with ammonia in a Strecker reaction to form glycine **Gly**-OH *via* α-aminonitrile. Photoreduction of **3** with H_2S as reductant, with CuCN gave glycoaldehyde **10**, the addition of cyanide gave cyanohydrin **35**, which is a precursor to serine **Ser**-OH. Cyanohydrin **35** was reduced to glyceraldehyde **10**, which can tautomerise to dihydroxyacetone **11**. Acetone **36** is formed by the reduction of **11**. Valine and Leucine amino acids can be formed through Kiliani-Fischer type cyanide homologation of acetone **36**. Addition of cyanide to acetone **36** forms cyanohydrin **37,** then a sequence of thiolysis, then a reduction of thioamide leads to the formation of the valine's cyanohydrin **38** which can form valine nitrile **Val**-CN upon addition of ammonia. Leucine is derived after another cycle of thiolysis, deoxygenation and reduction of

38.[28] Glycoaldehyde **9** is a key node in this network, Sutherland was able link the synthesis of pyrimidine ribonucleotides (Scheme 14) to the synthesis of amino acids through precursors such as glycolaldehyde **9** (Serine), cyanamide **22** (arginine); cyanoacetylene **15** (aspartate, asparagine) and glyceraldehyde **10** (valine, leucine) from dihydroxyacetone **11** and its reduction to acetone **36** (Scheme 13). Interestingly dihydroxyacetone **11** can be reduced to glycerol, which can be phosphorylated to form a lipid precursor, glycerol-1-phosphate.[28]
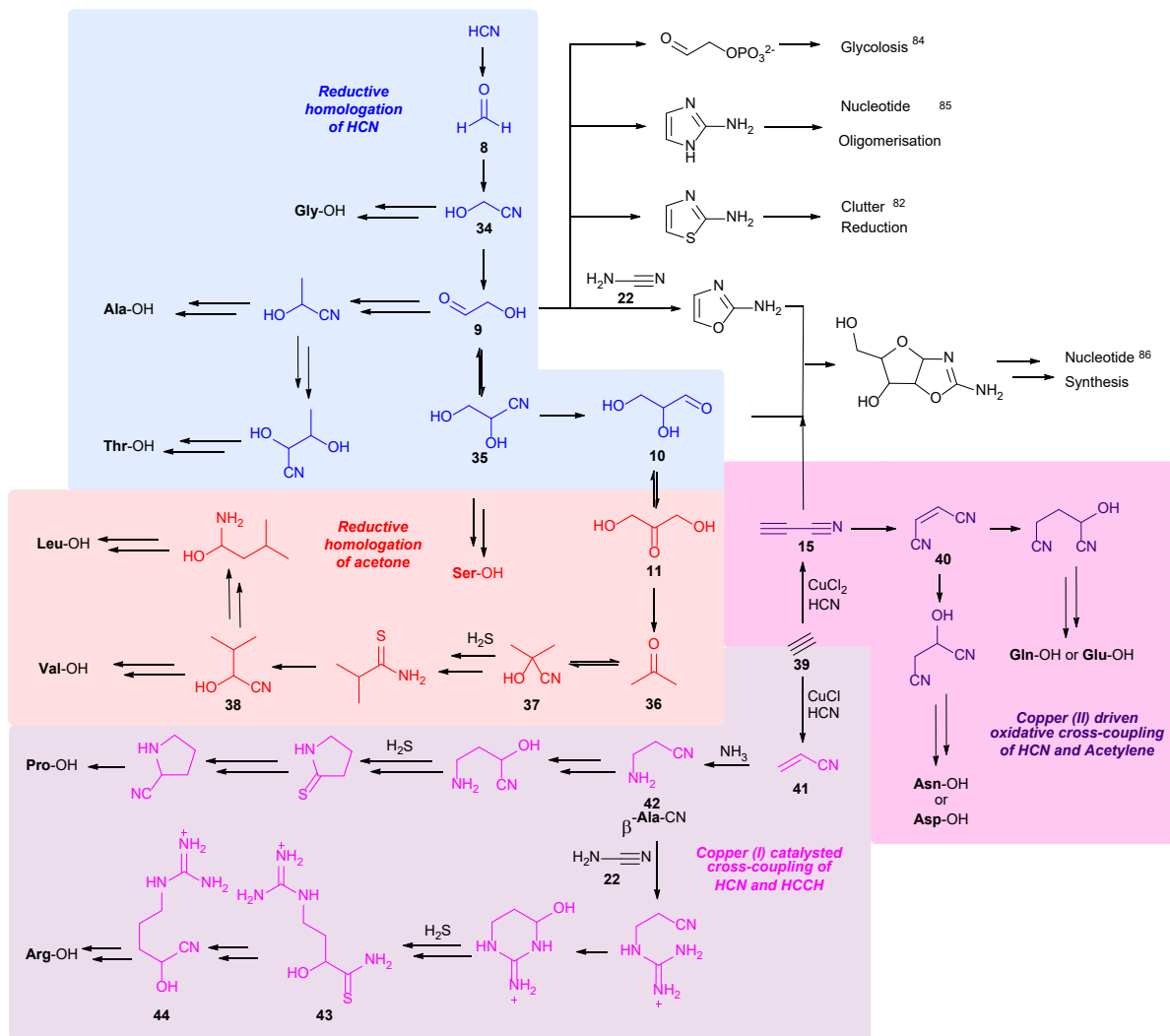
Scheme 14. An overview of the 'cyanosulfidic' systems chemistry network.

Some of the more esoteric chemistry can be derived from copper chemistry and the formation of useful prebiotic reagents from cyanocuprates. From acetylene **39**, Cu(II) oxidative-cross-coupling with cyanide yields cyanoacetylene **15**.

Cyanoacetylene **15** is a key component of pyrimidine synthesis. Addition of HCN to cyanoacetylene **15** gives maleonitrile **40** which can be directly reduced to aspartic acid or asparagine. Maleonitrile **40** can also undergo a succession of photoreduction with H_2S, then homologation with HCN to give either glutamic acid or glutamine.

Alternatively, acrylonitrile **41** can be formed from acetylene **39** and cyanide with Nieuwland's catalyst which is a mixture of Cu (I) salts dissolved in high concentration sodium or potassium chloride. Reaction of acrylonitrile **41** with ammonia produces β-alanine nitrile **42**. Although β-alanine nitrile is a non-proteinogenic amino acid in biology, it can undergo transamination with pyruvate to form malonate and alanine over several steps. The β-alanyl motif is also found in CoA a universal cofactor that is utilised in the Krebs cycle, the properties of nitrile **42** will enable it to be used to great effect in the prebiotic synthesis of pantetheine **66** (Chapter 2)

From β-alanine nitrile **42**, proline or arginine can be made. Arginine is formed from **42** *via* the reaction with cyanamide **22**, then guanylation, cyclisation, ring opening, thiolysis to thioamide **43**, reductive homologation and then finally hydrolysis to **44**. Scheme 11 represents a 'network' of reactions built up over many years by Sutherland and others. Notable additions include Powner's use of glycolaldehyde **9** to investigate triose glycolysis pathways or when Islam *et al* were able to reduce the "clutter" caused by non-canonical ketones through the selective precipitation of proteinogenic (Strecker) amino acid precursors.[28, 134, 103, 126, 135] Sutherland suggests that nature did not 'choose' specific amino acids *via* evolutionary selection, and that they were chemically predisposed instead.[28] It was remarkable that this work was able to bring together known nucleotide and amino nitrile synthesis in a feasible prebiotic systems chemistry network. Although this network enables controlled synthesis of 12 of the proteinogenic amino acids. This work is not perfect and many of the reactions require spatiotemporal separation to reach the desired amino acids without deleterious side reactions. The prebiotic plausibility of some of the products could be questioned. For example, there are 11 steps to get to cyanohydrin **38** which is a precursor of valine **Val**-OH, which seems unlikely when mutually reactive molecules would be concurrently fighting for chemical space in this system.

1.8. Peptides

Peptides are short chains of amino acids typically less than 20 amino acids residues and these peptides can combine to form bigger polypeptide called proteins. Proteins can have masses greater than 10,000 Da. Proteins are present in every organism and play a significant role in many physiological and biological processes. This ranges from molecule transport, cell structure, DNA replication and catalysing metabolic reactions.[136–138] Enzymes are large macromolecules comprising of one or more proteins, that can enact complicated chemical transformations *via* their specific folded tertiary structure. These enzymes often require cofactors or a co-enzyme such as CoA in order to perform their specific functions effectively.[138,139] Peptides are intertwined with the machinery of life and therefore it is one of the most pressing challenges in the field to understand the prebiotic synthesis of amino acids, the reason for their secondary structure (side chains), amino acid oligomerisation and the relation between the genetic code and the canonical amino acids.

1.8.1. The 'Protein World' Hypothesis

The difficulty of the syntheses of RNA has led many prebiotic scientists to believe that the RNA world theory may not be the best route to solve the origins of life. An alternative hypothesis proposes that proteins may have preceded or holistically co-evolved with RNA as the species with catalytic and genetic function.[140,141] This theory relies on the assumption that the amino acids would spontaneously assemble into peptides. Ikehara supported by Kurland proposed that the peptide gly-ala-asp-valine (GADV) could have provided the basis for the coding and replication of nucleotides.[142–144] Even if amide bond formation was favourable from a large pool of amino acids the random sequence of amino acid residues would not guarantee a significant number of proteins with useful functionality. For example, oxytocin is a small peptide with 9 amino acid residues H-Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-NH_2. To arrive at this particular sequence, assuming there are only 20 amino acids available, would require the "selective" synthesis of 1 peptide amongst $20^9 = 5.12 \times 10^{11}$ combinations of different peptide sequences.

1.8.2. Amide Bond Formation

Proteins and peptides play a significant role in the design of life. Currently bioactive peptides form a significant portion of new medicinal, agrochemical and biological molecules that have improved the lives of millions of people across the globe.[136] So much so that efficient amide bond formation with improved atom economy was voted one of the ACS top green challenges.[145–147] Despite the ubiquity of the amide bond in a diverse array of compounds, the efficient formation of amide bonds has proven a very difficult challenge to overcome in the synthesis of peptides.[148]

Despite the difficulty in peptide synthesis, scientists have attempted to make amide bonds since the late 19th century. One of the first attempts of dipeptide formation was by Curtius, who was able to synthesise the *N*-protected dipeptide, benzoylglycylglycine **46**, from the reaction of the silver salt of glycine **Gly**-O⁻Ag⁺ and benzoylchloride **45** (Scheme 15A).[136,149] Nineteen years later, Fischer reported the synthesis of glycylglycine **Gly**-**Gly**-OH, which was achieved by hydrolysis of the diketopiperazine **47** to glycine **5** (Scheme 15A).[136,150]

During this time Fischer first introduced the terminology 'peptides' and developed another method using acyl chlorides as 'activated' amino acids in coupling reactions (Scheme 15B). Using this method Fischer was able to synthesize peptides containing up to 18 amino acid residues.[136,151,152] Many of these early methods required the use of *N*-protecting groups and activation of the carboxylate to grow peptides monomerically.

However, progress was slow since there was a lack of easily removable protecting groups and not all the 20 proteinogenic amino acids were known until 1935 (threonine). It wasn't until 1957 with the introduction of the acid labile *tert*-butoxycarbonyl group by Albertson and McKay that the use of orthogonal protecting groups became widespread among peptide chemists.[136,153]
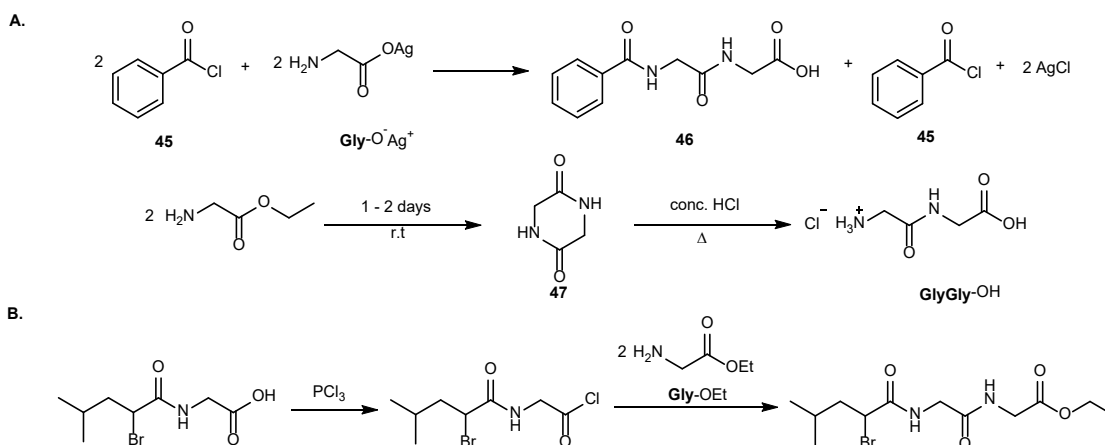
Scheme 15. (**A.**) The first ever dipeptide formation by Curtis in 1882 and the first unprotected peptide synthesis via diketopiperazine (DKP, **47**) intermediate by Fischer 1901.[149,150] (**B.**) Fischer's dipeptide synthesis using acyl chlorides.[151,152]

In water, peptide formation from uncharged amino acid monomers is thermodynamically favourable, however the rate of reaction is slow.[154] To overcome this problem high temperatures can be employed to increase the rate of amide formation and increase the yields in order to overcome the condensation between two amino acids in water.[155–157] In the process the harsh conditions also resulted in the hydrolysis/fragmentation of peptide products.[158]
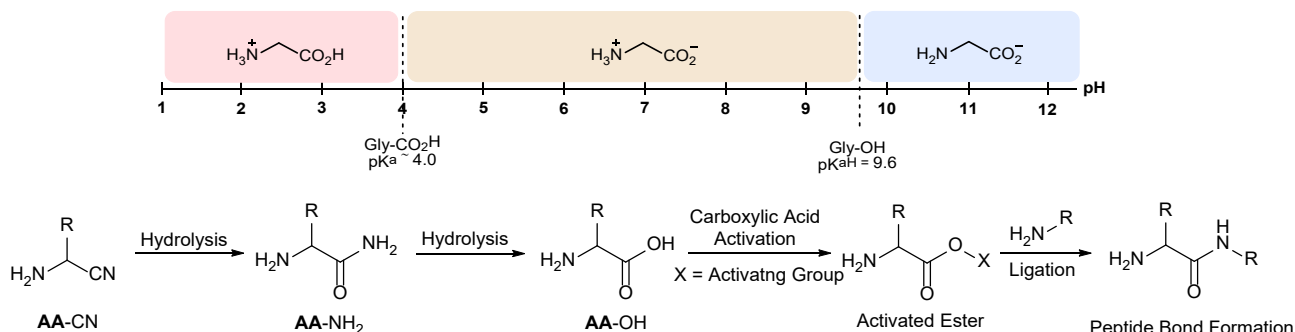
Scheme 16. (**Top**) pH scale representing the pKa's where glycine's amine is nucleophilic (blue) or the point where glycine's carboxyl is electrophilic (red). (**Bottom**) A general route to peptide formation from the hydrolysis of α-aminonitriles to amino acids and their ligation and coupling through carboxylate activation.[159]

The sluggish reactivity of each end of an amino acid can be explained by the pK_a mismatch. A low pH is required for the carboxylic acid ($pK_a \sim 4$) to be electrophilic, whereas the amine ($pK_{aH} \sim 9\text{-}10$) requires a high pH to be nucleophilic (Scheme 16).

Currently, solution-based peptide synthesis utilises orthogonal protecting groups on the amine and amino acid side chains, then activation of the carboxylic acid moiety with carbodiimide reagents such as N, N'-dicyclohexylcarbodiimide (DCC) or the hydrochloric acid salt of N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide (EDC·HCl). The reactive O-acylisourea intermediate can be coupled directly with the amine of an amino acid or be trapped as an activated ester with substrates such as pentafluorophenol (Pfp), N-hydroxysuccinimide (Su) or other derivatives. These active esters can be isolated, purified and have a definitive shelf life (Scheme 16).

The active esters of hydroxybenzotriazole (HOBt) or 1-hydroxy-7-aza-benzotriazole (HOAt) are used to suppress racemization by forming a stable activated ester insitu and preventing the rapid formation of an oxazolone intermediate by intramolecular cyclisation.   Uronium reagents such as **h**exafluorophosphate **a**zabenzotriazole **t**etramethyl **u**ronium (HATU) or phosphonium salts **b**enzotriazol-1-yl**o**xytri**py**rrolidino**p**hosphonium hexafluorophosphate (PyBOP) incorporate both a carboxylate activating unit and a nucleophilic component to form non-racemised activated esters *in situ* which is susceptible to nucleophilic attack to another amino acid to yield the desired peptide (Scheme 16).[137,148]

Unfortunately, the solution-based peptide synthesis strategy is grossly atom inefficient, often utilizes toxic solvents, results in the racemization of homochiral amino acids and is limited to small peptide fragments as solubility becomes an issue with increasing peptide length and multiple protecting groups.

In the 1970's Merrifield pioneered another strategy for the synthesis of longer peptides.[138,139,160–162]   Solid-state peptide synthesis (SSPS) enabled the rapid synthesis of peptides by building the peptide sequentially on an insoluble scaffold. The first step in this methodology involves the immobilisation of a Boc (Acid Labile) or Fmoc (Base labile) protected amino acid onto the resin. The resin bound amino acid is deprotected, then coupled with a N-protected amino acid at the N-terminus with an activating agent reagent.   The resin bound dipeptide can then be elongated sequentially with the desired N-protected amino acids until the required length is

achieved. At this point orthogonal deprotection of side chains can be attempted for side chain functionalisation or a global deprotection on the peptide can simultaneously remove all the protecting groups and cleave the peptide from the resin (Scheme 17). SSPS is advantageous because a large excess of reagents at high concentrations can be used to drive coupling reactions to completion and excess reagents and side products can easily be removed by filtration and washing steps after each coupling step. Unfortunately, even though the reaction conditions have been highly optimized and are quite efficient, if you get 98% of the coupled product at each step, after the addition of 30 amino acids only ~ 55% of your product will have the correct sequence. Therefore, only peptides containing less than 30 amino acids are normally synthesized using this method. It is common for longer sequences to be obtained through expression by bacterial cells such as *E. Coli* or by cysteine residues in native chemical ligation (Section 1.10.4)
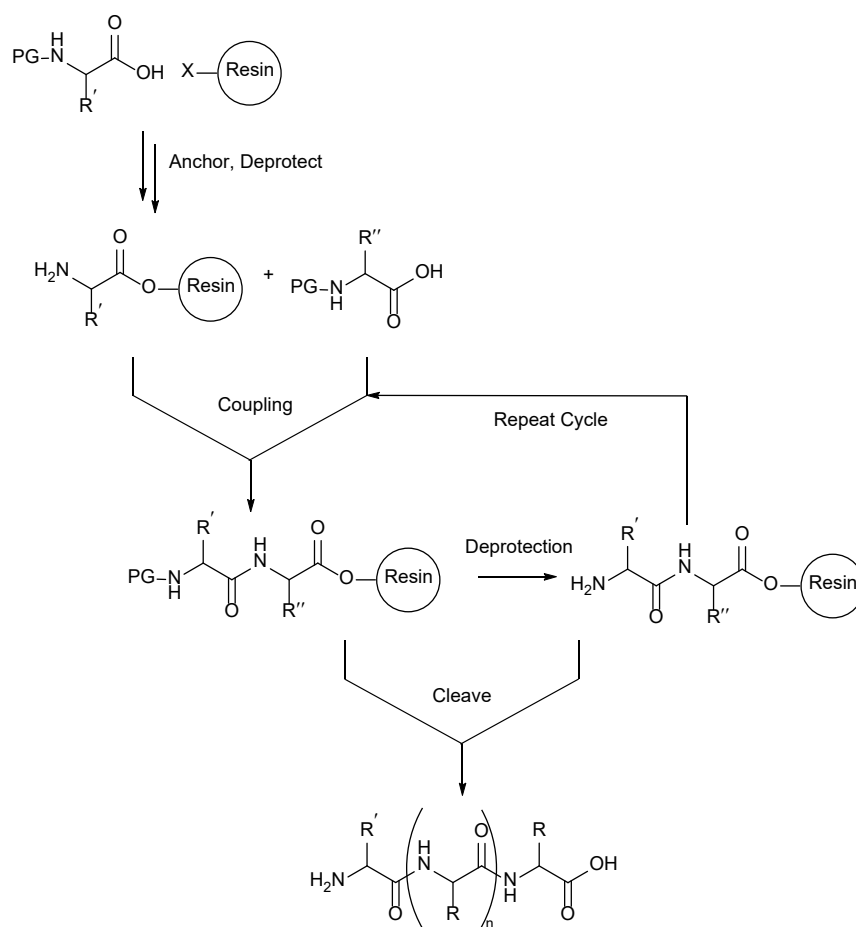
Scheme 17. Simplified depiction of solid-state peptide synthesis (SSPS) developed by Merryfield.

1.8.3. Prebiotic Peptide Synthesis

N-Carboxy amino acid anhydrides (**NCA** or Leuch's acid) have been known since the 20th century and their synthesis classically required the use of toxic compounds such as phosgene in anhydrous organic solvents.[95] Recently **NCA**'s have become a gained attention in the investigation of prebiotic peptide synthesis.
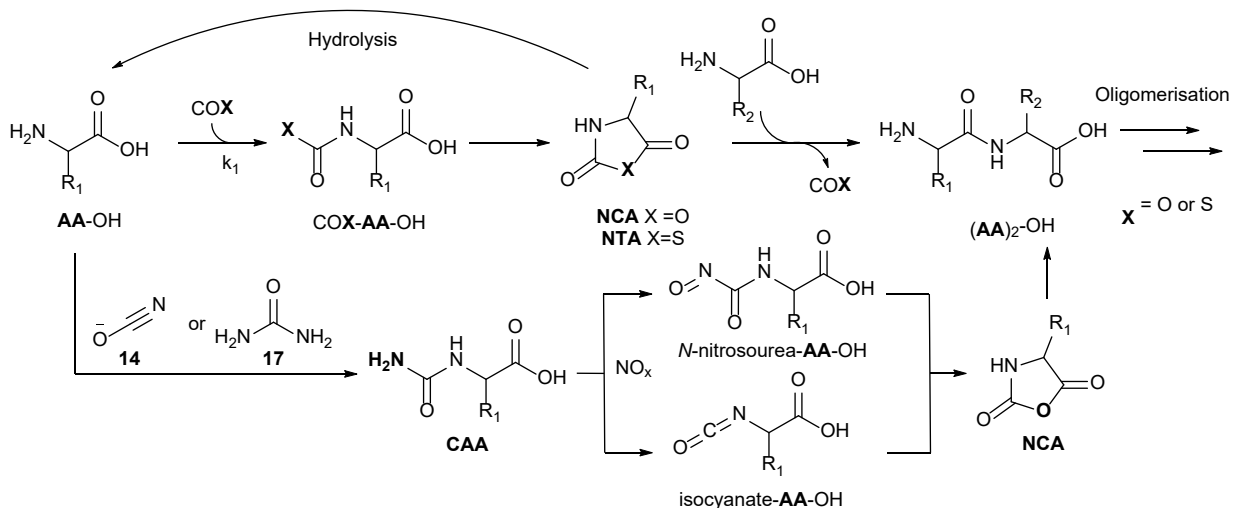
Scheme 18. Amino acid activation to NCA's from COX or NO_x.

In an analogous fashion to the Bücherer-Bergs reaction, **NCA**'s form from the reaction of CO_2 and the free amine of an amino acid forming a carbamoyl amino acid which undergoes intermolecular cyclisation.[93] The CO_2 concentration on the early Earth is believed to be 100-1000 times higher than at present and therefore carbamylation of amino acids in aqueous environments would be increased relative to the present day.[93,95] During their studies Hirshmann and co-workers recognised the potential for prebiotic synthesis of peptides from **NCA**'s in water.[163,164,165] Volcanic eruptions on the early Earth would have expelled significant quantities of H_2S and COS gas into the atmosphere which led Hirschmann and later Ghadiri to use the electrophilic COS to activate amino acids and form amino thiocarbamates.[164,95,93,166] The thiocarbamate could either cyclise to form **NCA** or *N*-thiocarboxyanhydride (**NTA**), both of which can be attacked by amino acid nucleophiles to yield dipeptides (**AA**)$_2$-OH, phosphate to form amino phosphate esters, 5' AMP to give aminoacyl adenylate and H_2S to yield amino thioacids.[167,168]

Additionally, Commeyras and co-workers found that the nitrosation of unreactive *N*-carbamoyl amino acids (**CAA**) with nitrous oxides produced from thermal or photochemical processes in the atmosphere could form **NCA**'s and therefore peptides (Scheme 18).[95,169]

When Hirschmann and co-workers were exploring the controlled synthesis of dipeptides in aqueous conditions they found problems with glycine **NCA**. They found for the glycine derived NCA that there was a lack of selectivity and amines could attack at either the C2 or C5 positions of glycine **NCA**'s. The **NCA** would go on to irreversibly form ureas/hydantoic acids **HDA** or hydantoins **HD** from the free amino terminus of the peptide, meaning that C to N peptide growth is incompatible with **NCA** and a free N-terminal of a peptide (Scheme 19).[163,165]
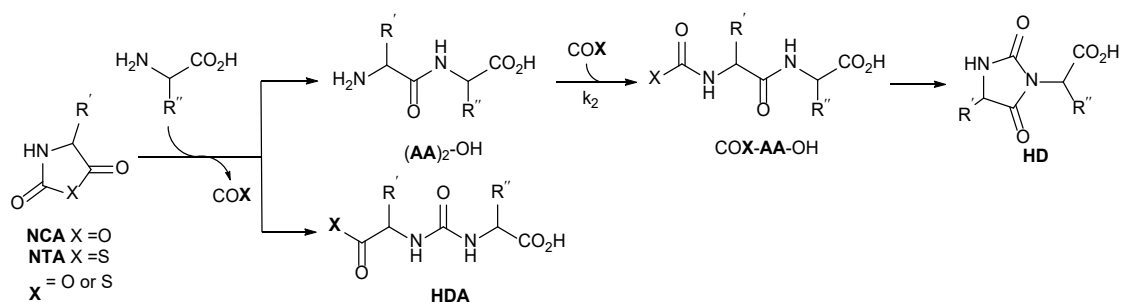
Scheme 19. Incompatible 'dead end' products that prevent the growth of peptides from NCA's.

The difficulty in prebiotically forming amide bonds led Krishnamurthy and Hud to postulate that α-hydroxy acids or thioesters could have formed polyesters before polypeptides on the early Earth.[170,171] Carboxylate activation of amino acids with α-hydroxy acids could form depsipeptides which are polymers that incorporate both amide and ester bonds into the polymer backbone. Hud and Krishnamurthy found that when lactic acid **48** is exposed to 4 wet/dry cycles (wet-dry cycles; wetted at 65°C, 5.5 h, then dried at 85°C, 18 h) polyesters are formed. When a 1:1 molar ratio of lactic acid **48** and glycine **Gly**-OH are mixed and subjected to the same wet-dry cycle conditions, depsipeptides of up to 10 residues are formed (Scheme 20).[170]
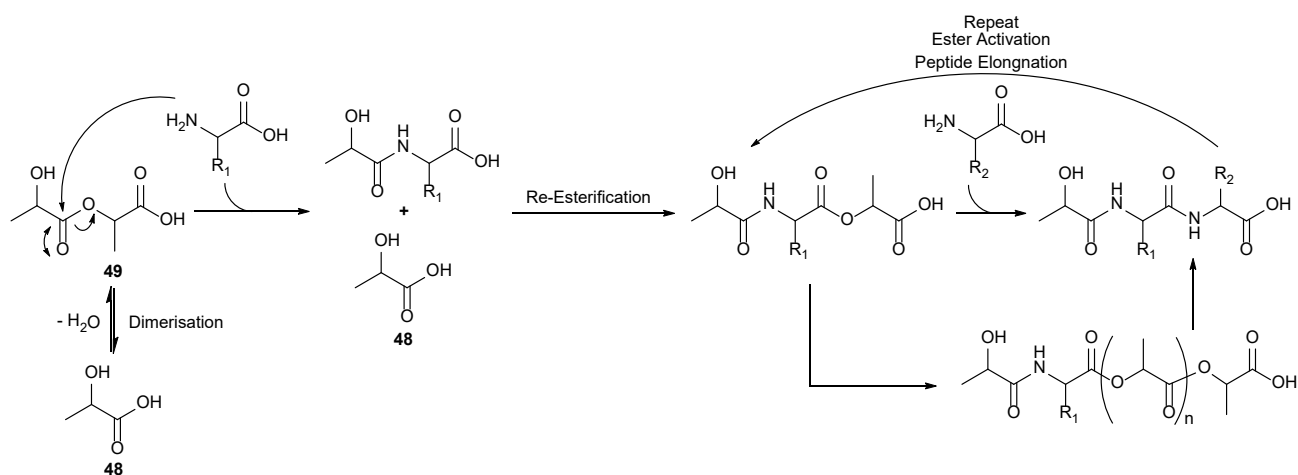
Scheme 20. Krishnamurthy's prebiotic depsipeptide synthesis.

The mechanism is thought to begin with nucleophilic attack of the lactic acid dimer **49** with glycine **Gly**-OH. The lactic acid *N*-protected glycine can undergo further esterification at the carboxylate group, thus activating the amino acid to further attack with either an amino acid or lactic acid. In the dry phrase solvent evaporation favors ester bond formation and then ester amide exchange. Then in the cool phase the reactants are dispersed, and the ester bonds are preferentially hydrolyzed. Chain elongation would occur over successive cycles and over time ester hydrolysis would out compete amide bond formation, this has been proposed to potentially result in peptides triumphing over depsipeptides but has yet to be demonstrated in practise.[170,171] These reaction conditions could represent a model of the weather or tidal forces found on the early Earth.

Additionally, thiodepsipeptide polymers could present an alternative backbone that would fit the 'Thioester World' proposed by de Duve (Section 1.10.3).[171,172] Although the thioesters are considered a labile species, this potential pitfall could be advantageous where this reactivity enables the dynamic sampling of sequence space.[95] Thiodepsipeptides may also facilitate peptide ligation in a manner similar to native chemical ligation.[95,171] Other groups have utilised wet-dry cycles in order to condense amino acids, however these methods are not reliable and suffer from poor reproducibility. As a result, these experiments result in low yields and/or uncontrolled polymerisation and give unclear conclusions.

1.8.4. Ribosomal Peptide Synthesis

Chemists can be inspired by nature to find alternative ways towards peptide synthesis. The discovery of the ribosome (Section 1.3.3) and the genetic code led to the elucidation of the mechanism for coded peptide synthesis from proteinogenic amino acids in biology.[9,65,173] Interestingly, ribosomal peptide coupling proceeds in a similar manner to synthetic carboxylic acid activation strategies, in this case the specific amino acid is activated by adenylation with ATP. The activated adenylated amino acid is recognised by their specific aminoacyl-tRNA synthetase and the amino acid is attached to their corresponding anticodon tRNA. This results in esterification at the 3' position of the anticodon loop. A small ribosomal subunit (30S) attaches itself to the 5' end of a mRNA strand and it moves along until it finds a start codon. The aminoacyl tRNA anticodon then binds to the A site with the larger ribosome subunit (50S) when it recognises the correct corresponding triplet codon of the of the mRNA strand at the A site. The *N*-terminus of these amino acylated anti codons react with the *C*-terminus to the growing peptide chain at the P site. The EG-F-GTP (prokaryote elongation factor G/translocase) protein complex binds to the A site to induce the translocation of the tRNA at the P and A sites to move the sequence along to the E and P sites. This pushes out the spent tRNA out of the ribosome at the E site. After the release of the EG-F-GTP complex the A site becomes free for another coded aminoacyl tRNA to repeat peptide elongation sequence again. A stop codon hydrolyses out the peptide releasing it from the ribosome, ready for to perform catalytic transformations (Figure 11).
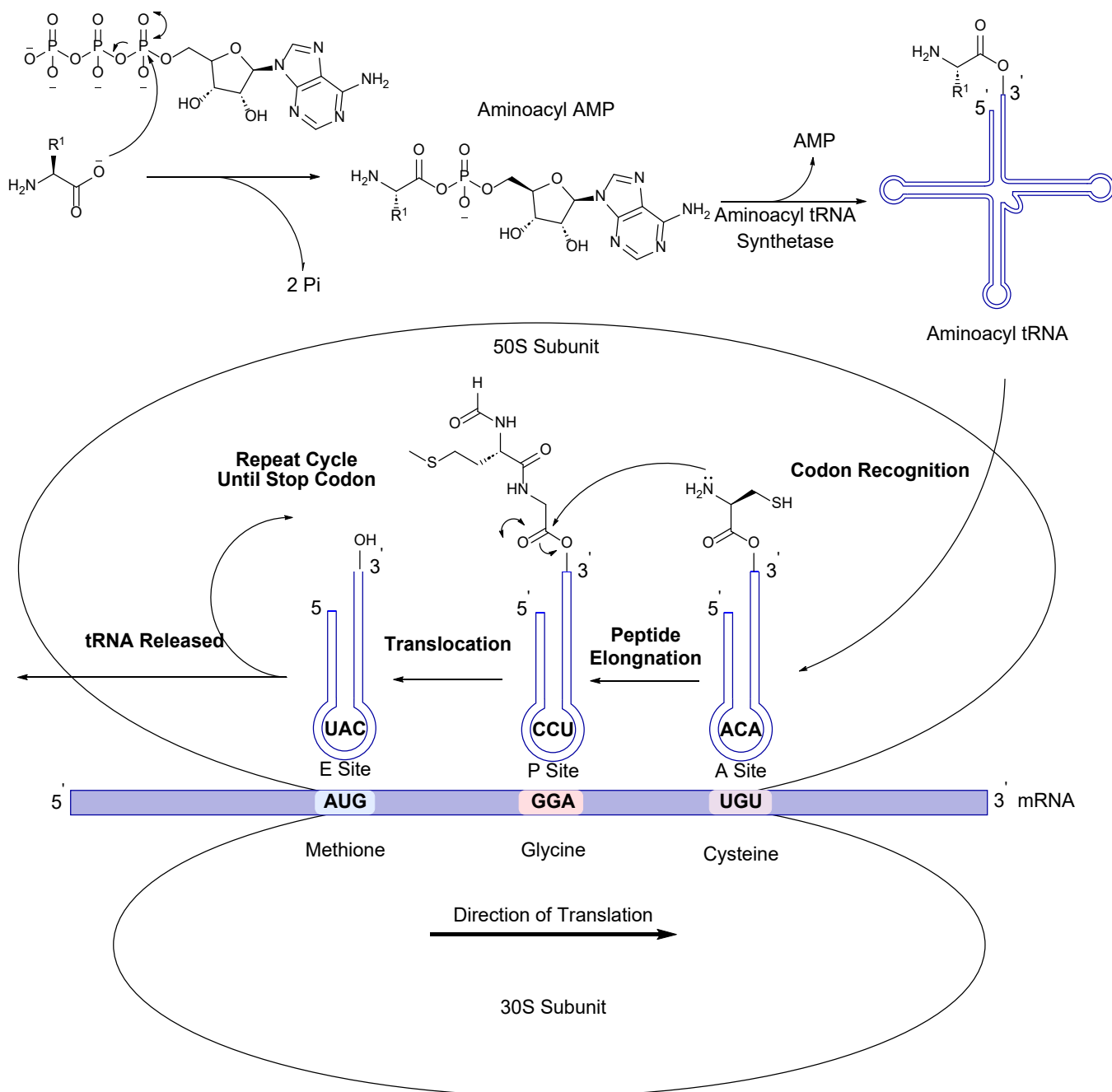
Figure 11. Diagram depicting the steps of peptide formation in the prokaryotic ribosome.[9,65,173]

1.8.5. Biomimetic Peptide Strategy

In classic organic chemistry the synthesis of peptides proceeds from the *C*-terminus to the *N*-terminus (peptide growth to the left). Biology on the other hand creates peptides in precisely the opposite way by elongating peptides to the right from the *N*

to *C*-terminus (ribosomal peptide synthesis (Section 1.3.3) and non-ribosomal peptide synthesis (Section 1.10.3)).[174]

Previously, Maurel and Orgel have suggested that thioacids would be effective activated carboxylic acids and could have been a plausible prebiotic alternative to biological thioesters on the early Earth.[175] Orgel found that potassium thioacetate is an efficient acylating agent in the presence of an oxidant in water.[176] It was found that thioacetate was able to form electrophilic dimers through oxidation, these dimers are really good leaving groups which expediently facilitate the formation of amide bonds in high yields.[176] Orgel was able to expand this process to enable the formation of peptides from α-glutamic thioacid **Glu**-SH with (Glu)$_{10}$ primer.[175] Despite this the ligation yields were low and over 500 equivalents of **Glu**-SH to the (Glu)$_{10}$ primer was required for ligation to occur. Interestingly, no extension of the primer occurred with just ferricyanide as an additive. The addition of bicarbonate increased the rate of the reaction and Orgel proposed that this was due to the activation of the amino acid *via* a reactive **NCA** intermediate (Scheme 21).[175]
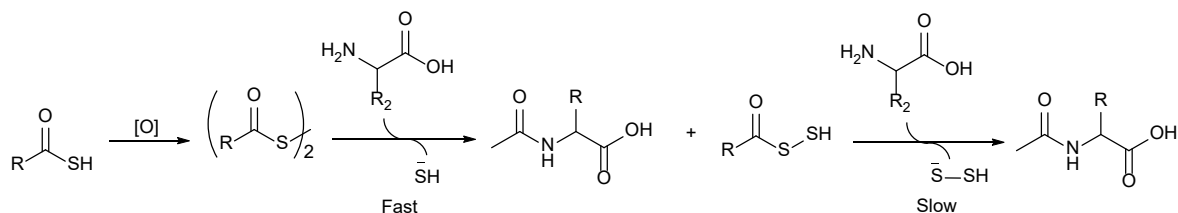
Scheme 21. Orgel's thioacid prebiotic peptide formation.

By exploiting the relative reactivity of hydrogen sulfide (H$_2$S) as a sulfur source with nitriles Canavelli *et al*. were able to discover an effective non-enzymatic peptide synthesis through a biomimetic ligation strategy.[174] This synthetic route was able to furnish *N*-acetylated amino thioacids Ac-**AA**-SH from *N*-acetylated α-aminonitriles Ac-**AA**-CN, through an acetylation (to initiate the cycle), thiolysis, hydrolysis and ligation cycle.[174] This occurs because the electron withdrawing nitrile moiety lowers the basicity of the amine functionality giving the pK_{aH} of an α-aminonitrile at approximately 5.3. This has two-fold effect, firstly the lower pK_{aH} amine of a α-aminonitrile is nucleophilic over a broader range of pH values than amino acids.

Secondly, the innate energy stored in the nitrile group unlocks the potential for other unexplored reactivity before the energy is dissipated when α-aminonitriles are hydrolysed to amino acids. The general procedure for the peptide ligation proceeds through the thiolysis of an *N*-acetylated aminonitrile under basic conditions to the *N*-acetyl thioamide Ac-**AA**-NH$_2$ in excellent yields for all proteinogenic amino acids. Prolonged heating at 60 °C in basic conditions of the thioamide gave the thioacid Ac-**AA**-SH. Sequential addition of ferricyanide oxidant and aminonitrile results in selective and fast ligation and the cycle can be repeated up to five times, fragment-based ligation was also possible by this process (Scheme 22).[174]
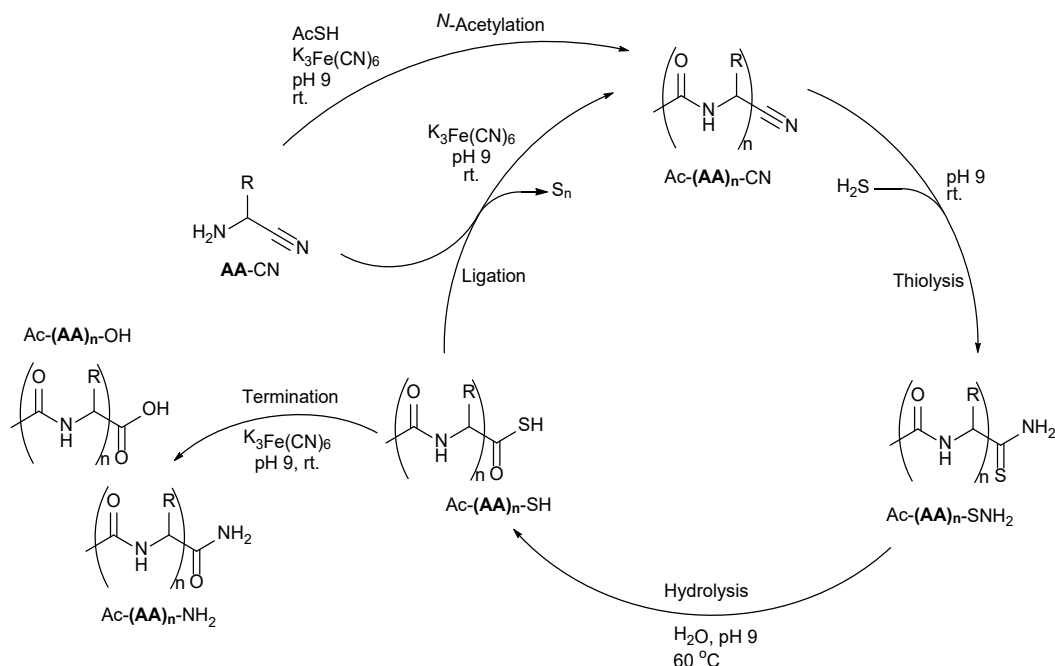
Scheme 22. Ligation cycle developed in the Powner group of the coupling of aminonitriles.[104]

This method overcomes derivatisation problems that occurred in dipeptide formation with COS gas (Section 1.8.3) and remarkably the ligation does not require protecting groups on any of the 20 proteinogenic amino acids. For example, Powner was able to selectively ligate lysine at the α position (>80:1). This is significant because the two amine functional groups make it notoriously difficult to ligate selectively at the α position over the ε position and this route presented the first non-enzymatic, protecting group free intermolecular lysine ligation.

Thoma *et al*. was able to build on the work of Canavelli *et al*. and demonstrate that lysine thioacids (Ac-**Lys**-SH, n = 3) undergo coupling with aminonitriles **AA**-CN in neutral water to afford peptides in near-quantitative yield, whereas non-proteinogenic lysine homologues, ornithine (Ac-**Orn**-SH, n = 2), and diaminobutyric acid (Ac-**Dab**-SH, n = 1) cannot form peptides due to rapid and quantitative cyclization to lactams that irreversibly blocks peptide synthesis (Scheme 23).[177] The experiments demonstrated that ornithine lactamization provides absolute differentiation of lysine and ornithine during (non-enzymatic) *N*-to-*C*-terminal peptide ligation. However, the shortest lysine homologue, diaminopropionic acid (Ac-**Dap**-SH, n = 0), also undergoes effective peptide ligation. Considering this work, a high-yielding prebiotically plausible synthesis of the diaminopropionic acid residue was developed, by peptide nitrile modification, through the addition of ammonia to a dehydroalanine nitrile residue. This work suggested that diaminopropinic acid could be an early precursor to lysine on the early Earth and lysine could have developed simultaneously with Dap or been a late addition to the genetically code by evolution.[177]
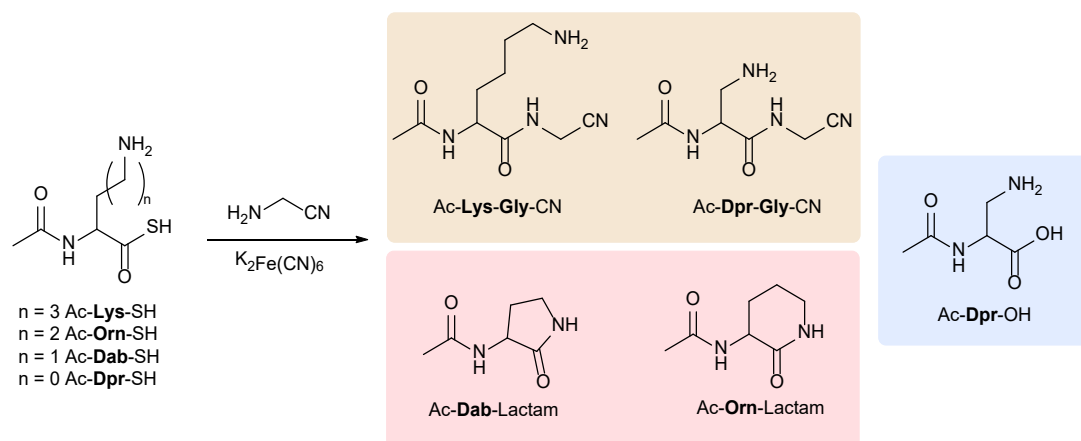
Scheme 23. Selective *N*-to-*C* terminal ligation of lysine peptides in water. The reaction of Ac-**AA**-SH (n = 0-3) with **Gly**-CN in neutral water yields ligation; Ac-**Lys**-**Gly**-CN and Ac-**Dpr**-**Gly**-CN-(green), hydrolysis; Ac-**Dpr**-OH (blue), or cyclization; Ac-**Orn** Lactam and Ac-**Dab** Lactam (pink).[177]

1.8.6. Conclusion

The prebiotic synthesis of amino acids has made major leaps over the past decade; however, some major amino acids are missing. The amino acids histidine, lysine, phenylalanine, tryptophan, cysteine have major catalytic and structural functions within proteins and cannot be ignored. Despite this many scientists speculate that these compounds are later additions through evolution and dedicate little time in understanding their origin. Peptide ligation was achieved with high yields and no protecting groups; however, this cannot be done in one-pot the authors recognised that the reduced gas H_2S and the $Fe(CN)_6^{-3}$ oxidant are mutually reactive feedstocks and must be added sequentially for ligation to occur. The precipitation of > 5 mer oligopeptides presents a double-edged sword where the solubility issues inadvertently purified peptides but also concurrently prevented longer chains in solution. The hydrolysis of thioamide to thioacid is not efficient and results in concurrent hydrolysis, thus limiting the yields.

One of most important unanswered questions that remains is the relation between RNA coding for amino acid sequences. The recent discoveries show that it is an exciting time in the field and there is an optimistic outlook that hurdles can be overcome and better route to peptides under prebiotically plausible conditions can be found.[174]

1.9. Compartmentalisation

Compartmentalization is of key importance in the quest towards the synthesis of the first cells.[178] There are three benefits to compartmentalisation; firstly, a semi-permeable membrane physically encloses biomolecules from the environment and increases their concentrations. This in turn enables the spatial coupling of information carriers with catalytic biopolymers. Encapsulation also prevents the dilution of important molecules to aqueous environments (rivers, lakes, oceans etc.). And finally, a semi-permeable membrane acts as a selective barrier and enables a cell to regulate its internal chemistry, which also protects against chemical parasites infecting the host.[179,180] Lipids are amphiphilic small molecules containing a hydrophilic polar head and a hydrophobic non-polar tail. Most modern cell membranes consist of closed spherical bilayers made of different families of lipids, such as phospholipids, glycolipids, and cholesterol in which transmembrane proteins with diverse functions are embedded.[181] The structure of these individual components is quite complex, and the consensus is that there is a low probability these molecules were formed on the early Earth. For these reasons, it is generally assumed that the composition of primordial membranes was much simpler than extant cell membranes.[181] Alkyl phosphates, fatty acids, sphingolipids and poly-prenyl chains have been proposed as possible constituents of early membranes (Figure 12).[181–184]
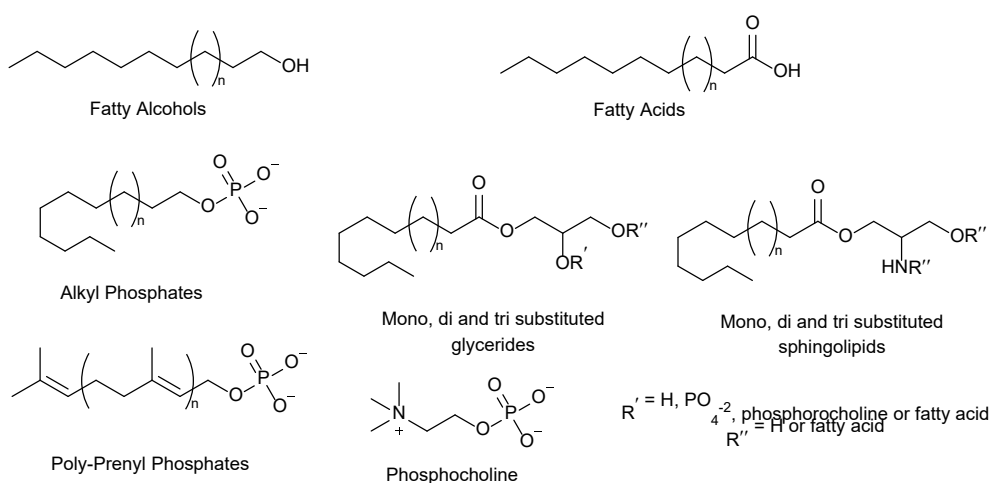
Figure 12. A set of possible amphiphilic molecules that could have formed the earliest membranes.

It is well documented that straight chain hydrocarbons can be synthesized efficiently by Fischer-Tropsch (FT) processes.[181,185,186] The FT process is a set of chemical reactions that converted syngas ($CO + H_2$) into long chain alkanes through the reduction of CO with H_2 in the presence of a metal catalyst (Fe, Co, Ni, Rh) at high temperatures and pressures (150-300 °C). As discussed previously, much of the hydrogen was outgassed hydrodynamically during the Haden era (Section 1.2.3). Due to the relative scarcity of hydrogen on Earth, the Kölbel-Engelhardt (KE) variant of FT appears more likely since this process utilises H_2O instead of H_2 to form long chain aliphatic alkanes, alcohols, and some carbonyl species under similar conditions.
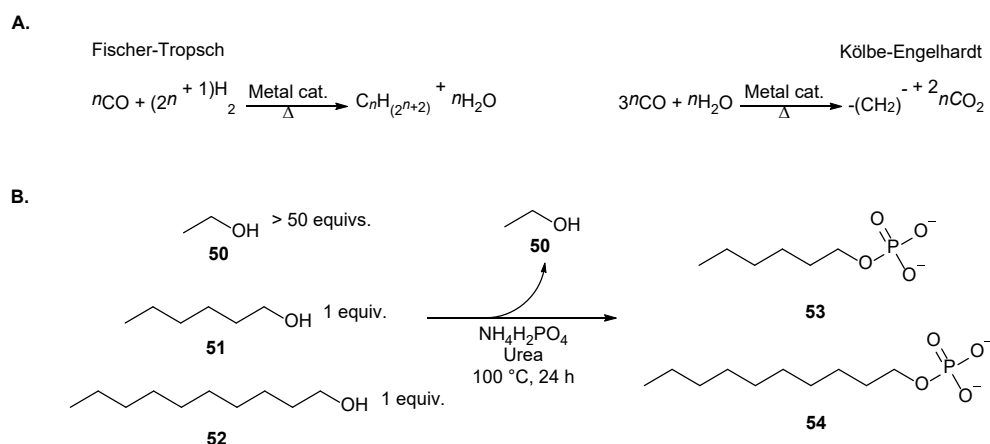
Scheme 24. (**A.**) A set of possible amphiphilic molecules that could have formed the earliest prebiotic membranes; (**B.**) The selective formation of long chain alkyl phosphates in the phosphorylation of mixed chain alkanols.

Work by Oró and co-workers were able to utilise FT conditions to synthesise characteristic saturated and unsaturated alkyl chains (C_{11}-C_{25}) in a closed system at temperatures of between 300-370 °C with either iron, meteoritic iron (Canyon Diablo meteorite or a Fe/Ni alloy.[187] More recently Simoneit *et al* conducted the same reactions by heating oxalic acid solutions at temperatures that simulate the conditions at deep-sea hydrothermal vents.[188] At the optimal temperature (150−250 °C), the lipid components ranged from C_{12} to more than C_{33}. Alkanoic acids increased in concentration while alcohols decreased when the temperature was raised above 200 °C. Phosphorylations have also been exploited to furnish phospholipid-type molecules selectively from a mixture of alcohols.[71] Dry-state

phosphorylation exploits evaporation to remove water and simple alcohols during these phosphorylation reactions. Long chain alcohols, such as decanol **52** and hexanol **51** can be phosphorylated in preference to shorter chain alcohols, such as ethanol **50**, to selectively yield decyl-phosphate **54** and hexyl-phosphate **53** in a 7:3 ratio distribution of products, by simple virtue of comparative volatility (Scheme 24).[71]
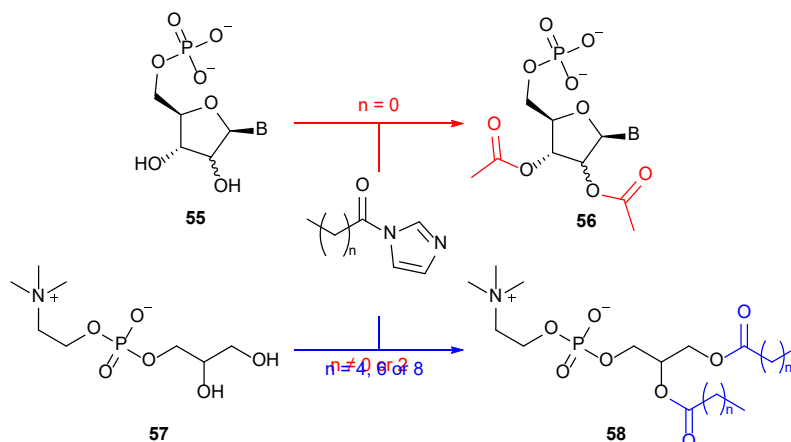
Scheme 25. Fernández-Garcia et al.'s selective acylation of **57** to **58** with activated *N*-imidazole long chain fatty acids (n = 4, 6 and 8) over acetylation of ribonucleotide **55**.

Glycerol-3-phosphate is a natural phospholipid precursor of dihydroxyacetone **11** can be derived from the cyanosulfidic protometabolic network of Patel and co-workers.[28]   Work by Fernandez-Garcia *et al*. found that **57** can be selectively acylated with activated long chain *N*-imidazole fatty acids (n = 4, 6 and 8) to lipids **58** whilst a ribonucleotide competitor **55** is concurrently acetylated to compound **56** (Scheme 25).[189]   These results could suggest that ribonucleotide activation with molecules such as amino acids could occur during lipid formation.

In extant biology, fatty acids are biosynthetically formed by enzymatically catalysed aldol reactions which is mediated by co-enzyme A (CoA).[190,191] In brief, fatty acid synthesis starts with the transfer of **Malonyl-CoA** and **Acetyl-CoA** to a 4'-pantetheine prosthetic arm on an acyl protein carrier (ACP). Then the **Malonyl-ACP** undergoes enzyme catalysed decarboxylation to an enolate species with liberation of CO_2. The aldol reaction between acetyl-ACP and the enolate forms the 3' hydroxy species which can be reduced to crotonyl species **59**. This species can be elongated

over several cycles to give unsaturated fatty acids or reduced to an alkyl residue **60**. This species can also undergo subsequent cycles up to n =16 (Scheme 26). Ketoacids and the cofactor CoA are key components in metabolic processes such as the Krebs cycle and fatty acid synthesis.[192] These classes of molecules are integral to metabolism, so how was their non-enzymatic synthesis achieved on the early Earth?
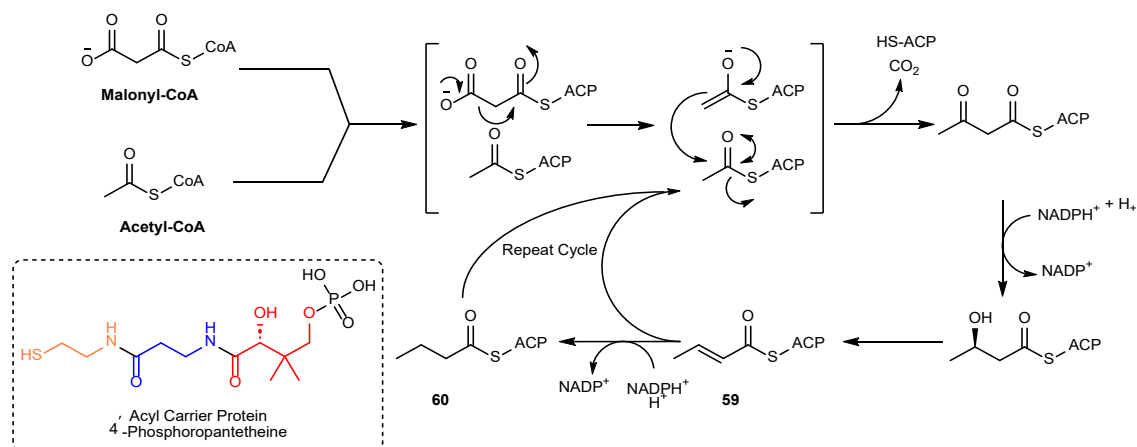
Scheme 26. The biosynthetic route of fatty acids from the CoA mediated aldol reaction of malonyl CoA and acetyl CoA in E. Coli.[190,191]

1.10. Metabolism

1.10.1.   Metabolism First?

Metabolism is the essential set of chemical reactions that cellular life utilises to function. There are three main properties of metabolism, firstly there is catabolism (e.g., Krebs Cycle, Glycolysis). These are the chemical reactions that breakdown complex molecules (food) to release the energy required for cellular processes and enable life to maintain an out of equilibrium state. Anabolism (e.g., Calvin Cycle, Reverse Citric Acid Cycle) is the building of the higher order molecules required by life built from ingested nutrients or smaller molecules that are released by catabolic pathways. The final function of metabolism is the removal of waste from the biological system.

The 'metabolism first' scenario proposed that life could have emerged from non-enzymatic chemical cycles that could have provided the energy required to synthesise heterotrophic precursor molecules e.g., nucleic acids and amino acids. This theory is conceptually attractive since a chemical system could in theory provide all the components required for life, which could then self-assemble into protocells.

Wächtershäuser was one of the first to propose that the earliest form of life originated in deep-sea volcanic hydrothermal vents. This hypothesis was informed by discovery of chemotrophic bacteria found in vents that could fix CO_2. This autotrophic origin of life was proposed to occur by a 'Iron-Sulphur world'.[193] It was thought that the oxidation of iron sulfide (FeS) to fool's gold (FeS_2) with H_2S would simultaneously reduce CO_2 from the atmosphere to formaldehyde and formate. From this point autocatalytic cycles would enable the synthesis of more complex molecules required for life. Although there have been significant contributions from the groups of Russel, Martin, the autotrophic theories origin of life has recently fallen foul of much criticism since there is little convincing experimental data to support the notion of autocatalytic metabolic cycles.

1.10.1.1 TCA Cycle

As discussed previously, metabolism is fundamentally the enzymatic synthesis of higher order molecules or the breakdown of complex biological molecules into smaller components. These processes normally require multistep pathways that rely on acetate, pyruvate, succinate, α-ketoglutarate, and oxaloacetate. These universal molecules are produced by the tricarboxylic acid (TCA) cycle (also known as the citric acid cycle or the Krebs cycle). The oxidative TCA cycle is catabolic that is, it breaks down metabolites into smaller units, in this case eventually to carbon dioxide (CO_2).

However, the reverse TCA (rTCA) cycle, which anabolically synthesises metabolites, is considered the most appealing to those who believe that CO_2-fixing metabolic cycles were an inevitable consequence of the geochemistry on early Earth. And as such is a fervent area of research in the field. The biological glyoxylate pathway is a shortcut within the oxidative TCA cycle that evades the CO_2-generating steps (isocitrate (C6) → α-ketoglutarate (C5) → succinate (C4)). The oxidative decarboxylations and α-ketoglutarate are bypassed through the splitting of isocitrate into succinate and glyoxylate. Succinate then continues in the cycle to generate oxaloacetate, while glyoxylate is intercepted by acetyl-CoA to feed back into the cycle (after hydrolysis) as malate, which can also continue in the oxidative direction to produce more oxaloacetate. The net benefit of the glyoxylate pathway is that it is anabolic and generates larger, complex organic molecules from acetate at a time when typical organic fuels (glucose, for example) are in short supply.

When run in reverse, a hypothetical reductive glyoxylate pathway is prebiotically appealing because it sidesteps the formidable reductive carboxylation steps of the rTCA cycle (succinate (C4) → α-ketoglutarate (C5) → isocitrate (C6). Simply reversing the steps, however, immediately causes a problem because the key carbon–carbon bond-forming step to generate isocitrate is inhibited by the inaccessibility of the carbon anion of succinate for reaction with glyoxylate.

1.10.2.   Thiols and the Origin of Life – A Prebiotic Synthesis of Cysteine

In biology cysteine is the primary source of sulfur and an essential feedstock of essential cofactors such as glutathione and CoA.[194] Cysteine is important in other biological process such as catalysis, electron transfer or can be traced back to ancient iron-sulphur proteins.[194,195,196,197] It is likely that cysteinyl thiol residues may have played a role in early biochemistry. However, attempts of the prebiotic synthesis of cysteine nitrile **Cys**-CN by the Strecker reaction from β-mercaptoaldehyde resulted in complex mixtures and a precipitate that was difficult to analyse and deconvolute the byproducts. A biomimetic strategy based on the biosynthesis of cysteine in plants or bacteria was proposed to overcome the challenges posed by the synthesis of cysteine nitrile **Cys**-CN through the Strecker reaction (Scheme 27).[198]
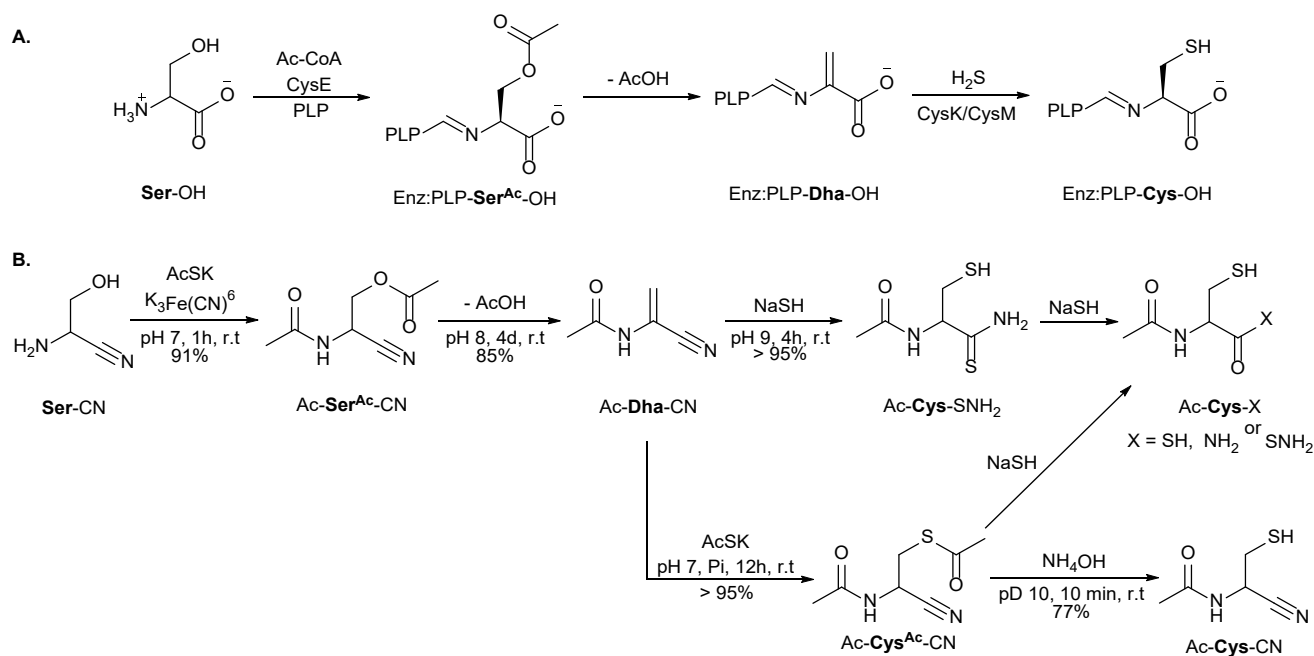
Scheme 27. (**A.**) Pyridoxal-5'-phosphate (PLP) dependant biosynthetic pathway to cysteine, (**B.**) The prebiotic synthesis of cysteinyl compounds from serine nitrile Ser-CN.[115]

The biosynthesis begins by the enzymatic acetylation of serine **Ser**-OH to *O*-acetylserine **Ser**^{Ac}-OH, this is followed by the β-elimination of acetic acid when **Ser**^{Ac}-OH is bound to PLP.[199] Adddition of H_2S or thiosulfate to PLP bound dehydroalanine residue Enz:PLP-**Dha**-OH results in a PLP bound cysteine Enz:PLP-

Cys-OH.[200] Powner and coworkers were able to exploit the biosynthesis of cysteine to find a plausible prebiotic synthesis of cysteine compounds from serine nitrile **Ser**-CN (Scheme 27).[198] The synthesis of *N*-acetyl cysteine nitrile Ac-**Cys**-CN began by the bis-acetylation of serine nitrile **Ser**-CN with potassium thioacetate **AcSK** and ferricyanide (Section 1.8.5) to give bis-acetyl serine nitrile Ac-**Ser**^{Ac}-CN in an excellent 91% yield. At pH 8 Ac-**Ser**^{Ac}-CN undergoes β-elimination to convert to *N*-acetyl dehydroalanine Ac-**Dha**-CN in 85% yield. The free amine of dehydroalanine **Dha**-CN had previously been synthesized by Eschenmoser but was found to be highly unstable.[201] Addition of H_2S quantitatively converted Ac-**Dha**-CN to the *N*-acetyl cysteine thioamide Ac-**Cys**-SNH$_2$. Alternatively, the addition of thioacetate at pH 7 conserves the nitrile functionality and forms the thioester Ac-**Cys**^{Ac}-CN. This intermediate could either undergo ammonolysis to the desired *N*-acetyl cysteine nitrile Ac-**Cys**-CN, which was unavailable by the Strecker reaction from β-mercaptoaldehyde. Or over 16 hours Ac-**Cys**^{Ac}-CN undergoes conversion to thioamide Ac-**Cys**-SNH$_2$ in 95% yield when incubated with H_2S. Further hydrolysis in the presence of H_2S results in the formation of *N*-acetyl cysteine thioacid Ac-**Cys**-SH 25%, *N*-acetyl cysteine amide Ac-**Cys**-NH$_2$ 12%, with 40% thioamide Ac-**Cys**-SNH$_2$ remaining unreacted in the reaction mixture after 6 days. This biomimetic strategy suggested that cysteinyl molecules may have been a secondary product from serine nitrile **Ser**-CN on the early Earth.

1.10.3.   CoA and the Thioester World

Cofactors such as Coenzyme A (CoA) play an integral part in metabolic processes and many cofactors have RNA relics incorporated into their structure (Section 1.3.3).[85] The reactivity of cofactors may indicate functionality that may have been lost as other molecular machinery took over. Today coenzymes are required for enzymes to function and increase their catalytic function in biochemical transformations. De Duve proposed that a 'thioester world' could have preceded genetically based life, where the latent energy stored in a thioester bond could at as a prebiotically plausible fore runner to high energy phosphorous ester bonds found in molecules like ATP.[172] The energy from these bonds could be used to drive

reactions to (heterotrophically) form complex molecules. CoA is involved with many enzymatic reactions and could be considered a node in metabolic systems.[202] It is according to Walsh one of the "big three" molecules of metabolism[202–205] In extant biology it plays a part in the non-ribosomal peptide synthesis, the Kreb's cycle, fatty acid biosynthesis and many more metabolic processes.[202,206] The reactivity of CoA is dictated by nucleophilic thiol cysteamine residue of the 'pantetheine arm' of CoA. Keefe and co-workers reported a plausible prebiotic synthesis of the pantetheine moiety of CoA from the heating of β-alanine, cysteamine and pantoyl lactone.[207] Despite this, the yields were poor (<1.8%) for a wide range of temperatures (60-118°C) with the best yield at less than 2% in sealed vials. No yield was observed if the vials were left open.[207] The limitations and the poor yields of this CoA synthesis will be explored further in the results and discussion of the prebiotic synthesis of pantetheine. Thioesters derived from metabolites such as CoA may have had an important part to play in the origins of life and the synthesis of peptides.
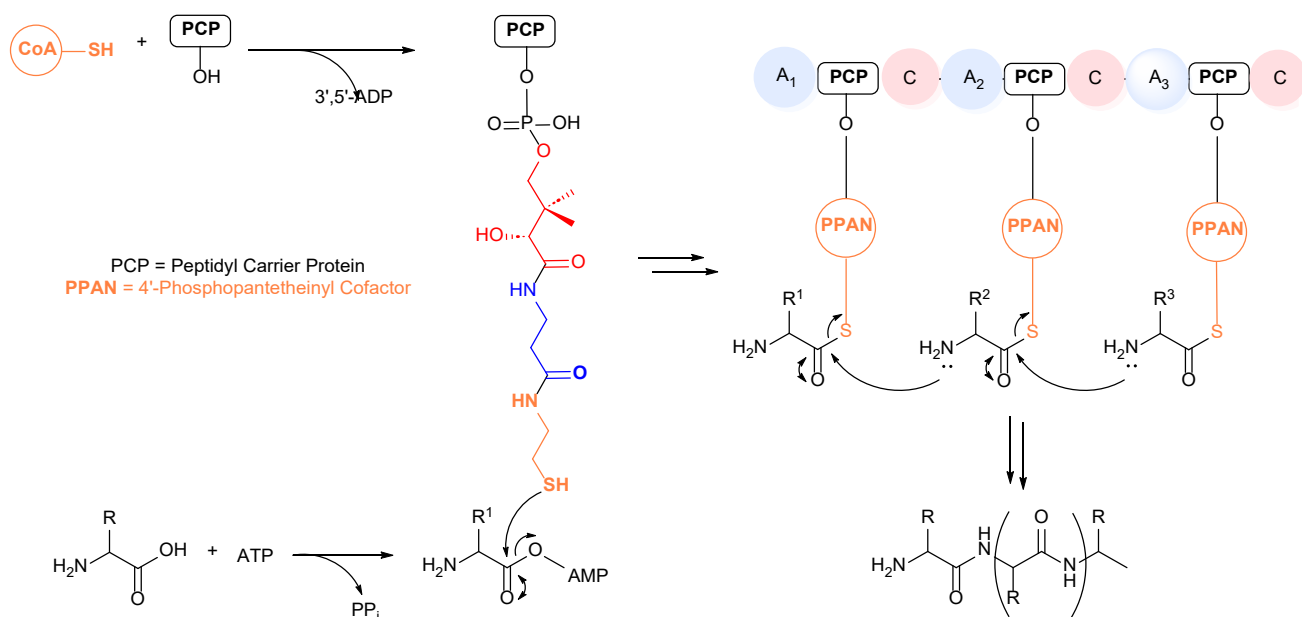
Figure 13. Schematic depicting the non-ribosomal peptide synthesis.

The 4'-phosphopantetheinyl arm derived from CoA binds to an AMP activated amino acid through the formation of a labile thioester bond. Whereas the route by ribosomal peptide synthesis proceeds by esterification onto the 3' position of the anticodon loop of tRNA of an AMP activated amino acid. Non-ribosomal peptide

synthesis starts with the linking of peptidyl carrier protein (PCP) linking with CoA to give a 4'-phosphopantetheinyl arm on the PCP. The newly incorporated PPAN thiol forms a thioester from ATP activated amino acids. A subsequent cascade of reactions takes advantage of the weak thioester bonds to form peptides.

1.10.4. Native Chemical Ligation

Modern peptide chemistry exploits the high reactivity of thioesters to ligate two peptide fragments together biomimetically in a similar fashion to non-ribosomal peptide synthesis in aqueous conditions (Scheme 28).
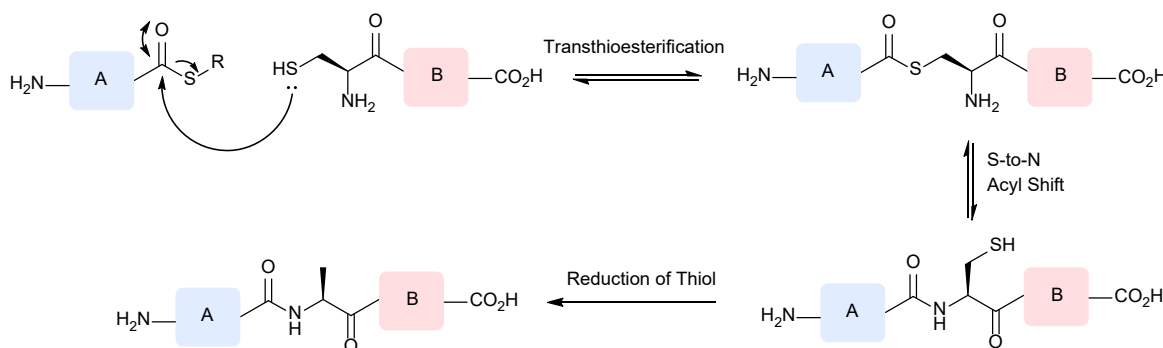
Scheme 28. Native Chemical Ligation.

Kent and coworkers exploited the in-built reactivity of a C-terminal thioester of one peptide chain so that it would undergo trans-thioesterifcation with the N-terminal cysteine residue of another peptide chain.[208] The intermediate thioester undergoes an intramolecular S to N acyl shift to form the amide bond and complete the ligation of two peptide fragments.[209] Today native chemical ligation is a routine procedure, but it requires the N-terminus to have a cysteine residue for ligation to occur and the yields decrease as the peptide fragment lengths increase. Recent advances have utilised reagents such Raney nickel or UV light to desulfurize the cysteine formed from NCL to form an alanine residue instead of a cysteine.[210] This procedure could be exploited in the prebiotic synthesis of peptides.

1.10.5. Catalytic Peptide Ligation with Thiols

Recently, Foden *et al.* reported a high-yielding, prebiotic synthesis of cysteine peptides. The work demonstrated a rare example of selective and efficient organocatalysis with *N*-acylcysteines. These thiols were able to catalyse peptide ligation, directly coupling kinetically stable -but energy-rich - α-amidonitriles to proteinogenic amines in water (Scheme 29).[198]
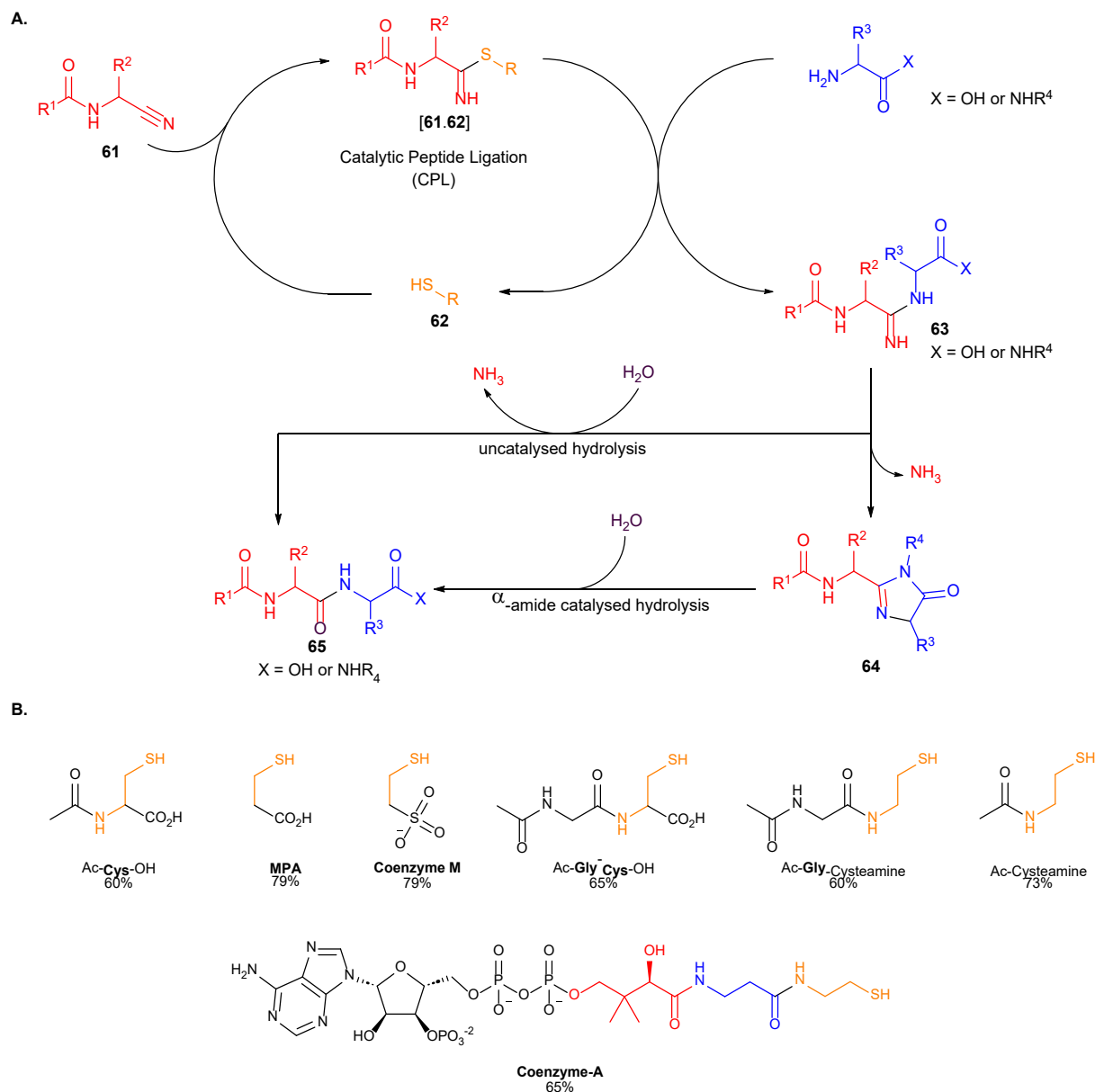
Scheme 29. (**A.**) The thiol catalysed dipeptide formation of Foden *et al.* and expanded on by Singh *et al.* (**B.**) A selection of high yielding thiols and their associated yields of Ac-**Gly^N**-**Gly**-OH from the reaction of Ac-**Gly**-CN (200 mM) and **Gly**-OH (1 equiv.) at pH 7 at 60 °C after 24h.[199,211]

This catalytic peptide ligation (CPL) is tolerant of amino acid side residues and utilized the kinetic stability and thermodynamic reactivity of α-aminonitrile **61** to deliver biomimetic *N* to *C* peptide growth without requiring any electrophilic activation. The amidine intermediate **63** observed during CPL suggested that the ligation occurs by the intermolecular addition of the nucleophilic coupling partner (amines) to a catalyst-bound thioimidate **61.62**. The amidine peptide eventually hydrolyses from cyclic intermediate **64** to peptide **65** under neutral conditions. It is notable that a single amino acid residue, cysteine, provided the catalysis for peptide ligation in water. The inherent catalytic activity of simple cysteinyl peptides made them excellent candidates as catalysts for protometabolic reactions in an abiotic environment. These chemical processes were ultimately inspired by non-ribosomal peptide synthesis (Section 1.10.3) and the chemistry of native chemical ligation (Section 1.10.4).

Sing *et al.* were able to build on this work to show that the outcome of ligation depends on the pH and that high pK_a primary thiols are the best catalysts.[199,211] Although the most rapid thiol catalyzed peptide ligation occurred at pH 8.5–9, the most selective peptide ligation, that was tolerant of all proteinogenic side chains, occurred at pH 7. Over the course of this study, the highly selective mechanism by which the intermediate peptidyl amidines undergo hydrolysis to α-peptides was discovered (Scheme 29). It was also found that the hydrolysis of peptidyl amidines with non-proteinogenic structures, such as β- and γ-peptides, had poor selectivity. This discovery enabled the highly α-selective protecting-group-free ligation of lysine peptides at neutral pH while leaving the functional ε-amine side chain intact.[211]

This ligation strategy further highlights how nitriles can serve as an early energy currency on the primordial Earth, perhaps acting as a forerunner to adenosine 5′-triphosphate and thiols (possibly thioesters as well) could have driven the first reactions to useful biopolymers before enzymatic process in cells could take over.

1.11.General Conclusion

In the 70 years since the Miller-Urey experiments there has been significant progress in unlocking the abiotic generation of biologically relevant molecules. The cumulative effort of many scientists has shed light on how the chemistry of the early Earth may have functioned, gives tantalising evidence on how life may have formed and why it takes its current form. The Powner group has contributed to this field by investigating the dynamic links and synergistic capacity of biological compounds through systems chemistry. Through the course of this general review, thiols such as coenzyme A and UV light has been shown to play a role in a plausible chemical network of the early Earth and the prebiotic syntheses of amino acids and nucleotides.

In this field, photochemistry has proven fertile ground for providing new routes to the synthesis of biological products and structurally complex molecules under geologically plausible mild conditions. However, photochemistry can suffer from unpredictable reactivity, substrate specificity and low conversions. Despite these shortcomings, photochemistry is still a powerful tool. For example, in industry the research into the sustainable synthesis of molecules has become a major stream so that there are greener alternative strategies over traditional synthetic techniques. As for origins of life scientists, photochemistry is especially relevant and may be a fundamental aspect to answering the question of "where do we come from?".

While the work described in this review has solved many problems, there still are many open questions. Whilst scientists have made in-roads towards linking the connections between nucleotides, peptides, and lipids. Work towards understanding the biologically universal cofactors is lacking. Finding prebiotic syntheses of universal cofactors is the beginning of the process to learning the relationship cofactors play in biology. Cofactors are an important source of biological data and the information gleaned in the study of cofactors could be triangulated with what has been learnt about prebiotic nucleotide and peptide chemistry to gain a deeper understand of the chemistry require to build a biological system.

2. Coenzyme A – Results & Discussion

2.1. Introduction

The chemistry that preceded life on Earth remains an open question and current efforts have focused on the origins of nucleotides and proteins (General Introduction). Enzymes are proteins that act as biological catalysts and can transform molecules into useful products that are required to sustain life. In many cases these proteins are inactive (apoenzymes) and require a cofactor to become an activated holoenzyme. Many cofactors are metal ions or small organic molecules that can either be bound permanently to a protein (prosthetic group) or transiently bound (co-substrate). In all cases cofactors act as helper molecules and facilitate important chemical transformations in extant biology (Figure 14).
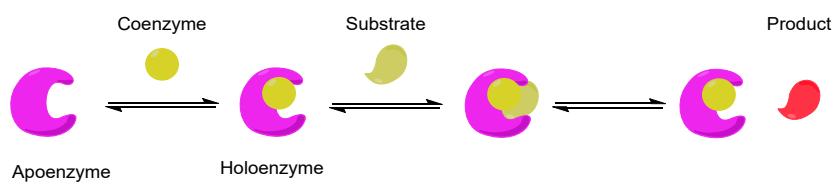
Figure 14. A simplified diagram showing how coenzymes participate in enzyme catalysis.

The prevalence of cofactors in biology suggests that they must have played a key role in the establishment of metabolic networks. Some scientists propose that cofactors may have formed before biopolymers and molecular replication took place. In this scenario, cofactors could have instigated protometabolic cycles and self-propagated through autocatalytic reactions. Another proposal asserts that cofactors exhibit poor catalytic properties by themselves and require enzyme-coenzyme complexes to be effective. Therefore, coenzymes would have to form in the presence of macromolecular templates such as ribozymes. For example, CoA is one of the smallest known ribozymes (or 'coribozymes'),[212] and it has recently been demonstrated that CoA can catalyse prebiotic peptide ligation in water (Section 1.10.5).[199,211] These 'holoribozymes' could catalyse the formation of more cellular building blocks thus increasing complexity rapidly. Despite some differences in how cofactors came into being on the early Earth, there is a broad consensus that

cofactors were important at the origin of life and therefore require further investigation.[85,103,105,172,181,195,213–215]

Coenzyme A (CoA) is a universal cofactor that is essential to all known life and is a cornerstone of metabolism. CoA is required for autotrophic carbon fixation pathways, fatty acid, polyketide, and non-ribosomal peptide syntheses.[192,216,217] CoA and its functional subunit, pantetheine **66**, are central to many proposed origin of life scenarios, but how and why these molecules emerged from prebiotic chemistry remains a mystery.[85,105,172,195,199,207,212,213,218] CoA contains two distinct fragments: pantetheine **66** and an adenosine nucleotide. Pantetheine consists of three fragments:[202,219] pantoic acid **67**,[220] β-alanine **68**,[28,58,59,90,220] and cysteamine **69** (Figure 15).[58,221–224]
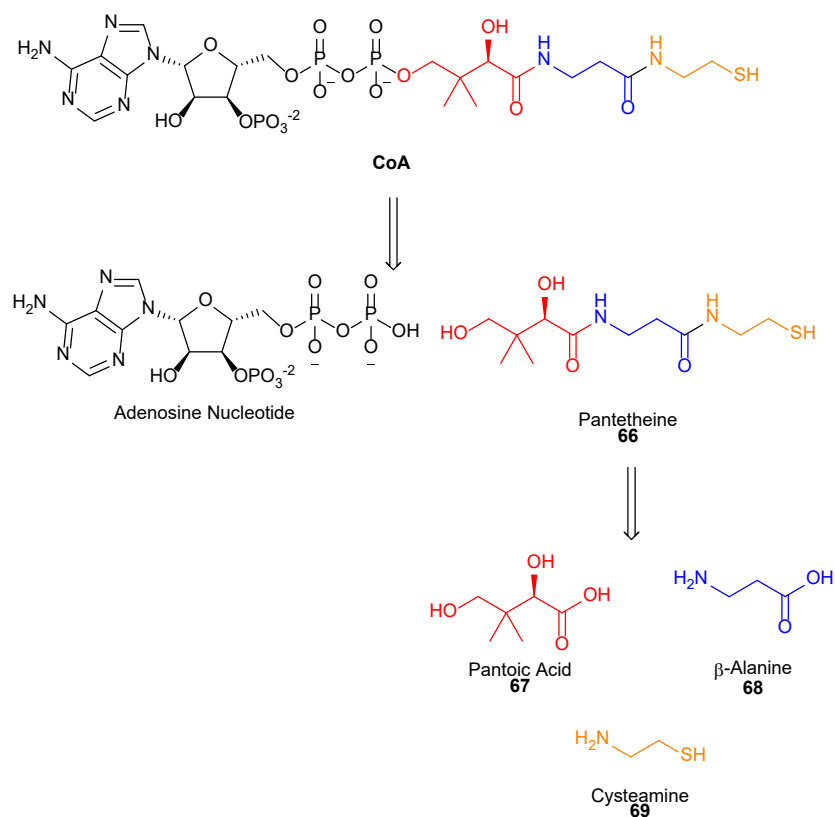
Figure 15. A retrosynthetic analysis of CoA. CoA can be fragmented into an adenosine nucleotide and pantetheine **66** moiety. Pantetheine **66** contains the reactive thiol functional group and can also be split into pantoic acid **67**, β-alanine **68** and cysteamine **69** fragments.

90

2.1.1. The Discovery of CoA

The origins for the discovery of CoA began in the late 1930s, when Williams and co-workers found a universal growth factor of universal biological occurrence in their studies with bacterial assays. Eventually Williams was able to verify the growth factor to be pantothenic acid.[225,226] Not long after, coenzyme A (the "A" stood for activated acetate) was identified by Lipmann in 1947.[213,227–229] Initially, Lipmann studied pyruvate oxidation in *Lactobacillus delbrueckii*, and found that acetyl phosphate could transfer phosphate to adenosine mono-phosphate (AMP) to form adenosine tri-phosphate (ATP).[230] Because of this work, Lipman began to think about the possibility that acetyl phosphate was an acetyl donor in the biosynthesis of essential metabolites and deduced the ability of ATP to function as an energy carrier. Lipmann initially intended to study acetyl transfer with acetal phosphate in pigeon liver. However, from these experiments he recognised that there was a unique factor that was not present in enzyme extracts but was evident in all organs of animals. Eventually, Lipmann was able to isolate and purify the factor from pig liver and discovered that its function was related to a coenzyme that was active in choline acetylation.[229,231] Pantothenic acid **70** was determined to be the central component of CoA by Guirard and Kaplan.[232,233] Lipmann won the Nobel prize in Medicine and Physiology for his discovery of coenzyme A along with Krebs for his discovery of the citric acid cycle in 1953. Eventually, Lynen was able to find that the active acetate was acetyl-CoA, a key intermediate in the metabolism of carbon compounds by all organisms. In 1964, Lynen was awarded the Nobel prize in Medicine and Physiology for his discovery of acetyl-CoA and its role in metabolism.

2.1.2. Biosynthesis of CoA

The biosynthesis of CoA in five steps from pantothenic acid **70** (vitamin B$_5$) is universal in both prokaryotes and eukaryotes. Animals and some microbes lack the capacity to form **70** and are totally dependent on external sources. However, most bacteria, plants and fungi can synthesise pantothenate **70** from aspartate and pantoic acid in another biosynthetic pathway. Initially the actual sequence of steps to produce CoA was unclear, Hoagland and Novelli suggested that pantothenic acid **70** reacts with cysteine to form pantothenoylcysteine, which is then decarboxylated to give pantetheine **66**. Subsequent phosphorylation of the primary alcohol with pantothenate kinase was proposed to form 4'-phosphopantetheine. Brown challenged this proposal and suggested an alternative route in which pantothenic acid **70** is phosphorylated to 4'-phosphor pantothenate first. Then the sequential condensation with cysteine which is then decarboxylated to yield 4'-phospho pantetheine. The final two steps were thought to be the same as those proposed by Hoagland and Novelli (Figure 16).[219,234,235]

The experimental evidence from studies on *Proteus morganni* supported Brown's hypothesis, which was verified with later studies with rat liver. In these studies, it was confirmed that the first step in the biosynthesis of CoA is the phosphorylation of pantotheic acid **70** to 4'-phosphopantothenic acid by pantothenate kinase (PanK). Then the condensation of 4'-phosphopantothenic acid with cysteine is followed by a decarboxylation to yield 4'-phosphopantetheine. These two transformations are catalysed by the 4'-phosphopantothenoylcysteine synthase (PPCS) and 4'-phosphopantothenoylcysteine decarboxylase (PPCDC) domains of a bifunctional enzyme in prokaryotes or by two distinct proteins in eukaryotes. Next, the AMP moiety is added to form dephospho CoA with ATP and phosphopanthetheine adenylyltransferase PPAT). Dephospho CoA is subsequently phosphorylated by dephospho CoA kinase (DPCK) at the 3' hydroxyl of the ribonucleotide to yield CoA.[219,234,235]
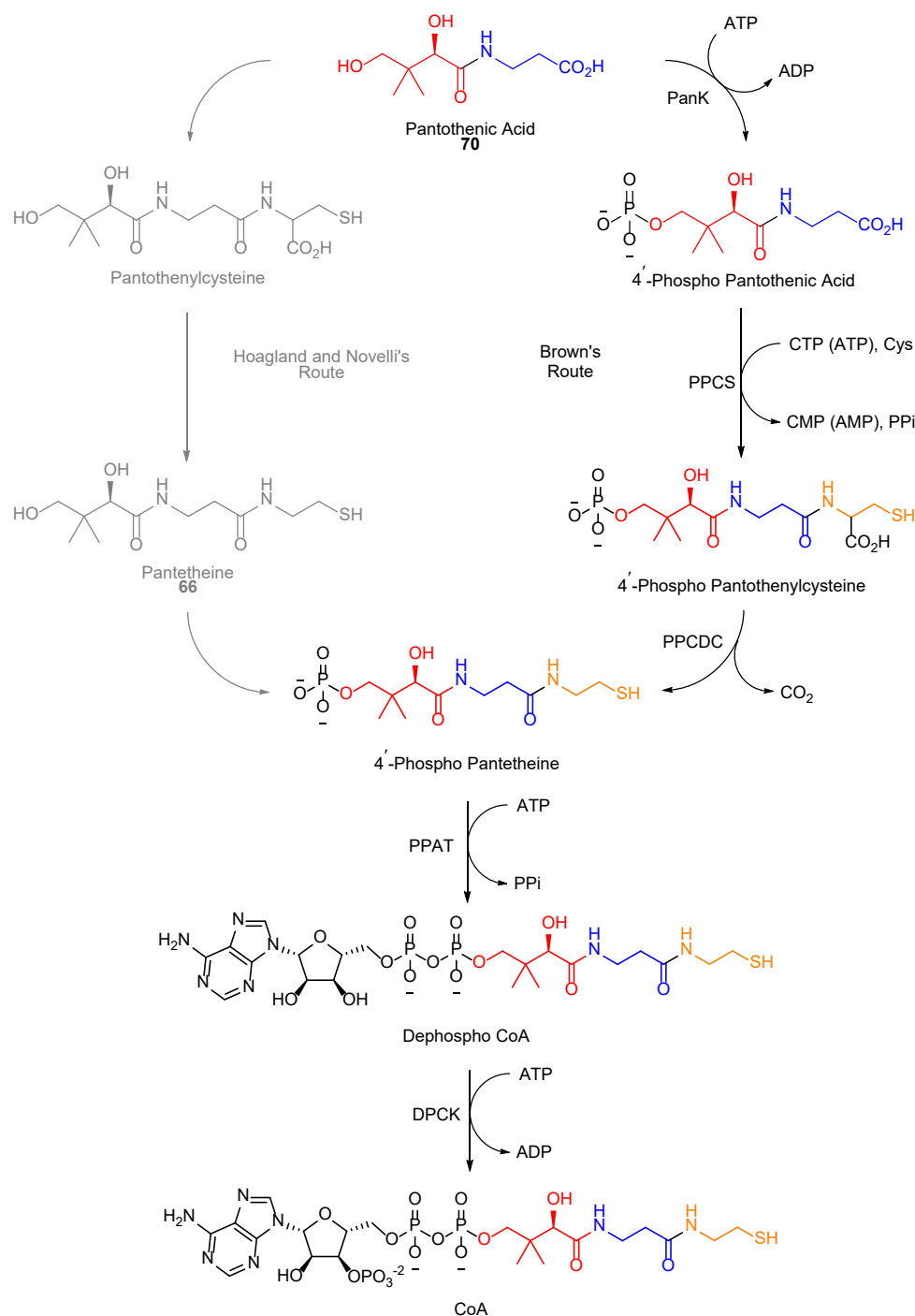
Figure 16. A simplified scheme depicting the proposed (greyed out) biosynthetic pathway of coenzyme A (CoA) by Hoagland and Novelli and the accepted biosynthesis by Brown *et al*. of CoA from pantothenic acid **70** in five steps.

CoA biosynthesis is a complex multistep pathway that consumes methylene tetrahydrofolate, nicotinamide dinucleotides, several nucleoside triphosphates, and requires pyruvoyl- and flavin-dependent decarboxylases.[202,219] These processes are

clearly a product of evolution and if CoA or pantetheine **66** were essential for a nascent metabolism then it must have been supplied by prebiotic chemistry.[236]

2.1.3. Project Aims – Why Pantetheine?

Although CoA contains an adenosine nucleotide moiety, the addition of the nucleotide to pantetheine **66** may have been a later evolutionary modification.[85,195] For example, nucleic acid cofactors, such as CoA, are incorporated into RNAs during the initiation of transcription in organisms,[237,238] and this may indicate that these cofactors are the ancient remnants of an RNA-based metabolism (Section 1.3.3).[85,105,212] However the ribozyme-catalysed incorporation of CoA into ribozymes could of also provided a mechanism by which pantetheine **66** was recruited by life prior to the advent of translation.[212,218,239]

Nevertheless, it is pantetheine **66** that contains the functional fragment that forms high-energy thioesters in enzyme active sites, whereas the nucleotide is a binding motif or lost during the transfer of **66** to enzymes (e.g Acyl Carrier Proteins).[172,216,217] Only the thiol-bearing pantetheine fragment is essential for catalytic activity, which led us to investigate the prebiotic origins of pantetheine **66**.[195,199]

Pantetheine **66** is a molecule of interest, since it is universally conserved across all living organisms; every genome sequenced requires pantetheine **66**.[203,205,240] Pantetheine is as highly conserved as DNA, RNA, and proteins, and is essential in biochemistry because CoA drives multiple anabolic (metabolic) pathways.[204] Thus, it must be an essential target for prebiotic chemistry, especially because the biochemistry of CoA (not other thiols) lies at the heart of many influential origins of life theories including the 'thioester world', the 'peptide world', and 'metabolism-first' concepts focused on replicating the (reverse)-TCA cycle or the acetyl-CoA pathway.[241–243] It was suspected that the structure of **66** and its relationship to RNA and proteins, if correctly analysed, would yield vital information about life's chemical origin and the interplay between biology's biopolymers.[28,103,135,174,199,215,244]

94

Previous attempts to condense pantoate, β-alanine and cysteamine fragments have until now produced near negligible pantetheine and were not able to provide a rationale for its unusual structure.[207,220] In this work these ineffective reactions were reinvestigated to verify the reasons why these reactions are poor and whether nitrile chemistry could be a superior pathway to pantetheine **66**. By building on the body of work developed by the group we show that chemoselective aldol and aminonitrile reactions are key to a selective synthesis of pantetheine **66**, because these reactions can selectively build panoate **67** and also exploit the intrinsic energy of nitriles whilst simultaneously by-passing the requirement for electrophilic activation of carboxylic acids to form amide bonds in water.[174,199,211] Therefore, in collaboration with Dr S. Islam and J. Singh the aim of the project was to investigate the synthesis of pantetheine **66** and discover selective chemical pathways to **66** by harnessing the unique reactivity of nitriles.[28,244]

2.2. Syntheses of Pantothenic Acid & Pantetheine in Water

2.2.1. Re-evaluating the synthesis of pantothenic acid and pantetheine through the reactions of carboxylic acids in water

Thirty years ago, Miller and coworkers attempts to uncover a prebiotic synthesis of pantetheine **66**.[207] Miller assumed that the direct condensation of pantolactone **71**, β-alanine **68**, and cysteamine **69** would generate **66**. Their studies were led by earlier observations that **68** and **69** could be accessed prebiotically.[207,220,223] For example, **68** and **69** have both been observed in a range of atmospheric spark-discharge and ionisation experiments[59,90,221,222,224,245] and **69** had also been accessed from the reaction of ammonia and thiirane[211,223] (Section 2.4.2). Building on these observations, the first step in Miller's synthetic route explored coupling pantolactone **71** and β-alanine **68** to generate the amide bond of pantothenic acid **71** (i.e., vitamin B$_5$).[207]

Lactone **71** is the simplest activated form of pantoic acid **67**, which is formally activated by dehydration. Pantolactone **71** is readily formed by the acid catalysed

cyclisation of **67** (≥ 88%, ≤ pH 3). Unfortunately, in neutral conditions the yields of **71** become negligible (< 4%, pH 6.1, 6 days) which may detract from the prebiotic plausibility of such a cyclisation forming, but also results in an incompatibility of the addition of high pK_{aH} amines such as β-alanine **68** to lactone **71** to achieve amide bond formation (Figure 17).
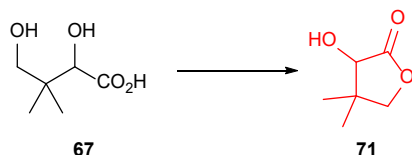
Scheme 30. The acid-catalysed cyclisation of pantothenic acid **70** to pantolactone **71**.

As such it was hypothesised that the reaction of pantolactone **71** with β-alanine **68** in water would produce only very low yields of pantothenic acid **6** due to preferential hydrolysis over coupling (Figure 17). We verified this by testing the reaction of pantolactone **71** in a wide range of concentrations (50 – 500 mM) with different stoichiometric quantities of β-alanine **68** (2 – 5 equiv.) from pH 7 to 10.
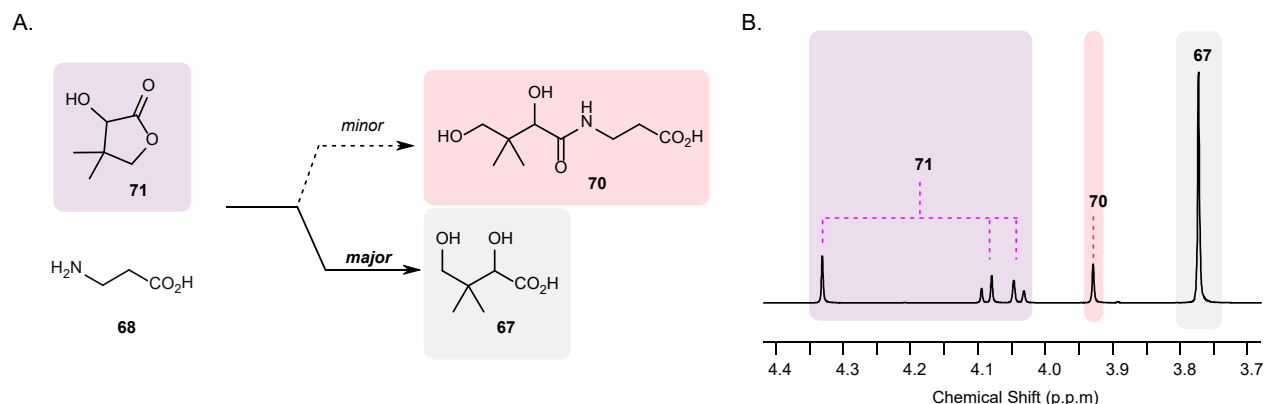
Figure 17. (**A.**) The reaction of pantolactone **71** with β-alanine **68** in water; (**B.**) ^{1}H NMR (600 MHz, H$_2$O/D$_2$O 99:1) spectrum to show the reaction of **71** (500 mM) in phosphate solution (pH 9; 500 mM) at 20 °C with β-alanine **68** (2 equiv.) yielding the hydrolysis product pantoate **67** (74%) and pantothenic acid **70** (12%) as the minor product after 7 days.

The results indicated that no reaction occurred at pH 7 under the conditions tested. At pH 9 with **71** (100 mM) resulted in low formation of **70**, even with 5 equivalents of β-alanine **68**. The best yield observed was that of pantolactone **71** (500 mM) and β-alanine **68** (1 M) which led to only 12% pantothenic acid **70** at pH 9 at 20 °C (Figure

17B). At pH 10, which is considered outside of prebiotically plausible conditions, the reaction of **71** (500 mM) with **68** (2 equiv.) resulted in a maximum 17% yield of **70**. This marginal improvement is likely observed because the pH begins to reach the the pK_{aH} of β-alanine (pK_{aH} = 10.5 at 20 °C).[246] Despite this result, the reaction was still dominated by hydrolysis of **71** to **67** in highly alkaline conditions. In summary, it was observed that the hydrolysis of pantolactone **71**, to yield pantoic acid **67**, outcompetes the coupling of β-alanine **68** with lactone **71** to diminish the yield of pantothenic acid **70** in water. The synthesis of **71** is incompatible with the synthesis of **70**, with the former requiring highly acidic conditions (pH < 3) to form pantolactone and the latter needing highly alkaline conditions (pH > 10) to form amide bonds.

Pantetheine universally carries a β-alanyl motif, therefore requires its pantoyl fragment to selectively couple with β-alanine (β-**Ala**-OH) **68** or a β-alanyl derivative. The reaction of pantolactone **71** with glycine **Gly**-OH was conducted to find out whether α-amino acids would also couple with **71** in water. In an analogous fashion to the reactions conducted with β-alanine **68** and lactone **71**, the hydrolysis of **71** outcompeted ligation with glycine **Gly**-OH at pH 9. The hydrolysis of **71** to pantoic acid **67** still also dominated at pH 10, however at pH 10 the pK_{aH} of **Gly**-OH (pK_{aH} = 9.6 at 20 °C)[159] is exceeded and incubation of pantolactone **71** with glycine **Gly**-OH at pH 10 resulted in appreciable formation of pantoyl-glycine **72** (31%) after 1 day.
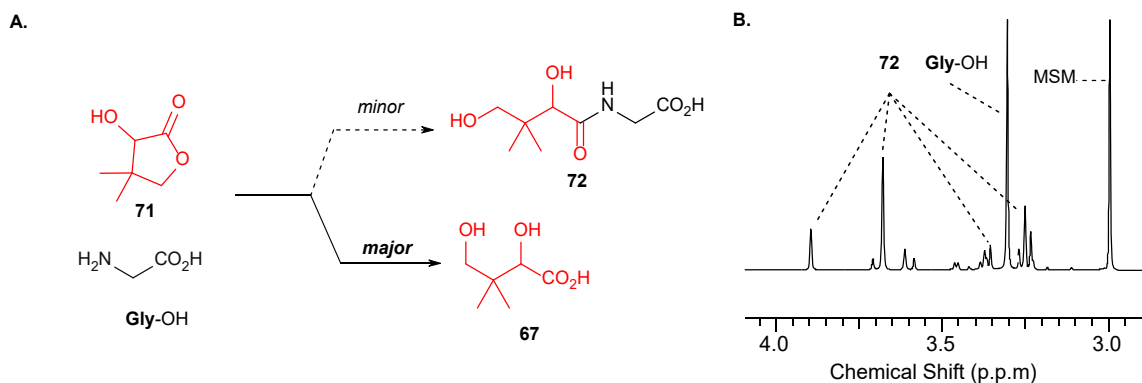
Figure 18. (**A.**) The reaction of pantolactone **71** with glycine **Gly**-OH in water; (**B.**) ^{1}H NMR (600 MHz, H_2O/D_2O 99:1) spectrum to show the reaction of **71** (100 mM) at pH 10 at 20 °C with glycine **Gly**-OH (2 equiv.) yielding the hydrolysis product pantoate **67** (69%) and pantoyl-glycine **72** (31%) as the minor product after 1 day.

Any synthesis of **66** must account for the canonical β-alanyl-structure that is a universal element of pantetheine **66**. However, no selectivity in favour of β-alanine **68** over other amino acids (e.g., glycine (**Gly**-OH), alanine (**Ala**-OH), or γ-aminobutyrate (**GABA**-OH)) has been reported for reaction with pantolactone **71**.[220] Indeed we observed the reverse selectivity with **72** being observed in higher yield than **70** under comparable conditions with **Gly**-OH and β-**Ala**-OH, respectively. With this in mind, the competition reaction of pantolactone **71** with β-alanine **68** and glycine **Gly**-OH was next tested.
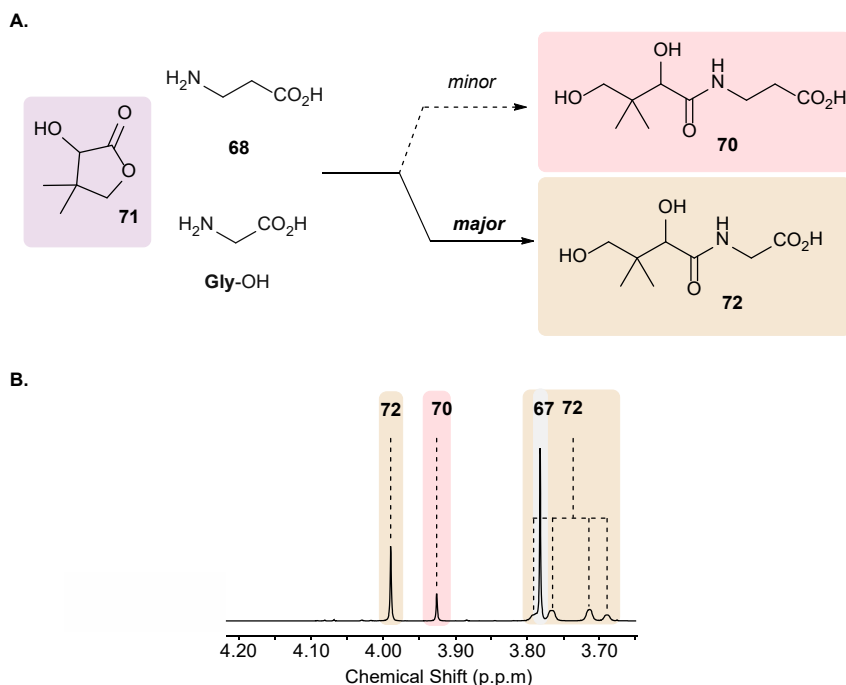
Figure 19. (**A.**) The reaction of pantolactone **71** with glycine **Gly**-OH in water; (**B.**) ¹H NMR (600 MHz, H₂O/D₂O 99:1) spectrum to show the reaction of **71** (100 mM) at pH 10 at 20 °C with glycine **Gly**-OH (2 equiv.) yielding the hydrolysis product pantoate **67** (69%) and pantoyl-glycine **72** (31%) as the minor product after 1 day.

At high concentrations the best result for conversion of **71** to amides was observed, for example the reaction of pantolactone **71** (500 mM), glycine (**Gly**-OH; 1 M), and β-alanine (**68**; 1 M), however this competition selectively yielding pantoyl-glycine **72** (29%) over the canonical pantothenic acid **70** (11%) and hydrolysis of **71** to **67** (59%) still dominated this reaction at pH 9. At lower concentrations of pantolactone **71** (100 mM) α-amino acids also react in preference to β-amino acids in competition

experiments however the total yield of amide was substantially suppressed relative to hydrolysis.

2.2.2. Conclusion

These results demonstrate that α-amino acids (e.g., glycine, **Gly**-OH) react selectively with pantolactone **71**, in preference over β-alanine **68** in solution. Therefore, the reaction of lactone **71** with amino acid mixtures would selectively yield non-canonical homologues of pantothenic acid **70** (Scheme 19). Thus, these results demonstrate the synthesis of pantetheine **66** is disfavoured from mixtures of α or β amino acids and so far, provide no compelling rationale for the observed canonical structure of **66**.[103,220,244]

2.2.3. An Investigation of Miller's Dry-State Reactions

Following on from their aqueous lactone experiments which suggested that pantolactone **71** selectively condenses with α-amino acids, in preference over β-alanine **68** in water,[220] Miller abandoned the synthesis of pantetheine **66** in water and attempted 'wet-dry' cycles.[207] The shift from water solutions to dry-heating was an attempt to drive forward the synthesis of amides from acids and amines; the rationale for wet-dry cycling given by the authors was that a '*drying lagoon*' or a '*hot, dry beach*' scenario would eliminate the detrimental effects of hydrolysis.[207,220] However, pantetheine **66**, and its purported precursor pantothenic acid **70**, have a λ-hydroxyl moiety which was anticipated to be predisposed to fragment these molecules irrespective of how "dry" the conditions are under which they are formed.

Miller also proposed that the poor selectivity of β-alanine **68** coupling with pantolactone **71** within mixtures of amino acids might be overcome by "drying conditions" as well.[220] He speculated that this was because β-alanine **68** '*is one of the most soluble amino acids*', **68** might separate from a mixture of other amino acids upon drying and subsequent increase in concentration. However, it remains untested whether the presence of other amino acids would interfere with pantetheine **66** synthesis under these 'dry' regimes, and separation has not been demonstrated

in support of this hypothesis. Our observations found that drying pantothenic acid **70** led to its degradation (rather than its synthesis) and means, like the original authors,[220] this thesis did not seek to test this secondary hypothesis of the spaciotemporal separation of α-amino acids and β-alanine **68**.

Previously reported attempts to synthesise **66** under conditions that would supposedly simulate a hot beach or a hot-dry 'lagoon margin' were carried out by concentrating aqueous mixtures of pantolactone **71**, β-alanine **68**, and cysteamine **69** at initial pH 7,[220] and then heating the residue in a sealed vessel. The yield of pantetheine **66** was stated as '*several percent*' at temperatures as low as 40ºC. However, the yield reported of **66** never went beyond 'permyriad' yields (i.e., 0.034% = 0.34‰) at 40 ºC. The maximum yield reported was no more than 1.7%, and this was only observed at >100ºC (Table 3). However, it should be noted that the reported 'yields' also fluctuated in an unexplained way over time.

Table 3. Yields of pantetheine **66** reported by Miller and co-workers[207] that after concentrating pantolactone **71** (1 mmol), β-alanine **68** (1 mmol), and cysteamine **69** (1 mmol) in 2 mL of water at pH 7 and then sealing the resultant residues (under an argon atmosphere) and heating at the specified temperature and for the specified time. [a]Numerical yields were not provided in reference 204 (Entries 3-11). These yields have been extrapolated from the graph in reference 204, figure 3).

Entry	Temperature (ºC)	Time (days)	Maximum yield of pantetheine **1** (%)
1	40	31	0.018
2	40	62	0.034
3[a]	60	35	0.25
4[a]	80	7	0.35
5[a]	100	1	0.40
6[a]	100	4	1.45
7[a]	100	6	1.25
8[a]	100	7	1.65
9[a]	118	1	1.45
10[a]	118	3	1.05
11[a]	118	4	1.10

In addition to the highly unsatisfactory (trace) and fluctuating yields of pantetheine **66**, the nature of the remaining material (98.3 – 99.9%) was not reported. Moreover, the only analysis carried out, was a subsequent product specific derivatisation of the

reaction mixture with monobromobimane,[207] this procedure did not address the by-products of this reaction. Therefore, the major outcome of these reactions was left unanswered, along with the fate of the unaccounted 98% of starting material. The reasons for such a remarkably low and fluctuating yield were also a puzzle. To understand and resolve these problems, these earlier studies were replicated with the advantage of modern high resolution analytical techniques (NMR) not available to or not used by Miller in the early 90s.[207]

To determine the fate of the > 98% of the material not reported in Miller's dry-heating reactions, next these reactions were repeated and analysed by NMR. The first set of reactions investigated the wet-dry cycling of pantolactone **71** (500 mM) and β-alanine **68** (500 mM) at pH 7, either by drying under a stream of air and then heating the residue in an open vessel at 100 °C for 24 hours. Or by drying down the solution at 100 °C for 24 hours after which the sample is analysed by NMR.
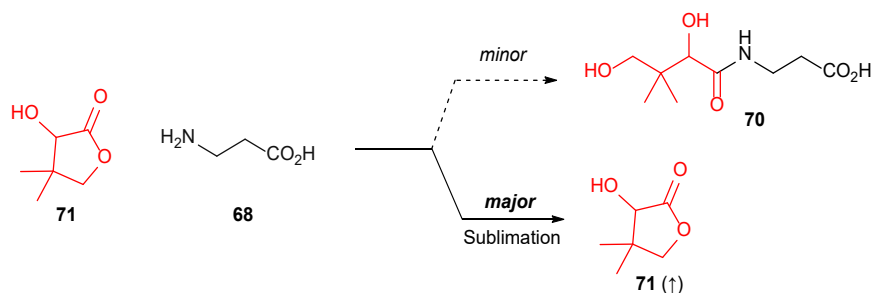
Scheme 30. The unsuccessful attempt at the synthesis of pantothenic acid **70** by the wet-dry cycling of pantolactone **71** and β-alanine **68**.

The results showed that pantolactone **71** undergoes facile sublimation in the dry state (Scheme 30 and Figure 20). For example, in the air-dried reaction of pantolactone **71** and β-alanine **68** at atmospheric pressure and 20ºC over 60 hours, the resultant dry-film was heated at 100ºC in an open system. After 24 hours, pantolactone **71** sublimed and was completely lost from the reaction vessel. No products derived from **71** were observed, only residual β-alanine **68** (Figure 20). Furthermore, directly heating a solution of **71** and **68** at 100ºC for 24 hours led to the majority of the initial pantolactone **71** (84%) being lost to sublimation; only 13% residual **71** was observed after 24 hours, alongside trace **70** that could not be

quantified by NMR spectroscopy (Figure 20). This reaction was clearly ineffective, and no further effort was expended quantifying these yields.
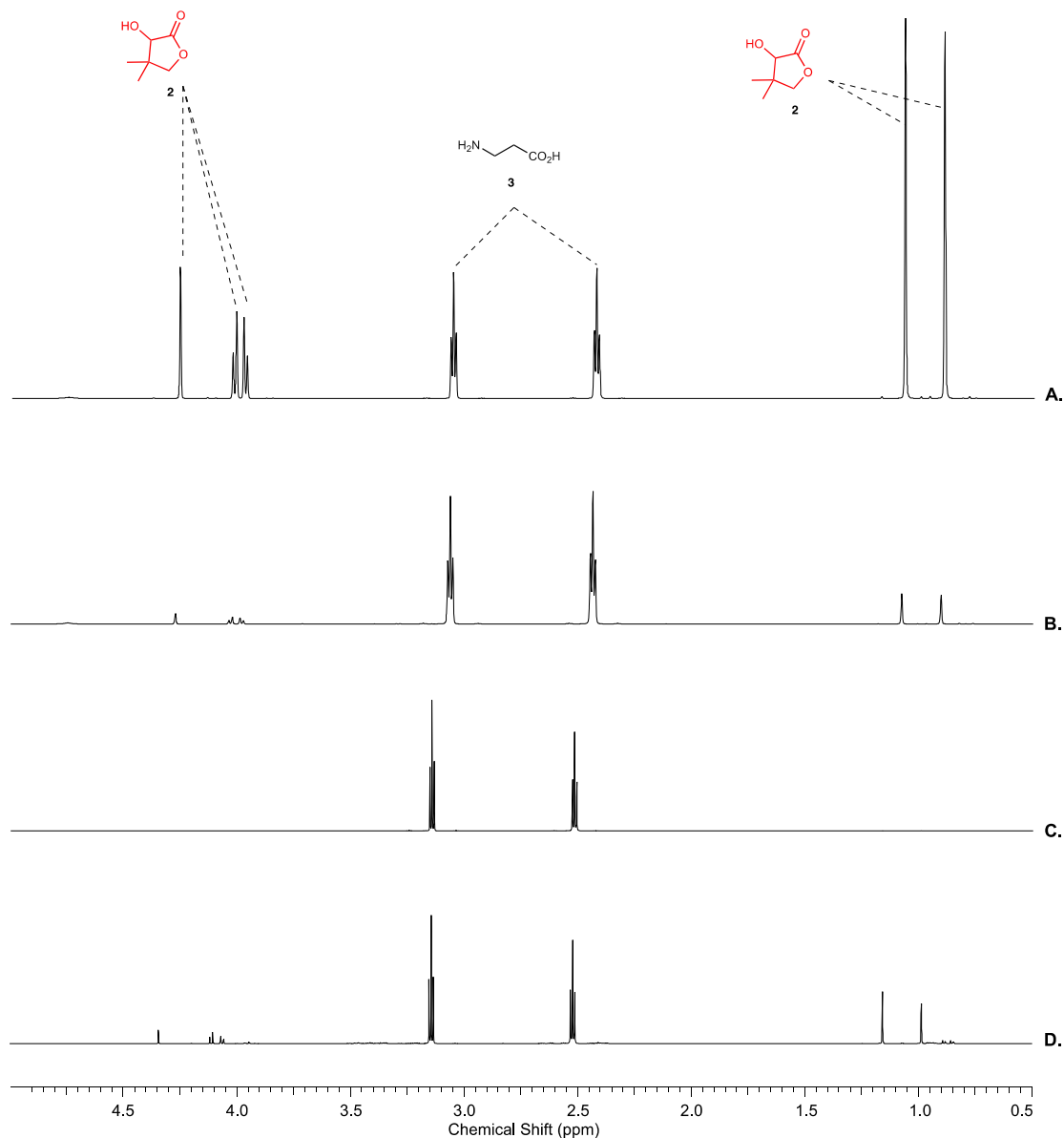
Figure 20. ^{1}H NMR (600 MHz, H$_2$O, noessygppr1d) spectra to show (**A.**) pantolactone **71** (500 mM) and β-alanine **68** (1 equiv.) at pH 7.0; (**B.**) after leaving pantolactone **71** (500 mM) with β-alanine **68** (1 equiv.) at pH 7.0 *in vacuo* (0.04–0.08 mbar) at 20 ºC for 2 hours; (**C.**) after leaving a sample of pantolactone **71** (500 mM) and β-alanine **68** (1 equiv.) at pH 7.0 at atmospheric pressure under a slow stream of air for 60 hours at 20 ºC, followed by heating of the resulting residue at 100 °C for 24 hours at atmospheric pressure; (**D.**) after heating a sample of pantolactone **71** (500 mM) with β-alanine **68** (1 equiv.) at pH 7.0 at 100 °C for 24 hours at atmospheric pressure.

Heating dried residues of pantolactone **71**, β-alanine **68** and cysteamine **69** in an open system also led to substantial (78%) loss of **71** by sublimation. Pantetheine **1**

was not observed in any open vessel reaction of **71**, **68** and **69**, only traces of pantetheine **66** (< 1%) were observed when the solutions were initially dried the to a residue at 20ºC and 220 mbar in a desiccator over P_2O_5, and then heated at 100 ºC in a closed system.[207] Whilst some pantolactone **71** was still lost (25%), it is likely that it was lost during the low temperature drying (open) phase of these reactions, most of pantolactone **71** (75%) was retained if the reactions were sealed before heating.   However, although sealing these reactions curtailed the loss of pantolactone **71**, only trace pantetheine **66** (< 1%) was observed, alongside pantolactone **71** (60%) and a complex mixture of by-products derived from β-alanine **68** and cysteamine **69**. Pantolactone **71** was, as expected internally sublimed within sealed vessels, phase-separating from **68** and **69**. These results indicated that this dry-heating strategy (even in a prebiotically impossible sealed vial) is a highly unsatisfactory method for the synthesis of **66**.
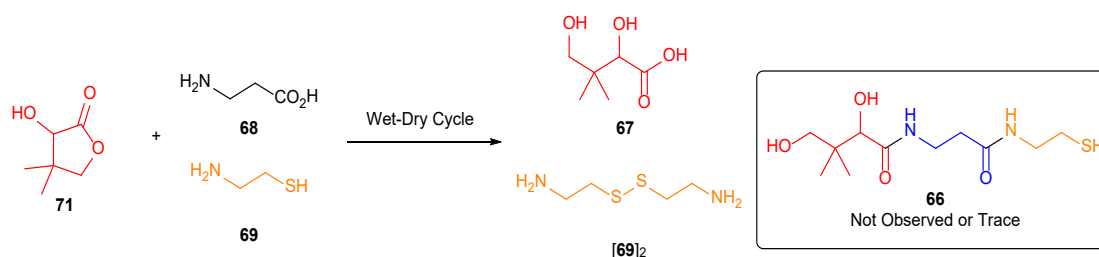
Scheme 31. The unsuccessful reaction of pantolactone **71**, β-alanine **68** and cysteamine **69** heating in wet-dry cycles to pantetheine **66**.

The stability of the desired product of of "dry-state" amide bond formation had not been reported by Miller and co-workers, therefore pantothenic acid **70** was also subjected to the conditions reported for the dry-state synthesis of pantetheine **66**.[207] We suspected that **70** would decompose under dry-state reaction conditions.[247] Upon heating pantothenic acid salts **70**-M^{n+} at 100 °C for 24 hours, the fragmentation of **70**-M^{n+} to β-alanine **68** and pantolactone **71**, as well as the hydrolysis of pantolactone **71** (or amide **70**-M^{n+}) to pantoate **67** was observed. After 24 hours only 46% of the hemi-calcium salt of pantothenic acid [**70**]$_2$$Ca^{2+}$ remained; the degradation products pantolactone **71** (3%), β-alanine **68** (51%) and pantoic acid **67** (51%) were observed. The sodium salt was observed to be slightly more stable under these conditions however, the sodium salt of **70**-Na^+ still fragmented to

pantolactone **71** (3%), β-alanine **68** (25%) and pantoic acid **67** (17%). These results demonstrate that the fragmentation and hydrolysis of **70** are substantially more effective than the synthesis of **70** under these 'hot and dry' reaction conditions (Scheme 32).
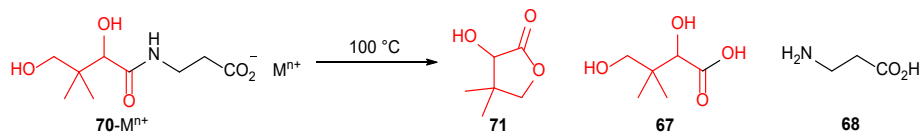
Scheme 32. Decomposition of pantothenic acid salts (Ca^{+2} and Na^{+}) under dry-heating conditions.

2.2.4. Conclusion

These results demonstrate that even pure β-alanine **68** and pantolactone **71** cannot be effectively coupled by drying/heating, and that the desired product of this coupling, pantothenic acid **70**, is not stable under these conditions. Pantothenic acid **70** decomposes to **67** and **68** under hot, dry conditions (Scheme 32). These results confirmed that wet-dry cycles are a highly unsatisfactory method for pantetheine **66** synthesis. Which led us to suspect that alternative substrates and mechanisms were needed for the effective synthesis of **66**.

Heating pantolactone **71**, β-alanine **68** and cysteamine **69** in wet-dry cycles formed pantetheine **66** in trace yield (< 1%) in agreement with Miller's results. However, **66** was only observed when heating was carried out in a closed system. The remarkably low yield of **66** in these experiments revealed two (previously unreported) problems that severely limit the synthesis of pantetheine **66** in the dry state. Firstly, pantolactone **71** sublimes, and secondly, the fragmentation of pantothenic acid **70** is more effective than its synthesis upon dry heating.

Ultimately, **66** was only detected if dry-heating reactions were carried out in an (artificially) closed system. Miller and co-workers placed their samples in glass vials and evaporated them *in vacuo* (vacuum pressure not reported). The vials were then evacuated, sealed, and then heated. It was stated that the reactions had been

sealed "*to avoid oxidation*".[207] As such, sealed vessel is an inaccurate model for an open system such as hot-dry '*beach*' or a '*lagoon margin*'. Only upon drying down and then sealing these reactions was even trace synthesis of **1** observed, therefore based on these results the major reaction pathways do not afford pantoyl products.

2.2.5. An Investigation of Pantetheine Synthesis with Electrophilic Activating Agents in Water

The synthesis of cyanamide **22** from ammonium cyanide or methane, ammonia and water mixtures under UV irradiation has been demonstrated under plausible conditions found on the primitive earth. Carbodiimides are currently used as amide coupling agents in chemical synthesis (Section 1.8.2) and it has been postulated that cyanamide **22** and its tautomer carbodiimide **72** is a plausible condensing agent for the polymerization of amino acids on the early Earth.[248]
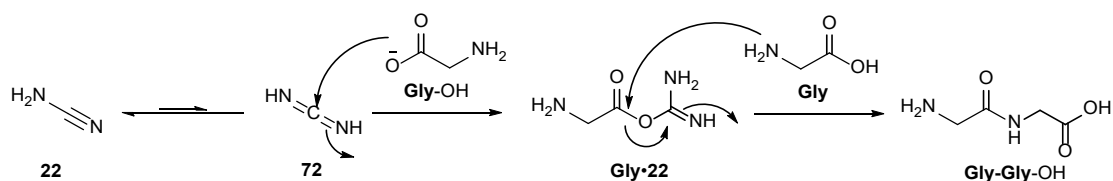
Scheme 33. The proposed formation of dipeptide **Gly-Gly**-OH with cyanamide **22** as an activating agent.

One of the first attempts at peptide formation with cyanamide **22** in solution was achieved by Peterson, who observed the formation of four dipeptides and tripeptide when an aqueous solution of glycine and leucine was exposed to UV light in the presence of cyanamide.[249] However, the yields of the peptides were low. Nooner and Oró were able to improve the yields of peptide formation with cyanamide **22** by drying down samples of glycine, isoleucine, or phenylalanine with adenosine 5'-triphosphate (ATP) and 4-amino-5-imidazolecarboxamide **AICA** additives at 90 °C over 24 hours. Peptides were formed in yields of 5%, 17% and 66%, respectively with glycine and phenylalanine producing mixtures of di-, tri- and tetrapeptides, while *L*-isoleucine gave only the dipeptide in detectable quantities.[250] The scope of amino acids was expanded under these conditions. Three histidyl peptides in yields of up

to 11% was observed when aqueous solutions of histidine, leucine, ATP, cyanamide **22** and MgCl$_2$ were evaporated and heated for 24 h at 80 °C. In addition, peptides of phenylalanine, leucine and alanine were formed in yields of up to 56%, 35%, and 21%, under the same conditions. The observation of amino acid-cyanamide adduct **Gly·20** as the major side product in this series of reactions also indicated the mechanistic pathways these reactions followed.[250]

More recently, Pascal *et al*. used *N*-acylated amino acids to extend sequentially the peptide chain with cyanamide in water.[251] Through the reaction of *N*-protected alanine **Ala**-OH cyanamide **22** and glycine **Gly**-OH, they observed oxazolone **73** which increased the rate of reaction over direct nucleophilic attack on activated amino acids.[251]
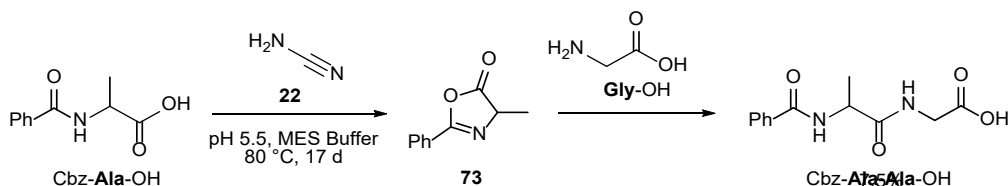
Scheme 34. Pascal and co-worker's peptide formation from *N*-acetylated amino acid Cbz-**Ala**-OH activated by cyanamide **22** to form oxazolone intermediate **73** which increases the rate of reaction to dipeptide Cbz-**Ala-Ala**-OH.[251]

Although electrophilic activating agents (e.g., anhydrides and carbodiimides) are often viewed as model prebiotic condensation reagents there has notable lack of selectivity and controlled polymerisation in the literature. In one of the few examples, Orgel reported that α-amino acids, unlike β-amino acids, are selectively oligomerised by electrophilic activating agents.[251–255] Since pantetheine **66** has an β-alanyl unit in its canonical structure, it is important to explore these potential selective synthetic pathways towards pantetheine **66** in water.[252]

To investigate this hypothesis, two different model prebiotic activating agents were tested in the initial attempts to synthesise pantetheine **66** in water. The first activating agent tested, acetyl phosphate **AcP**, has been viewed as an simpler analogue of adenosine triphosphate (ATP),[255] whilst the second 1-ethyl-3-(3-

dimethylaminopropyl)carbodiimide (EDC),[251–254,256,257] is considered a water-soluble electrophilic activating agent, and ostensibly a model of carbodiimide,[257] which can be considered a tautomer of cyanamide **22**.[28,135,244] Therefore, we surveyed the application of electrophilic carboxylic acid activation to the synthesis of pantetheine.

Acetyl phosphate **AcP** has been proposed as a 'primordial energy currency' in the absence of adenosine triphosphate (ATP) and a potential 'activating agent' for the polymerisation of amino acids in water.[255,258] Lane *et al*. were able to phosphorylate adenosine diphosphate (ADP) to ATP with acetyl phosphate **AcP** in a modest 2% yield, however **AcP** was observed to exclusively *N*-acetylate amino acids and not instigate polymerisation in water.[255,258] Following on from this work Pinna and co-workers were able to improve the phosphorylation of ADP to ATP in up to 20% yield by utilising Fe^{3+}.[255,258] However, they did not report the activation amino acids with ATP under their conditions for ATP synthesis. To investigate the application of **AcP** to the synthesis of pantetheine **66**, pantoic acid **67**, β-alanine **68**, cysteamine **69** and acetyl phosphate **AcP** was incubated in water. Only the chemoselective acetylation of cysteamine **69** to yield *N*-acetyl-cysteamine Ac-**69** was observed alongside hydrolysis of acetyl phosphate **AcP** to acetic acid and phosphate.
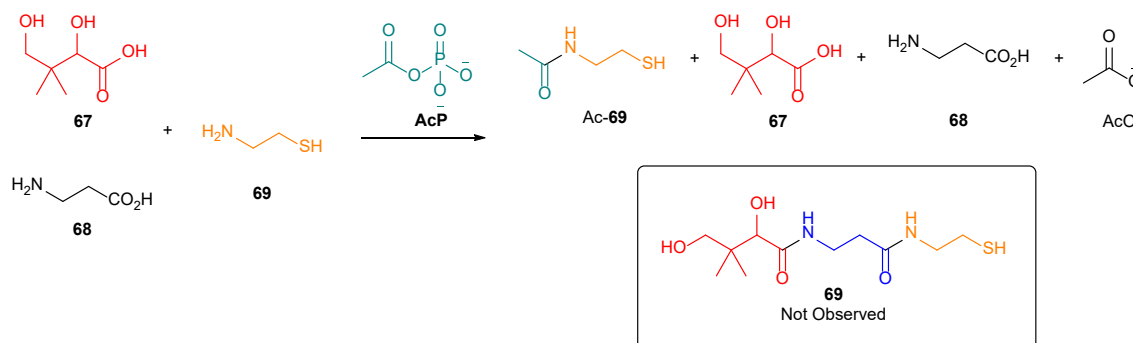
Scheme 35. Attempted synthesis of pantetheine **66** from pantoate **67**, β-alanine **68**, cysteamine **69** using acetyl-phosphate **AcP** as a potential activating agent.

Following on from these results, the reaction of pantothenic acid **70** (100 mM) and cysteamine **69** (1 equiv.) with acetyl-phosphate **ACP** (1.1 equiv.) at pH 7 resulted in the selective acetylation of cysteamine to *N*-acetyl-cysteamine Ac-**69** in 31%, while 40% of the acetyl-phosphate was hydrolysed to acetate. No pantetheine was detected in these sets of reactions. As an aside, acetyl-phosphate (1.1 equiv.) was

able to acetylate the amino acid precursor glycine nitrile **Gly**-CN (100 mM) to *N*-acetyl glycine nitrile Ac-**Gly**-CN (pH 7 = 15%, pH 9 = 63%) in phosphate buffer after 24 hours. These results demonstrate the incompatibility of acetyl-phosphate **AcP**, as an 'activating agent', with the synthesis of pantetheine **66**. Furthermore, no evidence for carboxylic acid activation by acetyl phosphate **AcP** was observed, and so EDC activations were investigated next.
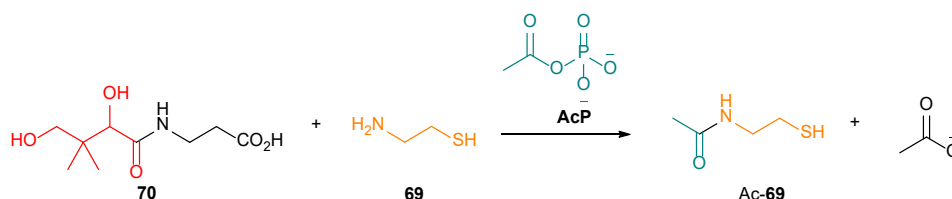
Scheme 36. Attempted synthesis of pantetheine **66** from pantothenic acid **70** using acetyl-phosphate **AcP** as a potential prebiotic activating agent. Only *N*-acetyl cysteamine Ac-**69** formation was observed (31%).

The reactions of EDC with pantoic acid **67**, β-alanine **68** or cysteamine **69** were studied to understand and formulate a strategy to synthesise pantetheine **66** using EDC activation in water. As will be discussed, adverse reactions of these substrates will prevent the synthesis of pantetheine **66**.

The reaction of pantoic acid **67** (100 mM) and a slight excess of EDC (1.1 equiv.) was incubated at pH 5.5-9 over 24 hours. In basic conditions cyclisation to pantolactone was poor with only 5% **71** observed and 95% of the starting material recovered. This could be explained by the concomitant formation and hydrolysis of pantolactone **71** at pH 9. When the reaction was repeated at pH 7, the yield of pantolactone **71** improved to 32% and almost 90% pantolactone **71** was formed with the same reaction at pH 5.5. In acidic conditions lactone **71** formation is favoured (Section 2.2.3), and in these conditions the hydrolysis to **67** is also suppressed. Although a stoichiometric quantity of EDC is expended to cyclise pantoic acid at pH 5.5, the lactone is considered an activated form of pantoic acid and available for nucleophilic addition with amines such as β-alanine **68**. Unfortunately, the pH would need to be high enough to overcome the pK_{aH} of β-alanine to couple to **71** (Section 2.2.1).
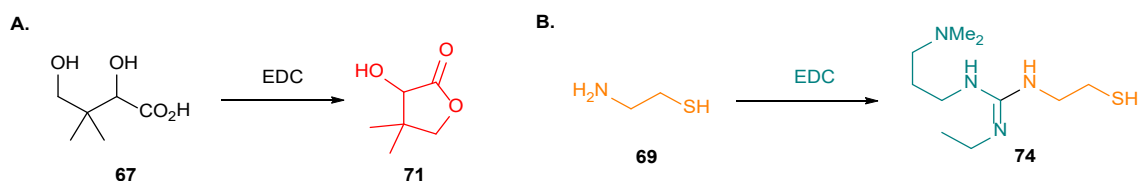
Scheme 37. (**A.**) The EDC induced cyclisation of pantoic acid **67** to pantolactone **71**; (**B.**) The EDC *N*-guanidylation of cysteamine **69**.

EDC activation of carboxylic acids was also found to be incompatible with cysteamine **69**. Incubation of cysteamine **69** (100 mM) with EDC (~ 1 equiv.) at either pH 7 or 9 resulted in the rapid (15 minutes) and selective *N*-guanidylation of **69** to compound **74** (Scheme 37B). This reaction prevented any onward reaction of **69**, and therefore, like acetyl phosphate **AcP**, the reactions with EDC would block the incorporation of cysteamine **69** into pantetheine **66**.
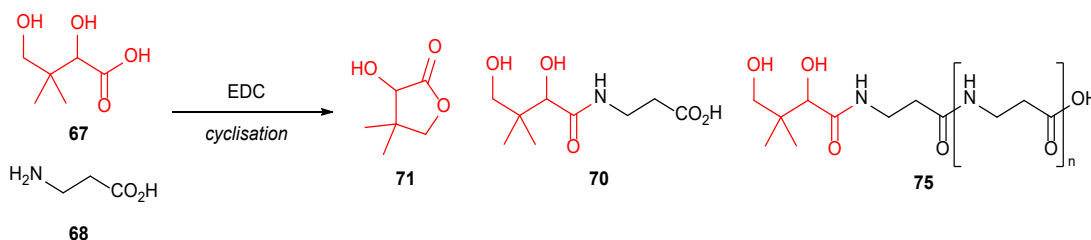
Scheme 38. The EDC induced cyclisation of pantoic acid **67** to pantolactone **71** over the coupling of pantoic acid **67** with β-alanine **68** to give pantothenic acid **70**.

The attempt to form pantothenic acid **70** from pantoate **67** and β-alanine **68** with by activation of carboxylic acids with EDC resulted in the preferential cyclisation of pantoic acid **67** (100 mM) to pantolactone **71** (pH 7 = 32%, pH 5.5 = 78%) instead. The EDC-activated coupling of **67** with β-alanine **68** only gave low yields of pantothenic acid **70** (pH 7 = 4%, pH 5.5 = 7%). At pH 5.5 new triplet resonances (2.50-2.44 p.p.m, *J* = 7.0 Hz) also suggested that β-alanine **68** underwent further EDC-mediated side reactions, including reaction with pantothenic acid **70** to form oligomeric products **75**. These oligomeric byproducts were not further characterised due to the complex overlapping NMR signals (Scheme 38). The minimal yields of **70** would prevent another round of EDC coupling with cysteamine **69** to form pantetheine **66** in any appreciable quantities. The availability of cysteamine for coupling to **70** would be also be limited due to guanidylation.
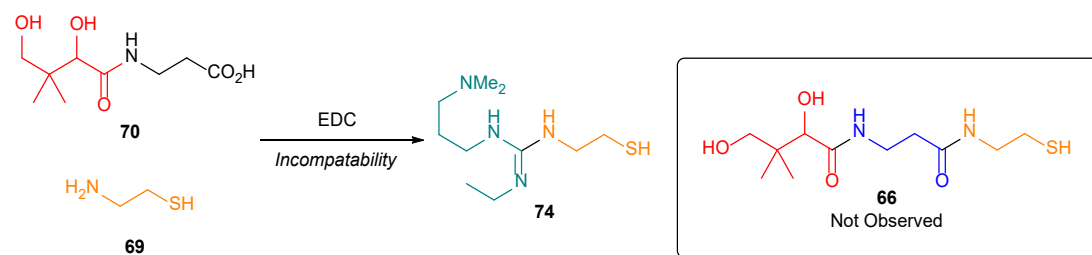
Scheme 39. The EDC-mediated *N*-guanidylation of cysteamine **69** to **74** over the coupling of pantothenic acid **70** with cysteamine **69** to give pantetheine **66**.

The ability of pantothenic acid **70** to couple with cysteamine **69** to form pantetheine **66** by coupling with EDC was tested. Unfortunately, guanidylation also occurred in the presence of pantothenic acid **70** (100 mM) and cysteamine **69** (1 equiv.) with EDC (1 equiv.). At pH 5.5-9 quantitative (> 95%) conversion of **69** to **74** was observed. The pantothenic acid **70** also remained unchanged, and the EDC was fully consumed during the transformation into guanidine **74**. The reactions were spiked with commercial pantetheine **66**, resulting in the appearance of new resonance peaks and this confirmed that the condensation of pantothenic acid **70** with cysteamine **69** had not occurred. EDC underwent preferential reaction with cysteamine **69** to form the guanidine **74** and the EDC was not available to fragment the pantothenic acid further (Scheme 39).
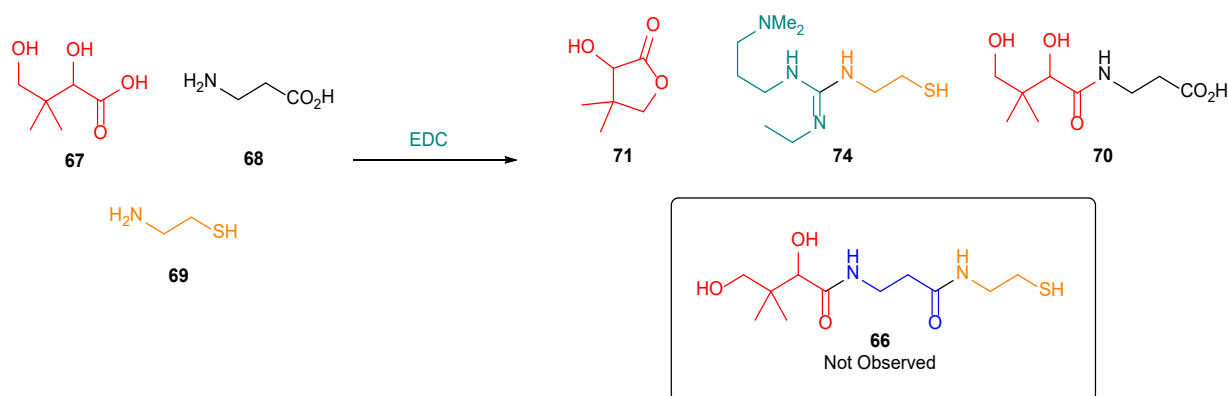
Scheme 40. The multicomponent reaction of pantoate **67** (100 mM), β-alanine **68** (100 mM), cysteamine **69** (100 mM) and EDC (2 x 1.1 equiv.), showing the favourable EDC-mediated cyclisation of pantoic acid **67** to pantolactone **71** and the EDC *N*-guanidylation of cysteamine **69** over the formation of pantothenic acid **70** and pantetheine **66** which was not detected in these reactions.

A multicomponent reaction of β-alanine **68** (100 mM), cysteamine **69** (100 mM), pantoic acid **67** (100 mM) and EDC was added in two batches, over two days (2 x

1.1 equiv., 2.2 equiv. total) resulted in the quantitative (> 95%) formation of guanidine **74** at pH 5.5-9. At pH 5.5, upon the second addition of EDC, pantoic acid **67** slowly yielded lactone **71** (35%) and outcompeted the synthesis of pantothenic acid **70** (< 6%). At pH 7 only 14% pantolactone was formed and at pH 9 no lactone was observed, only the quantitative recovery of pantoic acid **67** and β-alanine **68**.

These reactions demonstrate that the condensation of pantoic acid **67**, β-alanine **68** and cysteamine **69** to pantetheine **66** did not occur through EDC-mediated activation. Instead, EDC underwent preferential reaction with cysteamine **69** to form the guanidine **74**. Furthermore, EDC-mediated cyclisation of pantoic acid **67** to pantolactone **71** is preferential over the condensation of pantoic acid **67** with β-alanine **68**. These experiments confirmed the incompatibility of carbodiimide-mediated activation of pantothenic acid **70**, pantoic acid **67** and β-alanine **68** in the presence of a 1,2-aminothiol such as cysteamine **69** to form pantetheine **66** in one pot.

Importantly, as well as the poor yield of pantothenic acid **70** observed from the EDC-mediated coupling of **67** and **68**, EDC unexpectedly promoted the fragmentation of pantothenic acid **70**. The reaction of pantothenic acid **70** (100 mM) and EDC (1 equiv.) at pH 9 and 7 resulted in no fragmentation and quantitative (> 95%) recovery of the starting material. However, at pH 5.5 pantothenic acid **70** fragmented to pantolactone **71** (31%) and β-alanine **68** (24%) after 4 hours (Scheme 41A). In addition, pantothenic acid underwent EDC-mediated side reactions with β-alanine **68** to form 20% of oligomeric products **75**. During this study, transient species (**i-iv**) were detected in the early time points of the NMR time course. From this data the mechanism is thought to involve an intra-molecular cyclisation from activated species **i** to oxazolone **ii**, then the 4' hydroxy group of the pantoate residue 'bites back' to form bi-cyclic compound **iii**. The rearrangement of **iii** to imine **iv**, enables the fragmentation to pantolactone **71** and β-alanine **68** by hydrolysis (Scheme 41B). The control reaction of *N*-acetyl β-alanine and EDC resulted in no fragmentation to β-alanine and acetic acid, which indicated that a 4'-hydroxy was essential for fragmentation.
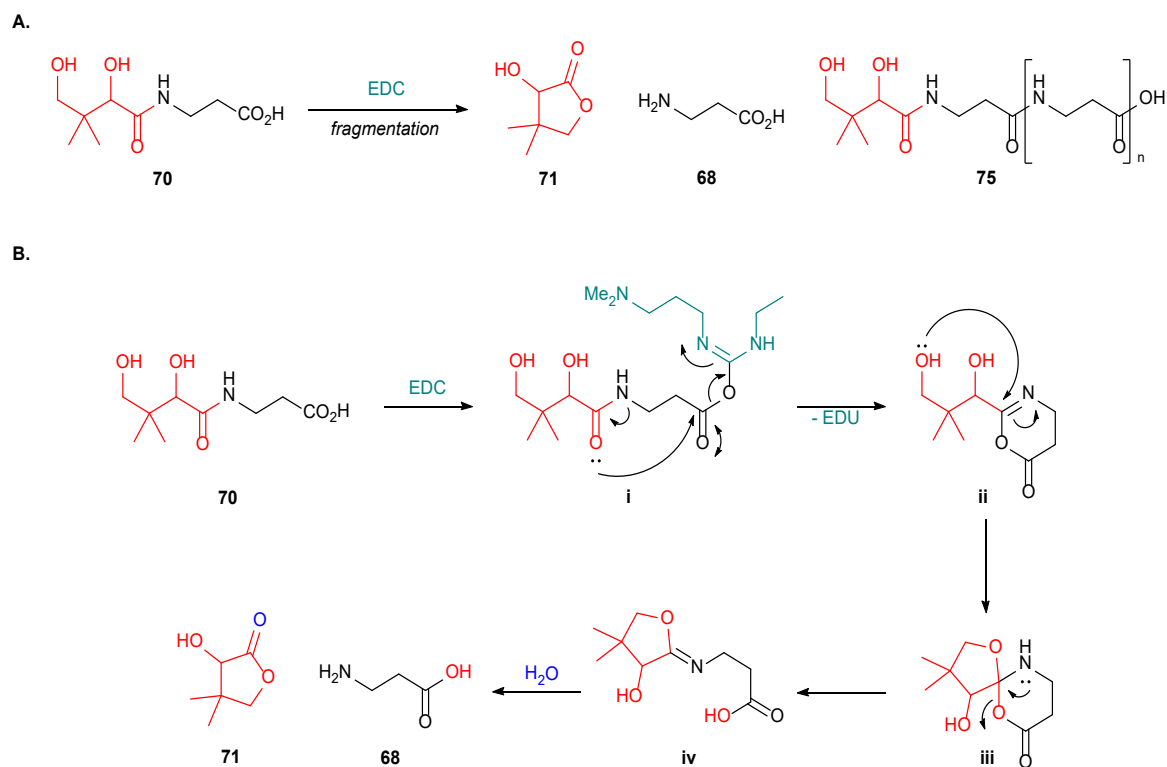
Scheme 41. (**A.**) The EDC-mediated fragmentation of pantothenic acid **70**; (**B.**) The proposed mechanism for the observed EDC-induced fragmentation of pantothenic acid **70** to pantolactone **71** and β-alanine **68**.

The fragmentation of pantoyl-glycine **72** (100 mM) also occurs upon the addition of EDC at pH 5.5. After the rapid consumption of EDC (~15 minutes), pantoyl-glycine **72** (< 5%), pantolactone **71** (78%), **Gly**-OH (51%), and pantoate **67** (16%) was detected. When this reaction was repeated in competition with pantothenic acid **70** (100 mM), β-alanine **68** (31%) and glycine **Gly**-OH (66%) was observed and indicated that the pantoyl-glycine **72** favorable fragmented over canonical **70**. This can be attributed to the formation of a 5-membered oxazolone ring for pantoyl-glycine rather than a six membered ring in the fragmentation of pantothenic acid **70** (Scheme 41B).
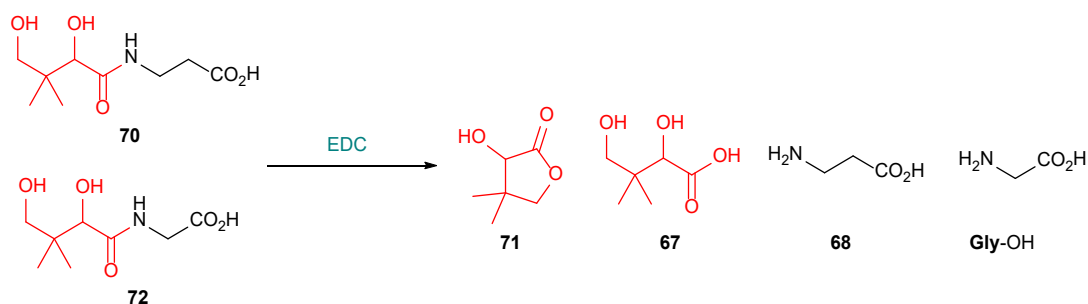
Scheme 42. The EDC-mediated fragmentation of pantothenic acid **70** and pantoyl-glycine **72** to form pantolactone **71**, pantoate **67**, β-alanine **68** and glycine **Gly**-OH.

In addition to the degradation of pantothenic acid **70** upon EDC activation, the reaction of pantothenic acid **70** (100 mM) with EDC (1.1 equivs.) and β-alanine nitrile **76** (1 equiv.) at pH 5.5 led to transamidation and the formation of pantothenic acid nitrile **77** (39%) from pantothenic acid **70** (Scheme 43). However, in stark contrast to the EDC-promoted degradation of pantothenic acid **70**, the EDC-promoted degradation of pantothenic acid nitrile **77** was not observed. In the next chapter pantothenic acid nitrile **77** will be shown to be a key intermediate in the synthesis of pantetheine **66** from nitrile substrates.
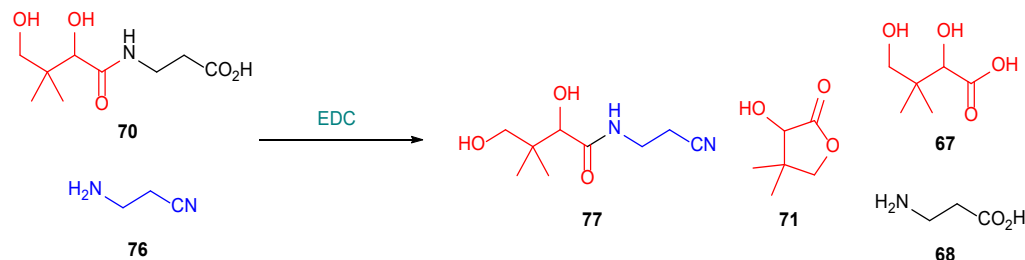
Scheme 43. The EDC-induced exchange of pantothenic acid **70** with β-alanine **68** to form pantothenic acid nitrile **77** and the fragmentation products pantolactone **71**, pantoate **67** and β-alanine **68**.

2.2.6. Conclusion

Our investigation of electrophilic carboxylic acid activation highlighted severe difficulties and problems associated with generating pantetheine **66** from carboxylic acids (i.e., pantoic acid **67** and β-alanine **68**). The results of the experiments indicated that each step towards pantetheine **66** from pantoic acid **67**, β-alanine **68**,

and cysteamine **69** was severely compromised by adverse side reactions that were induced by electrophilic activating agents (Scheme 44).
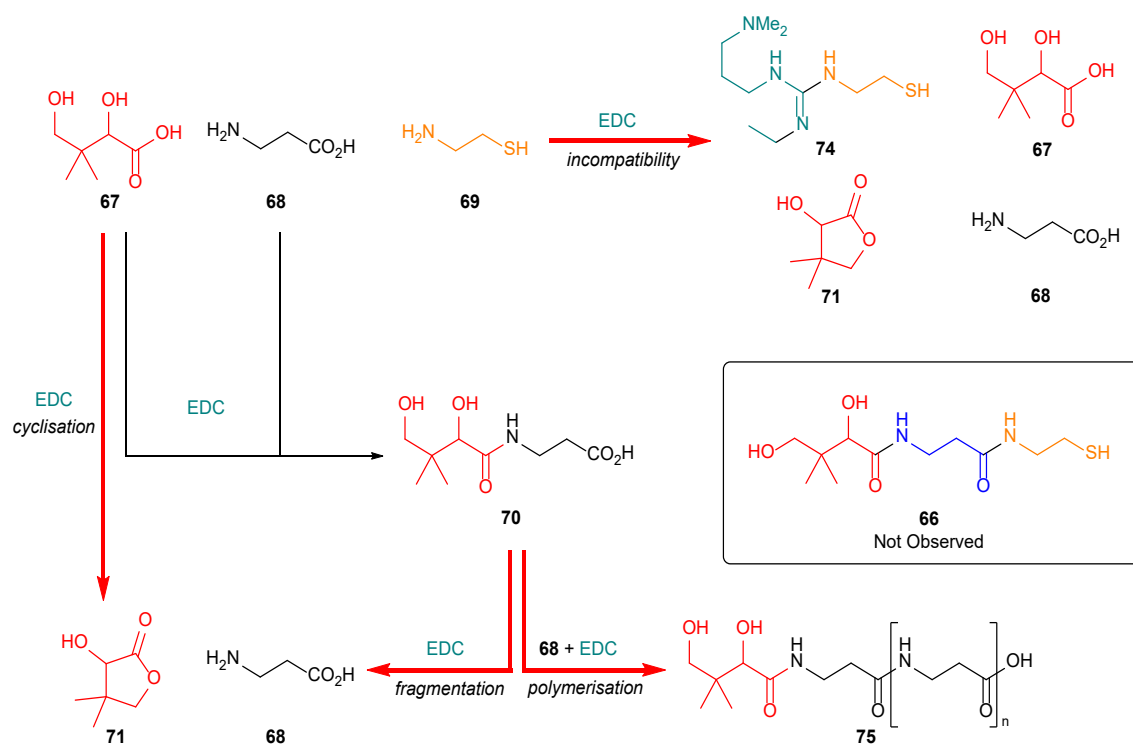
Scheme 44. A summary of the reactions that prevented the synthesis of pantetheine **66** using the carboxylic acid activating agent *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide (EDC) as a model.

Cyclisation of **67** to **71** dominated over coupling of **67** with **68** across the full pH range (5.5-9) where EDC was observed to favourably activate pantoic acid **67**. Therefore, only very low yields of pantothenic acid up to 7% could be observed (even in the absence of **69**). In addition to the cyclisation of **67**, oligomerisation of β-alanine **68** to **75** was also observed and thus further diminish the yields of **70**. The EDC-promoted fragmentation of **70** further inhibits the synthesis of **66** through electrophilic carboxylic acid activation. On top of this cysteamine **69** undergoes preferential *N*-guanydilation to compound **74** and this blocks further reactivity to pantetheine **66**. This demonstrates the incompatibility of carbodiimide-mediated activation of pantothenic acid **70** with 1,2-aminothiols such as cysteamine **69**.

The Powner group has recently demonstrated that amide bonds can be synthesised in water directly from amino- and amido nitriles, to completely by-pass nitrile

hydrolysis and the activation of free-amino acids.[174,177,199,211] These nitrile reactions exploit the energy of the nitrile moiety to drive amide bond formation, rather than dissipating the nitriles inherent energy by hydrolysis. These nitrile couplings are highly selective for proteinogenic α-peptides and overcame numerous identified problems for the direct ligation of α-amino acids through electrophilic activation (Section 1.8.5 and 1.10.5).

Given the detrimental effects observed upon electrophilic carboxylic acid activation, it is especially noteworthy that pantothenic acid nitrile **77** is intrinsically activated towards reaction with cysteamine **69**. The reaction of **77** with **69** does not requires exogenous electrophilic activation agents that would block the reactivity of cysteamine **69**, and therefore block the synthesis of pantetheine **66**. The incubation of pantothenic acid nitrile **77** with cysteamine **69** in water will be discussed in Chapter 2.4. However, an effective prebiotic synthesis of pantoyl-β-alanine nitrile **77** will be discussed first (Chapter 2.3).

2.3. Syntheses of Pantetheine using Nitrile Activation

2.3.1. Introduction

In the previous chapter the reactivity of pantolactone **71**, β-alanine **68**, cysteamine **69** was explored under various conditions following on from the work of Miller.[207,220,223] The results found that reactions of these reagents in solution, or by wet-dry cycling or with electrophilic activating agents were ineffective in the synthesis of either pantothenic acid **70** or pantetheine **66** (Section 2.2.3). Moreover, the non-canonical glycine derived analogues were shown to be favoured over the natural β-alanyl homologues.
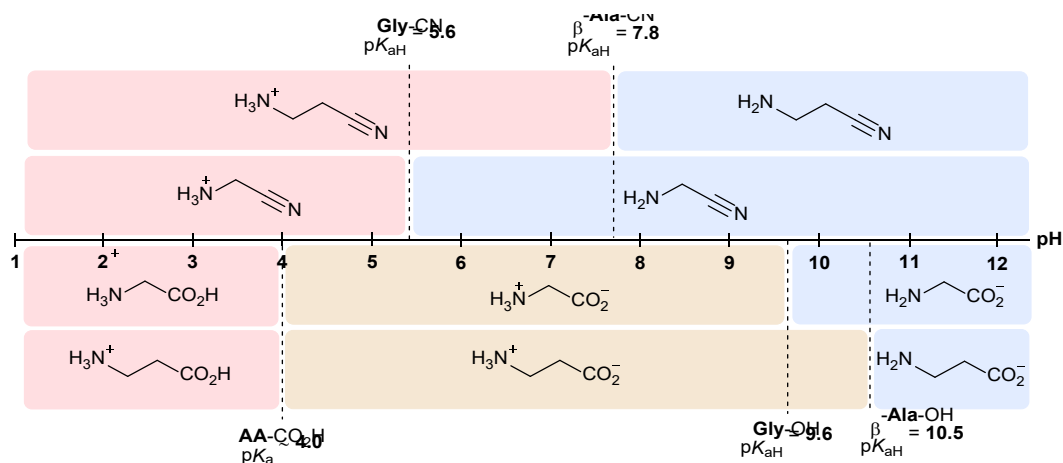
Figure 21. Figure depicting the protonation states and the pK_{aH} values of glycine **Gly**-OH, glycine nitrile **Gly**-CN, β-alanine β-**Ala**-OH and β-alanine nitrile β-**Ala**-CN.[132,159,246]

Aminonitriles formed by the Strecker reaction are precursors of amino acids (Section 1.5.1).[28,244] However, the hydrolysis of aminonitriles to amino acids dissipates the energy stored within the nitrile moiety.[174] By taking advantage of the latent nitrile activation inherent to aminonitriles, Canavelli *et al.* recently reported that the chemoselective synthesis of proteinogenic peptides can be achieved by coupling α-aminonitriles **AA**-CN in water (Section 1.8.5).[174,177] These mechanisms by-passed α-amino acids to generate α-peptides without electrophilic carboxylic acid activation, which is incompatible with various proteinogenic amino acid side chains (e.g., aspartic acid **Asp**-OH, glutamic acid **Glu**-OH).[256] This chemistry exploited the lower basicity of α-aminonitriles (**Gly**-CN; pK_{aH} = 5.4)[132] over α-amino acids which have a

higher pK_{aH} (**Gly**-OH; pK_{aH} = 9.6).[159] The uniquely low basicity of α-aminonitriles enabled their selective ligation in water by rendering them nucleophilic across a broader pH range than α-amino acids.

The pK_{aH} of β-alanine-nitrile **76** (pK_{aH} = 7.8)[159,246] is over 2 pka units lower than that of β-alanine **68** (pK_{aH} = 10.5)[246] at 20 °C, and so it was suspected that β-alanine-nitrile **76** would have a nucleophilic advantage when coupling with pantolactone **71** over β-alanine **68** to form pantothenic acid nitrile **77**. This is significant since pantothenic acid nitrile **77** would retain latent activation within its nitrile moiety. Pantothenic acid nitrile **77** is a key intermediate since this molecule's nitrile functional group enables it to react with cysteamine **69** and onwards to pantetheine **66** without requiring electrophilic activation. This latent nitrile activation was hypothesised to be important because attempts to synthesise pantetheine **66** from β-alanine **68**, cysteamine **69**, pantoic acid **67** and pantothenic acid **70** were thwarted by a myriad of detrimental reactions caused by electrophilic carboxylate-activating agents. Chief amongst these problems were the incompatibility of cysteamine **69** with electrophilic activation of carboxylic acids, and the fragmentation of **70** (Section 2.2.5).

2.3.2. Coupling of β-Ala-CN with Pantolactone

Pantothenic acid nitrile **77** had previously been synthesised by Hosokawa *et al.* by heating pantolactone **71** at 50 °C with in neat β-alanine nitrile (1 equiv.) overnight in 92% yield.[259]
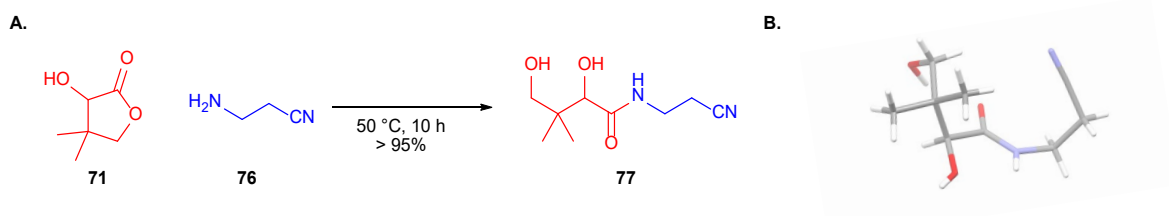
Figure 22. (**A.**) Synthesis of pantoyl-alanyl nitrile **77** from heating pantolactone **71** and β-alanine nitrile **76** neat at 50 °C for 10 hours. (**B.**) The single X-ray crystal structure of nitrile **77** recrystalised from EtOAc.

The procedure was repeated with slight modification to the original method (1.5 eq. β-**Ala**-CN, 50 °C, 10 h, > 95%). The resulting material was used as a synthetic standard for spiking and in subsequent experiments when exploring the reactivity of this class of compounds. Initially the product was isolated as an oil which on standing in the fridge became a white amorphous solid It was found that recrystallisation from ethyl acetate (EtOAc) gave colourless crystalline prisms which were of sufficient quality for X-Ray analysis (Figure 22B).

In the previous chapter (Section 2.2.3) it was shown that *D/L* pantolactone undergoes sublimation under heating or wet-dry cycles. On top of this Miller found that pantolactone **71** is unstable in basic conditions with a half-life measured to be 17.5 min at pH 10.5.[220] Considering that the coupling reaction of pantolactone **71** with neat β-alanine nitrile **76** gave near quantitative yields in dry conditions. Dry down reactions were initially tried to find out whether water was deleterious towards amide formation. Heating a 1 mL solution of pantolactone (500 mM) with β-alanine nitrile (1 equiv.) at 100 °C and pH 9, resulted in 36% yield of nitrile **77**. Although there was some **77** formed, heating also caused significant hydrolysis which curtailed the total yields of **77**. Nevertheless, encouraged by these preliminary results, attempts of the coupling of **71** and **78** were attempted under different conditions in water.

To test the efficacy of coupling, the first set of reactions investigated the reaction of lactone **71** (50 mM) and β-aminonitrile **76** (2 equiv.) at pH 7,8 and 9 in phosphate buffer (500 mM) until **71** was consumed. The yields were low with reactions conducted below pH 9. For example, this reaction at pH 7 only yielded 2% **77** after 6 days. Even at higher concentrations of **71** (500 mM) and super stoichiometric quantities of β-alanine nitrile **76** (5 equiv.) at pH 7, only 12% **77** was formed after 10 days. At low concentrations, hydrolysis outcompetes coupling especially at higher pH. As the concentration of lactone **71** is increased, the reaction with 2 equivalents of **76** yielded 37% and 50% of nitrile **77** for 50 mM and 100 mM pantolactone **71** respectively. High concentrations of pantolactone **71** (500 mM) were found to consistently give the highest yields of **77**, with the reaction with **76** (1 M) resulting in

a yield of 84% pantoyl-β-alanyl nitrile **77** after three days. Pantothenic acid nitrile **77** was able to be formed in the highest yield (94%) with 500 mM pantolactone **71** and 3 equivalents of β-alanine nitrile **76** at pH 9 over 12 hours (Figure 23). These results indicate that the coupling between lactone **71** and β-alanine nitrile **76** in water work best above the pK_{aH} of **76** at pH 9 and that higher concentrations gave higher yields in shorter times with minimal hydrolysis of **71** to pantoic acid **67**.
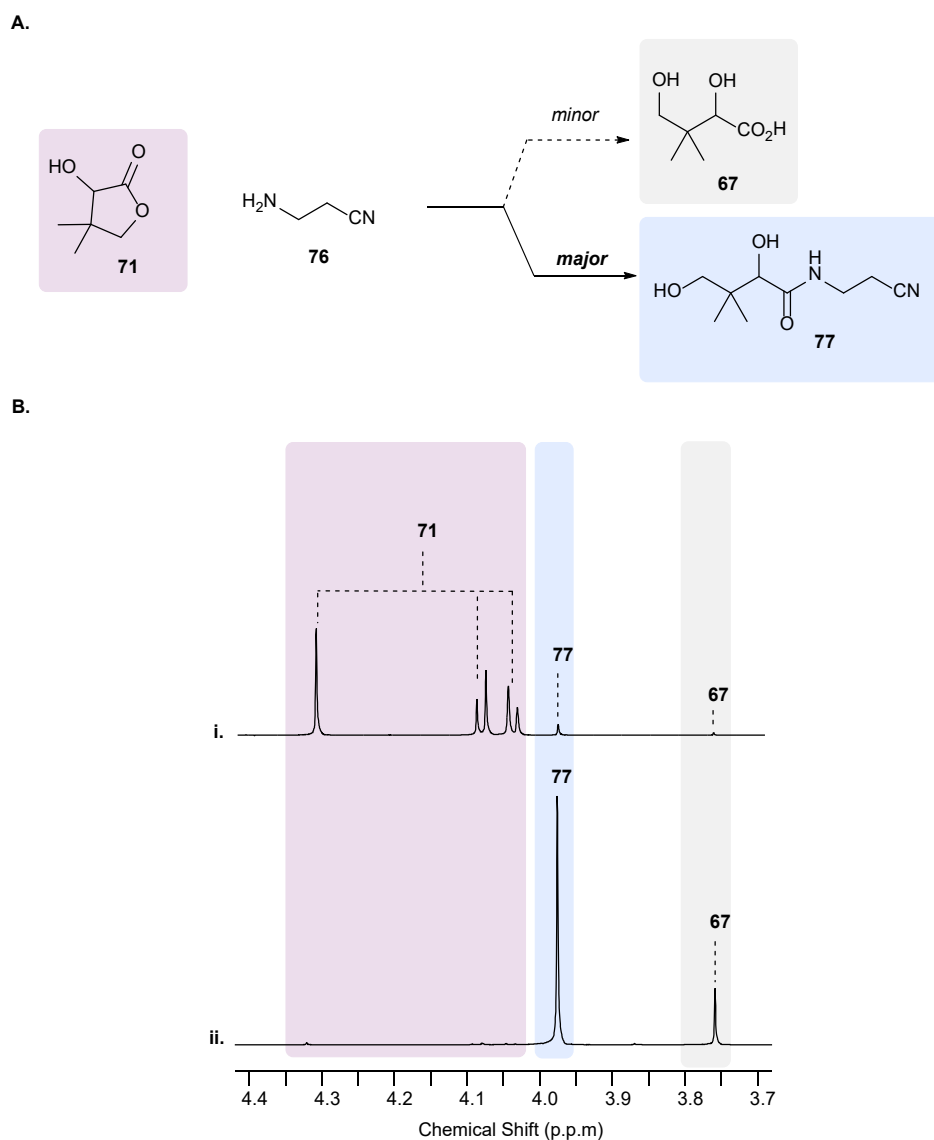
Figure 23. Lactone **71** undergoes highly effective coupling with weakly basic β-alanine-nitrile **76** (pK_{aH} = 7.8 at 20 °C, ref 38) in water. (**A.**) High-yielding coupling of β-alanine-nitrile **76** with lactone **71** to yield nitrile **77**. (**B.**) 1H NMR (600 MHz, H$_2$O/D$_2$O 99:1) spectrum to show the reaction of lactone **71** (500 mM) in phosphate solution (pH 9; 500 mM) at 20 °C with **76** (2 equiv.) (**i.**) Reaction after 10 mins; (**ii.**) Reaction after 3 days yielding nitrile **77** (83%).

2.3.3. Competition reaction of β-Ala-CN and β-Ala-OH with Pantolactone

Pantetheine **66** possesses a unique β-alanyl motif, so the next question was whether β-alanine-nitrile **76** could selectively react with lactone **71**. Specifically, whether **71** would discriminate β-aminonitrile **76** from the α-aminonitrile **Gly**-CN and λ-aminonitrile **79** homologues. Previous results showed that the hydrolysis of pantolactone **71** to pantoic acid **67** was favoured over the coupling of β-alanine **68** to give vitamin B$_5$. The superior reactivity of β-alanine-nitrile **76** with respect to β-alanine **68** was then confirmed by direct competition; the stoichiometric competition of β-alanine **68** (2 equiv.) and nitrile **76** (2 equiv.) with lactone **71** (500 mM) at pH 9 returned nitrile **77** in 84% yield after 2 days, alongside only a trace yield (3%) of acid **70**. The reaction was conducted at a lower concentration of **71** (100 mM), and again nitrile **76** (46%) was formed favourably relative to vitamin B$_5$ **70** (1%). However, hydrolysis to pantoic acid **67** (49%) was the major product of this reaction. The results show that **76** is vastly better nucleophile than the acid **67**.
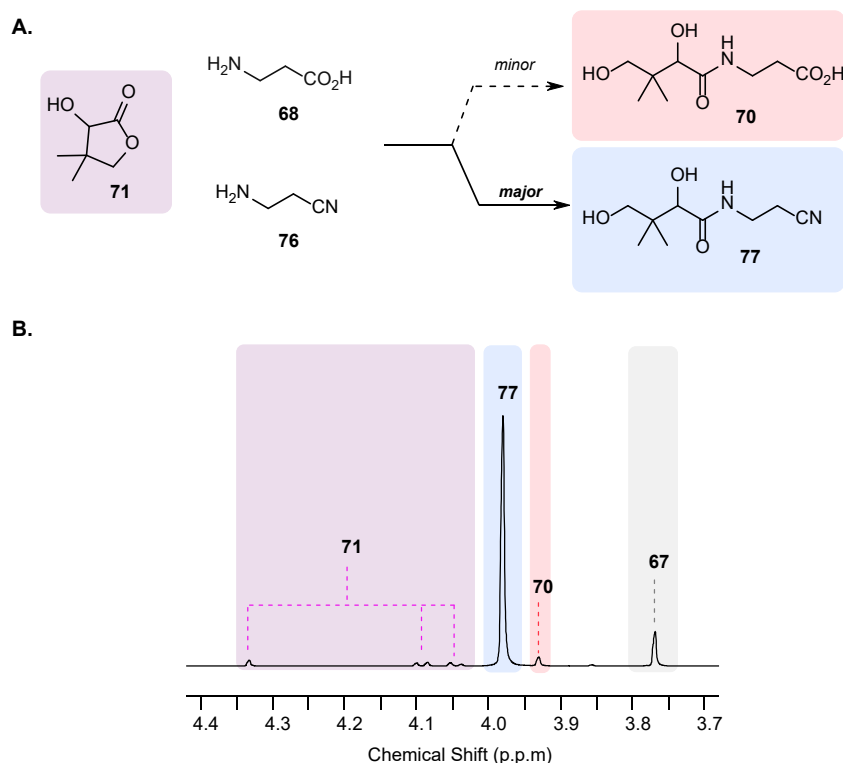
Figure 24. (**A.**) Selective coupling of β-alanine-nitrile **76** with lactone **71** in stoichiometric competition with β-alanine **77**. (**B.**) ^{1}H NMR (600 MHz, H$_2$O/D$_2$O 99:1) spectrum to show the reaction of lactone **71** (500 mM) in phosphate solution (pH 9; 500 mM) at 20 ºC with a stoichiometric mixture of β-alanine **68** (2 equiv.) and nitrile **77** (2 equiv.) leading to selective synthesis of nitrile **77** (84%) alongside trace **70** (3%) after 2 days.

2.3.4. Competition reaction of β-Ala-CN and Gly-CN with Pantolactone

Having demonstrated the selective incorporation of β-alanine-nitrile **76** into pantolactone **71** over β-alanine **68**, the compatibility of this coupling with peptide precursors **AA**-CN was investigated next. As discussed in this thesis, glycine nitrile **Gly**-CN is considered a precursor to amino acids resulting from the Strecker reaction and would have been present in the prebiotic milieu. As the shorter α-homologue of **76**, glycine nitrile **Gly**-CN was hypothesised to be significantly more reactive over a wider pH range due to its low basicity (pK_{aH} = 5.4).[132] The subsequent reaction of pantolactone **71** (100 mM) and glycine nitrile **Gly**-CN (2 equiv.) at pH 7 resulted in the surprisingly sluggish formation of pantoyl-glycine nitrile **78** in 18% yield over 7 days. Both reactions suffered from detrimental hydrolysis of **71** to pantoic acid **67**. Even at pH 9 the yield of **78** only reached 39% after 5 days. Given the improved reactivity of aminonitriles over higher pKa amines in previous aqueous coupling reactions,[174] this low yield was surprising. The reaction of pantolactone **71** with **Gly**-CN was therefore investigated in the presence of β-alanine-nitrile **76** to see if more could be learned from a competition reaction.
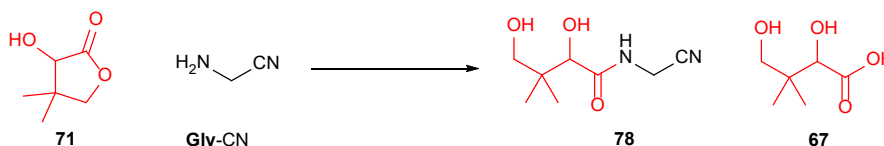
Scheme 45. Coupling of glycine nitrile **Gly**-CN (1-2 equiv.) with lactone **71** (100 mM) to yield nitrile **78** in phosphate solution (pH 7-9; 500 mM) at 20 ºC.

On mixing lactone **71** (500 mM) with glycine nitrile **Gly**-CN (2 equiv.) and β-alanine nitrile **76** (2 equiv.) nearly equal amounts of pantoyl-β-alanyl-nitrile **77** and pantoyl-α-glycyl-nitrile **78** (**77** = 42%, **78** = 37%) were produced after 6 hours. However, upon further incubation an unanticipated equilibration leading to the formation of pantoyl-β-alanyl-nitrile **77** (72%) as the major product and **78** in 8% yield (**77/78**; >5:1) after 6 days (Figure 25) was observed. This reaction was also performed with lower concentrations pantolactone **71** (100 mM) and stoichiometric quantities of **76** and **Gly**-CN, (2 equiv.) at pH 9. The transamidation also occurs giving similar yields but takes a longer period to equilibrate to the desired nitrile **77**. This reaction

appeared to be occurring due addition of **Gly**-CN to lactone **71** being reversible while addition of **76** was irreversible. To test this hypothesis an authentic sample of **78** was prepared.
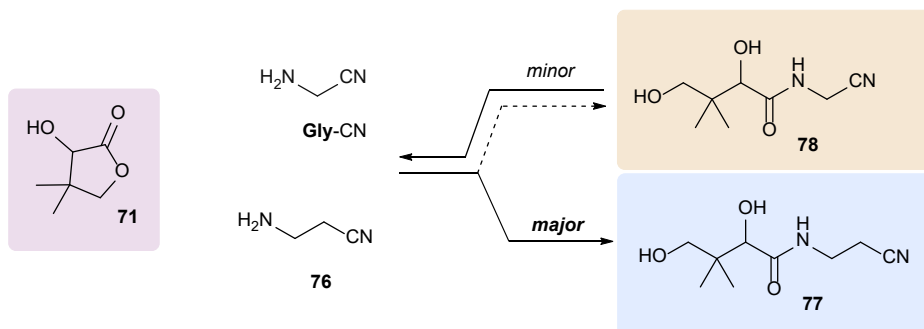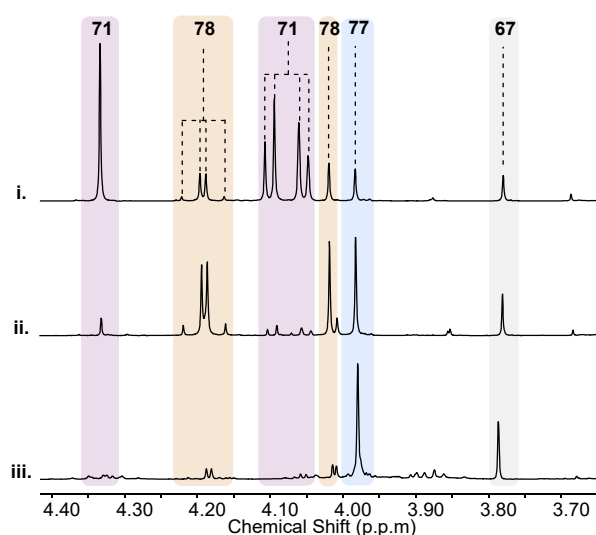
Figure 25. Selective coupling of lactone **2** with a β-alanyl nucleophile is essential for pantetheine **66** synthesis. (**A.**) Incubation of **71** with a mixture of β-alanine nitrile **76** and glycine nitrile **Gly**-CN resulted in an initial 1.1:1 mixture of β-nitrile **77** and α-nitrile **78**. However, prolonged incubation depleted the shorter (non-canonical) α-homologue **78** through transamidation by β-alanine nitrile **76**, leading to the canonical β-homologue **77** as the major product. (**B.**) ^{1}H NMR (700 MHz, H_2O/D_2O 99:1) spectra to show the reaction of lactone **71** (500 mM) with **76** (2 equiv.) and **Gly**-CN (2 equiv.) after (**i**) 10 mins, (**ii**) 6 hours, and (**iii**) 6 days, yielding **77** as the major product (**77/78**; > 5:1). All reaction were carried out in phosphate solution (pH 9; 500 mM) at 20 °C.

Pantothenic acid nitrile **77** had previously been synthesised by heating pantolactone **71** and β-alanine nitrile **76** neat (Section 2.3.2) and the synthesis of nitrile **78** was investigated to spike experiments with authentic samples and for use in the next set of experiments. The procedure was repeated with a modified method of Hosokawa, by heating pantolactone **71** and glycine nitrile **Gly**-CN in triethylamine (NEt₃) as a

base to neutralise the HCl salt and form the free base.[259] As the free amine, **Gly**-CN was able to couple with **71** to form pantoyl-glycine nitrile **78** as a golden oil in 59% yield (Scheme 46).
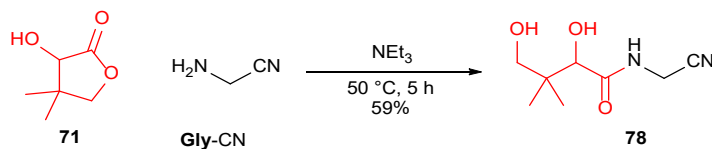
Scheme 46. Synthesis of pantoyl-glycyl nitrile **78** from heating pantolactone **71** and glycine nitrile **Gly**-CN in neat NEt₃ at 50 °C for 5 hours.

This dynamic reactivity was confirmed by incubating isolated pantoyl-glycine nitrile **78** (500 mM) with β-alanine nitrile **76** (1-5 equiv.), which yielded pantothenic acid nitrile **77** in up to 93% yield after 19 days at the expense of pantoyl-α-nitrile **78** (Scheme 47A). Conversely, the incubation of 500 mM pantothenic acid nitrile **77** with glycine nitrile **Gly**-CN, even with a 5-fold excess, returned only 3-7% pantoyl-glycine nitrile **78** (Scheme 47B). The specific structure of pantoate therefore triggers a reversibility to ligation which causes an unexpected β/α selectivity favouring the natural structure.
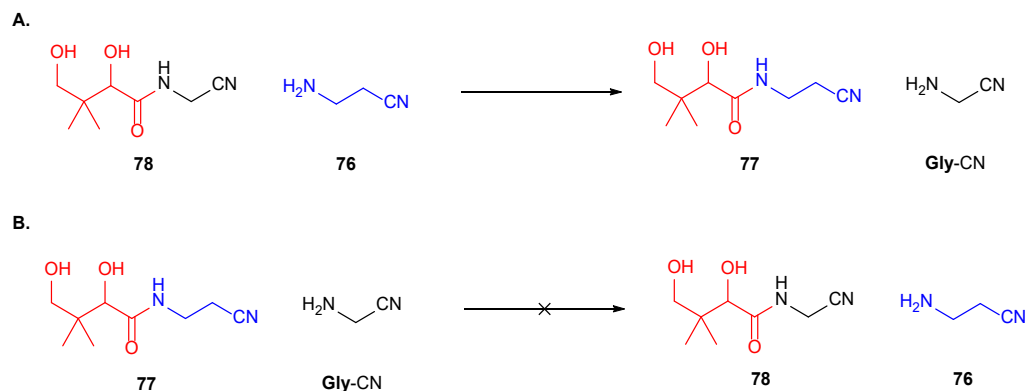
Scheme 47. (**A.**) The transamidation from the reaction also occurs from synthetically prepared pantoyl-glycine nitrile **78** and β-alanine nitrile **76** in water at pH 9. (**B.**) The reaction to show that the transamidation is favoured for pantoyl-β-alanyl nitrile **77** and is non-reversible from the reaction of **77** and glycine nitrile **Gly**-CN at pH 9.

2.3.5. The reactions of pantolactone and γ-aminobutyronitrile in water

Gamma-aminobutyric acid (GABA) is an amino acid and a homologue of β-alanine and glycine with three CH_2 units in its backbone. In extant biology GABA is formed from the decarboxylation of glutamic acid by a decarboxylase enzyme with pyridoxal-5'-phosphate (PLP) (Section 1.5.1) and GABA is used as an inhibitory neurotransmitter in mammals. GABA has been detected in repeats of Millers spark-discharge and it is possible that the nitrile precursor γ-aminobutryonitrile **79** (**GABA**-CN) could have been present on the early Earth.

A.
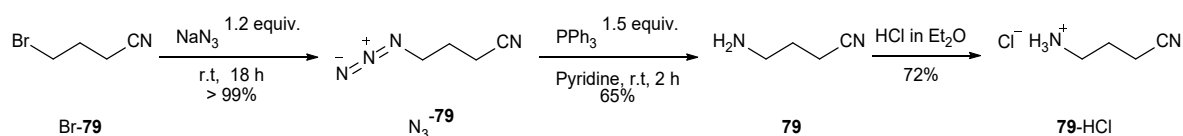
B.
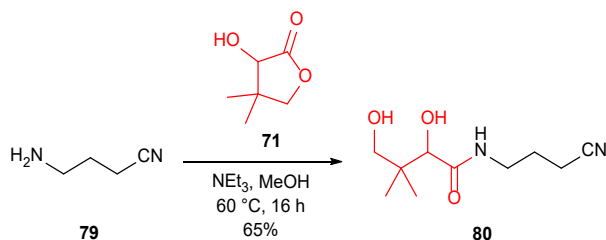
Scheme 48. (**A.**) The three-step synthesis of **GABA**-CN HCl salt from the addition of an azide to 4-bromobutanenitrile Br-**79**, then the reduction of azide N₃-**79** with triphenylphosphine results in the formation of γ-aminobutyronitrile **79**. (**B.**) The synthesis of pantoyl-γ-butyronitrile by heating pantolactone **71** and **GABA**-CN **79** overnight in NEt₃ and MeOH at 60 °C.

Following a procedure by Capon *et al.*, γ-aminobutyronitrile **79** was synthesised over three steps.[260] The first step involved the quantitative (> 95%) addition of an azide by nucleophilic substitution to 4-bromobutanenitrile Br-**79** in dimethylformamide. Then the Staudinger reaction using triphenylphosphine in pyridine reduced the azide N₃-**79** to an amine group after an aqueous workup in 65% yield. Studies by Capon showed that the free amine undergoes intramolecular cyclisation to cyclic amidine **81** and the HCl salt was more bench stable. Therefore, final step converted the free base to the HCl salt by precipitating out **79** with 2N HCl in diethyl ether in 72% yield (Scheme 48A). The resulting material was recrystalised from ethanol to give pure **79**-HCl, which was used for subsequent reactions. Pantoyl-γ-butyronitrile **80** was

synthesised in a modified procedure to pantoyl-β-alanyl nitrile **77** synthesis. Pantolactone **77** was heated overnight at 60 °C with **79**-HCl, triethylamine (NEt$_3$) and methanol to give a yield of 65%, NEt$_3$ was used to neutralise the HCl salt and form the free base of **79** and enable coupling to **71**. This compound was subsequently used as a standard for spiking and further experiments. The pK_{aH} of GABA is 10.8 at 20 °C,[246] and is similar to the pK_{aH} of β-alanine (pK_{aH} = 10.5).[246] Because of this, it was suspected that the pK_{aH} of γ-aminobutyronitrile **79** would also be high. The pK_{aH} of **79** was found to be 10.2 at 20 °C by measuring the pH titration with ^{1}H NMR. Therefore, it was hypothesised that **79** would not couple effectively with pantolactone **71** in a similar fashion to the other high pK_{aH} amines (e.g β-**Ala**-OH, **Gly**-OH etc.) tested.
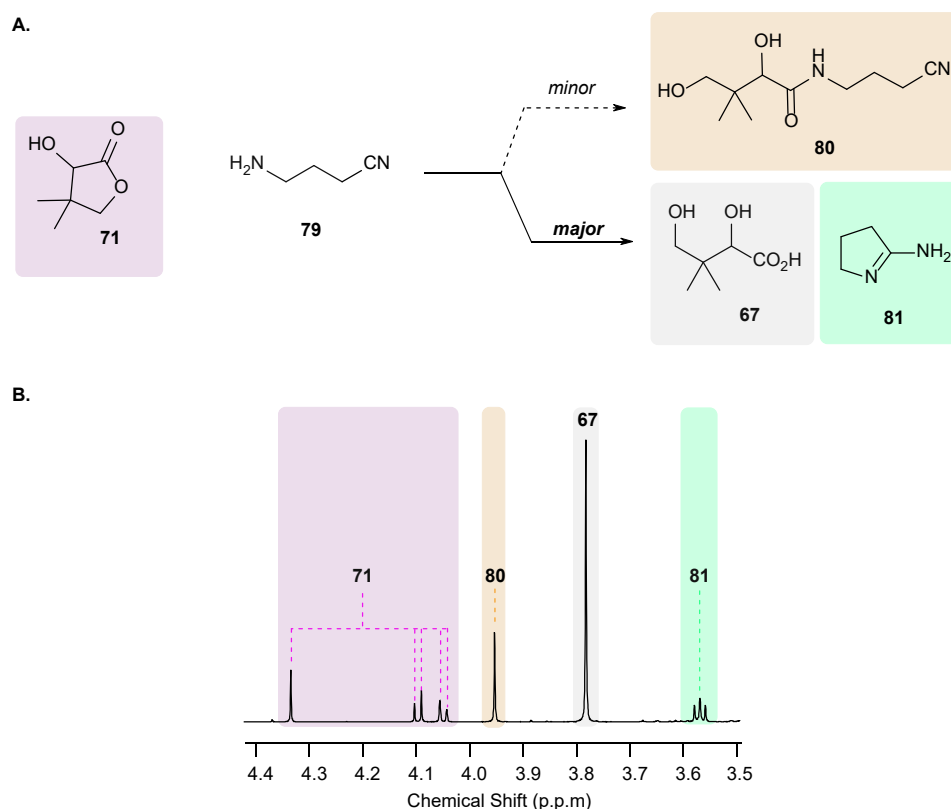
Figure 26. γ-Aminobutyronitrile **79**, due to its high bascicity (pK_{aH} = 10.2 at 20 °C), does not couple effectively with lactone **71** in water. (**A.**) Incubation of lactone **71** with γ-aminobutyronitrile **79** leads to predominantly to the hydrolysis of **71** to pantoic acid **67**. Nitrile **79** also slowly cyclises to yield amidine **81**.[260] (**B.**) ^{1}H NMR (700 MHz, H$_2$O/D$_2$O 9:1) spectrum to show the reaction of lactone **71** (500 mM) in phosphate solution (pH 9; 500 mM) at 20 °C with γ-aminonitrile **79** (2 equiv.) yielding **67** (68%), **80** (21%) and **81** (11%) after 6 days.

The results showed that the reaction of γ-aminonitrile **79**, which is substantially more basic than β-aminonitrile **76** and did not effectively couple with lactone **71** to yield γ-amidonitrile **80**. Instead, **71** predominately underwent hydrolysis, whilst **79** slowly cyclised to yield cyclic amidine **81** (Figure 26).[260] When incubating lactone **71** (500 mM) with stoichiometric β-aminonitrile **76** and γ-aminonitrile **79** in competition at pH 9 the pantoyl-β-alanyl-nitrile **77** (71%) is selectively produced, alongside only 10% of its longer γ-homologue **80** and 11 % of the cyclic amidine **81** (Figure 27).
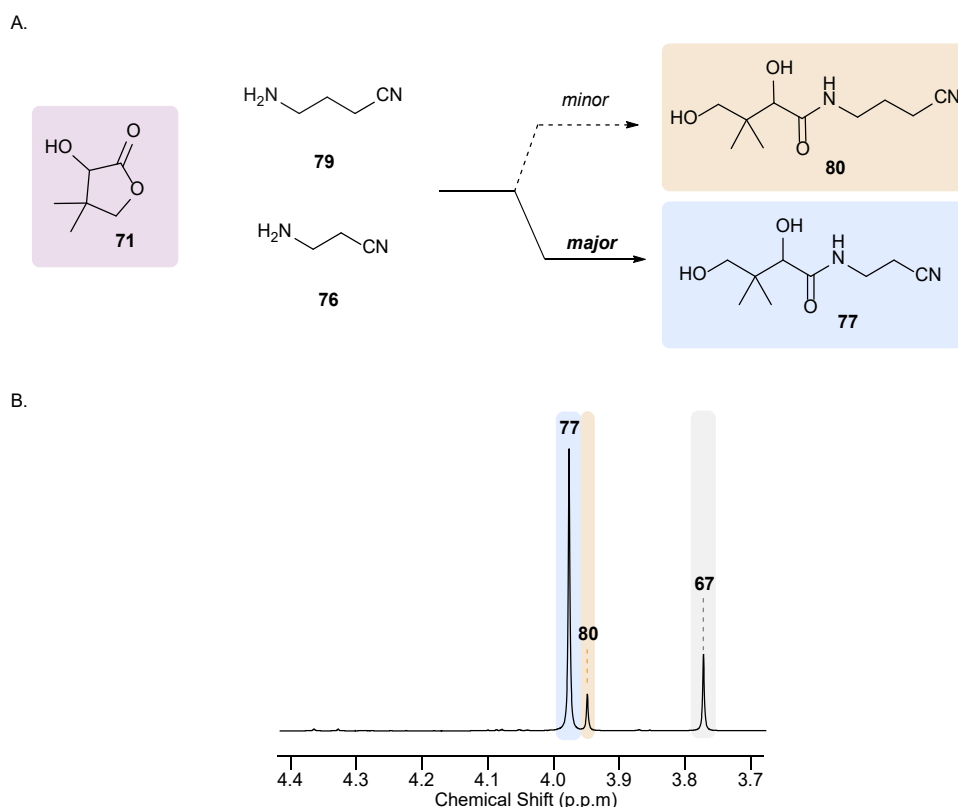
Figure 27. (A.) Incubation of pantolactone **2** with stoichiometric β-alanine **8** and γ-nitrile **11** yields pantoyl-alanyl-nitrile **9** as the major amide-product. **(B.)** ¹H NMR (700 MHz, H₂O/D₂O 9:1) spectrum to show the reaction of lactone **2** (500 mM) with a stoichiometric mixture of β-aminonitrile **8** (2 equiv.) and γ-aminonitrile **11** (2 equiv.) in phosphate solution (pH 9; 500 mM) at 20 ºC to yield nitrile **9** (71%), **5** (19%), **12** (10%) and **13** (11%) after 2 days.

2.3.6. Conclusion

These results demonstrate the reactivity of β-alanine-nitrile **76** markedly favours the synthesis of the canonical structure of pantetheine **66** over both shorter and longer homologues in water. In the previous chapter it was found that the hydrolysis of pantolactone **71**, to yield pantoic acid **67**, outcompetes the coupling of β-alanine **68** with pantolactone **71** to diminish the yield of pantothenic acid **70** in water (Section 2.2.1).[207] In stark contrast, under the same conditions pantolactone **71** reacted with β-alanine nitrile **76** to give pantothenic acid nitrile **77** in excellent yields (Section 2.3.2). At lower concentrations, pantolactone **71** hydrolysis to pantoic acid **67** in the presence of β-alanine **68** was observed at pH 9 and 20 ºC (Section 2.3.2). On the other hand, β-alanine nitrile **76** was observed to couple effectively, albeit sluggishly, with pantolactone **71** under the same conditions, yielding 37% pantothenic acid nitrile **77** after 5 days (Section 2.3.2). This shows that coupling of β-alanine nitrile **76** is markedly better than the analogous reaction with β-alanine **68**.

Any synthesis of **66** must account for the canonical β-alanyl-structure which is a universal element of pantetheine **66**. However, no selectivity in favour of β-alanine **68** over other amino acids (e.g., glycine **Gly**-OH, alanine **Ala**-OH, or γ-aminobutyrate **GABA**-OH) has been reported for reaction with pantolactone **71**.[207,220] Instead, problematically, it has been reported that pantoyl glycine **72** must be expected to form in higher quantities than pantothenic acid **70**, and the reaction of a mixture of amino acids must result in mixtures of non-canonical structures in preference to canonical pantetheine **66**.[207] These statements were based upon comparative rates of individual, separate amino acid reactions (not direct competition reactions). Moreover, the predictions were based upon experimental results obtained at high pH which were then extrapolated to lower pH's; no data was reported for competition reactions.[207,220]

In contrast, the experiments conducted in these chapters now show that α-amino acids react in preference to β-amino acids with pantolactone **71** (Section 2.2.1). These results conclusively demonstrate that the synthesis of pantetheine **66** is

disfavoured from mixtures of amino acids. On the other hand, competition experiments with nitriles show that a mixture of pantolactone **71**, β-alanine nitrile **76**, and glycine nitrile **Gly**-CN selectively yields pantothenic acid nitrile as the major product, and only a minor yield of pantoyl-glycine nitrile **78** (Section 2.3.4). Unexpectedly the pantoyl glycine nitrile **78** formed during the initial stages of this reaction underwent efficient transamidation with β-**Ala**-CN **76** to yield nitrile pantoyl-β-alanine nitrile **77**. The selective transamidation severely limits the synthesis of non-canonical **78** and favours (for the first time) the canonical structure of pantetheine **66**. Furthermore, the high pK_{aH} of γ-aminobutyronitrile **79** (**GABA**-CN) also severely limits coupling with **71**, and the competition reaction with β-alanine nitrile **76** resulted in a 7:1 ratio in favour of the natural β-alanyl structure **77** over the non-canonical homologues.

2.4. Reactions of Pantothenic Acid Nitrile with Cysteamine

2.4.1. Introduction

The cysteamine residue in pantetheine **66** contains the thiol residue that is responsible for the functionality of CoA in biology.[195,199] Cysteamine is constitutionally the simplest stable aminothiol, and as such, a prebiotic synthesis of cysteamine have been attempted to varying levels of success in the past.
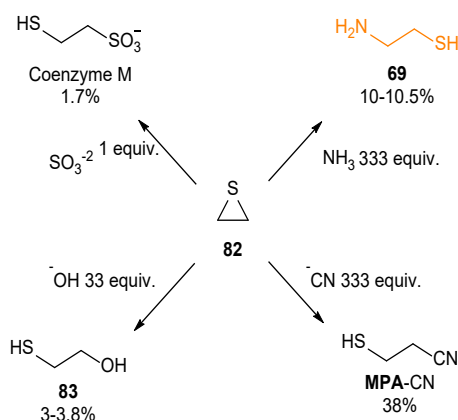
Scheme 49. Miller's prebiotic synthesis of cysteamine and other simple thiols from the addition of nucleophiles to thiirane **82** in water.[223]

Cysteamine has been produced in low yields from the electron irradiation of CH_4, NH_3, H_2O, and H_2S.[221] Cysteamine was also produced when H_2S, N_2 and acetylene are subjected to X-rays.[223] Miller did not detect cysteamine in the original spark discharge experiments, however Miller repeated these experiments with H_2S gas. Reanalysis of old samples of these detected cysteamine and other sulphur compounds from these experiments. Also, Miller was able to form cysteamine (~ 10%) in water from the reaction of thiirane **82** and ammonia over 2 weeks.[223] In addition, Miller was able to form other interesting thiols from different nucleophiles. For example, coenzyme M (~ 2%) was detected from the reaction of **82** and sulphate. Unfortunately, some of these reactions used up to 333 equivalents of nucleophile and are all conducted at pH 11, which cannot be considered prebiotically plausible.[223] In addition, there is currently no prebiotic synthesis of thiirane **82**, with Miller speculating that **82** could be formed from the reaction of alkenes such as

ethylene and electronically excited sulphur atoms. The source of these sulphur atoms was also vague, with Miller stating that they could be formed from the photolysis of COS and CS_2 gases expelled from volcanos. Nevertheless, a successful prebiotic route to thiirane **82** might result in a plausible path to cysteamine **69** under mild condition in appreciable quantities.

Recently, Singh *et al.* were able to build upon the catalytic peptide ligation strategy found by Foden and co-workers (Section 1.10.5).[211] In the course of their studies the use of 2-mercaptoethanol **83** as a catalyst resulted in comparatively more hydrolysis of Ac-**Gly**-CN to Ac-**Gly**-NH$_2$ and the loss of the catalyst **83**, the subsequent formation of thiirane **82** was also observed in the ^{1}H NMR spectra. Thiirane **82**, was confirmed by spiking with an authentic sample and the mechanism to **82** was thought to begin with the formation of thioimidate **84** and then an *S*-to-*O* acyl transfer. This intermediate collapses to Ac-**Gly**-NH$_2$ and liberates thiirane **82** in the process (Scheme 50).[211]
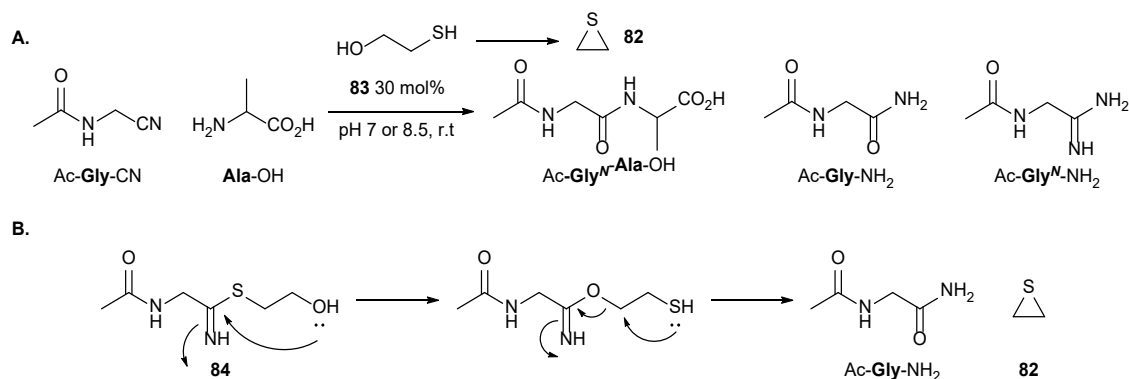
Scheme 50. (**A.**) Figure depicting the catalytic ligation of Ac-**Gly**-CN and **Ala**-OH with 2-mercaptoethanol **83** and its conversion to thiirane **82**; (**B.**) The proposed mechanism of thiirane **82** formation.[211]

Earlier reports on prebiotic pantetheine **66** syntheses did not address the condensation of cysteamine **69** with pantothenic acid **70** in water.[207,220] In section 2.2.5 it was shown that **4** is incompatible with electrophilic activation of **6** using either acetyl phosphate (**AcP**) or a water-soluble carbodiimide (EDC). It was observed that **69** undergoes preferential reaction with 'carboxylic acid' activating agents, and that these reactions block the amine of **69**, which is necessary (but now no longer

available) to form the final β-alanylcysteamine amide bond of **66**. These reactions demonstrated that electrophilic activating agents block the synthesis of pantetheine **66** from cysteamine **69** and pantothenic acid **70**.

Given the detrimental effects observed upon electrophilic carboxylic acid activation, it is especially noteworthy that pantoyl-β-alanine nitrile **77** is intrinsically activated towards reaction with cysteamine **69**. The reaction of nitrile **77** with **69** does not requires exogenous electrophilic activation agents that would block the reactivity of cysteamine **69**, and therefore block the synthesis of **66**. Therefore, the synthesis of the thiazoline intermediate **85** and its subsequent hydrolysis to pantetheine **66** were investigated to complete the stepwise synthesis.

2.4.2. Thiazoline Formation

Thiazoline **85** and pantetheine **66** had also been synthesised by Hosokawa *et al.* by heating nitrile **77** and cysteamine **69** (1.1 equiv.) in sodium ethoxide solution at reflux for 6 hours to give **85** (68%).[259] They also synthesized pantetheine **66** by stirring thiazoline **85** in acetic acid at room temperature for 24 hours in 98% yield.[259]
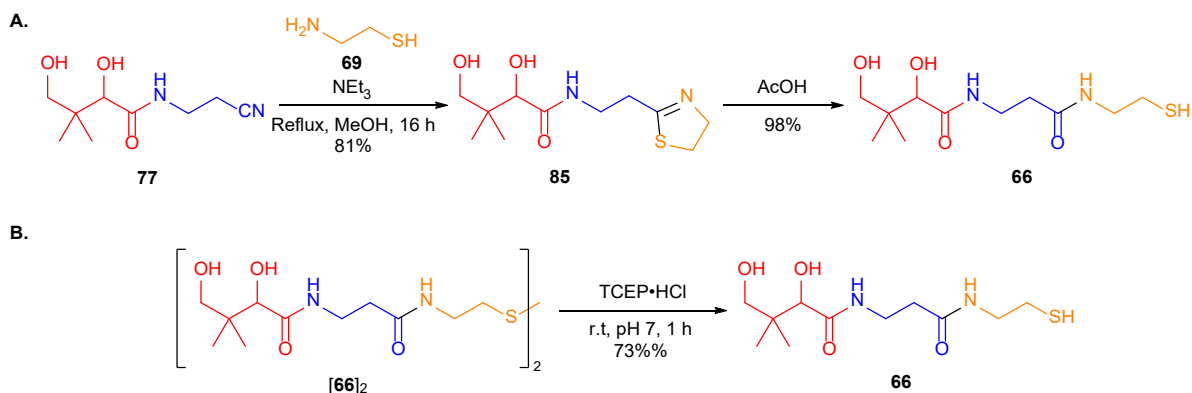
Scheme 51. (**A.**) Synthesis of thiazoline **85** from heating nitrile **77** and cysteamine **69** in NEt₃ and MeOH at reflux overnight. (**B.**) Synthesis of pantetheine **66** from the reduction of commercial pantethine [**66**]₂ with TCEP a pH 7 for 1 hour.

The procedure was repeated with slight modification to the original method, pantoyl-β-alanine nitrile **77** was heated at reflux with cysteamine **69** and triethyl amine (NEt₃) in methanol for 16 hours in 81% yield. The resulting material was an oil that was

difficult to dry and was stored under inert atmosphere to prevent hydrolysis. This sample was used as a synthetic standard for spiking the multicomponent reactions and subsequent experiments exploring the hydrolysis of **85** to pantetheine **66**. The method by Hosokawa was not used to synthesise pantetheine, although hydrolysis of **85** to **66** at pH < 4, then subsequent extraction and purification to **66** had been successfully attempted. Instead, commercial pantethine [**66**]₂ was reduced to pantetheine **66** with (tris(2-carboxyethyl) phosphine hydrochloride (TCEP·HCl) at pH 7, then extraction and purification by column chromatography to give **66** as a wax in 73% yield. This was stored in the fridge under an inert atmosphere to prevent oxidation and hydrolysis/degradation (Scheme 51).
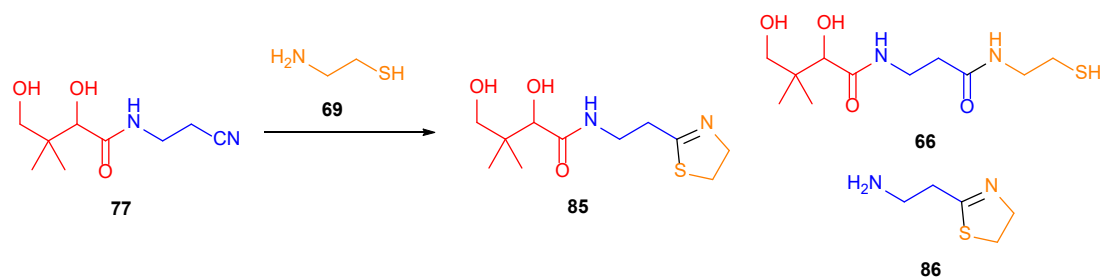
Scheme 52. The reaction of nitrile **77** and cysteamine **69** at pH 7-9 results in thiazoline **85** formation. At pH 7 concomitant thiazoline formation and hydrolysis to pantetheine **66** is observed over long time periods.

The reaction of cysteamine **69** with nitrile **77** were investigated under different conditions. Previous work in the group had observed β-amidonitriles to be resistant to reactions with sulfur nucleophiles.[174,199] It was thought that the reactions should be conducted above the pK_{aH} of the thiol of cysteamine (pK_{aH} = 8.37)[261] to encourage fast high yielding coupling to thiazoline **85**. Upon incubating nitrile **77** (100 mM) with **69** (1.1 equiv.) the formation of thiazoline **85** (63%) was observed after 7 days. The efficacy of the reaction was tested at lower concentrations. The yields of **85** were modest (45%) as the concentration of **77** was decreased down to 25 mM of **9**. At a concentration of 3.1 mM of **77** the yields of **85** sharply decreased (15%, 8 days). Increasing the concentration to 500 mM **77** and doubling the equivalents of cysteamine resulted in 84% yield of thiazoline **85** after 2 days. Slight degradation of **85** to pantoic acid **67** and 2-ethylaminothiazoline **86** (8%) was also observed at pH

9. It was hypothesised that at neutral pH the decomposition would be reduced and so the coupling of **77** with cysteamine was investigated at pH 7 (Figure 28 and 29).
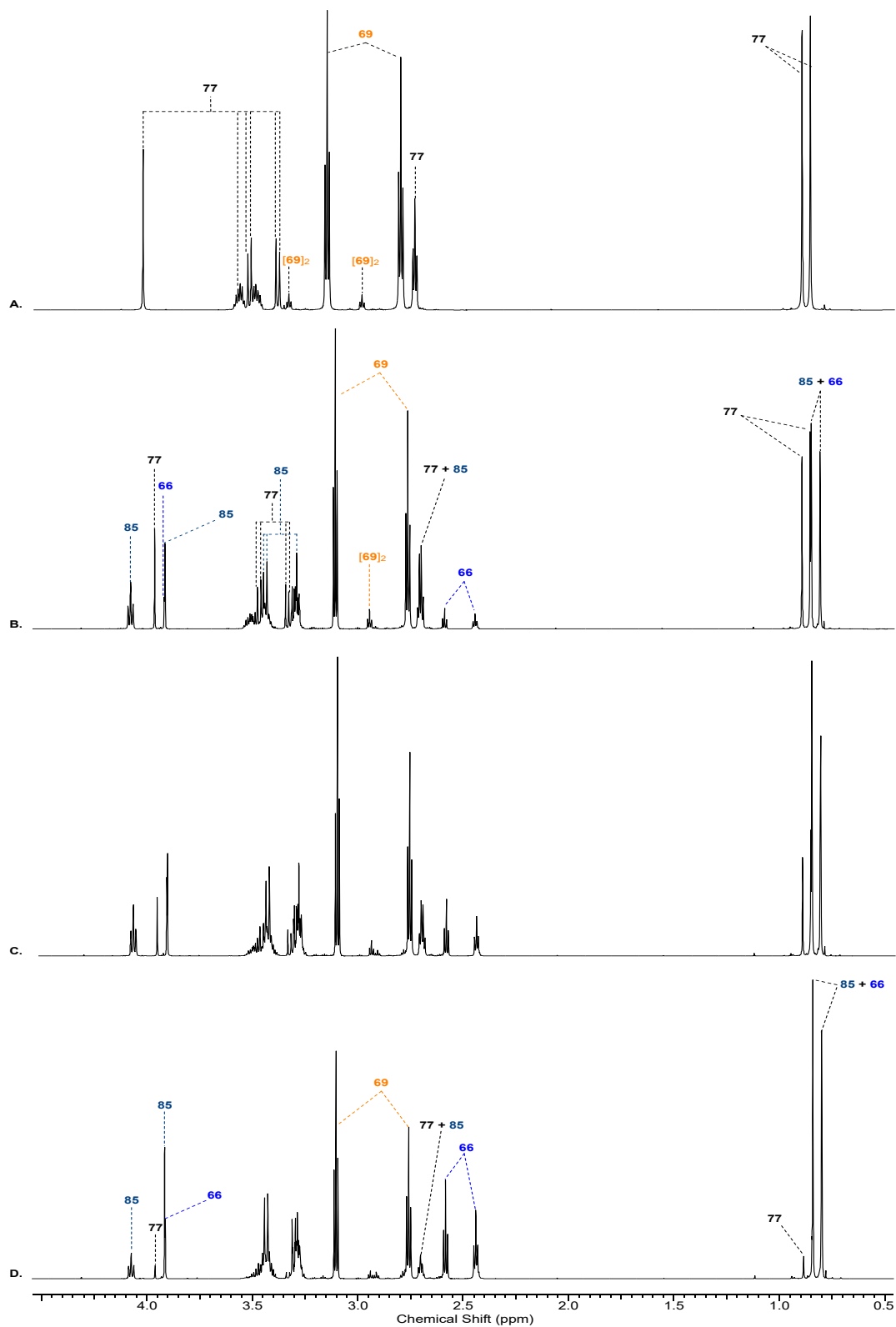
Figure 28. ^{1}H NMR (700 MHz, H_2O/D_2O 99:1, noesygppr1d) spectra to show the reaction of pantothenic acid nitrile **77** (500 mM) with cysteamine **69** (2 equiv.) at pH 7.0 and 20 °C. after: (**A.**) 10 min; (**B.**) 7 days; (**C.**) 14 days; (**D.**) 30 days, where pantetheine **66** (62% and thiazoline **85** (28%) were the major species. See Figure 2. for the NMR spectra for the continuation of the reaction from 36 days (**66** [77%]; **85** [18%]) to 60 days (**66** [93%]; **85** [4%]).
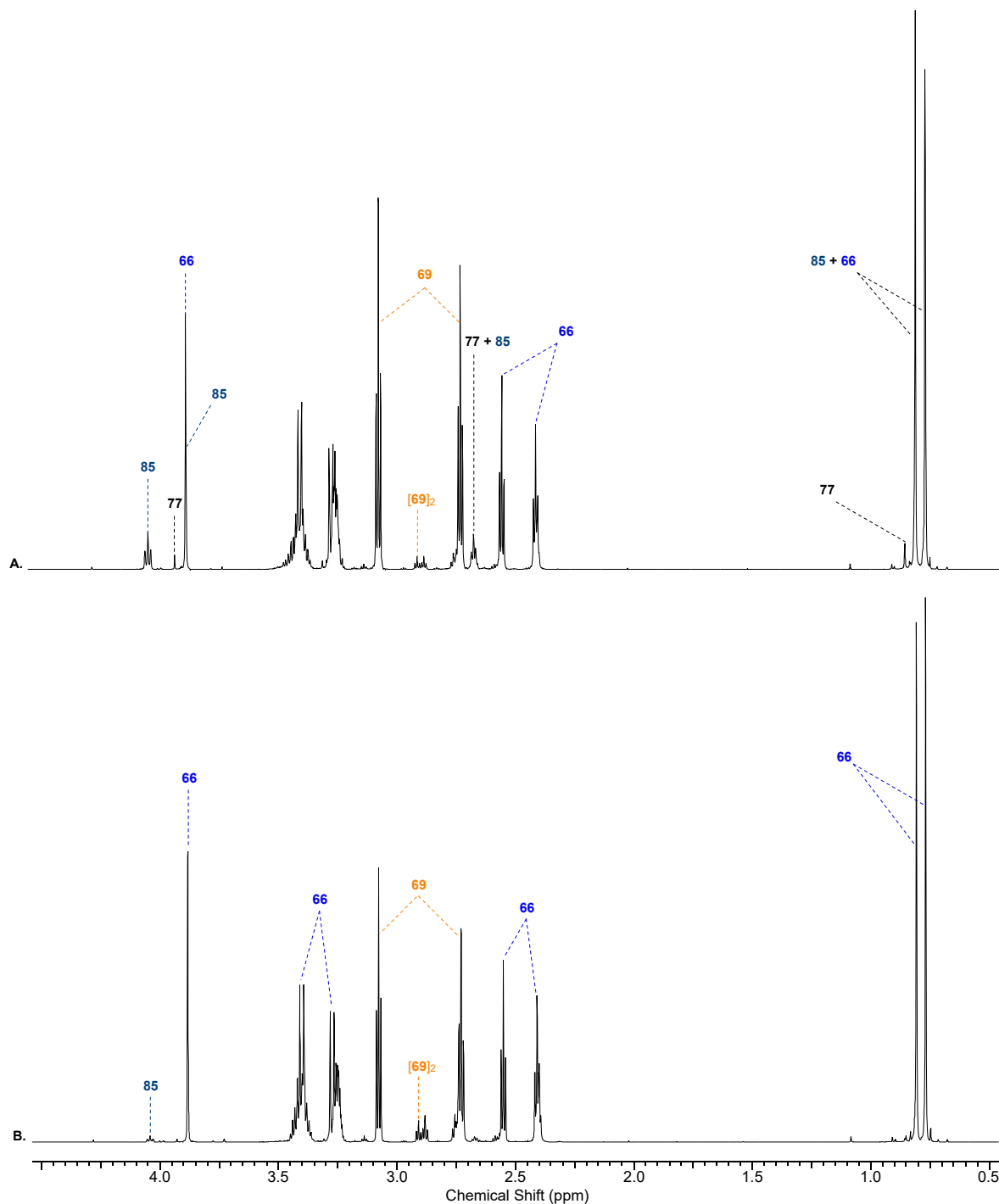
Figure 29. ^{1}H NMR (700 MHz, H_2O/D_2O 99:1, noesygppr1d) spectra to show the reaction of pantothenic acid nitrile **77** (500 mM) with cysteamine **69** (2 equiv.) at pH 7.0 and 20 °C. after: (**A.**) 36 days, where pantetheine **6** (77%) and thiazoline **85** (18%) were observed, and (**B.**) after 60 days to give pantetheine **66** (93%) and thiazoline **85** (4%).

134

2.4.3. Thiazoline Hydrolysis

In the previous section it was shown that the enhanced reactivity of cysteamine **69** is likely due to its ambident nucleophilicity, and as an aminothiol. Cysteamine **69** is primed to undergo irreversible thiazoline formation, unlike monodentate thiols like Ac-**Cys**-OH or MPA which could instigate catalytic peptide ligation. The thiazoline can then ring open in acidic conditions to form a highly stable amide bond and liberate the free thiol.

During the thiazoline formation experiments, it was observed that incubating nitrile **77** (500 mM) with cysteamine **69** (2 equiv.) at neutral pH led to the *in-situ* formation and concomitant hydrolysis of thiazoline **85** to furnish pantetheine **66** in 93% yield over 60 days. Although this was a long reaction, locking pantetheine **66** as a thiazoline resulted in little degradation or oxidation of pantethine [**66**]$_2$ at pH 7. Also, by thoroughly degassing the water negligible oxidation of **66** to [**66**]$_2$ was observed. Therefore, under anoxic conditions of the early Earth, pantetheine **66** can be synthesised in three high-yielding activating-agent-free steps. Pantetheine **66** is produced in water through the remarkable nucleophilicity and latent electrophilicity of β-alanine nitrile **68** and therefore underpins the role that nitriles could have had on the primordial Earth.
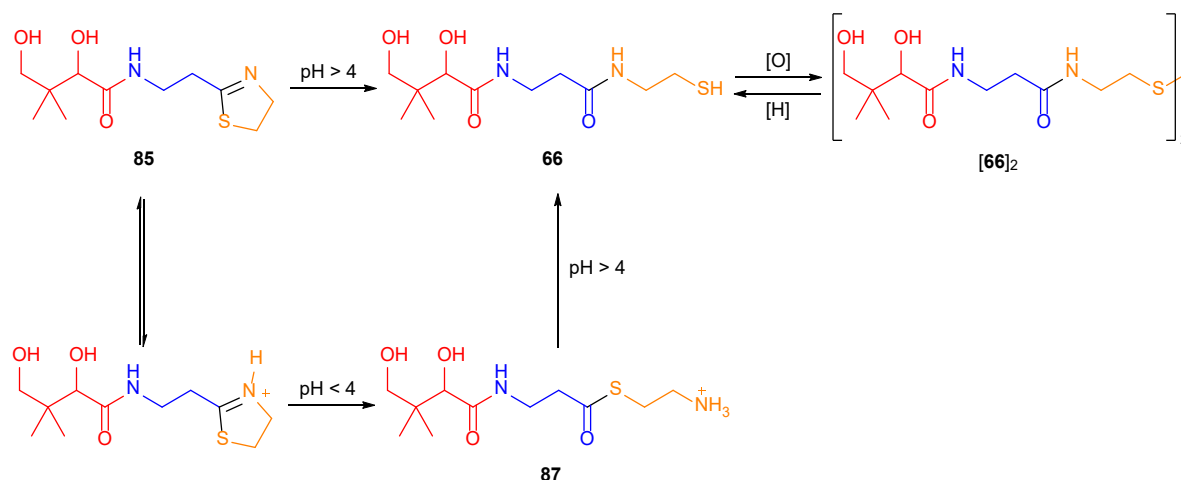
Scheme 53. Rapid hydrolysis of thiazoline **85** occurs close to the pK_{aH} of thiazoline **85**. The pH modulates the direction the thiazoline ring opens during hydrolysis.

The concomitant formation of **85** and hydrolysis to **66** indicate that thiazoline **85** hydrolysis was acid catalysed. The hydrolysis of **85** was repeated at pH 7, with pure synthetic thiazoline and slowly hydrolysed to furnish pantetheine **66** in near-quantitative (> 95%) yield over 43 days (Figure 30). The hydrolysis of **85** was measured across a broad pH range (pH 3–9) and was most rapid at pH 4 which is near the pK_{aH} of **85**. The pK_{aH} of thiazoline **85** was measured by pH titration with ^{1}H NMR and found to be pK_{aH} = 4.1.

At low pH (pH < 3) thioester **87** was observed up to 48% yield at pH 1, however leaving these sample over a prolonged period (7 days) resulted in the degradation into species such as pantoic acid **67**, pantolactone **71** and β-alanyl species. Bringing the pH back to neutral conditions resulted in the rapid *S*-to-*N* transfer back to pantetheine **66**.
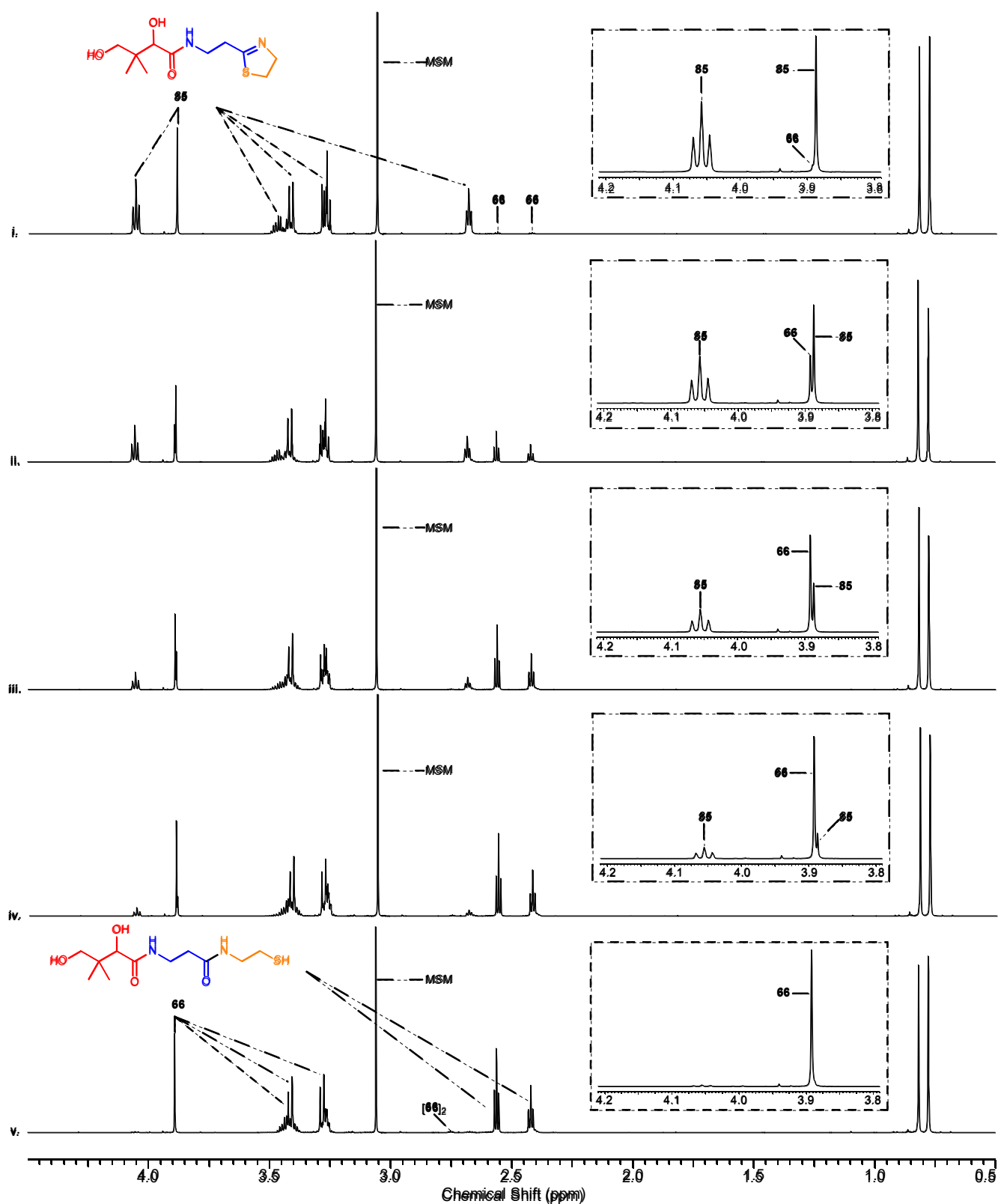
Figure 30. ^{1}H NMR (700 MHz, H_2O/D_2O 9:1, noesygppr1d) spectra to show the hydrolysis of thiazoline **85** (500 mM) in phosphate buffer (pH 7; 500 mM) at 20 °C to yield pantetheine **66**: (**i.**) 10 mins; (**ii.**) 4 days; (**iii.**) 12 days; (**iv.**) 21 days; (**v.**) 43 days, yielding pantetheine **66** (95%), alongside pantetheine [**66**]$_2$ (4%).

2.4.4. Three Component One-Pot Synthesis of Pantetheine

In the previous sections stepwise addition of reagents were used to control the reactivity. Such a constraint would seem to be less prebiotically realistic than the coexistence of all reagents in one location at the same time. In a mixed system, multiple reactions including the desired synthesis of **66** have the potential to occur from mutually reactive feedstocks. These reactions would form molecules in the wrong order and possibly not produce pantetheine **66**. For example, cysteamine **69** could react more rapidly with **71** than β-alanine nitrile **76** could with pantolactone **71** to form pantoyl-cysteamine **88** (Scheme 54). This was therefore experimentally tested with a 3-component reaction of pantolactone **71** (500 mM) with varying stoichiometric quantities of β-alanine nitrile **76** and cysteamine **69** at different pH values.
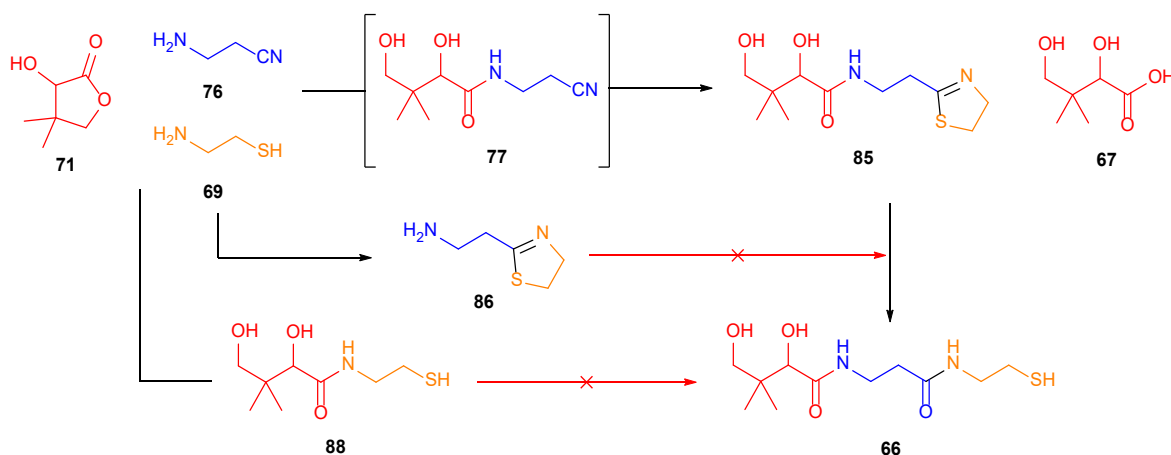
Scheme 54. The 3-component reaction of pantolactone **71**, β-alanine nitrile **76** and cysteamine **69** at pH 9 results in the formation of nitrile **77**, which then reacts with cysteamine **69** to form thiazoline **85**. Simultaneous reactions of β-alanine and cysteamine formed 2-aminoethylthiazoline **86** and the reaction of pantolactone with cysteamine also formed pantoyl-cysteamine **88**. These reactions curtailed the maximum possible yield of thiazoline **85**. Adjusting the pH down to 4 resulted in the hydrolysis of the thiazoline to pantetheine **66**.

At pH 7 the reactivity was dominated by the formation of 2-aminoethylthiazoline **86** from **76** and **69**, with little or no pantoyl β-alanyl nitrile **77** being detected. This can be rationalised by the pH being below the pK_{aH} of **76** and so the amine of β-alanine nitrile **76** is protonated and has diminished nucleophilicity and increased electrophilicity – cysteamine **69** reacts faster with **76** than β-alanine nitrile **76** does with lactone **71**. This selectivity would be expected to reverse above the pK_{aH} of **76**.

Incubating lactone **71** (500 mM) with cysteamine **69** (2 equiv.) and β-aminonitrile **76** (2 equiv.) at pH 9 furnished thiazoline **85** after 3 days, which was then quantitatively hydrolysed at pH 4 to furnish pantetheine **66** (57% yield over 3 steps in one pot from **71**; Figure 31). Each of these mechanistic steps were inferred by the intermediates found in the linear synthesis of **66** as discussed in previously.

To minimise the formation of 2-aminoethylthiazoline **86**, the reaction of **71** (500 mM) with 2 equivalents of β-**Ala**-CN **76** and only sub stoichiometric equivalent of cysteamine **69** (1 equiv.) were attempted. The reaction produced thiazoline **85** and nitrile **77**, but the reaction stalled as the depletion of 1 equivalent of cysteamine prevented the maximum conversion of pantoyl-β-alanine nitrile **77** through to thiazoline **85**. A sequential addition of another equivalent of cysteamine to this reaction mixture enabled the reaction to continue, through to thiazoline **85**. Unfortunately, the same reaction with 4 equivalents of cysteamine resulted in the β-alanine nitrile being consumed rapidly to 2-aminoethylthiazoline **86** which appeared unable to add to pantolactone **71** at pH 9.

The 3-component reactions also worked with lower concentrations of pantolactone (100mM) as the limiting reagent, however hydrolysis of lactone **71** to pantoic acid **67** dominated, with low yields of **85** (< 35%) being observed. Therefore, the best yields of thiazoline **85** (and pantetheine **66** from the hydrolysis of **85**) came from the reaction of lactone **71** (500 mM), and β-alanine nitrile **76** (2 equiv.) and cysteamine **69** (2 equiv.) at pH 9.
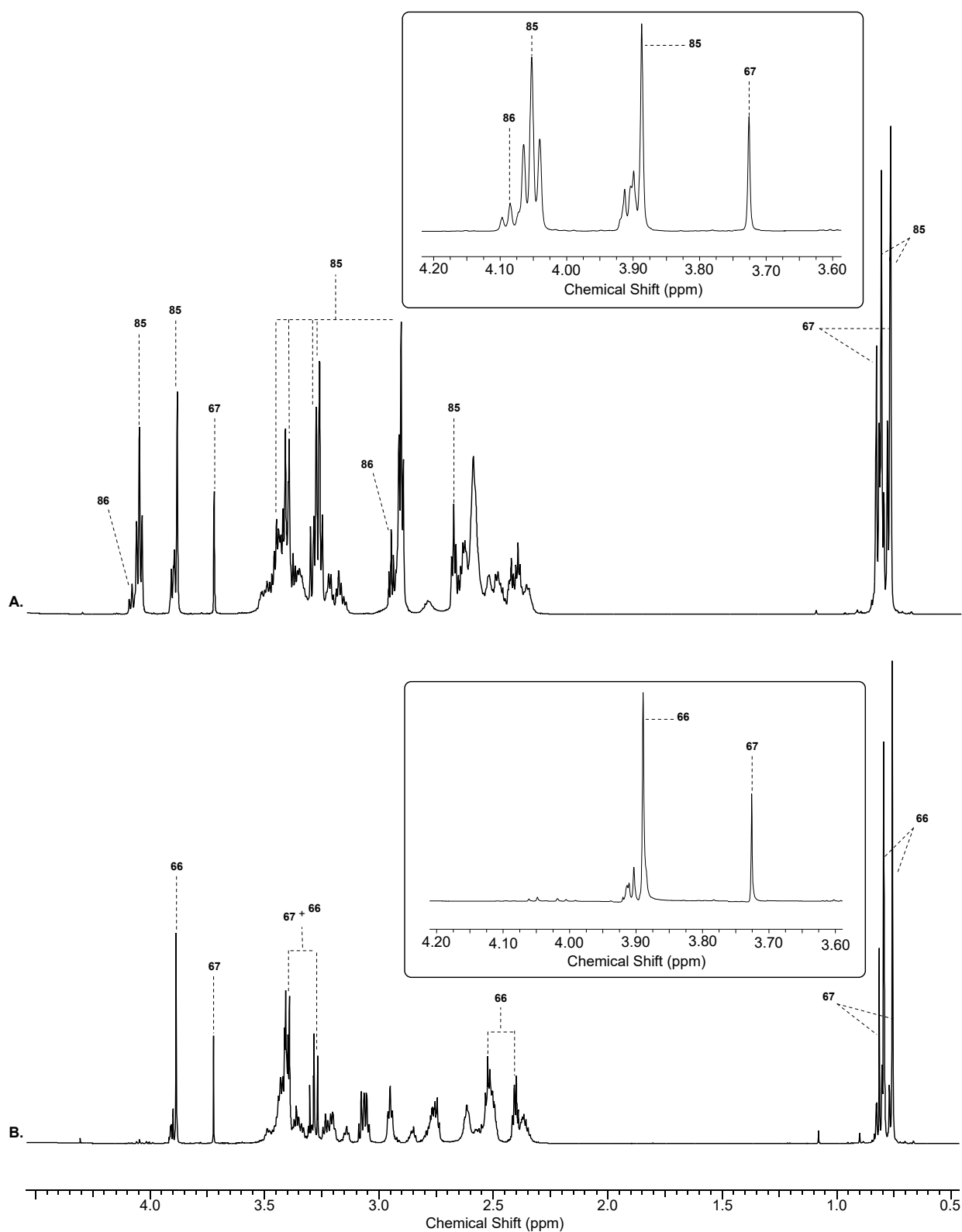
Figure 31. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d) spectra to show the acid-catalysed hydrolysis of thiazoline **85** to pantetheine **66**. (**A.**) The multicomponent one-pot synthesis of **85** by reaction of pantolactone **71** (500 mM), cysteamine **69** (2 equiv.) and β-alanine nitrile **76** (2 equiv.) in phosphate buffer (pH 9, 500 mM) at 20 ºC after 4 days; (**B.**) pantetheine **66** formation after incubating the crude reaction mixture at pH 4 and 20 ºC for 1 day, followed by NMR analysis at pH 9. Inset: Expanded ^{1}H NMR spectral region between 3.6–4.2 ppm.

141

2.4.5. Four Component One-Pot Synthesis of Thiazoline

Finally, the 4-component reaction of pantolactone **71**, β-alanine nitrile **76**, cysteamine **69** and glycine nitrile **Gly**-CN was investigated using the best results from the 3-component reaction. In the previous section, the 3-component reaction demonstrated that the desired synthesis of **66** could be achieved in good yields despite the unwanted side reactions from mutually reactive starting materials. Glycine nitrile is an α-aminonitrile and a precursor to prebiotic peptide synthesis (Sections 1.8.5 and 1.10.5). Glycine nitrile and α-aminonitriles are highly reactive by virtue of the low pK_{aH} of their amine and their electrophilic nitrile group. The high reactivity of α-aminonitriles with either pantolactone **71** or cysteamine **69**, could result in deleterious side reactions that could prevent a reasonable prebiotic synthesis of pantetheine **66**. The presence of α-aminonitriles in a multicomponent reaction must be taken account if a synthesis of pantetheine **66** would remain plausible in the prebiotic milieu.

Incubating lactone **71** (500 mM) with cysteamine **69** (2 equiv.), α-aminonitrile **Gly**-CN and β-aminonitrile **76** also furnished thiazoline **85** (33%) at pH 9 after 5 days as the major pantoyl-amide product (Figure 32). During this four-component reaction, the rapid reaction of glycine nitrile **Gly**-CN with cysteamine **69** suppressed the reaction of the α-aminonitrile with lactone **71** and depressed the reaction of β-alanine nitrile **76** with cysteamine **69** as well. Consequently, this happily favoured the addition of β-aminonitrile **76** to lactone **71** over the other reagents.
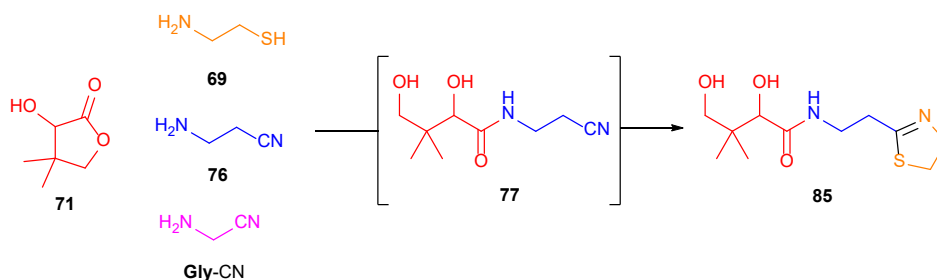
Scheme 55. The four-component reaction of pantolactone **71**, β-alanine nitrile **76**, glycine nitrile **Gly**-CN and cysteamine **69** at pH 9 results in the formation of nitrile **77**. The reaction also results in the rapid trapping of cysteamine **69** by glycine **Gly**-CN. This prevents coupling of glycine nitrile **Gly**-CN with lactone **71** and over time cysteamine **69** leaches back into the reaction mixture which then reacts with **77** to form thiazoline **85**.
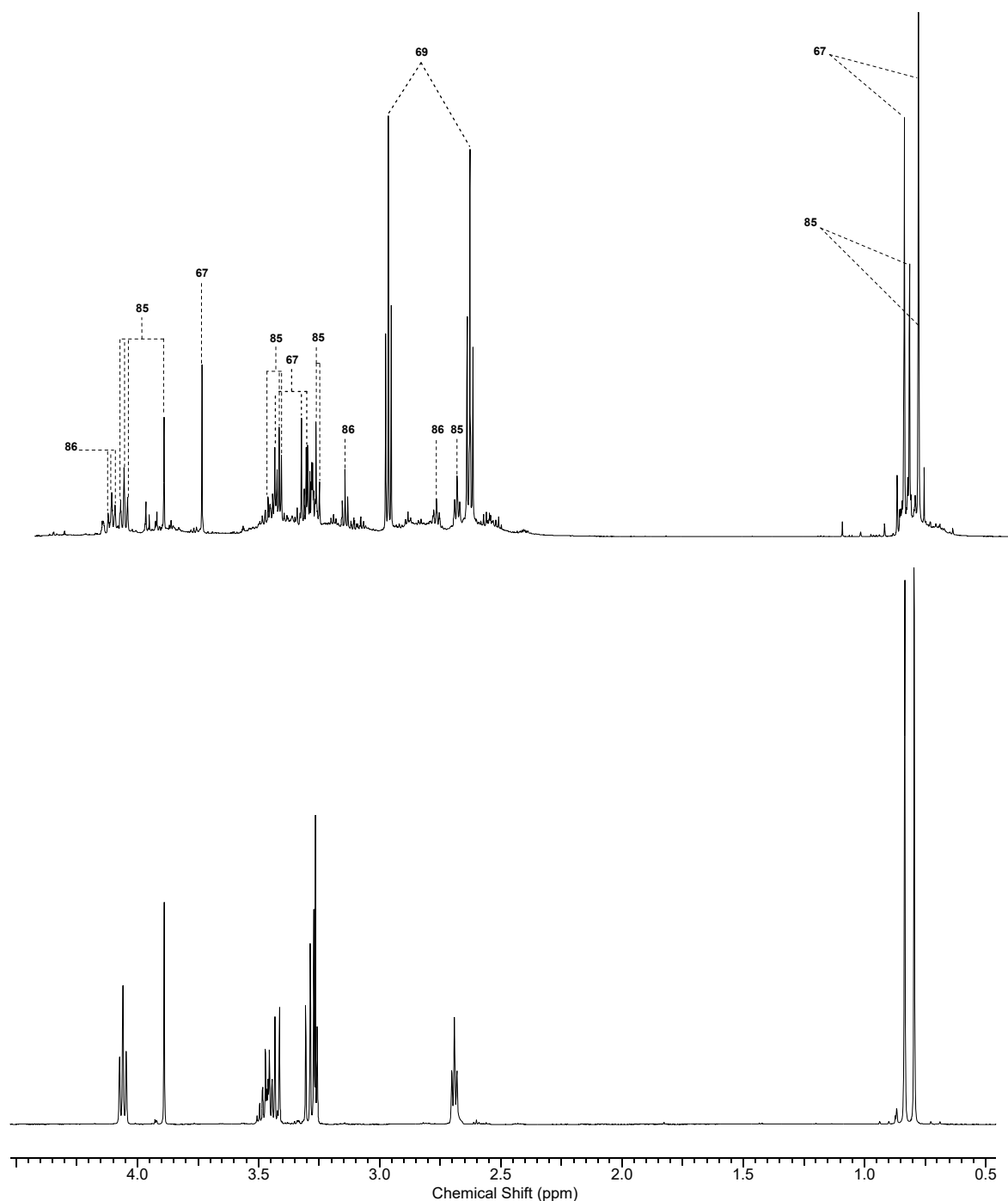
Figure 32. ^{1}H NMR (600 MHz, H_2O/D_2O 99:1, noesygppr1d) spectra to show the 4-component one-pot synthesis of thiazoline **85** by reaction of pantolactone **71** (500 mM), cysteamine **69** (2 equiv.), β-alanine nitrile **76** (2 equiv.) and glycine nitrile **Gly**-CN (2 equiv.) in phosphate buffer (pH 9; 500 mM) at 20 °C after: (**A.**) 5 days; and (**B.**) compared to authentic thiazoline **85**.

Remarkably in the 4-component reaction, the concentration of cysteamine **69** exhibited unusual behavior after observing its concentration by analysis of ^{1}H NMR and ^{1}H-^{13}C HMBC data over time. The data indicated that cysteamine was consumed early in the reaction within a few hours and thereafter was slowly released back into solution over 5 days from species presumably formed from the reaction of cysteamine **69** with α- and β-aminonitriles. Unfortunately, signal overlap prevented accurate quantification of cysteamine and the transient species by ^{1}H NMR. To test this hypothesis a control experiment of a sample of β-alanine nitrile **76** (500 mM), glycine nitrile **Gly**-CN (500 mM) and then the addition of cysteamine **69** (2 equiv.) into the reaction mixture at pH 9 was monitored by ^{1}H NMR. Cysteamine **69** depleted over 6 hours (36%) and then the concentration of **69** gradually increased (73%, 4 days). At this point > 50% of the total material was not accounted for by qNMR against an internal standard and an insoluble red/orange suspension/oil was observed in the reaction vessel. Heating the sample in boiling concentrated HCl resulted in > 90% recovery of pantoic acid **67**, β-alanine **68** and cysteamine **69**. This experiment demonstrated that the material was not lost indefinitely and that cysteamine can be recovered from this class of reactive multicomponent systems. These results show for the first time that the synthesis of thiazoline **85** (and **66**) can occur in appreciable quantities just by mixing the components of pantetheine **66** and even the reactive α-aminonitrile **Gly**-CN at pH 9. This again demonstrates the privileged role that α- and β-aminonitriles have in the formation of prebiotically relevant molecules.

Figure 33. General Summary of Results: (**A.**) β-Alanine-nitrile **76** facilitates high-yielding, activating-agent-free pantetheine **66** synthesis. (**i**) Lactone **71** reacts effectively with low pK_{aH} amine **76** in water to yield pantothenic acid nitrile **77**. Nitrile **77** reacts with cysteamine **69** to yield thiazoline **85**, which quantitatively hydrolyses to **66** in neutral water. (**ii**) One-pot multicomponent reaction of lactone **71**, β-alanine-nitrile **76**, and cysteamine **69** yields pantetheine **66** in water. (**B.**) Glycine nitrile **Gly**-CN reacts with lactone **71** to give pantoyl-glycine nitrile **78**. However, **78** equilibrates with **76** to yield **77** as the major product. The dynamic reaction of **71**, **Gly**-CN, and **76** favours the synthesis of the canonical β-alanyl structure of pantetheine **66**. (**C.**) ^{1}H nuclear magnetic resonance (^{1}H NMR; 700 MHz, H$_2$O/D$_2$O (99:1)) spectra showing the reaction of: (**i-ii**) lactone **71** (500 mM) and β-aminonitrile **76** (2 equiv.) in phosphate (Pi) solution (pH 9, 500 mM) at 20 ºC after: (**i**) 10 mins, and (**ii**) 3 days, yielding nitrile **77** (83%); (**iii**) nitrile **77** (500 mM) and cysteamine **69** (2 equiv.) at pH 9 yielding thiazoline **85** (84%) after 3 days; and (**iv**) thiazoline **85** (500 mM) in phosphate solution (pH 7, 500 mM) at 20 ºC, yielding pantetheine **66** (> 95%) after 43 days. (**D.**) ^{1}H NMR (700 MHz, H$_2$O/D$_2$O (99:1)) spectra showing the one-pot multicomponent reaction of: (**v**) lactone **71** (500 mM), **69** (2 equiv.), and **76** (2 equiv.) in phosphate solution (pH 9, 500 mM) at 20 ºC that yields **85** (57%) after 4 days; and (**vi**) the in-situ hydrolysis of **85** to pantetheine **66** (57% from **71**). ◆ = 2-Aminoethylthiazoline **86**.

2.4.6. Conclusions

The experiments in this chapter demonstrate that β-alanine nitrile **76** facilitates a high yielding, activating agent- and protecting group-free strategy to pantetheine **66** under mild conditions in water. The linear synthesis starting with the reaction of pantolactone 7**1** and β-alanine nitrile **76** forms pantoyl-β-alanyl nitrile intermediate **77** in high yields at pH 9. Competition reactions with aminonitriles indicate that the β-alanyl homologue is favoured over glycyl- and γ-amino butyl residues. Concomitant thiazoline **85** formation and hydrolysis to pantetheine **66** can be achieved at pH 7 in near quantitative yields over 60 days. These reactions can also be done in a 3-component one pot synthesis, yielding up to 57% thiazoline **85** after 4 days from pantolactone **71**, β-alanine **76** and cysteamine **69**. Quantitative hydrolysis of thiazoline **85** to **66** resulted in a 57% total yield over two steps. The multicomponent reaction is also compatible with glycine nitrile **Gly**-CN with 32% thiazoline **85** observed. The results are summarised on the next page (Figure 33).

2.5. Chemoselective aldol synthesis of hydroxypivaldehyde and pantolactone at neutral pH

2.5.1. Introduction

Pantothenic acid is comprised of a β-alanine and pantoic acid residue and in a seminal series of papers Williams was able to elucidate that pantothenic acid was a universal growth exponent of universal biological occurrence from studies on *Saccharomyces cerevisia*e. From the laborious and difficult extraction of pantothenic acid from sheep liver, Williams was able to isolate impure pantothenic acid (250 Kg of sheep liver yields 3g of pantothenic acid **70**, 40% w/w crude). At this time the structure was not known and could only be characterised by its activity as a growth factor in bacterial assays. Initially, pure samples were hard to come by from the extraction process, which hampered effective analysis of this compound. Despite these initial problems, Williams and Elvehjem simultaneously began to systematically elucidate the structure of pantothenic acid **67**. In the absence of modern techniques, such as NMR and HPLC, empirical testing, including combustion analysis, degradation studies, and comparison with synthetic samples eventually led to the structure being found. One of the most important findings was that pantothenic acid was a peptide with a β-alanine residue linked together by an amide bond to pantoic acid **67**. Pantoic acid was found to be a α, λ-dihydroxy acid which readily undergoes lactonization to pantolactone **71**.
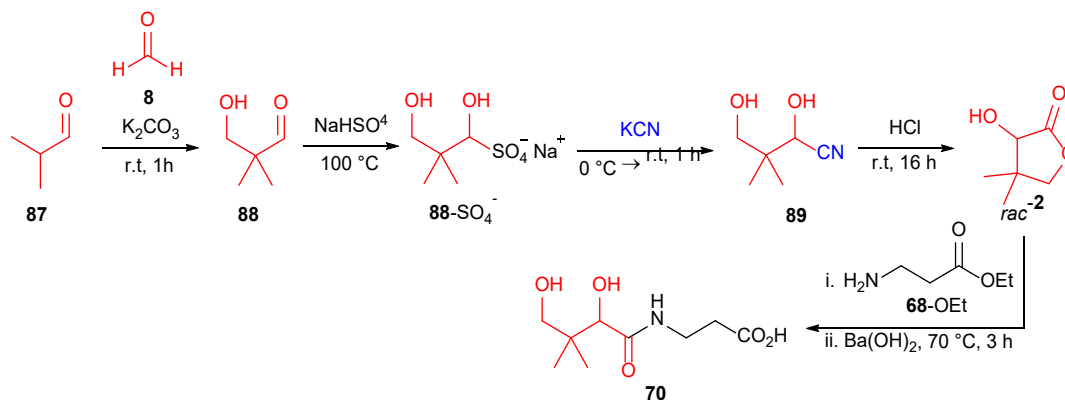
Scheme 56. The first synthetic route to pure pantothenic acid **70** via pantolactone **71** formation.[220,262]

The structure of pantoic acid **67** was confirmed by Snell in a partial synthesis of pantothenic acid **70**. Impure pantolactone **71** formed from the acid hydrolysis of pantothenic acid **70**, that was extracted from sheep liver, was heated (65-75 °C) with β-alanine methyl ester **68**-OEt for 30 minutes to give a sample of **70**.[263] Unfortunately, they did not isolate the compound and a team at Merck led by Folkers followed up with a total synthesis of pure pantothenic acid.[262] The aldol reaction of isobutylaldehyde **87** and formaldehyde **8** to hydroxypivaldehyde **88** had been investigated by Wessely, Glaser, Köhn and Neustädter in the early 1900's.[264] Inspired by this synthesis, Folkers was able to form aldehyde **88** from formaldehyde **8** and isobutyraldehyde **87** under alkaline conditions, and the onward reaction of the bisulfate salt of **88**-SO_4^- with HCN transiently furnished cyanohydrin **89**. From this point cyanohydrin **89** can form pantoic acid **67** at alkaline pH. However, pantoic acid **67** can only yield lactone **71** under acidic conditions (Section 2.2.1).[220,262]

From the synthesis of pure pantolactone **71**, the group at Merck were able to resolve its enantiomers by forming quinine salts. The synthetic (*R*)-(-) form was identical to the lactone formed from the hydrolysis of canonical pantothenic acid. (*R*)-(+) Pantothenic acid was successfully synthesised from (*R*)-(-) pantolactone and its biological activity matched extracted pantothenic acid. Racemic and (*S*)-(-) pantothenic acid had 50% and 0% growth stimulation activity of (*R*)-(+) pantothenic acid respectively.[262]

2.5.2. The Aldol Reaction between Formaldehyde and Isobutylaldehyde

Lactone **71** can be synthesised from hydroxypivaldehyde **88** and HCN,[220,262] but these methods required sequential (highly) alkaline and acidic conditions. Having demonstrated a nitrile-directed synthesis of pantetheine **66** in Chapter 2.4, the origins of pantolactone **71** in mild conditions were investigated.

General acid-base catalysis was thought to provide a route to aldehyde **88** in mild neutral pH conditions.[134,265,266] Initial experiments of isobutyraldehyde **87** (1.6–50 mM) and formaldehyde **8** (2.7–85 mM; 37% wt.% in water) in water or phosphate

buffer (16–500 mM) were tested at room temperature across a wide pH range. Addition of NaCN resulted in the simplification of the spectra from a mixture of aldehydes and their hydrates to their respective cyanohydrins, which enabled accurate quantification of the species formed. This would become essential later in these studies with mixed aldehyde systems. The solution pH was maintained at pH 7 on addition of NaCN to prevent further reactions to other useful species, to allow (at this point of the investigation) the cross-aldol reaction to be investigated and optimized.
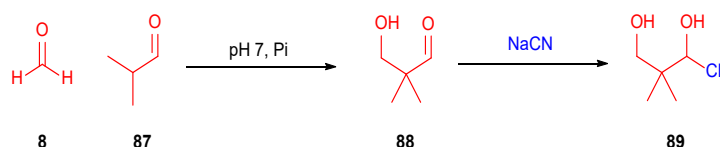
Scheme 57. The phosphate catalysed cross-aldol reaction of isobutylaldehyde **87** and formaldehyde **8** to give hydroxypivaldehyde **88** under neutral conditions.

The crossed aldol reaction of **8** with **87** to hydroxypivaldehyde **88** was observed to be sluggish at room temperature, for example 55% yield of hydroxypivaldehyde **88** was observed after 22 days from isobutylaldehyde **87** (50 mM) and formaldehyde **9** (~1.6 equiv.) at pH 7 in phosphate buffer (500 mM). In the absence of phosphate buffer, yields of **88** were considerably lower under neutral (pH 7, 2% yield), acidic (pH 4.3, 6% yield) or alkaline (pH 9.8, 15% yield) conditions over 7 days. Heating the Pi buffered reactions at 40 or 60 °C, significantly improved the rate of reaction, even for low concentrations of isobutylaldehyde **87**. At the lowest concentration of isobutylaldehyde (1.6 mM) a yield of 78% **88** was observed over 28 days at 60 °C. Whilst at a concentration of 50 mM of **88** a near-quantitative (94%) yield was reached after 2 days at 60 °C. These reactions were the most effective with phosphate buffer (10 equivs.) and reactions without buffer only gave hydroxypivaldehyde **88** in between 15-21% yield after one week at 60 °C (Scheme 57).

2.5.3. Chemoselective Aldol Reactions

Pantoic acid nitrile **89** could be formed quantitatively from hydroxypivaldehyde **88** and NaCN at pH 7. This allowed us to test of the selectivity of the cross-aldol reaction in the presence of other proteinogenic aldehydes, as well as investigating the stability of pantoic acid nitrile **88** in water. To test the selectivity of this crossed aldol reaction, aldehydes **8**, **87**, and another enolisable aldehyde, acetaldehyde **90** (alanine's Strecker precursor), were incubated in pH 7 phosphate solution at 60 °C over 48 hours. Again, the formation of hydroxypivaldehyde **88** (94%) was observed after 2 days. In addition to **88**, quantitative recovery of acetaldehyde **90** was now observed (Scheme 58) which did not form cross-aldol products with **8** or **87**.
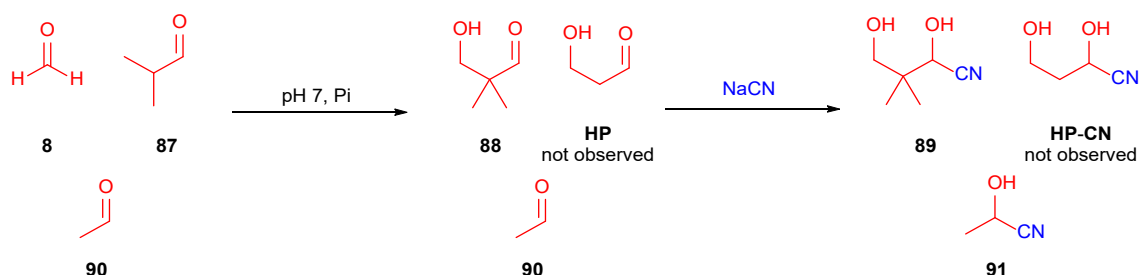
Scheme 58. The selective aldol reaction of isobutylaldehyde **87** and formaldehyde **8** in the precence of acetaldehyde **90**. Quantitative formation of **88** and the full recovery of acetaldehyde **90** was observed.

Finally, incubation of 20-156 mM formaldehyde **8** (glycine's Strecker precursor), 17 mM isobutylaldehyde **87** (valine's Strecker precursor), 23-31 mM mM acetaldehyde **90** (alanine's Strecker precursor) and 20-22 mM glycolaldehyde **9** (serine's Strecker precursor) at pH 7 in phosphate buffer (500 mM) at room temperature, 40 °C or 60 °C. All these reactions yielded hydroxypivaldehyde **88** as the major aldol product (Figure 34), with excellent recovery of **90**. Interestingly, alongside the synthesis of **88**, partial conversion of **9** to dihydroxyacetone **11**[23,35] from the cross-aldol and Lobry de Bruyn–Alberda van Ekenstein transformation of glycolaldehyde **10** and formaldehyde **8** was also observed to deliver up to 24% yield of **11**. Subsequent addition of HCN led to quantitative conversion of aldehydes **8**, **87**, **90** and **9** to their respective cyanohydrins **34**, **35**, **38**, **89** and **91**. As in previous experiments, aldehyde **88**, formed *in-situ*, was transformed into pantoic acid nitrile **89**. At pH 9.8,

the cross-aldol still occurred between **9** and **87** to give **88** (41%, 3 days, 20 °C), however a complex mixture was also observed, with the complete loss of glycolaldehyde **9** and substantial loss of **9**, **87** and **90**.
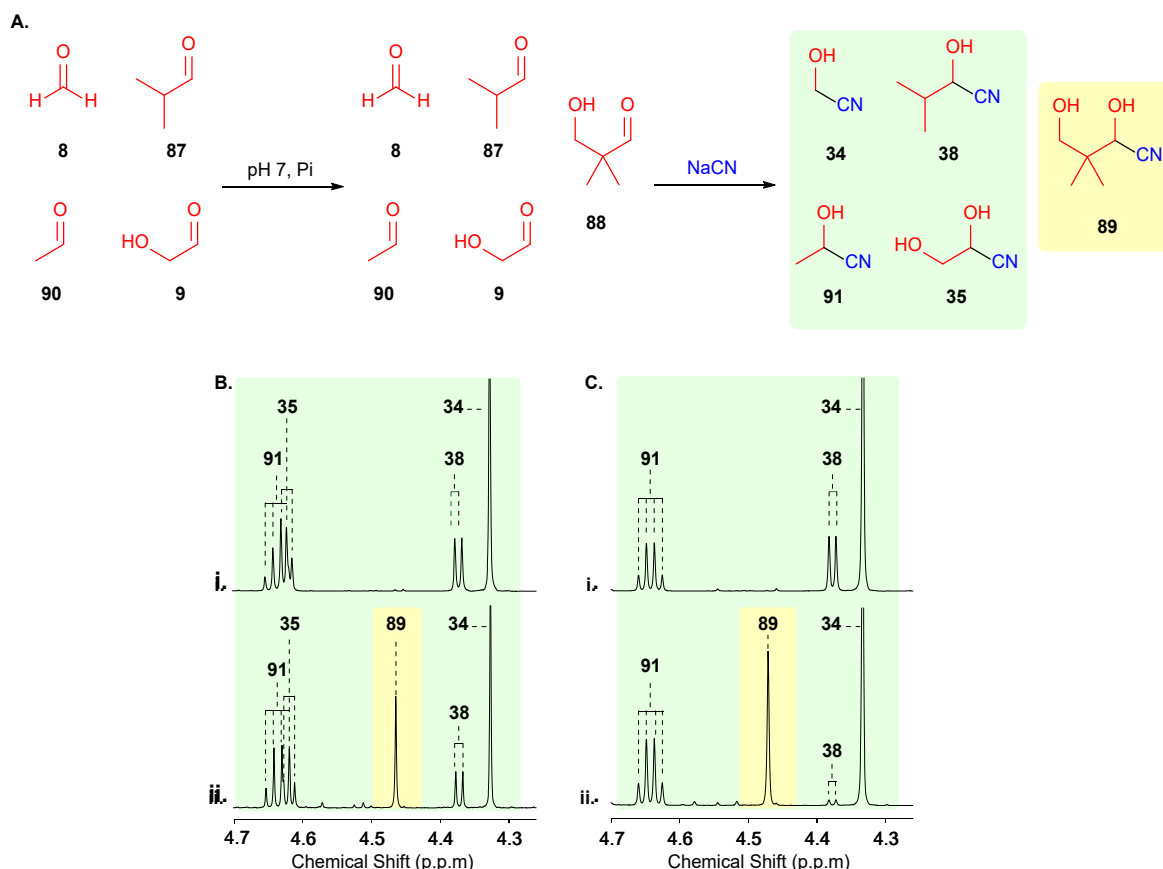
Figure 33. (**A.**) Phosphate-catalysed aldol condensation of formaldehyde **8** and isobutyraldehyde **87** generates hydroxypivaldehyde **88** in water at pH 7, even in the presence of acetaldehyde **90** and glycolaldehyde **9**. Aldehydes **9**, **10**, **87**, **88**, **90** react quantitatively with HCN to yield cyanohydrins **34**, **35**, **38**, **89**, **91**. (**B.**) ^{1}H NMR (600 MHz, H_2O) spectra to show the reaction of: (**i**) **8** (22 mM), **87** (17 mM), **90** (22 mM), **9** (22 mM) and NaCN (300 mM) after 20 mins at 20 °C; (**ii**) **8** (22 mM), **87** (17 mM), **90** (22 mM) and **9** (22 mM) after 1 day at 60 °C, followed by NaCN (300 mM) addition at 20 °C, yielding **89** (57%). (**C.**) ^{1}H NMR (600 MHz, H_2O) spectra to show the reaction of (**i**) **8** (100 mM), **87** (32 mM), **90** (50 mM) and NaCN (200 mM) after 20 mins at 20 °C; (**ii**) **8** (100 mM), **87** (32 mM) and **90** (50 mM) after 2 days at 60 °C, and addition of NaCN (200 mM) at 20 °C, yielding **89** (94%).

2.5.4. Investigating the Reactivity of Pantoic Acid Nitrile

Incubation of pantoic acid nitrile **89** in concentrated HCl resulted in 66% conversion to pantolactone **71** at room temperature over 7 days. While the result is good, the same reaction of pantoic acid nitrile at 60 °C resulted in the rapid quantitative (> 95%) lactonization in 2 hours. Although formation of pantolactone **71** can be formed

under these Fischer esterification conditions, these harsh conditions cannot be considered prebiotic. Moreover, the coupling of amines in the subsequent steps towards pantetheine **66** synthesis would be impossible under these conditions owing to protonation of the required amines below their pK_{aH} and this mismatched reactivity would be compounded by the concurrent hydrolysis of amide bonds in concentrated HCl, rendering the desired product unstable. It must be noted that leaving the reaction for prolonged periods or by heating for longer, resulted in the reaction hydrolysing out to inert pantoic acid **67**. This disperses the latent reactivity of the intermediates, which is required for onward reaction to pantetheine.
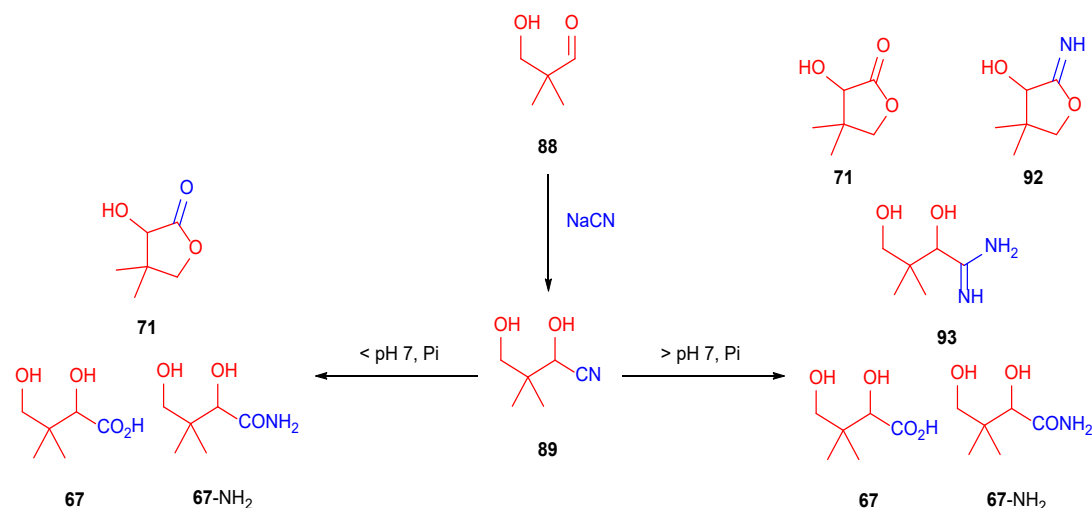
Scheme 59. The reactivity of pantoic acid nitrile **89** can be modulated according to the pH conditions. Under neutral to acidic conditions **71** is favoured. Whereas in basic conditions the reactive species, imidolactone **92** transiently dominates the chemical space. Total hydrolysis to **67** and **67**-NH$_2$ dominates over longer time periods.

Therefore, the next set of experiments tested efficacy of pantoic acid nitrile **89** hydrolysis under milder conditions (pH 7-9, 20 °C). The reaction of **89** at pH 7 and room temperature proved very promising, with a yield of 45% pantolactone **71** forming over 10 days. Increasing the pH to pH 7.5 improved the rate of reaction and the yield of pantolactone **71** reached 51% after only 5 days, with hydrolysis (**67** = 13%, **67**-NH$_2$ = 13%) also observed. When **89** is incubated under Strecker conditions (≥ pH 9) the reactivity is switched to favour a combination of *in situ* ammonolysis and hydrolysis products. For example, at pH 9 after pantolactone **71** was the minor product, observed in only 14% yield after 2 hours. Instead, the major products were

imidolactone **92** (35%) and amidine **93** (14%) being formed at high pH. Recovery of 41% of the starting material **24** (41%) was also observed after 2 hours. Increasing the solution pH to 9.5 increased the observed maximum yield of imidolactone **92** to 63%, with only 2% amidine **93** and 4% starting material being detected after 2 hours. The simultaneous hydrolysis of these reactive intermediates to **67** and **67**-NH₂ was then observed as the reaction was left for longer periods. However, this experiment demonstrated that iminolactone **92** could not only be readily formed in good yield (up to 65%) in water but that **26** could also react *in-situ* with ammonia to afford amidines and amides. These results suggested that trapping **92** with amines might be possible. This would enable the latent reactivity of cyanohydrin **89** to be utilized in conjunction with the intramolecular catalysis afforded by the γ-hydroxyl moiety to give an effective pantetheine **66** syntheses in water.
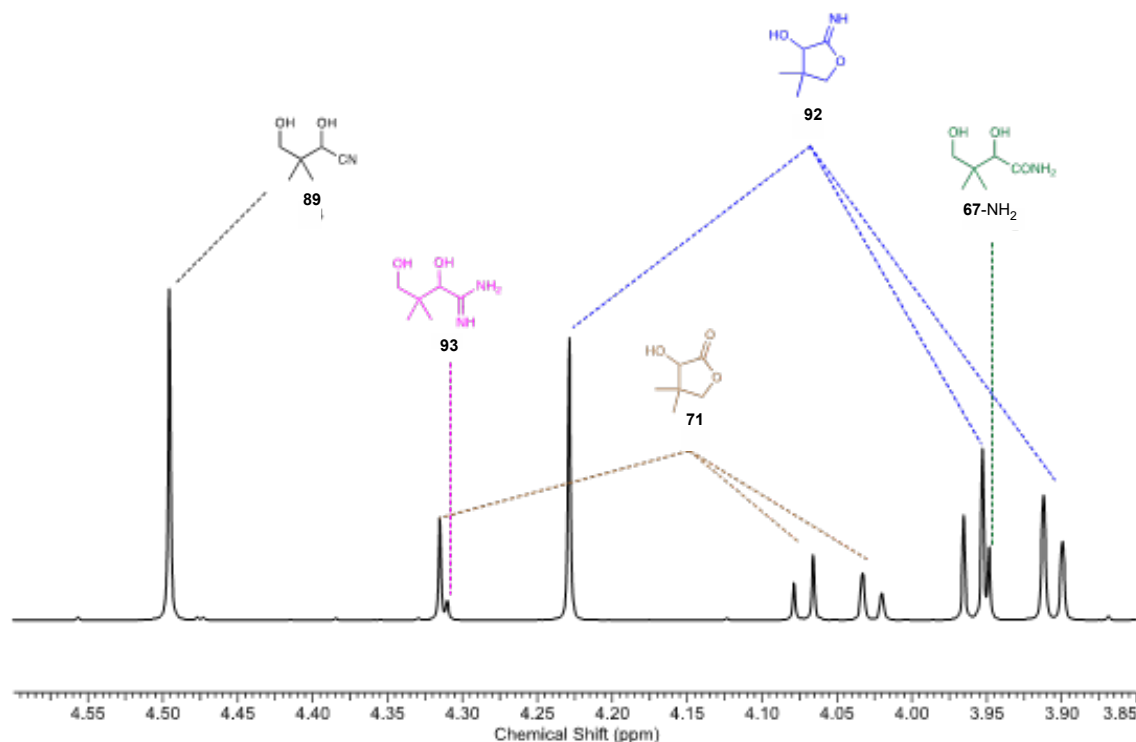
Figure 34. Expanded ¹H NMR (700 MHz, H₂O, noesygppr1d) spectrum after 2 hours of incubation of the pantoic acid nitrile **89** in phosphate buffer (pH 9.0; 500 mM) at 20 ºC. A mixture of pantoic acid nitrile **89**, pantoamidine **93**, iminopantolactone **92**, pantolactone **71** and pantoamide **67**-NH₂ were detected.

2.5.5. Conclusions

Over the course of these studies, it was found that the conversion of formaldehyde **8** and isobutylaldehyde **87** to hydroxypivaldehyde **88** is highly effective at neutral pH and general acid-base catalysed by phosphate. This work demonstrates the mild and selective conditions required to synthesise pantoic acid nitrile **89** in the presence of proteinogenic aldehydes. Subsequent experiments under the same conditions also provided a mild and direct, one step conversion of **89** to lactone **71** (Figure 36). Encouraged by the facile synthesis of lactone **71**, which was a key intermediate in the synthesis of pantetheine **66** (Section 2.5.4). The next step was to investigate the chemoselectivity required to integrate lactone **71** and utilise the by-products from the hydrolysis of cyanohydrin **89** in basic conditions towards the addition of amines such as β-alanine nitrile **76**. From this point the synthesis of pantetheine **66** with can then be explored.[174,177,199,211,244]
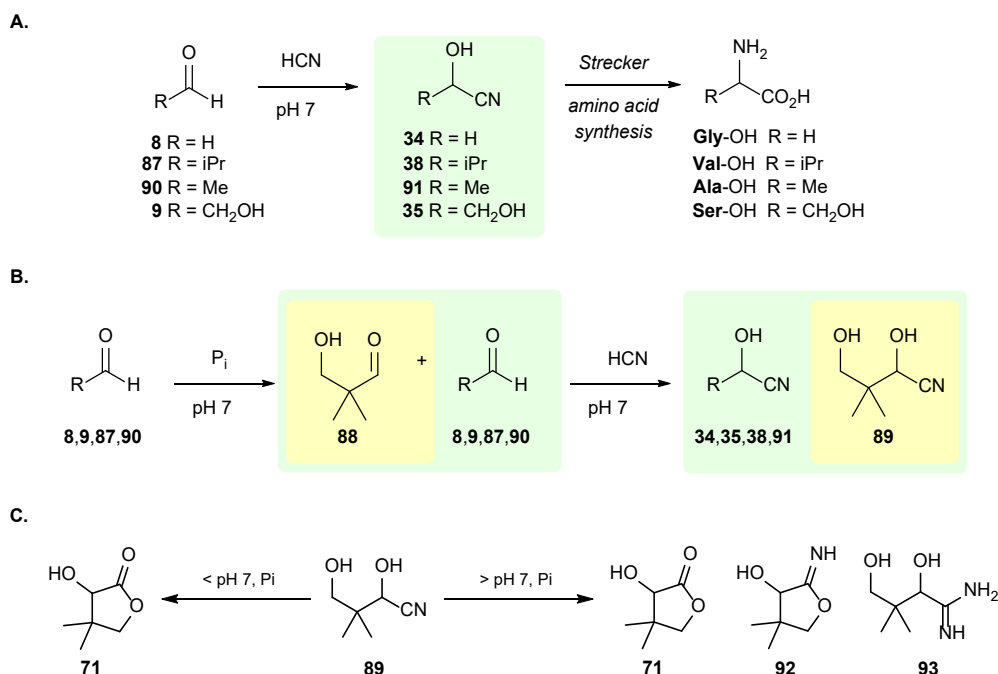
Figure 35. (**A.**) Strecker synthesis of amino acids from aldehydes. (**B.**) Phosphate-catalysed aldol condensation of formaldehyde **8** and isobutyraldehyde **87** to hydroxypivaldehyde **89** in water at pH 7. The reaction is selective even in the presence of acetaldehyde **90** and glycolaldehyde **9**. Aldehydes react quantitatively with HCN to yield their corresponding cyanohydrins **34**, **35**, **38**, **89**, **91**. (**C.**) The hydrolysis pathway for pantoic acid nitrile **89**, at pH < 7 pantolactone formation favoured, when the pH > 7 a spread of pantolactone **71**, iminolactone **92** and amidine **93** are formed.

2.6. Reactions of Hydroxypivaldehyde with Amines

2.6.1. Introduction

In the previous chapter pantoate synthesis had been achieved through exploiting the reactivity of formaldehyde **8** and isobutylaldehyde **87** whilst leaving the Strecker aldehyde precursors of proteinogenic amino acids intact.[28,244] However, selective concurrent synthesis of the pantoate moiety and proteinogenic α-aminonitriles **AA**-CN selectively in neutral water is more challenging. In this concurrent synthesis, the pantoate precursor, aldehydes **9** & **87**, should not form an aminonitrile; the aminonitriles of **15** could react with hydroxypivaldehyde **89** and form an α-amine, not the canonical pantoate α-hydroxyl moiety under Strecker conditions. Conversely, at the same time and under the same conditions, aldehydes, must also form appreciable quantities of proteinogenic α-aminonitriles **AA**-CN so that these substrates can be used in peptide ligation (Section 1.8.5).

2.6.2. Differentiation of Pantoic Acid Nitrile from Proteinogenic α-Aminonitriles

Remarkably, when incubating aldehydes **8**, **9**, **87**, **88**, and **90** under Strecker conditions with cyanide and ammonia led to the complete differentiation of peptide and pantoate precursors (Figure 37). Chemoselective proteinogenic α-aminonitrile glycine **Gly**-CN, valine **Val**-CN, alanine **Ala**-CN, and serine **Ser**-CN formation were observed from aldehydes **8**, **9**, **87**, **88**, and **90**. However, the non-proteinogenic α-aminonitrile synthesis from aldehyde **88** is blocked by rapid iminolactone **92** formation. Iminolactone **92** reacted with NH_3 to yield α-hydroxy amidine **93**, which contains the complete pantoyl carbon and oxygen framework essential to the synthesis of pantetheine **66**. It will be demonstrated in this chapter that amidine **25** can undergo transamidation with β-alanine-nitrile **93** even in the presence of excess NH_3 (Section 2.6.5). This spontaneous differentiation of proteinogenic α-amino acid and pantoate synthesis demonstrates the required reactivity to selectively deliver the α-hydroxyl moiety of pantetheine **66**.
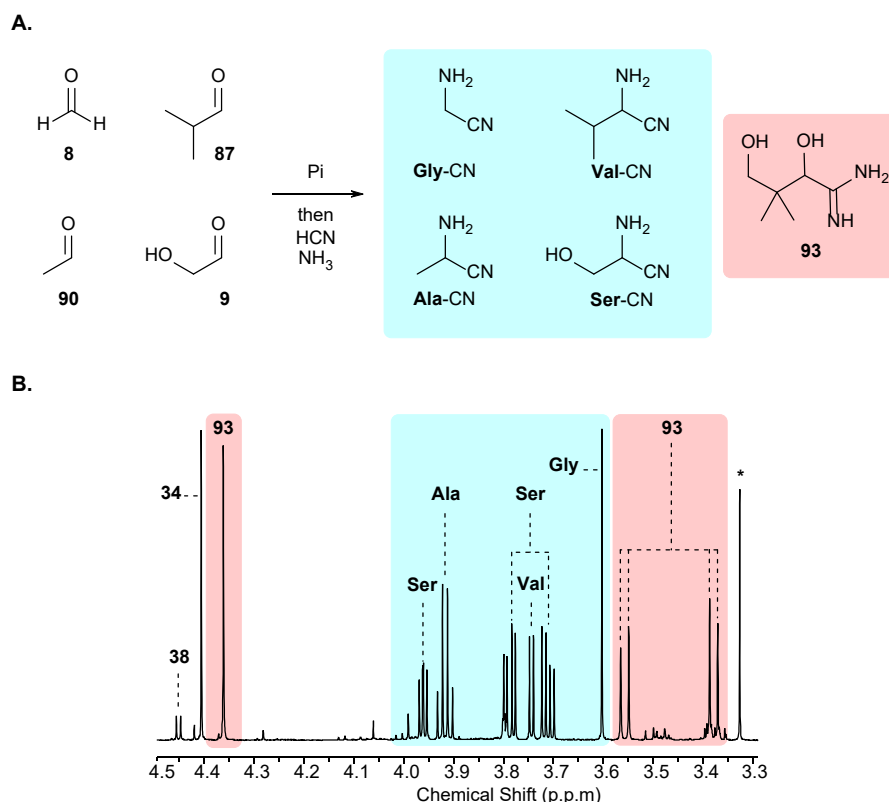
Figure 36. (**A.**) Phosphate-catalysed aldol condensation of formaldehyde **8** and isobutyraldehyde **87** selectively yields hydroxypivaldehyde **88** at neutral pH, even in the presence of other enolisable aldehydes (i.e., acetaldehyde **90** and glycolaldehyde **9**; (**B.**) ^{1}H NMR (700 MHz, H$_2$O) spectrum to show the reaction of aldehydes **88** (20 mM), **8** (20 mM), **87** (17 mM), **90** (31 mM), **9** (20 mM), NH$_3$ (500 mM) and NaCN (150 mM) after 1 day in phosphate (Pi) solution (pH 9.5, 500 mM) at 20 °C, yielding α-aminonitriles **Gly**-CN, **Ala**-CN, **Val**-CN, and **Ser**-CN and α-hydroxy amidine **93**. * = methanol.

2.6.3. Synthesis of Pantothenic Acid by reaction from Hydroxypivaldehyde, β-Alanine and Cyanide

The observed differentiation of hydroxypivaldehyde **88** from amino acid precursors implicated iminolactone **92** as a useful intermediate. Earlier reports suggested that **92** has 'extraordinarily high sensitivity to hydrolysis'[267] but this is inconsistent with these results. We observed iminolactone **92** in up to 63% yield during the reaction of **88** with HCN in water (Section 2.5.4). However, it was anticipated that the addition of cyanide to a mixture of aldehyde **88** and β-aminonitrile **68** would initiate a reaction cascade that would generate **92**, which would then be intercepted by β-alanine nitrile **76** to yield pantothenic acid nitrile **77**. This multicomponent reaction would streamline

pantetheine **66** synthesis by creating new carbon-carbon and amide bonds in one pot. This would also bypass the less electrophilic pantolactone **71**, in a single step.

To test this proposal, the cascade reaction was tested on β-alanine **68** for the synthesis of pantothenic acid **70** at concentrations where β-alanine **68** failed to react with lactone **71** (Section 2.2.1). Upon incubating aldehyde **88** (50 mM), cyanide (1.5 equiv.) and β-alanine **68** (2 equiv.) the formation of pantothenic acid **70** in up to 51% yield was observed (Figure 38).
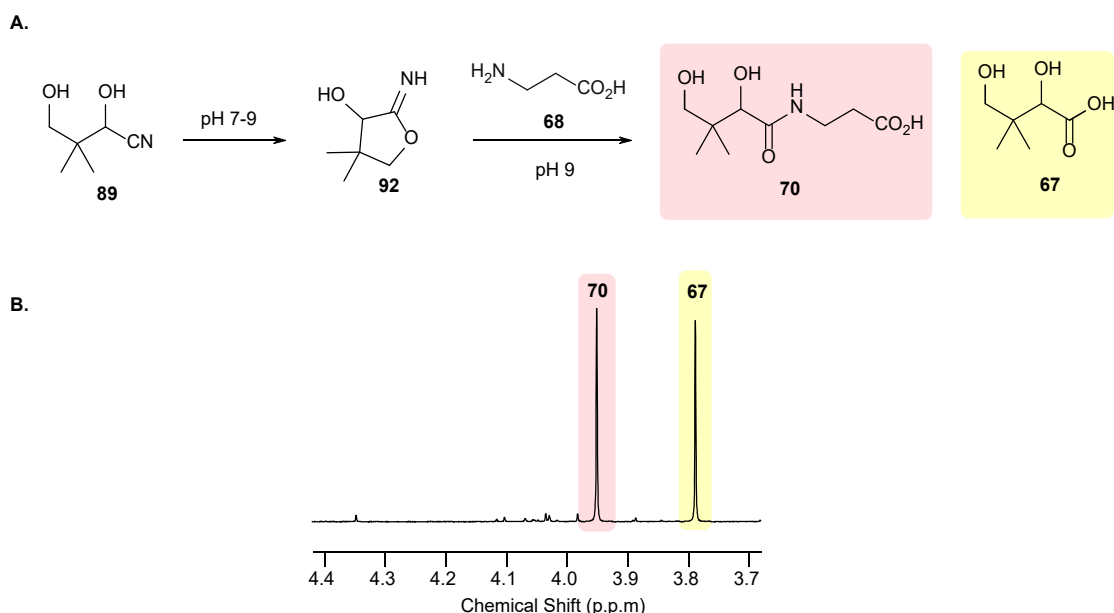
Figure 37. By-passing pantolactone using intramolecular nitrile activation allows pantothenic acid synthesis in dilute aqueous solution. Iminolactone **92** is a superior electrophile than lactone **71**, and **92** is sufficiently stable to form amide bonds in dilute aqueous solutions. (**A.**) Iminolactone **92**, generated *in-situ* through cyclisation of pantoic acid nitrile **89** (< 50 mM), reacts with β-alanine **68** (2 equiv.) to give pantothenic acid **70** at pH 7.5-9.0 in up to 51% yield. (**B.**) ^{1}H NMR (700 MHz, H$_2$O) spectrum to show the reaction of aldehyde **88** (12.5 mM), **68** (2 equiv.) and NaCN (1.5 equiv.) in phosphate solution (pH 9.0; 125 mM) yielding **70** (51%) and **67** (44%) after 7 days at room temperature.

The enhanced electrophilic reactivity of iminolactone **92** provides a mechanism for pantoyl-amide bond formation that is effective even at high dilution. The reaction with β-alanine nitrile **76** was attempted at low concentrations to bypass the potential Strecker reaction that could occur with this amine. Following a similar protocol to the synthesis of pantothenic acid with the iminolactone, aldehyde **88** (3.1 mM) was

incubated with β-alanine nitrile **76** (2 equiv.) and HCN (1.5 equiv.). Remarkably this reaction returned nitrile **77** in 44% yield after 6 days (Figure 39).
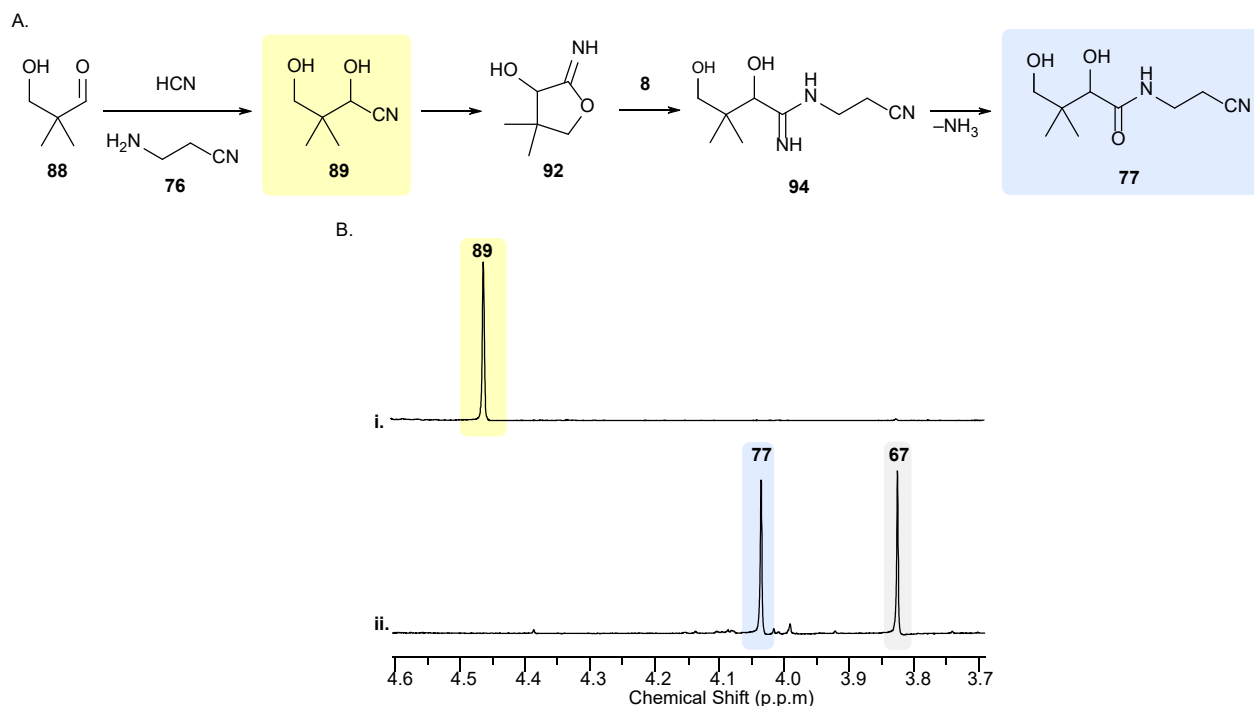
Figure 38. (**A.**) One-pot multicomponent synthesis of pantothenic acid nitrile **77** through the selective reaction of aldehyde **88**, HCN and low pK_a amine, β-alanine-nitrile **76**. (**B.**) ^{1}H NMR (700 MHz, H$_2$O) spectra to show the one-pot multicomponent reaction of **88** (3.1 mM), **76** (6.3 mM), and NaCN (3.4 mM) in phosphate (Pi) solution (pH 9.0, 31 mM) after: (**i**) 10 mins, showing pantoic acid nitrile **89**, and (**ii**) 6 days, yielding pantothenic acid nitrile **77** (44%) and pantoate **67** (41%).

2.6.4. Competition Reaction of β-Alanine Nitrile and Glycine Nitrile with Hydroxypivaldehyde and Cyanide

The synthesis of pantothenic acid nitrile **77** *via* iminolactone **92** suggested a novel and inherent mechanism to block the synthesis of (non-canonical) α-homologues of pantetheine **66**. In this cascade reaction, the initial β-alanine nitrile **67** addition product, β-amidine-nitrile **94**, slowly hydrolysed to **77** (Figure 39). However, we predicted that α-amidine-nitrile **95** from **Gly**-CN addition would undergo rapid cyclisation to aminoimidazole **96** and irrevocably block onward reaction with cysteamine **69** and formation of non-canonical analogues of pantetheine **66**. In line with our prediction, we observed that the reaction of **88**, cyanide and glycine nitrile **Gly**-CN furnished aminoimidazole **96** 42% (Figure 40). Struck by this specific mechanism to block the synthesis of non-canonical pantetheine homologues,

158

reactions with α-aminonitrile **Gly**-CN and β-aminonitrile **76** were carried out in anticipation that pantothenic acid nitrile **77** would emerge as the only pantoyl-amide capable of onward reactivity with cysteamine **69**. Pleasingly, the reaction of **Gly**-CN, **76**, **88** and cyanide resulted in a highly chemoselective formation of pantoyl-β-alanine nitrile **77** (44%) alongside only traces of aminoimidazole **96** (< 4%), and again pantoyl-α-glycine nitrile **78** was not detected (Figure 40).

Figure 39. Chemoselective pantothenic acid nitrile synthesis by a multicomponent reaction cascade activated by cyanide in dilute aqueous solution. (**A.**) Addition of HCN to aldehyde **88**, β-aminonitrile **76**, and α-aminonitrile **Gly**-CN initiates a reaction cascade that begins with quantitative formation of pantoic acid nitrile **89**. Intramolecular cyclisation of **89** generates iminolactone **92** which reacts chemoselectively with β-aminonitrile **76** to furnish amidine **94**, which undergoes hydrolysis to pantothenic acid nitrile **77**. The shorter homologue **77** is not observed; its synthesis is blocked by cyclisation of amidine **95**. (**B.**) ^{1}H NMR (700 MHz, H_2O) spectra to show the one-pot multicomponent reaction of **88** (3.1 mM), **76** (6.2 mM), **Gly**-CN (6.2 mM) and HCN (4.7 mM) in phosphate solution (pH 9, 31 mM) at 20 °C after: (**i**) 10 mins, and (**ii**) 3 days, yielding pantothenic acid nitrile **77** (44%).

These results demonstrate that the reaction of aminonitriles with iminolactone **92** favours the synthesis of the canonical pantothenic acid nitrile **77**. The reaction of α-aminonitriles with **92** also irrevocably blocks the onward synthesis of non-biological pantetheine analogues by a mechanism that is unique to aminonitriles. Therefore, nitrile reactivity provides the selectivity essential for pantetheine **66** synthesis by routes that unequivocally account for the chemical basis of the β-alanyl fragment of pantetheine **66**.

2.6.5. Synthesis of Pantothenic Acid Nitrile under Strecker Conditions

Finally, having observed the rapid conversion of aldehyde **88** into α-hydroxy amidine **93** during the differentiation of aldehyde **88** from the Strecker (aldehyde) precursors of proteinogenic amino acids (Section 2.6.2), it was essential to establish whether conditions for α-aminonitrile synthesis are compatible with pantothenic acid nitrile **77** formation. The synthesis of amidine **93** appeared to be an irrevocable by-product that could, in principle, completely block synthesis of **77**. However, upon the addition of β-alanine-nitrile **76** to amidine **93**, pantoyl-alanyl nitrile **77** (up to 64%) was observed. This is thought to occur through intramolecular γ-hydroxyl-catalysed transamidation, even in the presence of a large (5–25 equiv.) excess of ammonia (Figure 41). The subsequent addition of cysteamine **69** resulted in the *in-situ* conversion of nitrile **77** to pantetheine **66** in 51% overall yield from aldehyde **88**.
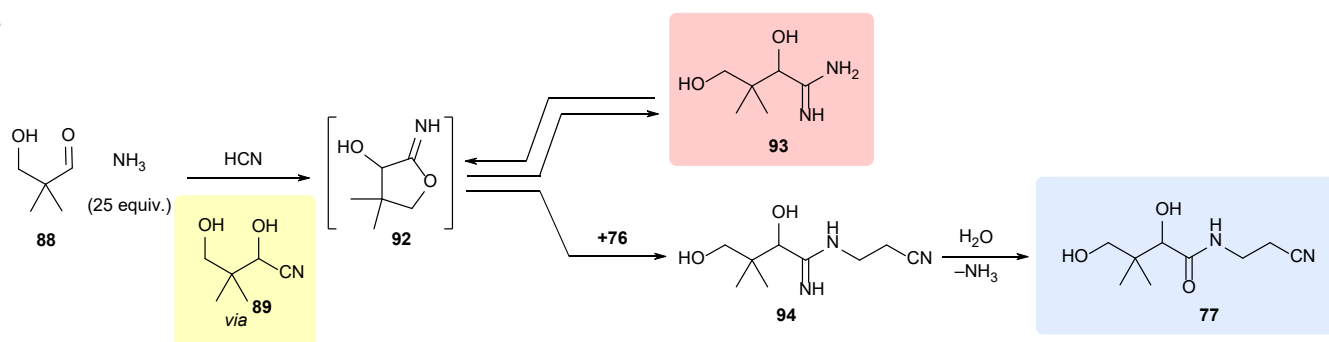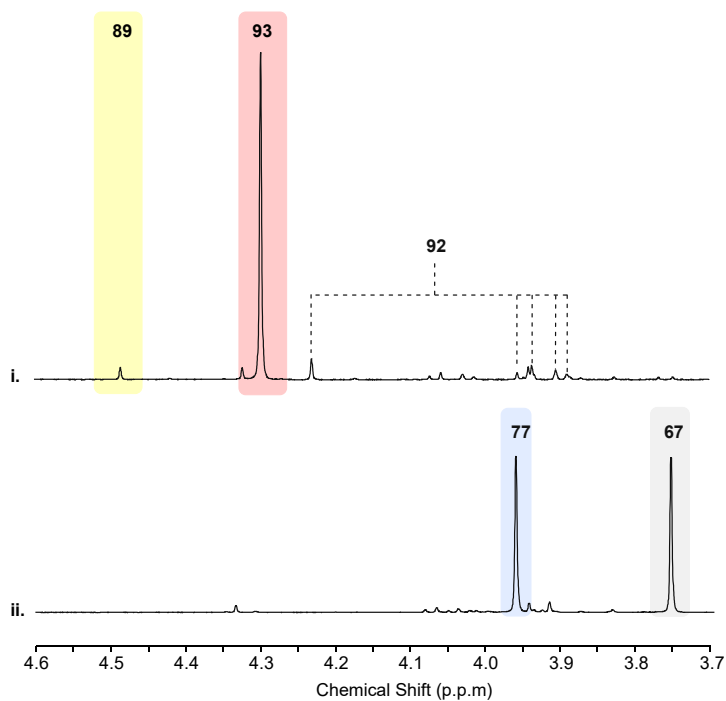
Figure 40. Chemoselective pantothenic acid nitrile synthesis by a multicomponent reaction cascade activated by cyanide in dilute aqueous solution. (**A.**) The reaction of HCN and ammonia (NH_3) with **88** furnishes α-hydroxy amidine **93** in near-quantitative yield under Strecker conditions. However, **93** equilibrates with iminolactone **92**, and therefore subsequent addition of β-aminonitrile **76**, even with a very large excess of NH_3, leads to pantothenic acid nitrile **77** as the only major pantoyl-amide product. (**B.**) ^{1}H NMR (600 MHz, H_2O) spectra to show the one-pot synthesis of pantothenic acid nitrile **9** *via* α-hydroxyamidine **93** under Strecker conditions by reaction of **88** (20 mM), NaCN (30 mM), NH_3 (500 mM) in phosphate solution (pH 9.5, 500 mM) at 20 ºC, after: **(i)** 4.5 hours, yielding **93** (82%), followed by **(ii)** subsequent addition of **76** (40 mM) to yield **77** (46%).

2.7. Conclusion

Pantetheine **66** is a structurally unique and essential biological β-amide, and its (prebiotic) synthesis must be reconciled with the concurrent generation of proteinogenic α-peptides. In this thesis, we have discovered a series of reactions that are exclusive to nitrile chemistry and contribute to the synthesis of **66** with unprecedented selectivity over non-biological homologues. Furthermore, by querying the chemical relationship of **66** to Strecker precursors of α-peptides, we have discovered the phosphate-catalysed reaction of formaldehyde **8** and isobutyraldehyde **87**, the Strecker precursors of **Gly**-OH and **Val**-OH, yields hydroxypivaldehyde **88** in the first step towards **66**. This aldol condensation is highly effective at neutral pH, even within mixtures that include other enolisable aldehydes. Interestingly, with respect to concurrent prebiotic sugar, amino acid and pantetheine synthesis, we also observed slow but selective conversion of C2-aldose glycolaldehyde **9**, to C3-ketose dihydroxyacetone **11**. Dihydroxyacetone was observed under the same reaction conditions that furnished aldehyde **88**. It is therefore of note that we have previously demonstrated the sequential accumulation of C2 and C3 aldose sugars, in the order required, for prebiotic nucleotide synthesis, from complex sugar mixtures by coupling the phosphate-catalysed isomerisation of C3 sugars with C3-aldehyde selective co-crystallisation.[35] Using the same mechanism, we have additionally demonstrated aldehyde co-crystallisation separates ketones (e.g., dihydroxyacetone) from Strecker aldehydes, en-route to selective proteinogenic amino acid synthesis.[35] These observations, and the common origins of these nucleotide and peptide syntheses with those reactions reported here for the synthesis of pantetheine **66**, warrant further investigation by simultaneous (physicochemical directed) synthesis of life's core metabolites from the aldehyde products of cyanide reduction.[28,244]

The reaction of hydroxypivaldehyde **88**, with HCN and β-alanine-nitrile **76** selectively generates pantothenic acid nitrile **77**, even in direct competition with α-aminonitriles **AA**-CN. Furthermore, no activating agents are required for the synthesis of pantetheine **66** by nitrile chemistry, latent nitrile-activation is installed into pantoic

acid nitrile **89** and pantothenic acid nitrile **77**. Moreover, pantoyl-amide formation requires no external catalysis; the ideally poised γ-hydroxyl moiety of nitrile **89** is an intramolecular nucleophilic catalyst for amide bond formation. The γ-hydroxyl of **89** also blocks α-aminonitrile synthesis, and thus provides a highly selective mechanism to differentiate proteinogenic α-aminonitriles from γ-hydroxy-pantoate derivatives. This mechanism endows iminolactone **92** with a sufficient lifetime and reactivity in water to be intercepted by β-alanine-nitrile **76**, whilst blocking pantetheine analogue synthesis from α-aminonitriles **AA**-CN. Although **92** is transformed into amidine **93** under Strecker conditions that produce α-aminonitriles **AA**-CN, the γ-hydroxyl of amidine **93** catalyses transamidation of ammonia with β-alanine-nitrile **76** to yield pantothenic acid nitrile **77**, even in the presence of excess ammonia. The additional discovery of the dynamic reactions of α- and β-aminonitriles and cysteamine **69** with lactone **71** (a product of iminolactone **92** hydrolysis) allowed for direct one-pot multicomponent, activating-agent-free synthesis of pantetheine **66** from pantolactone **71**, cysteamine **69**, and β-aminonitrile **76** even in direct competition with α-aminonitrile **Gly**-CN. Collectively, our results suggest that the chemical origins of pantetheine **66** are best rationalised through nitrile, not carboxylic acid, chemistry.

The selective syntheses of **66** in water challenges the persisting dogma that, despite it being the 'solvent of life',[268] water is problematic (or even a 'poison') for prebiotic chemistry.[268] However, we observed highly effective nitrile-activated amide bond formations in water, even below physiological CoA concentrations within modern cells.[269] This chemistry not only favours the canonical structure of **66** but also closely aligns with previously reported prebiotic pathways to α-peptides, RNA, and lipids.[28,135,177,199,211,222,244] Therefore, our results suggest that **66** would have been a product of cyanosulfidic reaction pathways prior to the emergence of life on Earth. Once available, it is simple to envisage how **66** could have been deployed at the origins of life, for example as a (nucleotide-coded) organocatalyst or cofactor to enhance the functional limitations of early ribozymes[85,212,239] or peptide catalysts,[199] mirroring its essential role augmenting the functional repertoire of enzymes in extant biochemistry.[85,105,202]

3. Photochemistry of Thioacids – Results & Discussion

3.1. Introduction

As described previously (Section 1.2), the early Earth was not screened from UV light, and this may have enabled new pathways to prebiotic molecules to be unlocked. It was previously discussed in this review that UV light could be utilised to give desired nucleotides or amino acids through the cyanosulfidic protometabolic network (Section 1.7). Here the potential for photochemical acylation will be discussed.

In another example, Hagan Jr. was able to use UV light to acetylate phosphate to synthesise the metabolite acetyl phosphate with potassium thioacetate and uracil as a catalyst in mild conditions at pH 7.[270] Acetyl phosphate is a precursor to phosphate esters and pyrophosphates and could have been an important component of the first metabolic cycles.[255,258,270]

In Powner's peptide ligation of α-aminonitriles in water (Section 1.8.5), the cycle requires the sequential addition of ferricyanide $Fe(CN)_6^{-3}$ and H_2S. This poses a potential problem in prebiotic chemistry, since this ligation cycle fluctuates between reducing and oxidising conditions with reagents that are mutually reactive.[174] To overcome this spatiotemporal problem, the first aim of the project envisages simulating the Earth's day/night cycle to potentially acetylate α-aminonitriles with potassium thioacetate by irradiation of UV-light in a light phase. This would eliminate the need for an oxidant and allow for the subsequently activated acyl aminonitrile to undergo thiolysis with H_2S in a dark phase. The thioacid formed could potentially ligate with α-aminonitriles when irradiated with UV light in the day and so the ligation cycle could continue without the need for the $Fe(CN_6)^{-3}$ oxidant. Therefore, project aims to find conditions where UV light could form dipeptides from Ac-**AA**-SH and **AA**-OH.

Recent work by Tan found that amide bonds could be formed by the irradiation of thioacetate and substituted anilines with visible light, in air and in the presence of a

(prebiotically implausible) photoredoxcatalyst (Ru(bpy)$_3$Cl$_2$) in acetonitrile (Scheme 60).[271] They suggest that this approach mimics acetyl-CoA to selectively acetylate amines. They were able to explore substrate scope through a wide range of thioacids for example, the acetylation of the methyl esters of alanine, phenylalanine, and tyrosine in good yields. They also performed a ligation with *N*-boc alanine thioacid with tyrosine methyl ester.[271]
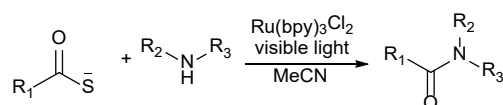
Scheme 60. Tan's acylation of amines with photocatalyst and visible light.

Further work by Li expanded on the work by Tan to form amide bonds with thioacids and Mes-Acr-MeBF$_4$ as a photocatalyst in blue light. They were able to ligate anilines with *N*-protected thioacids of glycine, alanine, proline, methionine, threonine and phenylalanine. Whilst Ru(Bpy) is a redoxphotocatalyst and are not prebiotically plausible, acetone **36** is known to be a triplet photosensitiser and it is thought that photosensitisers could enable the reactivity to be tailored to the desired pathway during the early Earth's day/night cycle. This is important since acetone **36** is a node in the 'cyanosulfidic' protometabolic network (Section 1.7) to amino acids leucine and valine.[28]   Additionally, acetone **36** is formed from the reduction of dihydroxyacetone **10**, which could also be a photosensitising agent.   These molecules can be considered prebiotically plausible since, dihydroxyacetone **10** can be formed from the reductive homologation of HCN from formaldehyde **8** over several steps in the 'cyanosulfidic' protometabolic network or by its synthesis in the Formose or aldol reactions. If photo-ligation is successful, a screen of other potential photocatalysts such as uracil (Hagan Jr.)[270] will be investigated.

Recent work in the Powner group has found that thiols such as *N*-acetyl cysteine are good catalysts in the formation of dipeptides and larger peptides by fragment ligation (Section 1.10.5).[199]   In biology cysteine is the primary source of sulfur and an essential feedstock of essential cofactors such as glutathione and CoA. Cysteine is important in other biological process such as catalysis, electron transfer or can be

traced back to ancient iron-sulphur proteins.[199,272] It is likely that cysteinyl thiol residues may have played a role in early biochemistry (Section 1.10) and in the biosynthesis of pantetheine (Chapter 2)

Through the dethiocarboxylation of *N*-acetyl amino thioacids Ac-**AA**-SH, it was hypothesised that *N*-acetyl cysteine thioacid, Ac-**Cys**-SH[199] would also undergo de-thiocarboxylation to *N*-acetyl cysteamine **27** (Scheme 61).
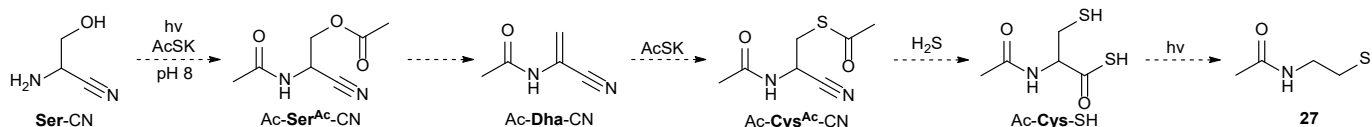
Scheme 61. The proposed prebiotic syntheses of *N*-acetyl cysteamine **27** from serine nitrile **Se**r-CN using UV light.

 In principle, the 'pantetheine arm' of CoA could be synthetized by the formation of the dipeptide *N*-acetyl β-alanine-cysteine Ac-β-**Ala**-**Cys**-SH from the combination of chemistry discussed previously. When Ac-β-**Ala**-**Cys**-SH is irradiated with UV light, de-thiocarboxylation would hopefully give *N*-acetyl-β-alanine cysteamine, a significant proportion of the pantetheine unit of CoA (Chapter 2). This could be a route to improve on Keefe's prebiotic synthesis of CoA (Section 2.2), by finding efficient prebiotic or biomimetic routes to add the pantoic acid, diphosphate and 3'-phosphoadenosine units of CoA. This could pave the way to other CoA analogues that may have been present in rudimentary protometabolism and thus participate in biological processes such as fatty acid synthesis or thiol-catalysed peptide bond formation. The project will investigate the role of UV light in prebiotic synthesis of thiols such as CoA. Then assess their catalytic properties in peptide formation (Figure 41).
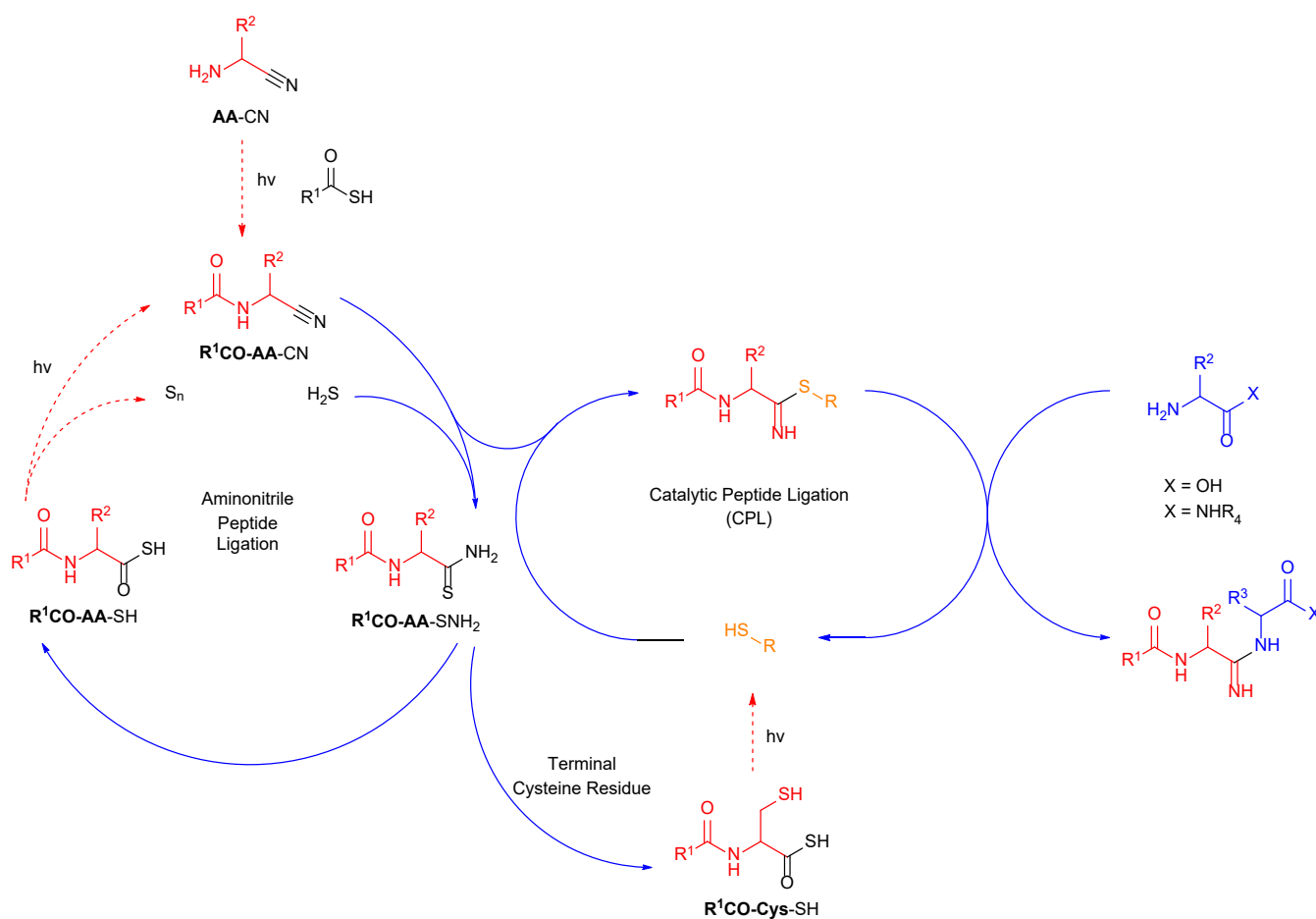
Figure 41. The proposed peptide ligation cycle using UV light to replicate the Earth's day/night cycle (red = light phase, blue = dark phase) linked to the alternate peptide ligation pathway by thiol catalysts possibly synthetized by de-thiocarboxylation.

3.2. UV/Vis Spectra of Thioacids

The UV/Vis spectra of thioacids were studied to investigate the photochemical properties of thioacids. UV/Vis spectra was acquired for the simplest thioacid potassium thioacetate, which exhibited strong absorbance at 245 nm with a extinction co-efficient of 7308 $M^{-1}cm^{-1}$. This compared favourably with Hagan's experimental values of a λ_{max} of 245 nm and a ε of 7308 $M^{-1}cm^{-1}$ for thioacetate at pH 7.[270] This would indicate that irradiation of this compound at 254 nm would be enough to excite the thioacetate to a higher energy level and to undergo photochemical reactions.

The UV/Vis spectra for α-amino thioacids such as alanine thioacid **Ala**-SH (λ_{max} =

168

230-240 nm) had been measured by Wieland in 1954. However, the physiochemical properties of *N*-acetylated amino thioacids have been extensively studied. The UV/Vis spectra were measured to find the absorbance required to maximise photochemical transformations. The λ_{max} and the associated extinction coefficients ε were measured for a representative set of synthesised *N*-acetylated proteinogenic α-amino thioacids. These were: of glycine Ac-**Gly**-SH, valine Ac-**Val**-SH, and phenylalanine Ac-**Phe**-SH. The results showed that the α-amino thioacids exhibited a strong absorbance in a range of 246-251 nm and an extinction coefficient of ε between 8036-7823 $M^{-1}cm^{-1}$ (Figure 43)
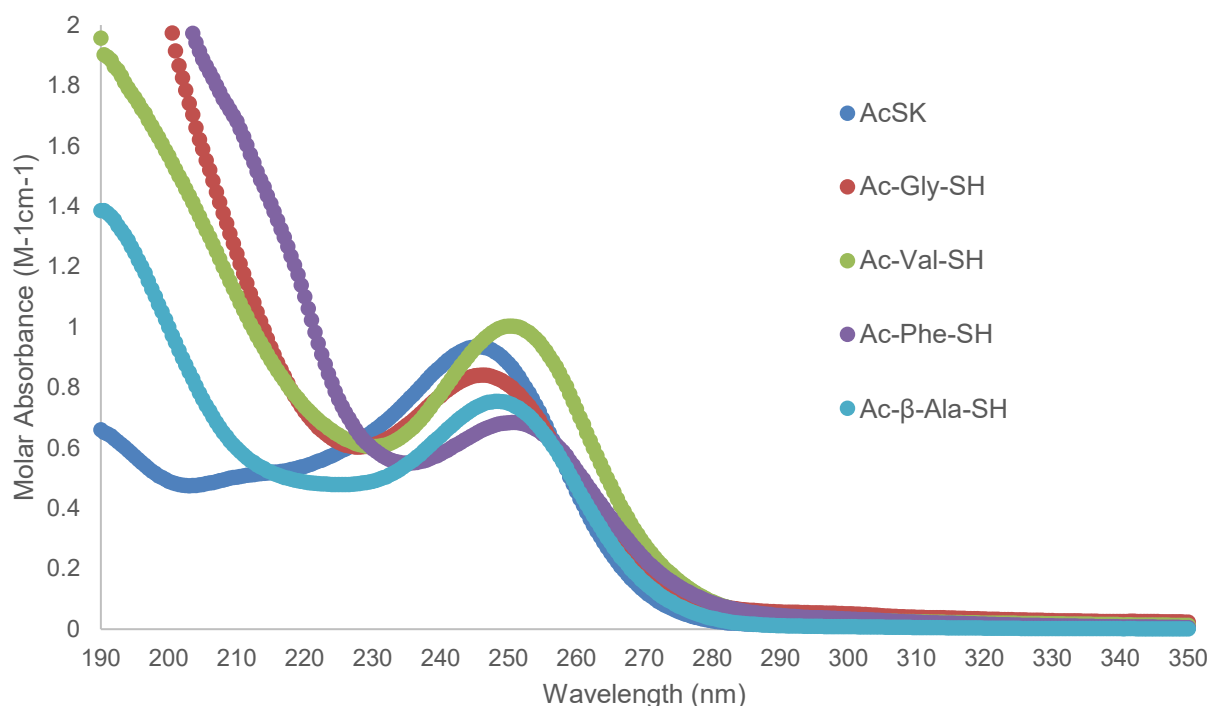
Figure 42. UV/Vis spectra (190-350nm, H_2O) to show the λ_{max} of the thioacids AcSK, Ac-**Gly**-SH, Ac-**Val**-SH, Ac-**Phe**-SH and Ac-β-**Ala**-SH.

3.3. Acetylation of Glycine Nitrile with UV Irradiation

Acetylation of α-amino nitriles **AA**-CN with thioacetate has previously been achieved by oxidation (Section 1.8.5) or with UV light with a photosensitiser (Section 3.1). Initial experiments examined the acetylation of **Gly**-CN with excess thioacetate (3 eq.) and 254 nm UV light at pH 7. Analysis confirmed *N*-acetyl glycine nitrile Ac-

Gly-CN (85%) and *N*-acetyl glycine-thioamide Ac-**Gly**-SNH₂ (15%) were formed, and all **Gly**-CN was consumed after 6 h (Figure 44).
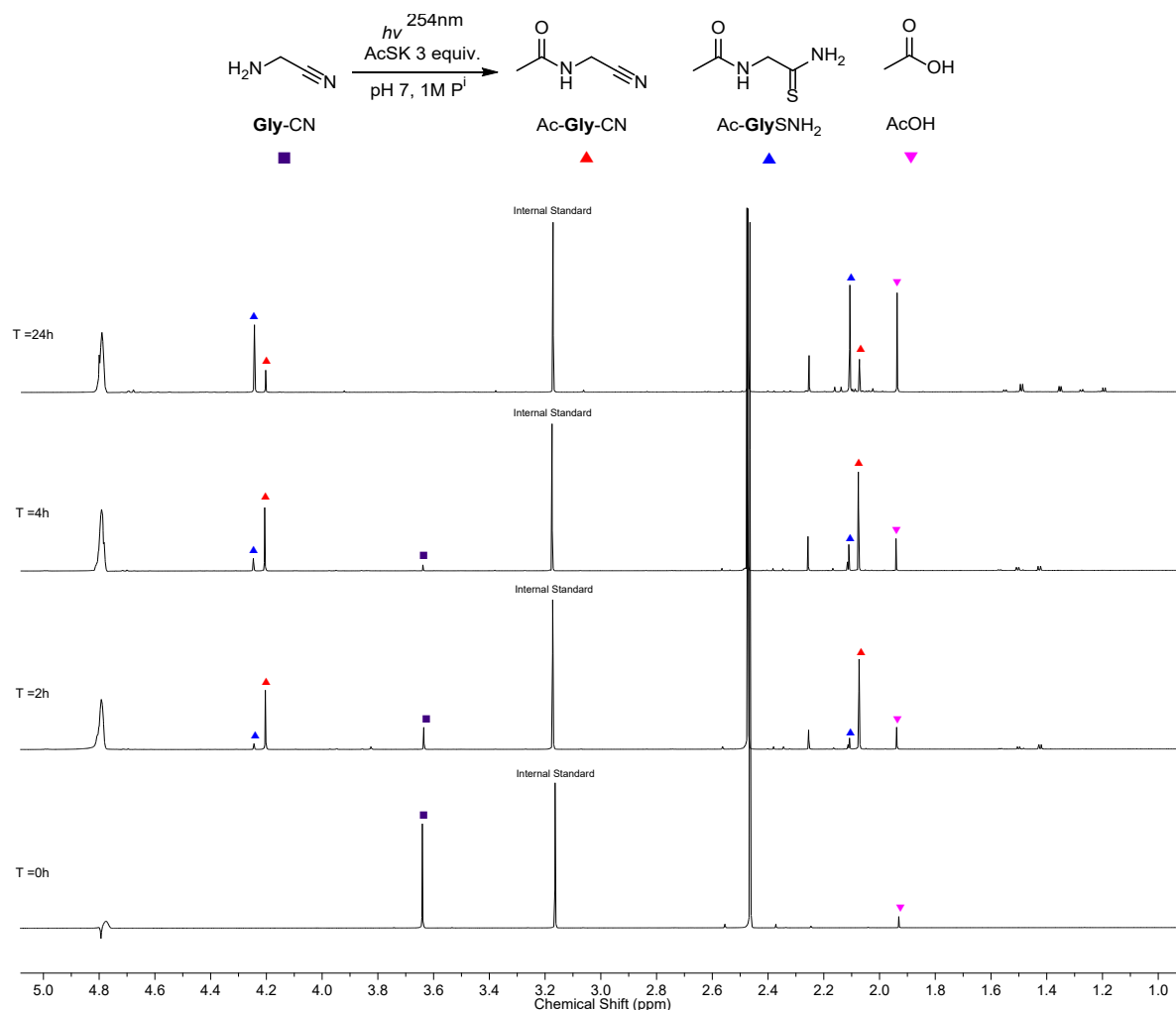
Figure 43. Spectrum to show the acetylation of glycine nitrile **Gly**-CN with thioacetate AcSK at pH 7, with 254 nm UV light over 24 h.

Acetylation of α-amino nitriles **AA**-CN with thioacetate was next repeated in triplicate in buffered solutions at pH 5, 7 and 9 (acetate, phosphate, and borate, respectively) to quantify the yield of the products and investigate the pH dependence of this coupling. Figure 45 illustrates the total acetylated products of the photochemical reaction of **Gly**-CN and thioacetate at pH 5, 7 and 9.
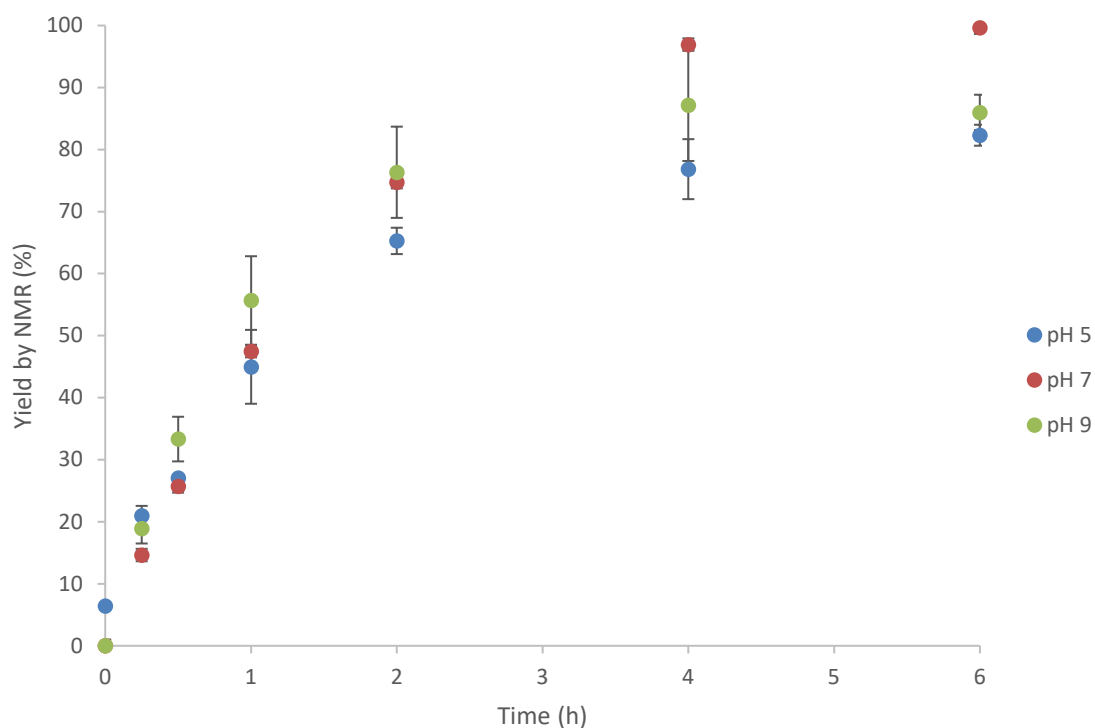
Figure 44. Graph to show the total yields of Ac-**Gly**-CN and Ac-**Gly**-SNH$_2$ by the acetylation of **Gly**-CN (50 mM) with thioacetate (3 eq.) at 5, 7 and 9 in 100 mM buffer, with 254 nm UV irradiation over 6 h.

Marginal differences in the acetylation yield were observed at different pH values. The starting material was consumed 6 h into irradiation and, at pH 7 quantitative acetylation was observed. It was also found that with prolonged irradiation Ac-**Gly**-SNH$_2$ became the major product after 24 h, which was likely caused by the release of sulfide species (such as ⁻SH or ⁻SSH) during the UV light mediated acetylation. The identity of the reaction products was confirmed by a spiking the 6 h timepoint with a pure synthetic sample of Ac-**Gly**-CN and Ac-**Gly**-SNH$_2$.

Synthetic Ac-**Gly**-CN was prepared from the reaction of acetic anhydride with the hydrochloride salt of glycine nitrile (**Gly**-CN·HCl) at pH 9. Extracting the product with EtOAc from the residue resulted in pure Ac-**Gly**-CN in excellent yield (Scheme 62). Ac-**Gly**-CN was thiolysed with NaSH (8.8 equiv.) in pH 9 1M borate buffer over one day to yield pure Ac-**Gly**-SNH$_2$ after purification by column chromatography in a moderate yield of 54% (Scheme 64).
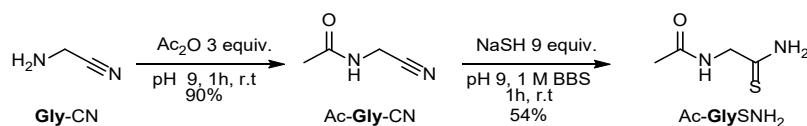
Scheme 62. The acetylation of **Gly**-CN at pH 9 with acetic anhydride to give Ac-**Gly**-CN.

To determine whether UV light caused acetylation, control reactions of **Gly**-CN with AcSK were prepared at pH 3, 5, 7, 9 and 11 and left in the 'dark' (without UV light) and monitored by ^{1}H NMR (Table 2). Gratifyingly, between pH 7 and 11 only trace quantities of Ac-**Gly**-CN were detected by NMR after 24 h. This trace acetylation is likely due to oxidation by oxygen, despite thoroughly degassing the water by freeze pump thaw method.
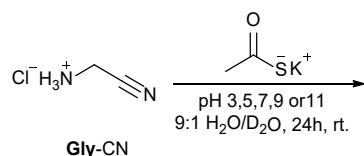
Table 2. Table to show the total yields of Ac-**Gly**-CN and Ac-**Gly**-SNH$_2$ by the acetylation of **Gly**-CN with thioacetate (3 eq.) at pH 3, 5, 7, 9 and 11 over 24 h.

Entry	pH	Ac-**Gly**-CN (%)	Ac-**Gly**-SNH$_2$ (%)	Ac-**Gly**-NH$_2$ (%)
1	3.0	52	n.d	n.d
2	5.0	47	3	n.d
3	7.0	<5	n.d	n.d
4	9.0	<5	n.d	n.d
5	11.0	n.d	n.d	24

Unsurprisingly at pH 11 Ac-**Gly**-CN began to hydrolyse to the amide Ac-**Gly**-NH$_2$ (24%) over 24 h. Unexpectedly, however, acetylation in significant quantities was observed at room temperature under acidic conditions (Table 2). At pH 3 over 24 h, there was 52% conversion to Ac-**Gly**-CN and at pH 5 over the same period there was conversion of 47% Ac-**Gly**-CN and 3% Ac-**Gly**-SNH$_2$.

Using UV light to acetylate **Gly**-CN with thioacetate has provided a unique opportunity to investigate amide bond formation. With the success of the initial reactions, simple thiolate substrates were synthesised to investigate substrate scope of acetylation. One of the traditional routes to the synthesis of protected α-amino

thioacids is through the activation of the carboxylate functional group, then bubbling H_2S gas through the sample or by the addition of sulfide metal salts. The synthesis of thioacids can present many challenges since they can be sensitive to oxidation and hydrolysis. Consequently, they can be hard to purify or isolate in good yields. An acid work-up and extraction into organic solvents completes the isolation of the desired product without need for further purification.

Synthesis of a selection of alkyl thioacids was undertaken to explore the scope of the UV light-initiated acetylation. Ac-**Gly**-SH was synthesised from Ac-**Gly**-OH by EDC activation and trapping with pentaflourophenol (Pfp). The Pfp esters were selected as an intermediate because they are relatively easy to purify and can be stored for long periods cold. Additionally, upon addition of sodium hydrosulfide (NaSH) the pentafluorophenol by-product is easily washed away by trituration with organic apolar solvents such as Et_2O. Sodium propanethioate **97**-SH, was made in excellent yield (>99%) from propionic acid **97** *via* its Pfp ester **97**-Pfp (Scheme 63).
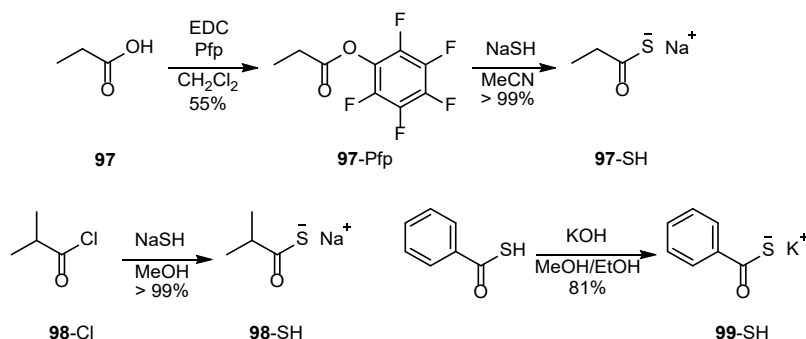
Scheme 63. Synthesis of thioacids for acylation of **Gly**-CN with UV irradiation.

2-Methylpropane thioacid **98**-SH was formed from the reaction of isobutyryl chloride **98**-Cl and sodium hydrosulfide in anhydrous MeOH. Whilst benzoic thioacid is commercially available it is insoluble in water, and so was converted to potassium benzothioate salt **99**-SH with KOH in MeOH/EtOH. With these thioacids in hand, acylation of these substrates and UV light at pH 7 was attempted. Generally, it was found that the alkyl thioacids gave good yields of acylated **Gly**-CN, however, acylation was observed to be slower than photochemical reaction with thioacetate, and **Gly**-CN was not fully consumed over a longer irradiation period with bulkier

thioacids.
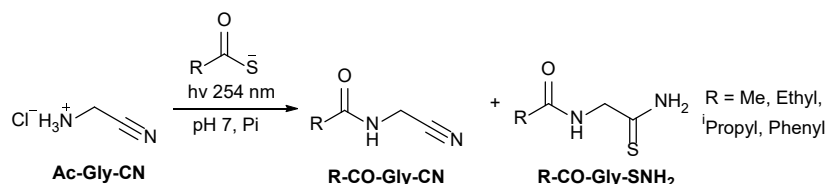
Table 3. Table to show the yields of acylated **Gly**-CN from the photochemical reaction of **Gly**-CN and acylating agent (3 eq.) under 254 nm UV irradiation at pH 7.

Entry	Acylating Agent	R-CO-**Gly**-CN (%)	R-CO-**Gly**-SNH$_3$ (%)	Total Acylation (%)	Time (h)
1	30	65	9	74	8
2	32	66	10	76	20
3	34	32	7	39	20
4	Thioacetamide	n.d	n.d	-	-

Future work would look to investigate the acylation of proteogenic α-aminonitriles which could be formed prebiotically from the Strecker reaction. In addition, it must be shown the UV light acetylation is chemoselective and robust. This could be achieved by stochiometric competition reactions between amines such as ammonia, glycine, β-alanine nitrile with glycine.

3.4. UV Light Mediated Peptide Bond Formation

Encouraged by the successful acetylation of glycine nitrile with UV irradiation, the investigation into the photochemistry of thioamides was essential to establishing whether a photoswitchable reaction could be possible. From previous work in the group, the thiolysis of nitrile terminated peptides with H$_2$S readily occurs at pH 9, however the photochemistry of amino thioamides has not been extensively studied. The synthesis of Ac-**Gly**-SNH$_2$, from thiolysis of Ac-**Gly**-CN in 1 M BBS at pH 9 over 24 h proved successful with > 96% conversion by NMR and isolation of 54% of Ac-**Gly**-SNH$_2$ after column chromatography (Scheme 64).

NaSH•xH$_2$O 10 eq.

pH 9, 1 M BBS, r.t, 24h
54%
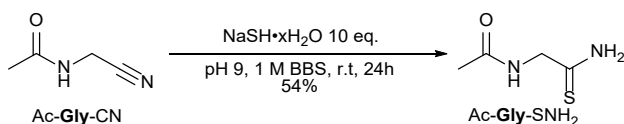
Ac-**Gly**-CN → Ac-**Gly**-SNH$_2$

Scheme 64. Thiolysis of Ac-**Gly**-CN with NaSH at pH 9 to give pure thioamide Ac-**Gly**-SNH$_2$

With Ac-**Gly**-SNH$_2$ in hand, photochemistry of the thioamide with 254 nm UV light was investigated. The irradiation of 50 mM Ac-**Gly**-SNH$_2$ at pH 5, 7 and 9 displayed little reactivity after 8 h (Table 4).

hv, 254 nm

pH 5, 7 or 9
8h

Ac-**Gly**-SH → Ac-**Gly**-**Gly**-OH + Ac-**Gly**-OH + Ac-**Gly**-SH
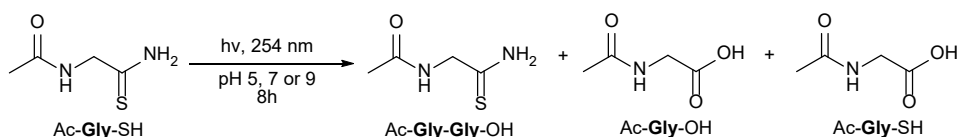
Table 4. Table of the results of the irradiation at 254 nm of Ac-**Gly**-SNH$_2$ at pH 5, 7 and 9 over 8h.

Entry	pH	Starting Material	Products	
		Ac-**Gly**-SNH$_2$ (%)	Ac-**Gly**-OH (%)	Ac-**Gly**-SH (%)
1	5	>99	<1	n.d
2	7	94	6	n.d
3	9	73	4	6

The only notable reactivity observed in 1D and 2D NMR was a trace (< 6%) hydrolysis of Ac-**Gly**-SNH$_2$ to Ac-**Gly**-OH at all pHs. At pH 9 hydrolysis of the thioamide to Ac-**Gly**-SH was also observed in a low quantity 6%, this is unsurprising since previous work found Ac-**Gly**-SH was formed by heating Ac-**Gly**-SNH$_2$ at 60 °C and pH 9. Slightly elevated temperatures (~ 30-35 °C) from the Rayonet reactor was caused by heat from the mercury lamps. Over the course of these studies control reactions were performed at 40 °C in the 'dark' to ensure that heating does not cause the reactivity reported. The work described previously, found that the conversion of glycine nitrile **Gly**-CN to N-acetyl glycine nitrile Ac-**Gly**-CN with thioacetate can be achieved in good yields when irradiated with UV light. Further irradiation or heat enabled Ac-**Gly**-CN to undergo thiolysis to form N-acetyl glycine thioamide Ac-**Gly**-SNH$_2$. Our result indicate that this compound is photostable, however hydrolysis of the thioamide can give N-acetyl glycine thioacid Ac-**Gly**-SH. Ac-**Gly**-SNH$_2$ was found to be stable under UV light in the presence of **Gly**-CN, and therefore not amenable to direct photochemical ligation. Due to the previous UV light acetylation

reactions, we wondered whether the direct ligation of Ac-**Gly**-SH and **Gly**-OH would result in dipeptide formation under UV light.
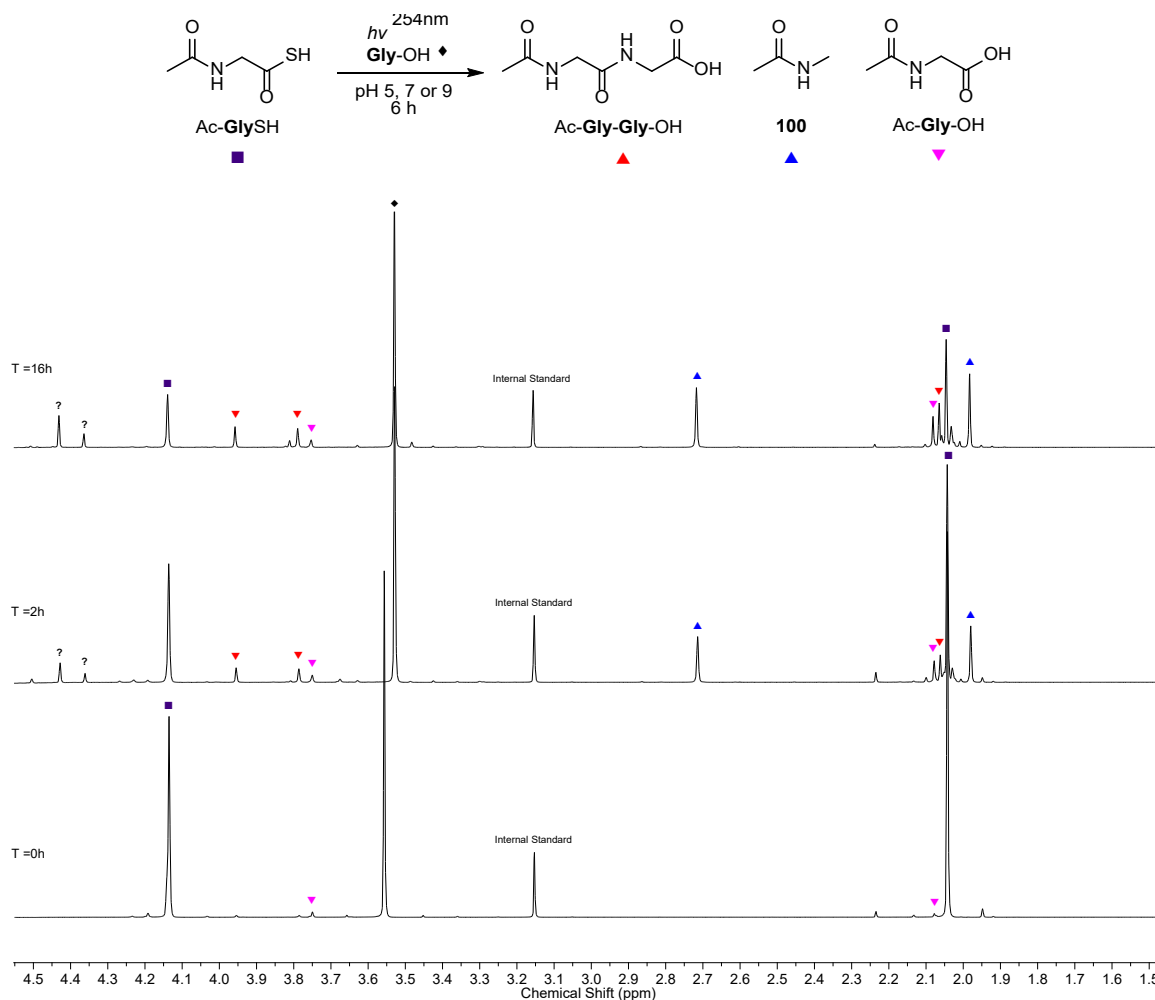
Figure 45. Spectrum to show the photochemical reaction of glycine **Gly**-OH (50 mM) and Ac-**Gly**-SH (50 mM) with 254 nm UV light at pH 7 over 16 h.

The irradiation of Ac-**Gly**-SH (50 mM) and **Gly**-OH (50 mM) in degassed 9:1 H_2O/D_2O with 254 nm light over 16 h formed the dipeptide Ac-**Gly**-**Gly**-OH across pH 5-9 (Figure 46 & Table 5). The acquisition of 2D NMR data added further evidence of dipeptide formation with the observation of the α-protons coupling to the carbon signals of both the adjacent amide and carboxylate groups. No dipeptide was formed in the absence of UV light. Spiking with synthetically prepared Ac-**Gly**-**Gly**-OH prepared from **Gly**-**Gly**-OH and Ac_2O, confirmed the structure of the product (Scheme 65).

176
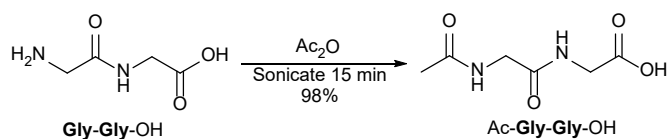
Scheme 65. Synthesis of pure Ac-**Gly-Gly**-OH, from acetylation of **Gly-Gly**-OH and Ac₂O.

Unfortunately, the ligation of Ac-**Gly**-SH with **Gly**-OH to form the Ac-**Gly-Gly**-OH with 254 or 300 nm UV light resulted in low yields (5-13%). Irradiation unexpectedly resulted in the formation of methyl acetamide **100** as a by-product, which accounted for the yield lost.
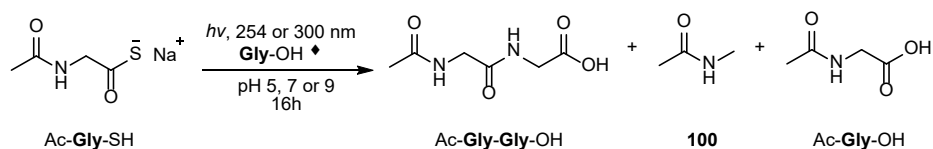
Table 5. The dipeptide yields of the UV irradiation (254 nm or 300 nm) of Gly-CN with Ac-Gly-SH at pH 5, 7 or 9.

Entry	Wavelength (nm)	pH	Yield (%)		
			Ac-**Gly-Gly**-OH	**100**	Ac-**Gly**-OH
1		5	6	60	4
2	254	7	9	45	4
3		9	13	56	6
4		5	6	11	10
5	300	7	5	11	9
6		9	10	9	6

With the discovery of this surprising chemical transformation, investigations to ascertain whether these dethiocarboxylation reactions were general for Ac-**AA**-SH were started. A range of *N*-acetylated thioacids were prepared. The synthesis of Ac-**Gly**-SH had already been accomplished through activation of Ac-**Gly**-OH and thiolysis with H₂S. The sodium salts of aminothioacids Ac-**AA**-SH were prepared from oxazolone **101**, which were then thiolysed to Ac-**AA**-SH with H₂S (Scheme 66). The photochemistry of these Ac-AA-SH was investigated.
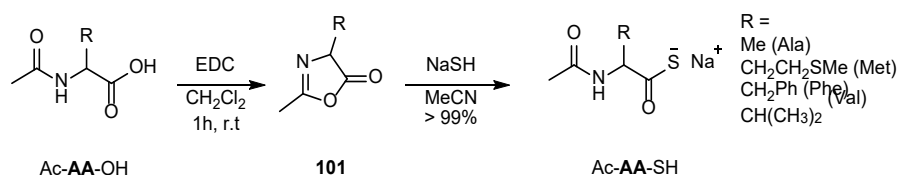
Scheme 66. Yields of the preparative synthesis of Ac-**AA**-SH from Ac-**AA**-OH via oxazolone intermediate **101**, Ac-**Ala**-SH, Ac-**Met**-SH, Ac-**Phe**-SH and Ac-**Val**-SH were prepared.

Aminothioacids Ac-**AA**-SH (50 mM) were then irradiated at 254 nm and the progress of the reaction was tracked by NMR (Figure 47). Each of the aminothioacids Ac-**AA**-SH underwent a reductive dethiodecarboxylation to their respective alkyl acetamides under UV irradiation (Table 6).
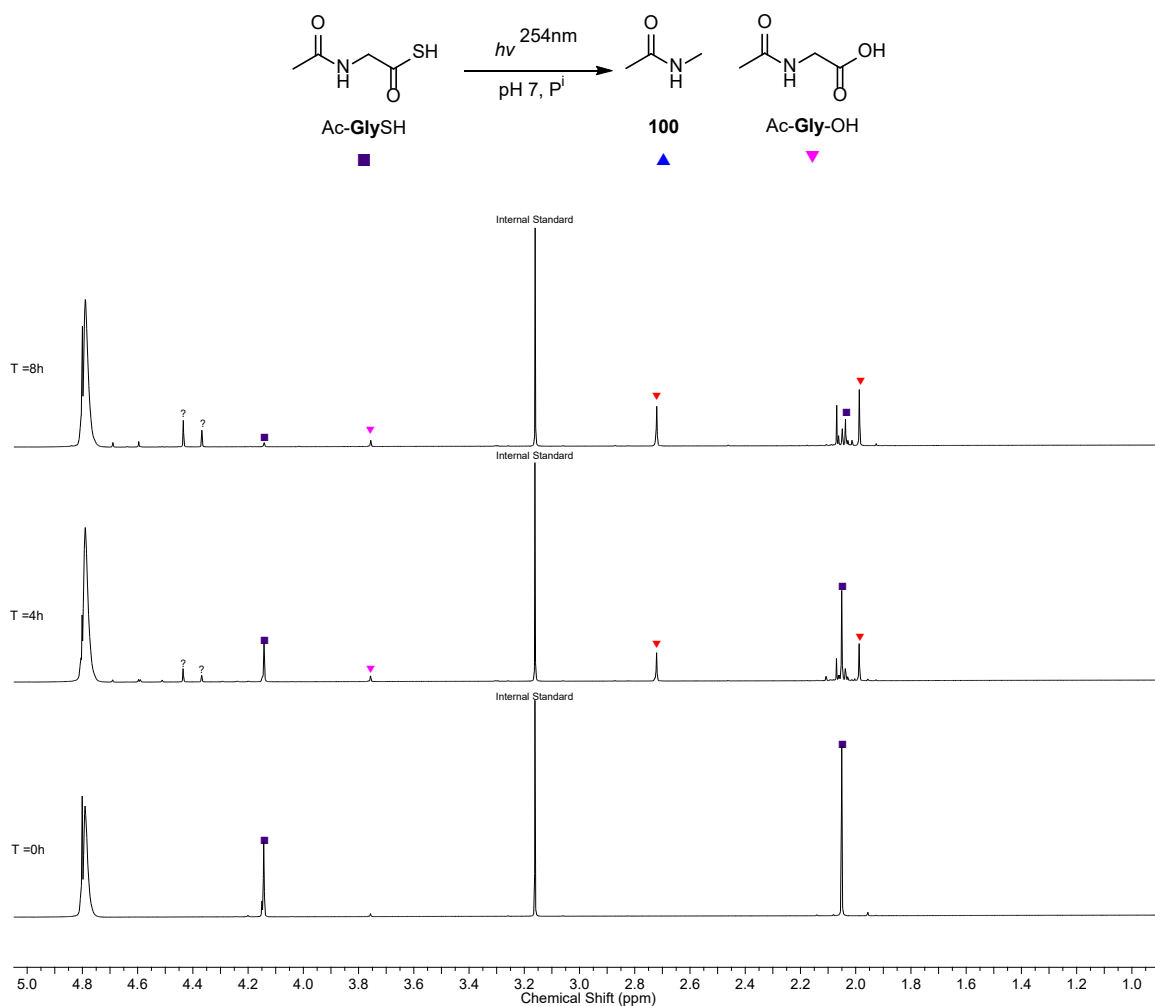
Figure 46. Spectrum to show the dethiocarboxylation of Ac-**Gly**-SH with 254 nm UV irradiation at pH 7 over 8 h.

Confirmation of the products were achieved by spiking with analytically pure methylacetamide **100** or phenyethylacetamide **102** synthesised by acetylation of the alkyl amines with acetic anhydride in excellent yields. In some reactions, bubbles of gas were observed evolving from the solution. This could be COS gas being released during irradiation. As one of the many gases detected from volcanic explosions, COS is an interesting compound since Ghadri implicated its role in formation of NCA's from amino acids.[167] NCA's are thought to be a valid prebiotic route to peptides since they are activated toward nucleophilic attack to give dipeptides in the presence of amines.
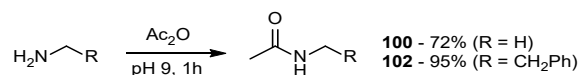
Scheme 67. Acetylation of alkyl amines to spike the photochemical dethiocarboxylation of Ac-**AA**-SH-
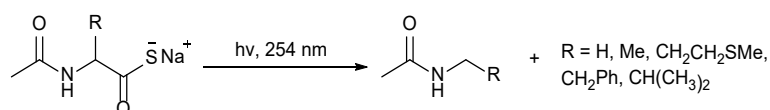
Table 6. Yields of the dethiocarboxylation of Ac-**AA**-SH with UV irradiation at 254 nm. Time (h) is the time it takes for all the starting material to be consumed.

Entry	Ac-**AA**-SH	pH	R	Yield (%)	Time (h)
1	Ac-**Gly**-SH	5	H	39	8
2	Ac-**Gly**-SH	7	H	42	8
3	Ac-**Gly**-SH	9	H	33	8
2	Ac-**Ala**-SH	7	Me	64	4
3	Ac-**Met**-SH	7	CH_2CH_2SMe	73	6
4	Ac-**Phe**-SH	7	CH_2Ph	89	2
5	Ac-**Val**-SH	7	$CH(CH_3)_2$	95	6

The mechanism for the dethiocarboxylation of Ac-**AA**-SH was proposed to proceed by the homolytic cleavage of the thiol with UV light to radical **38** (Scheme 68). This intermediate could degrade into alkyl acetamide radical **106** and release COS gas. The thiyl radical **103** can form diacyl intermediate **104** and could be the reason why a small amount of dipeptide formation is observed when these irradiations are conducted in the presence of **Gly**-OH. Intermediate **104** could also undergo

photodegradation to give COS and radical **103**which can photodegrade to **100**. Finally, alkyl acetamide radical **105** can abstract a hydrogen from the thiol of Ac-**AA**-SH, propagating a chain reaction. At higher pH the thioacid would be deprotonated and so the mechanism may proceed *via* a radical anion in a similar manner in scheme 28, but with an e⁻ released into solution to propagate the reaction.
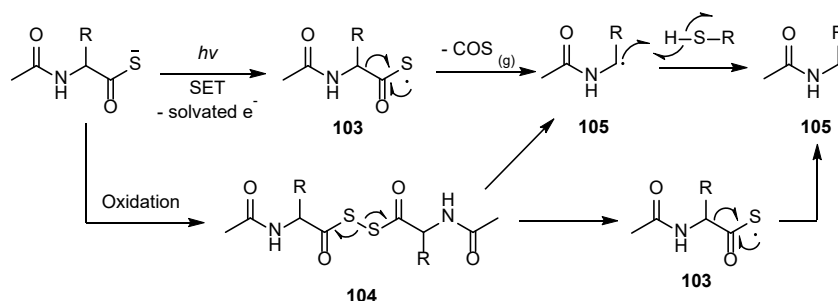
Scheme 68. Proposed mechanism for the photochemical dethiocarboxylation of Ac-**AA**-SH.

Future work in collaboration with the Sasselov group will hopefully seek to explore further the nature of this reaction. During this study it was found that alkyl thioacids such as thioacetate, propanethioate and 2-cyclopropylethanethioate did not dethiocarboxylate when irradiated at 254 nm at pH 7. The synthesis of zwitterionic **Gly**-SH was successfully achieved from Boc-**Gly**-OH in good yields (Scheme 69). In future work, irradiation of **Gly**-SH will determine whether *N*-acetylation predisposes Ac-**AA**-SH to thiodecarboxylation.
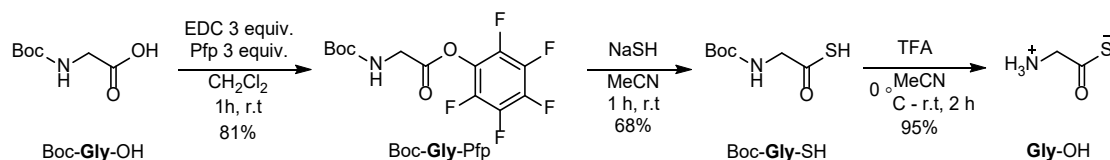
Scheme 69. The synthesis of zwitterionic **Gly**-SH from Boc-**Gly**-OH.

Acetone has been used in many photochemical reactions as a photosensitiser. It was thought that acetone could enable a different route to the excitation of Ac-**Gly**-SH and enable access to an excited state that is inaccessible by direct absorption. Alternative excited states may enable ligation by UV light to form the desired peptides.

180
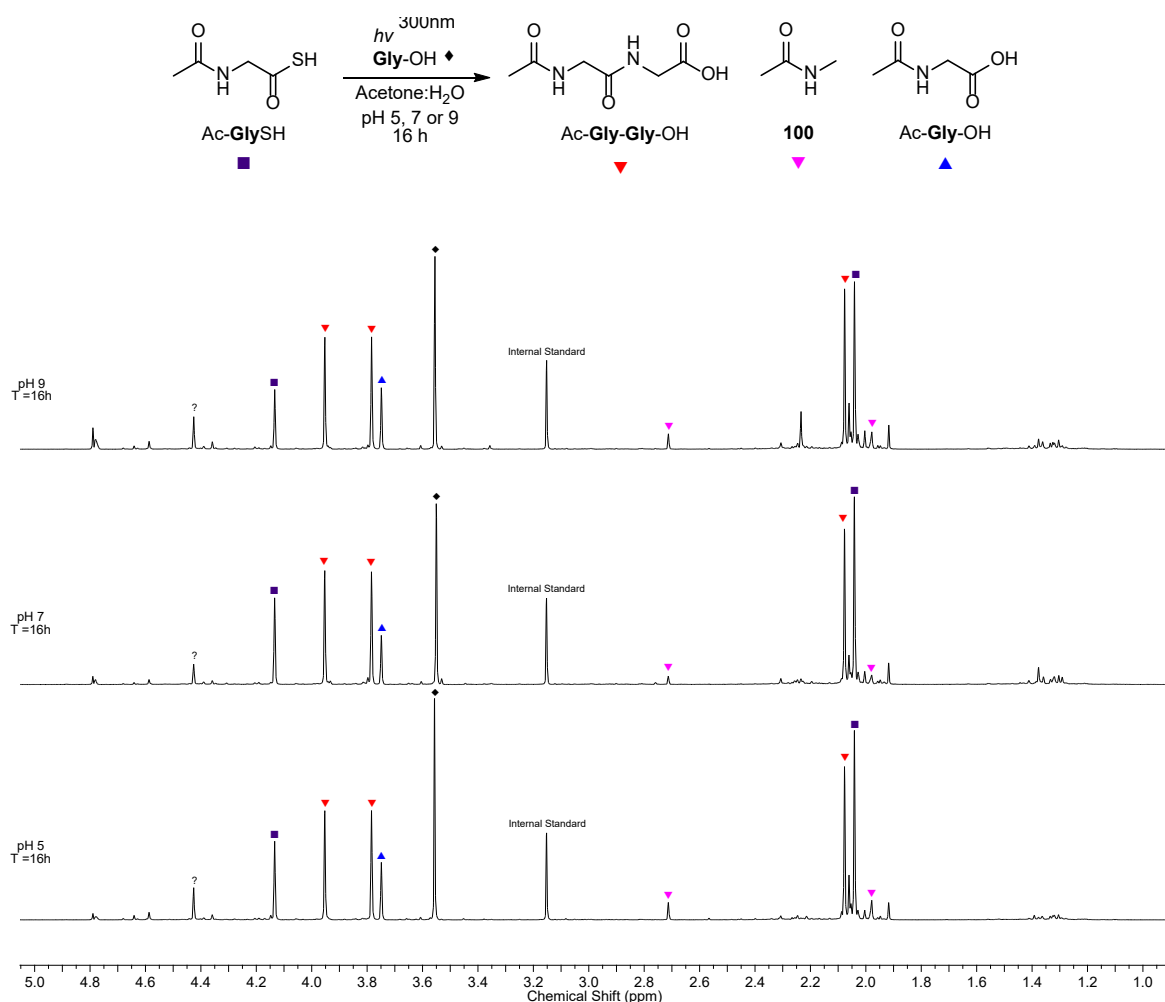
Figure 47. Spectrum to show the photochemical ligation of **Gly**-OH (50 mM) and Ac-**Gly**-SH (50 mM) under UV light 300 nm and acetone as a triplet sensitiser at pH 7.

A set of irradiations at pH 5, 7 and 9 in either 254 nm or 300 nm UV light was carried out on solutions of Ac-**Gly**-SH (50 mM) and **Gly**-OH (50 mM) in thoroughly degassed acetone/H₂O 1:1. Encouragingly, the ligation of Ac-**Gly**-SH with **Gly**-OH under 254 or 300 nm UV light to Ac-**Gly**-**Gly**-OH was observed in significant quantities (Figure 47 & Table 7). Irradiation at 254 nm resulted in the formation methyl acetamide but dethiocarboxylation was supressed at 300 nm, indicating that the energy of the light at this wavelength is not enough to excite the thioacid Ac-**Gly**-SH. With the success of these preliminary results, future work will look to optimise the reaction, decrease the acetone loading and investigate the scope of the reaction. Investigations into using **Gly**-CN as a nucleophile will seek to find a pH window where Commeyras'

ketone/aldehyde assisted hydrolysis of aminonitriles can be avoided and ligation can be maximised.[93] The UV photosensitised reactions with other aldehydes, ketones and hydroxy keto acids would enable prebiotic peptide synthesis with UV light.
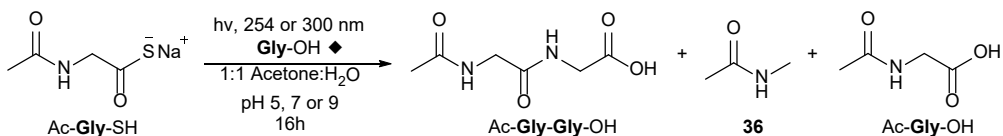
Table 7. Table of the yields of Ac-**Gly-Gl**y-OH formed from **Gly**-OH (50 mM) and Ac-**Gly**-SH (50 mM) at either 254 nm or 300 nm at pH 5, 7 or 9 over 16 h with acetone acting as a triplet sensitizer.

Entry	Wavelength (nm)	pH	Yield (%)		
			Ac-**Gly-G**ly-OH	**100**	Ac-**AA**-OH
1	254	5	26	13	23
2	254	7	28	12	21
3	254	9	37	13	33
4	300	5	44	7	5
5	300	7	44	4	19
6	300	9	43	7	24

3.5. The Prebiotic Synthesis of CoA/Pantetheine Analogues with UV Irradiation

Although β-alanine (β-**Ala**-OH) is not a proteinogenic amino acid, it is found in many biologically relevant molecules such as carnosine or in vitamin B_5. Vitamin B_5 is also incorporated into the pantetheine motif in CoA. As discussed in this thesis, CoA has many roles in metabolic processes, such as fatty acid synthesis or non-ribosomal peptide synthesis. The prebiotic syntheses of CoA are sparse, and experimental conditions are not mild and reported yields are low.[207] One of the aims of the project is to prebiotically make precursors to CoA that could catalyse peptide bond formation with its thiol group. Many routes to the prebiotic synthesis of biological compounds exhibit little selectivity and often have a plethora of similar, but non-biological molecules. UV light has often been used to explain why extant biology is as it is today. UV light can impart a selection pressure on molecules to photochemically destroy and 'filter' molecules. So far, this project has found that the UV light

instigates the acetylation of **Gly**-CN with alkyl thioacids, the dethiocarboxylation of *N*-acetyl-α-amino thioacids Ac-**AA**-SH over alkyl thioacids and the acetone triplet sensitised ligation between α-Ac-**AA**-SH and **Gly**-OH. This reactivity may indicate that a *N*-acetyl-β-amino thioacids may have desired reactivity predisposed to CoA synthesis.
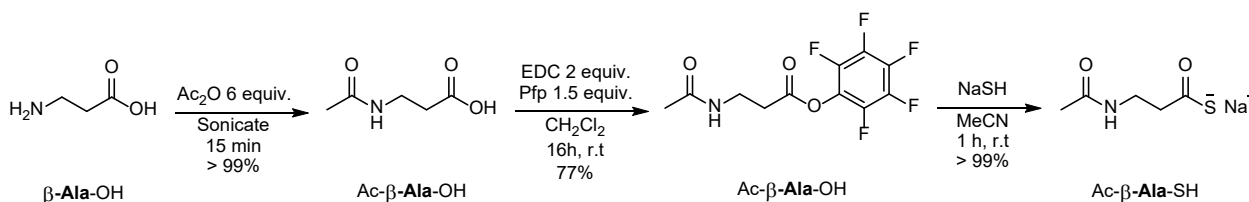
Scheme 70. Synthesis of Ac-β-**Ala**-SH from β-**Ala**-OH.

To begin this task the syntheisis of Ac-β-**Ala**-SH was carried out. β-Alanine was acetylated with Ac$_2$O to give Ac-β-**Ala**-OH in quantitative yield. Pfp activation of Ac-β-**Ala**-OH and thiolysis resulted in Ac-β-**Ala**-SH (> 99%) (Scheme 70). The irradiation (254 nm) of 50 mM Ac-β-**Ala**-SH in buffered solutions at pH 7 and 9 only resulted in the slow hydrolysis of the thioacid to its equivalent carboxylate in 6 h (Table 8). This displays similar reactivity to the alkyl thioacids, such as thioacetate AcSK, which hydrolysed to their carboxylate in the absence of an amine under UV irradiation. This might indicate that Ac-β-**Ala**-SH is a good acylating agent.
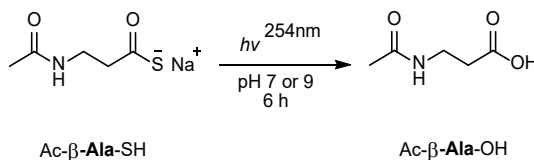
Table 8. Table of the irradiation with 254 nm UV light of Ac-β-**Ala**-SH.

Entry	pH	Starting Material	Products	
		Ac-β-**Ala**-SH (%)	Ac-β-**Ala**-OH (%)	AcNHEt
1	7	62	29	n.d
2	9	80	22	n.d

Irradiation of Ac-β-**Ala**-SH (50 mM) with either **Gly**-CN or cysteamine at pH 7 over 8 h resulted in significant ligation (Table 9). The yields were good but, the synthesis of a standard Ac-β-**Ala**-cysteamine is required to confirm the identity of the product.

Ac-β-**Ala**-cysteamine is a major component of the pantetheine arm of CoA and this reaction may allow the selective synthesis of CoA analogues, providing a (photo)chemical rationale for the incorporation of the β-**Ala** moiety. However more experiments to determine selectivity and the photostability of the thiols will be performed to determine the robustness of this reaction.
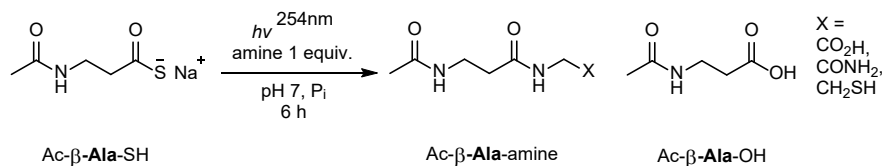
Table 9. Table of the yields of dipeptide formed from Ac-**β-Ala**-SH and an amine at pH 7 under 254 nm UV irradiation for 6h.

Entry	Amine	Dipeptide Product	Yield (%)		
			Dipeptide	Thioamide	Ac-β-**Ala**-OH
1	Gly-OH	Ac-**β-A**la-**Gly**-OH	9	-	37
2	Gly-CN	Ac-**β-Ala-Gly**-CN	56	9	31
3	Cysteamine	Ac-**β-Ala**-Cysteamine	48	-	38

Dethiodecarboxylation with UV irradiation could also play an important role in the prebiotic synthesis of the cysteamine residue in CoA. During the prebiotic synthesis of Ac-**Cys**-CN, the Powner group found that the thiolysis of Ac-**Cys**-CN to the thioamide Ac-**Cys**-SNH$_2$ could be achieved with NaSH. Hydrolysis subsequently formed the cysteine thioacid in 25% yield. Dethiocarboxylation of Ac-**Cys**-SH with UV irradiation could result in the formation of Ac-cysteamine motif in the pantetheine structure in CoA (Scheme 71).
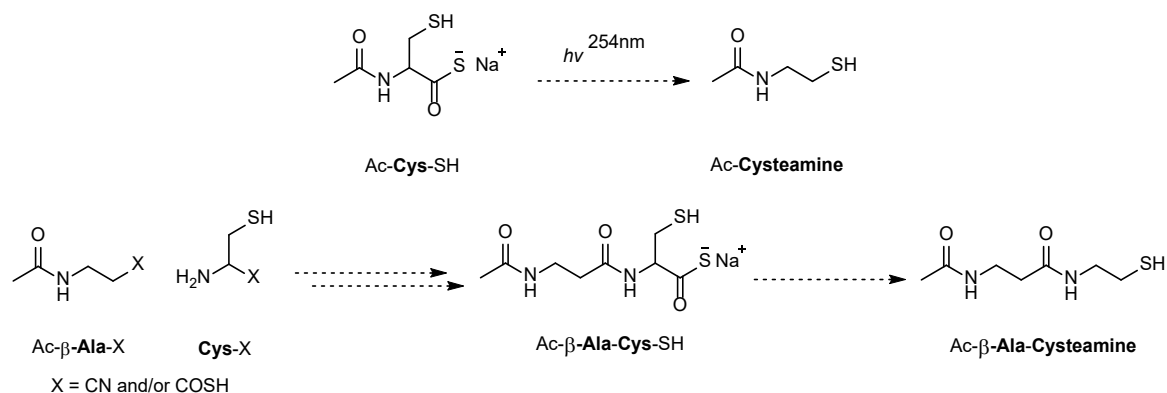
Scheme 71. Proposed Ac-**Cys**-SH targets for dethiocarboxylation to Ac-Cysteamine residues as thiol catalysts for peptide ligation.

To test this theory out work began on a synthetic sample of Ac-**Cys**-SH. Initial attempts focused on a protecting group strategy, the thiol of Ac-**Cys**-OH would be protected to Ac-**Cys**^{PG}-OH and then formation to oxazolone **42** would be performed with EDC as an activating agent. Oxazolone **42** would be thiolysed with NaSH to precipitate the sodium salt of the protected cysteine thioacid Ac-**Cys**^{PG}-SH. Global deprotection of Ac-**Cys**^{PG}-SH would yield the free thiol Ac-**Cys**-SH for irradiation.
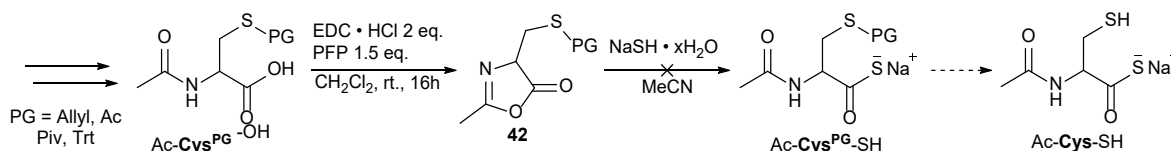
Scheme 72. Attempted synthesis of Ac-**Cys**-SH using thiol protecting groups, oxazolone formation and thiolysis.

Unfortunately, this strategy was unsuccessful due to either elimination to dehydroalanine residues or to the formation of multiple by-products when the thiol is protected with either an acetyl, pivaloyl, allyl or trityl protecting groups. This required a new route to be found to Ac-**Cys**-SH.
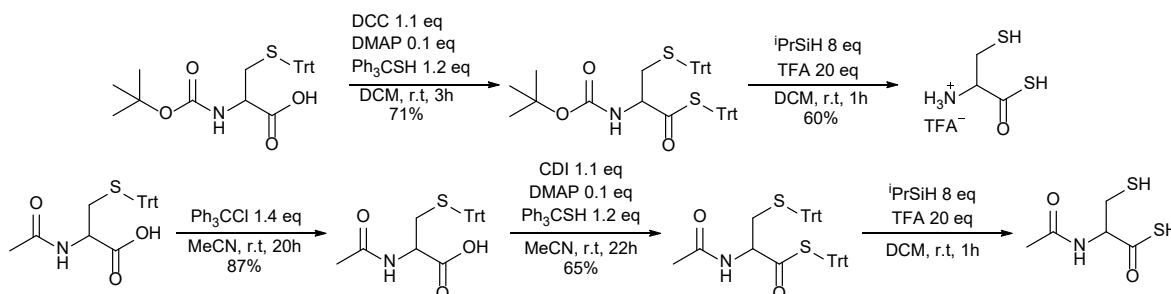
Scheme 73. The successful synthesis of **Cys**-SH and Ac-**Cys**-SH by protecting group and trityl thioester strategy.

Fortunately, upon switching NaSH thiolysis to trityl-SH thiolysis the synthesis of Ac-**Cys**-SH was successful (Scheme 73). The synthesis of Ac-**Cys**-SH was attempted with the same methodology. Ac-**Cys**-SH was successfully observed by NMR; however, isolation of the compound has proven tricky due to its instability in acidic conditions, work continues to complete this target and irradiate under UV light.

4. Acid Catalysed Acetylation and Peptide Ligation with N-Acetyl α-Aminothioacids – Results & Discussion

4.1. Introduction

During the investigation into the UV light mediated acetylation of glycine nitrile, control reactions demonstrated that acetylation of **Gly**-CN could occur at low pH's *via* a specific acid catalysed mechanism (Section 3.3). Acid peptide ligation has previously has previously gained attention in combination with methylisonitrile **35** as a potential prebiotic activating agent for amino acids (methylisonitrile **35** is also reported to activate nucleotides – albeit at high pH).[273,274] Recently, the Sutherland group had shown that amide bonds could be formed at low pH when the carboxylate moiety is activated by methylisonitrile **35**.
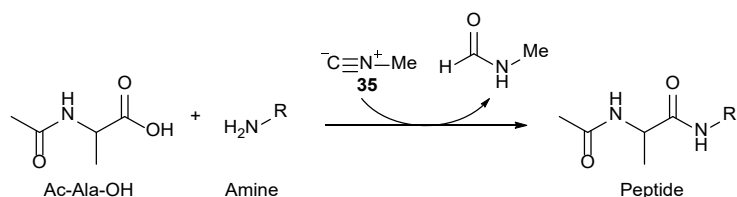
Table 10. Sutherland's yields for the formation of amide bonds from the reaction of different amine nucleophiles and methyl isonitrile **35** at pH 3, 4 and 5.[273]

Entry	Amine	Peptide	Yield of Peptide (%)		
			pH 5	pH 4	pH 3
1	**Gly**-OH	Ac-**Ala-Gly**-OH	62	58	46
2	**β-Ala**-OH	Ac-**Ala-β-Ala**-OH	81	47	5
3	**Gly-Gly**-OH	Ac-**Ala-Gly-Gly**-OH	72	19	n.d
4	**Gly**-NH$_2$	Ac-**Ala-Gly**-NH$_2$	69	20	n.d
5	**Gly**-CN	Ac-**Ala-Gly**-CN	86	>99	30
6	MeNH$_2$	Ac-**Ala-Gly**-NHMe	47	n.d	n.d
7	NH$_3$	Ac-**Ala**-NH$_2$	17	n.d	n.d

Although these low pH conditions could be considered low in the context of prebiotic chemistry, Sutherland rationalises these by suggesting that the geochemical lagoon scenario proposed in the cyanosulfidic protometabolic system (Section 1.7) could have had regional pockets of acidic environs during the early Earth.[28] Although Orgel

had successfully acetylated amino acids with potassium thioacetate in the presence of an oxidant (Section 1.8.5); the reactivity of α-amino thioacids at different pH was not extensively was not explored.[252] Soon after Orgel also showed that the elongation of a decaglutamic acid primer with α-thioglutamic acid could also be also achieved under oxidative conditions.[175] However, Kajihara and co-workers later exploited the oxidation potential of α-amino thioacids to form electrophilic diacylsulfide disulphide moiety to oligomerise α-amino thioacids in acidic pH's (concentrate HCl) by oxidation.[275]
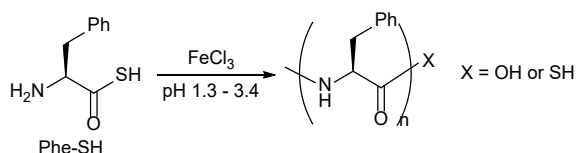
Scheme 74. Kaihara's uncontrolled oxidative polymerization of **Phe**-SH with FeCl$_3$ in acidic conditions (pH 1.3-3.4).[275]

In contrast to the results described here, Kajihara's control reactions of α-amino thioacids at low pH (with no oxidant) led only to trace amounts of peptide formation and this peptide was attributed to residual oxygen in the reaction.[275] Although, α-amino thioacids were effective for the oxidative regioselective α-peptide bond formation, the reaction is uncontrolled and appears to require oxidation. For example, phenylalanine thioacid, **Phe**-SH forms peptides under these conditions after 5 mins and after 2 h an aggregate precipitated out containing 7-12 mer phenylalanine residues. It is likely that the limitation of Kaihara's reaction is due to the protonation state of the amine. At low pKa phenylalanine or phenylaminethioacid will be completely protonated and therefore it's nucleophilicity quenched. It is therefore essential for their low pH coupling to processed via an intramolecular mechanism to overcome this kinetic block. This is likely true for Sutherland low pKa amino acid coupling also; the Sutherland group propose an anhydride intermediate that undergoes intramolecular acylation. However, this problem can be resolved by the low pK_{aH} of aminonitriles (see Table 10, entry 5). Therefore, despite the previously observed limitations of amino thioacid **AA**-SH ligations, preliminary results show that controlled peptide formation can be achieved by controlling the pH and pK_{aH} of the amine substrate. Thus, it is envisioned that under acidic conditions,

glycine nitrile **Gly**-CN (pK_{aH} 5.4)[132] can be acetylated by thioacetate which protects the amine and activates the nitrile towards thiolysis. Then a pH 'swing' from acidic to basic conditions and the addition of NaSH would enable thiolysis to the thioacid, which in the presence of more **Gly**-CN in acidic conditions would ligate to form dipeptide that could go through sequential rounds of elongation (Figure 48).
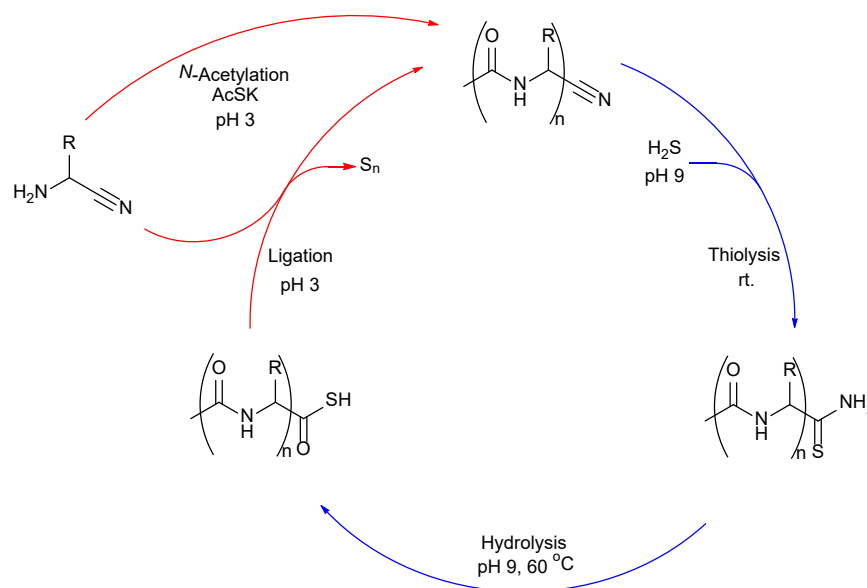
Figure 48. Diagram depicting the pH mediated ligation cycle, where amide bond formation can occur in acidic solutions (pH 3) and the thiolysis of Ac-**AA**-CN to thioacid Ac-**AA**-SH in basic conditions. The cycle can repeat to form polypeptide.

4.2. Acid Catalysed Ligation

This route maybe advantageous since oxidant is not required to form peptides and thus negating the need for mutually incompatible reductive and oxidative conditions. Secondly, the reactivity can be controlled, and amino acid residues added sequentially. From a synthetic point this is very useful to drive each step to completion before the next polymeric unit is added. Initially, the behaviour of **Gly**-CN with thioacetate at different pH was investigated as significant acetylation occurred at pH 3 (in "dark" control reaction for photochemical activation of thioacids). However, the reaction was observed to plateau after 24 h. It was also observed that the pH of the reaction mixture increased and so it was reasoned that this pH shift had terminated the reaction. To maintain stable pH, the reactions were repeated in

buffer solutions (Figure 50).
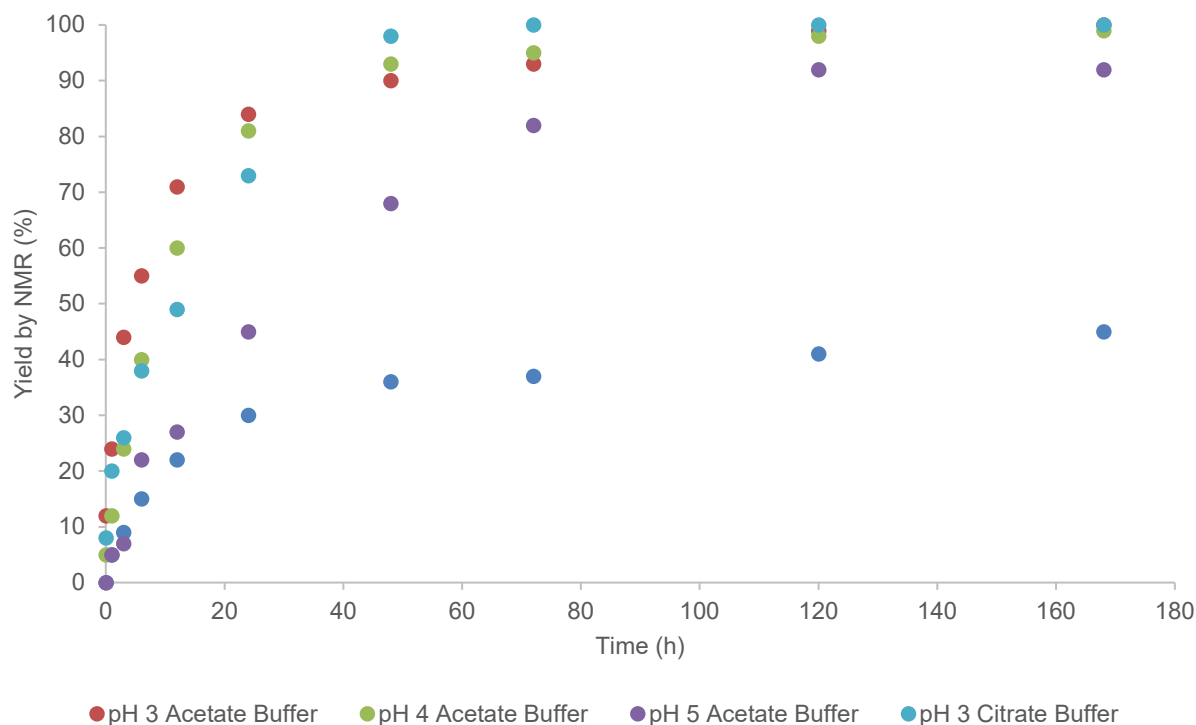
Figure 49. Graph depicting the total acetylation of **Gly**-CN with AcSK at pH 2, 3, 4 & 5, over 7 d.

Excellent reactivity of potassium thioacetate and **Gly**-CN was observed at pH 3-4. This effective coupling may be rationalised by consideration of the pK_{aH} values of aminonitriles and thioacids which presents a pH 'window' where amide bond formation is favoured. **Gly**-CN has a pK_{aH} = 5.4[132] which is significantly lower than glycine pK_{aH} = 9.8.[159,246] At neutral or slightly acidic pH the amine is almost all free-base and therefore nucleophilic. A thioacid such as thioacetate (pK_a ~ 3.4) is partially protonated at pH < 4 and the thiol (SH) moiety is a good leaving group when protonated, thus 'activating' the thioacetic acid to nucleophilic attack from amines. Interestingly, despite the low pH and volatility of H_2S, the released sulfide which was also able to thiolyse some of the nitriles present, which is the next step of the prebiotic proposed pH-controlled ligation cycle (Table 11).

Table 11. The acetylation yields of **Gly**-CN and thioacetate at rt. In buffered solutions pH 2, 3 4 and 5.

Entry	pH	Buffer	Ac-**Gly**-CN (%)	Ac-**Gly**-SNH$_2$ (%)	Total Acetylation (%)	Time (d)
1	2	N/A	45	4	49	7
2	3	Acetate	99	3	> 99	5
3	3	Citrate	94	9	> 99	2
4	4	Acetate	95	3	98	3
5	5	Acetate	83	9	92	5
6	7	Phosphate	< 5	n.d	< 5	7

In buffered solutions **Gly**-CN was completely converted to Ac-**Gly**-CN over the course of a week at room temperature. Upon heating at 60°C, acetylation of **Gly**-CN occurred cleanly in excellent yields over 24 h. For example, at pH 3 or 5 with 3 eq. of thioacetate quantitative acetylation was observed (Table 12).

Table 12. The acetylation yields of **Gly**-CN at pH 3 or 5 with 1.5 or 3 eq. of thioacetate at r.t or at 60 °C over 24 h.

Entry	AcSK (eq.)	pH	Buffer	Ac-**Gly**-CN (%)	Ac-**Gly**-SNH$_2$ (%)	Total Acetylation (%)
1	1.5		Acetate	78	5	82
2	3	3.0	Acetate	93	9	> 99
3	1.5		Citrate	94	7	> 99
4	3		Citrate	91	10	> 99
5	1.5	5.0	Acetate	51	22	73
6	3		Acetate	62	41	> 99
7	1.5	7.0	Phosphate	12	5	17
8	3		Phosphate	14	11	25

The Strecker reaction (Section 1.5.1) has long been recognised as a prebiotic route to amino acids *via* the hydrolysis of α-aminonitriles. The acetylation of different **AA**-CN substrates was explored to find out whether acidic coupling reaction was general for the proteinogenic acids. α-Aminonitriles were synthesised by the reacting the appropriate aldehyde (e.g., acetaldehyde **90** ≡ **Ala**-CN) with NaCN and NH₄Cl and

isolated as their hydrochloric acid salts by precipitation with ethereal HCl. The aminonitriles exhibited no degradation over a long period when stored in the freezer under argon. With aminonitriles in hand, acetylation with AcSK in acetate buffer at pH 3 was carried out (Table 13).
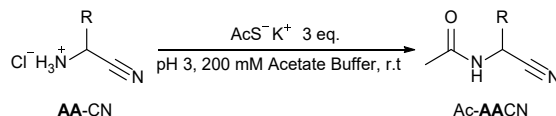
Table 13. The acetylation yields of different **AA**-CN substrates with thioacetate (3 eq.) at r.t and pH 3.

Entry	**AA**-CN	Ac-**AA**-CN (%)	Time (d)
1	**Gly**-CN	> 99	5
2	**Ala**-CN	68	8
3	**Met**-CN	> 99	8
4	**Phe**-CN	84	8
5	**Ser**-CN	91	8
6	**Val**-CN	69	7
7	β-**Ala**-CN	44	7

All aminonitriles acetylated cleanly in good yields. Only alanine-nitrile was sluggish with a 67% yield over 8 d and even β-alanine nitrile acetylated warranting further investigation of selectivity between the aminonitriles. A competition reaction between the α- and β-aminonitriles will be carried out to investigate how the pK_{aH} difference of α-aminonitriles (pK_{aH} ~ 5.4)[132] and β-aminonitriles (pK_{aH} ~ 7.8)[159,246] affects reactivity. With the success of the acetylation of **AA**-CN at pH 3 it was thought that the formation of dipeptides could be achieved with Ac-**Gly**-SH in a similar process. In previously published work the group had already observed the hydrolysis of N-acetyl glycine thioamide Ac-**Gly**-SNH$_2$ to N-acetyl glycine thioacid Ac-**Gly**-SH (81%, pH 9, 60°C, 1 d). Hydrolysis of Ac-**AA**-SNH$_2$ to other N-acetylated proteogenic thioacids were also successful with yields between 51-81%. Therefore, a synthetic sample of Ac-**Gly**-SH was synthetised from Ac-**Gly**-OH in good yield *via* Pfp activation (Scheme 75).
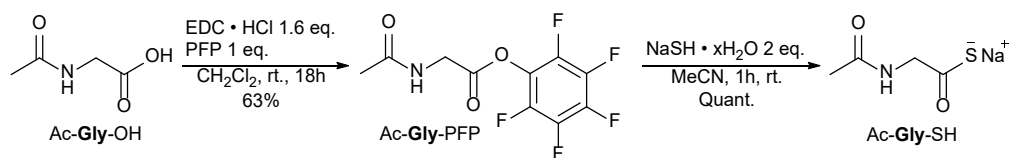
Scheme 75. Synthesis of Ac-**Gly**-SH from the activated ester Ac-**Gly**-Pfp and NaSH.
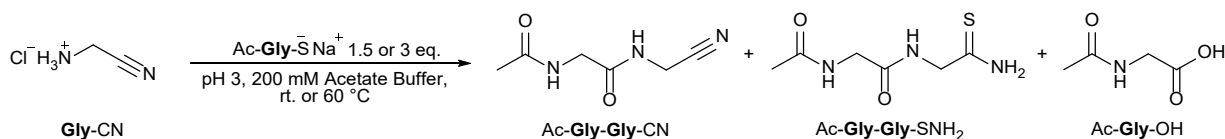
Table 14. The dipeptide yields of the reaction of **Gly**-CN with either 1.5 or 3 eq. Ac-**Gly**-SH at rt. Or heating at 60 °C, in acetate buffer at pH 3.

Entry	Ac-**Gly**-SH (eq.)	Ac-**Gly-Gly**-CN (%)	Ac-**Gly-Gly**-SNH$_2$ (%)	Total Dipeptide (%)	Temperature (°C)	Time
1	1.5	35	2	37	rt.	5 d
2	3	56	3	59	rt.	5 d
3	1.5	62	7	69	60	24 h
4	3	83	8	91	60	24 h

A series of ligation reactions of Ac-**Gly**-SH (75 or 150 mM) and **Gly**-CN (50 mM) was attempted at pH 3 with acetate buffer (Figure 51 & Table 14). Gratifyingly the dipeptide was observed in all cases. Analysis of the HMBC data showed the characteristic signature of amide bond coupling, the α-proton of **Gly**^{2} coupled to both a nitrile and amide carbon signal. A synthetic standard of Ac-**Gly-Gly**-CN was prepared from Ac-**Gly**-Pfp and **Gly**-CN to spike the samples (Scheme 76).
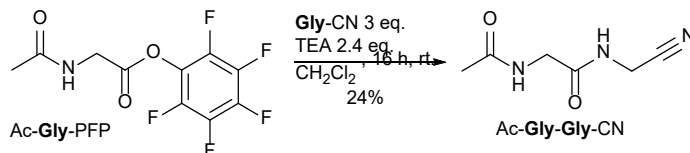
Scheme 76. Synthesis of dipeptide Ac-**Gly-Gly**-CN from the activated ester Ac-**Gly**-Pfp and **Gly**-CN.
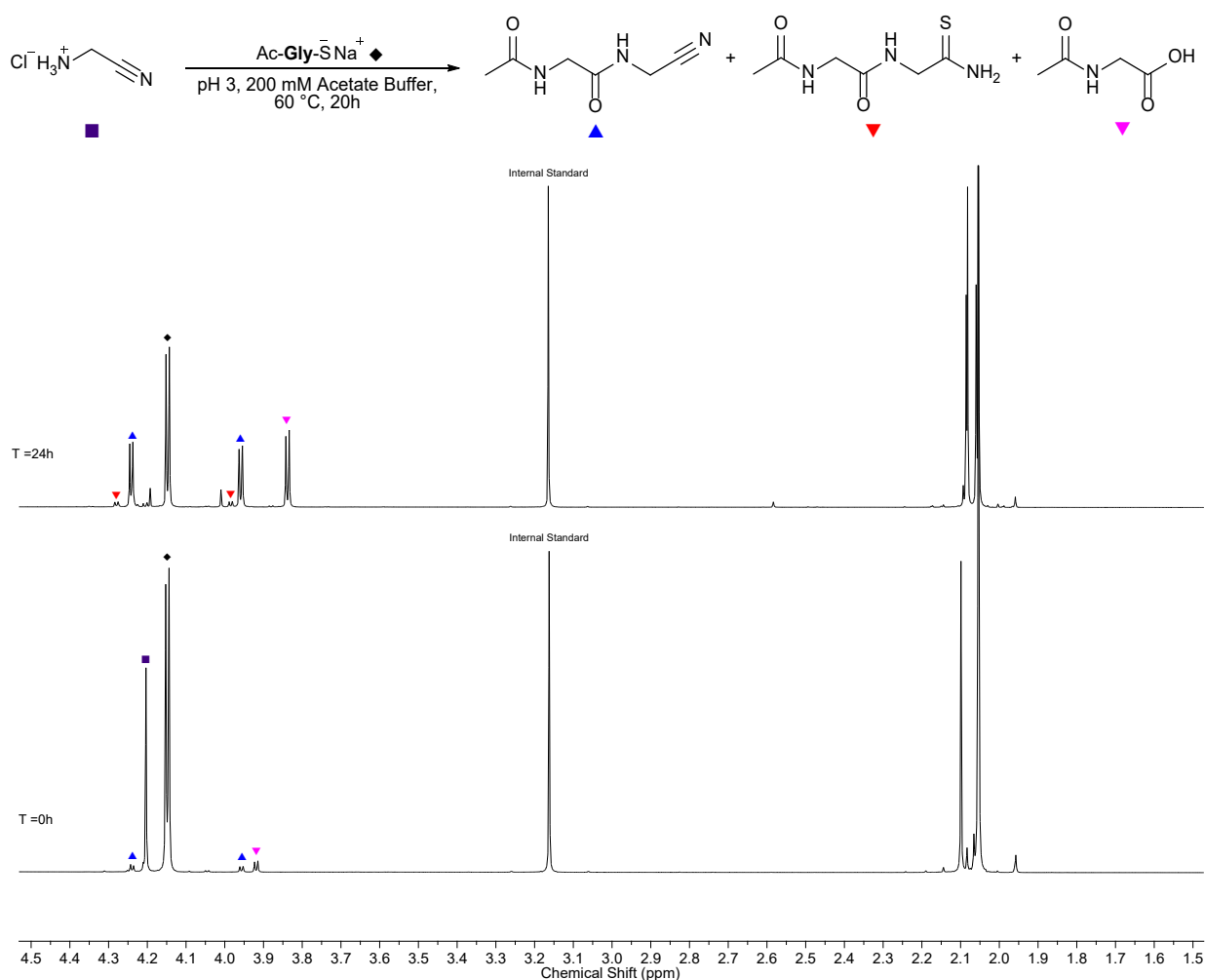
Figure 50. Spectrum to show the ligation of **Gly**-CN (50 mM) with Ac-**Gly**-SH (150 mM) at pH 3 over 24 h.

Spiking the sample with pure Ac-**Gly**-**Gly**-CN confirmed the identity of the dipeptide and the rate of reaction was affected by the temperature and quantity of Ac-**Gly**-SH used. With three equivalents of Ac-**Gly**-SH at 60 °C the **Gly**-CN was almost all consumed in 24 h and the total yield of dipeptide was 91%. In contrast the reactions at room temperature were sluggish requiring 5 d to reach 59% dipeptide with 3 eq. Ac-**Gly**-SH. Encouragingly, 1.5 eq. of Ac-**Gly**-SH also gave a good (69%) yield over 24 h.

4.3. Conclusion

The investigation into the role UV light could have had in prebiotic peptide formation has been very fruitful. Acetylation of **Gly**-CN with thioacetate under UV irradiation was achieved in almost quantitate yields between pH 5-9. Other thioacids were found to be effective acylating agents in UV light too, with comparable yields to acetylation with thioacetate. Control reactions found that UV photo-acetylation occurred at pH > 7 and acidic conditions (pH 3-5) were observed to enabled acetylation to also occur in the 'dark'. The reactions were optimised to complete conversion and the scope of the reaction was explored. The ligation of Ac-**Gly**-SH and **Gly**-CN was successfully achieved in very good yields at pH 3. UV irradiation was not effective in promoting ligation between Ac-**Gly**-SH and **Gly**-OH. However, acetone as a triplet sensitiser was effective in dipeptide formation with UV irradiation. Further investigation is underway to investigate the scope and nature of this reaction. Irradiation of *N*-acetyl aminothioacids Ac-**AA**-SH resulted in dethiocarboxylation to alkylacetamides and work is underway to investigate the dethiocarboxylation of Ac-**Cys**-SH to Ac-cysteamine for peptide catalysis. The thioacid Ac-**β-Ala**-SH did not dethiocarboxylate under UV irradiation and was found to be an excellent acylating agent. Ac-β-**Ala**-SH ligated with both **Gly**-CN and cysteamine under UV light, with the dipeptide Ac-β-**Ala**-cysteamine comprising the pantetheine moiety on CoA. These experiments may indicate that UV irradiation is important for creating selection pressure for specific classes of molecules being incorporated in biology.

5. References

1. Whitesides, G. M. Reinventing chemistry. *Angewandte Chem. International Edition* **54**, 3196–209 (2015).

2. Naddaf, G. *The Greek Concept of Nature* (State Universty of New York, 2005).

3. Gregory A. *Anaximander: A Re-assessment* (Bloomsbury Academic, New York, 2016).

4. Zwier, K. R. Methodology in Aristotle's theory of spontaneous generation. *J Hist Biol* **51**, 355–386 (2018).

5. Malik, A. H., Ziermann, J. M. & Diogo, R. An untold story in biology: the historical continuity of evolutionary ideas of Muslim scholars from the 8th century to Darwin's time. *J Biol Educ* **52**, 3–17 (2018).

6. Cavaillon, J.-M. & Legout, S. Louis Pasteur: Between myth and reality. *Biomolecules* **12**, 596 (2022).

7. Peretó, J., Bada, J. L. & Lazcano, A. Charles Darwin and the origin of life. *Origins of Life and Evolution of Biospheres* **39**, 395–406 (2009).

8. Wöhler, F. Ueber künstliche bildung des harnstoffs. *Ann Phys* **88**, 253–256 (1828).

9. Cech, T. R. The ribosome is a ribozyme. *Science* **289**, 878–879 (2000).

10. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).

11. Noffke, N., Christian, D., Wacey, D. & Hazen, R. M. Microbially induced sedimentary structures recording an ancient ecosystem in the ca. 3.48-billion-year-old dresser formation, Pilbara, Western Australia. *Astrobiology* **13**, 1103–1124 (2013).

12. Ruiz-Mirazo, K., Peretó, J. & Moreno, A. A universal definition of life: autonomy and open-ended evolution. *Origins of Life and Evolution of the Biosphere* **34**, 323–346 (2004).

13. Fleischaker, G. R. Origins of life: An operational definition. *Origins of Life and Evolution of the Biosphere* **20**, 127–137 (1990).

14. Szostak, J. W. Attempts to define life do not help to understand the origin of life. *J Biomol Struct Dyn* **29**, (2012).

15. Cleland, C. E. & Chyba, C. F. Defining 'life'. *Origins of life and evolution of biospheres* **32**, 387–393 (2002).

16. Deamer, D. W. & Fleischaker, G. R. *Origins of Life: The Central Concepts*. (Jones and Bartlett Publishers, Boston, 1994).

17. Sutherland, J. D. Opinion: Studies on the origin of life — the end of the beginning. *Nat Rev Chem* **1**, 0012 (2017).

18. Sutherland, J. D. Opinion: Studies on the origin of life — the end of the beginning. *Nat Rev Chem* **1**, 0012 (2017).

19. Cadet, J., Sage, E. & Douki, T. Ultraviolet radiation-mediated damage to cellular DNA. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **571**, 3–17 (2005).

20. Cadet, J. & Douki, T. Formation of UV-induced DNA damage contributing to skin cancer development. *Photochemical & Photobiological Sciences* **17**, 1816–1841 (2018).

21. Gargaud, M., Martin, H., López-García, P., Montmerle, T. & Pascal, R. *Young Sun, Early Earth and the Origins of Life*. *Young Sun, Early Earth and the Origins of Life* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012).

22. Williams, J. P. The astrophysical environment of the solar birthplace. *Contemp Phys* **51**, 381–396 (2010).

23. Bonanno, A., Schlattl, H. & Paternò, L. The age of the Sun and the relativistic corrections in the EOS. *Astron Astrophys* **390**, 1115–1118 (2002).

24. Hanslmeier, A. The Faint Young Sun Problem. in *Solar Magnetic Phenomena* 267–270 (Springer Netherlands, Dordrecht, 2005).

25. Ribas, I., Guinan, E. F., Gudel, M. & Audard, M. Evolution of the solar activity over time and effects on planetary atmospheres. I. high-energy irradiances (1–1700 A). *Astrophys J* **622**, 680–694 (2005).

26. Rimmer, P. B. *et al.* Timescales for prebiotic photochemistry under realistic surface ultraviolet conditions. *Astrobiology* **21**, 1099–1120 (2021).

27. Sutherland, J. D. The origin of life — out of the blue. *Angewandte Chemie International Edition* **55**, 104–121 (2016).

28. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat Chem* **7**, 301–307 (2015).

29. Ranjan, S. & Sasselov, D. D. Constraints on the early terrestrial surface UV environment relevant to prebiotic chemistry. *Astrobiology* **17**, 169–204 (2017).

30. Goldreich, P. & Ward, W. R. The formation of planetesimals. *Astrophys J* **183**, 1051–1061 (1973).

31. Orgel, L. E. The origin of life — a review of facts and speculations. *Trends Biochem Sci* **23**, 491–495 (1998).

32. Canup, R. M. & Asphaug, E. Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* **412**, 708–712 (2001).

33. Bell, E. A., Boehnke, P., Harrison, T. M. & Mao, W. L. Potentially biogenic carbon preserved in a 4.1-billion-year-old zircon. *Proceedings of the National Academy of Sciences* **112**, 14518–14521 (2015).

34.   Mojzsis, S. J., Harrison, T. M. & Pidgeon, R. T. Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* **409**, 178–181 (2001).

35.   Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* **435**, 466–469 (2005).

36.   Abramov, O. & Mojzsis, S. J. Microbial habitability of the Hadean Earth during the late heavy bombardment. *Nature* **459**, 419–422 (2009).

37.   Zahnle, K. J., Gacesa, M. & Catling, D. C. Strange messenger: A new history of hydrogen on Earth, as told by Xenon. *Geochim Cosmochim Acta* **244**, 56–85 (2019).

38.   Zahnle, K., Kasting, J. F. & Pollack, J. B. Mass fractionation of noble gases in diffusion-limited hydrodynamic hydrogen escape. *Icarus* **84**, 502–527 (1990).

39.   Hunten, D. M., Pepin, R. O. & Walker, J. C. G. Mass fractionation in hydrodynamic escape. *Icarus* **69**, 532–549 (1987).

40.   Oro, J. The origin and early evolution of life on Earth. *Annu Rev Earth Planet Sci* **18**, 317–356 (1990).

41.   Kasting, J. Earth's early atmosphere. *Science* **259**, 920–926 (1993).

42.   Olson, K. R. & Straub, K. D. The role of hydrogen sulfide in evolution and the evolution of hydrogen sulfide in metabolism and signaling. *Physiology* **31**, 60–72 (2016).

43.   Rasmussen, R. A., Khalil, M. A. K., Dalluge, R. W., Penkett, S. A. & Jones, B. Carbonyl sulfide and carbon disulfide from the eruptions of Mount St. Helens. *Science* **215**, 665–667 (1982).

44.   Sagan, C. The early faint Sun paradox: Organic shielding of ultraviolet-labile greenhouse gases. *Science* **276**, 1217–1221 (1997).

45.   Rimmer, P. B. *et al.* The origin of RNA precursors on exoplanets. *Sci Adv* **4**, (2018).

46.   Pavlov, A. A., Kasting, J. F., Brown, L. L., Rages, K. A. & Freedman, R. Greenhouse warming by CH_4 in the atmosphere of early Earth. *J Geophys Res Planets* **105**, 11981–11990 (2000).

47.   Sagan, C. & Chyba, C. The early faint Sun paradox: organic shielding of ultraviolet-labile greenhouse gases. *Science* **276**, 1217–1221 (1997).

48.   Kasting, J. F. Earth's early atmosphere. *Science* **259**, 920–926 (1993).

49.   Toner, J. D. & Catling, D. C. Alkaline lake settings for concentrated prebiotic cyanide and the origin of life. *Geochim Cosmochim Acta* **260**, 124–132 (2019).

50.   Todd, Z. R., Lozano, G. G., Kufner, C. L., Sasselov, D. D. & Catling, D. C. Ferrocyanide survival under near ultraviolet (300–400 nm) irradiation on early Earth. *Geochim Cosmochim Acta* **335**, 1–10 (2022).

51.   Keefe, A. D. & Miller, S. L. Was ferrocyanide a prebiotic reagent? *Origins of Life and Evolution of the Biosphere* **26**, 111–129 (1996).

52.   Sasselov, D. D., Grotzinger, J. P. & Sutherland, J. D. The origin of life as a planetary phenomenon. *Sci Adv* **6**, (2020).

53.   Ranjan, S. & Sasselov, D. D. Influence of the UV environment on the synthesis of prebiotic molecules. *Astrobiology* **16**, 68–88 (2016).

54.   Cleaves, H. J. & Miller, S. L. Oceanic protection of prebiotic organic compounds from UV radiation. *Proceedings of the National Academy of Sciences* **95**, 7260–7263 (1998).

55.   Lazcano, A. Historical development of origins research. *Cold Spring Harb Perspect Biol* **2**, a002089–a002089 (2010).

56.   Ehrenfreund, P. & Cami, J. Cosmic carbon chemistry: From the interstellar medium to the early Earth. *Cold Spring Harb Perspect Biol* **2**, a002097–a002097 (2010).

57. Kvenolden, K. *et al.* Evidence for extraterrestrial amino-acids and hydrocarbons in the Murchison meteorite. *Nature* **228**, 923–926 (1970).

58. Parker, E. T. *et al.* Primordial synthesis of amines and amino acids in a 1958 Miller H_2S-rich spark discharge experiment. *Proceedings of the National Academy of Sciences* **108**, 5526–5531 (2011).

59. Johnson, A. P. *et al.* The Miller volcanic spark discharge experiment. *Science* **322**, 404–404 (2008).

60. Niemann, H. B. *et al.* The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* **438**, 779–784 (2005).

61. Hazen, R. M. Paleomineralogy of the Hadean eon: A preliminary species list. *Am J Sci* **313**, 807–843 (2013).

62. Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol* **39**, 99–123 (2004).

63. Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231–236 (2013).

64. Wacey, D., Kilburn, M. R., Saunders, M., Cliff, J. & Brasier, M. D. Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nat Geosci* **4**, 698–702 (2011).

65. Ramakrishnan, V. Ribosome structure and the mechanism of translation. *Cell* **108**, 557–572 (2002).

66. Moore, P. B. & Steitz, T. A. The ribosome revealed. *Trends in biochemical sciences (Amsterdam. Regular ed.)* **30**, 281–283 (2005).

67. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).

68.	Peretó, J. Out of fuzzy chemistry: from prebiotic chemistry to metabolic networks. *Chem Soc Rev* **41**, 5394–5403 (2012).

69.	Sutherland, J. D. & Whitfield, J. N. Prebiotic chemistry: A bioorganic perspective. *Tetrahedron* **53**, 11493–11527 (1997).

70.	Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* **409**, 387–390 (2001).

71.	Powner, M. W. & Sutherland, J. D. Prebiotic chemistry: a new modus operandi. *Philosophical Transactions of the Royal Society B.* **366**, 2870–2877 (2011).

72.	Franklin, R. E. & Gosling, R. G. Molecular configuration in sodium thymonucleate. *Nature* **171**, 740–741 (1953).

73.	Clark, D. P., Pazdernik, N. J. & McGehee, M. R. Protein Synthesis. in *Molecular Biology* 397–444 (Elsevier, 2019).

74.	Crick, F. Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).

75.	Brenner, S., Benzer, S. & Barnett, L. Distribution of proflavin-induced mutations in the genetic fine structure. *Nature* **182**, 983–985 (1958).

76.	Nirenberg, M. *et al.* RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proceedings of the National Academy of Sciences* **53**, 1161–1168 (1965).

77.	Nirenberg, M. Historical review: Deciphering the genetic code – a personal account. *Trends Biochem Sci* **29**, 46–54 (2004).

78.	Khorana, H. G. Synthesis in the study of nucleic acids. The Fourth Jubilee Lecture. *Biochem J* **109**, 709–25 (1968).

79.	Crick, F. H. C. The origin of the genetic code. *J Mol Biol* **38**, 367–379 (1968).

80.	Orgel, L. E. Evolution of the genetic apparatus. *J Mol Biol* **38**, 381–393 (1968).

81.	Nissen, P. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science* **289**, 920–930 (2000).

82. Wimberly, B. T. *et al.* Structure of the 30S ribosomal subunit. *Nature* **407**, 327–339 (2000).

83. Yusupov, M. M. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896 (2001).

84. Nissen, P. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).

85. White, H. B. Coenzymes as fossils of an earlier metabolic state. *J Mol Evol* **7**, 101–104 (1976).

86. Wentrup, C. Origins of organic chemistry and organic synthesis. *European J Org Chem* **2022**, e202101492 (2022).

87. Schummer, J. The notion of nature in chemistry. *Studies in History and Philosophy of Science Part A* **34**, 705–736 (2003).

88. Strecker, A. Ueber die künstliche bildung der milchsäure und einen neuen, dem glycocoll homologen körper; *Annalen der Chemie und Pharmacie* **75**, 27–45 (1850).

89. Enders, D. & Shilvock, J. P. Some recent applications of α-amino nitrile chemistry. *Chem Soc Rev* **29**, 359–373 (2000).

90. Miller, S. L. A production of amino acids under possible primitive Earth conditions. *Science* **117**, 528–529 (1953).

91. Miller, S. L. Production of some organic compounds under possible primitive Earth conditions 1. *J Am Chem Soc* **77**, 2351–2361 (1955).

92. Miller, S. L. & Urey, H. C. Organic Compound Synthes on the Primitive Eart: Several questions about the origin of life have been answered, but much remains to be studied. *Science* **130**, 245–251 (1959).

93. Pascal, R., Boiteau, L. & Commeyras, A. From the Prebiotic Synthesis of α-Amino Acids Towards a Primitive Translation Apparatus for the Synthesis of

Peptides. in *Prebiotic Chemistry* 69–122 (Springer-Verlag, Berlin/Heidelberg, 2005).

94. Pascal, R., Boiteau, L. & Commeyras, A. From the Prebiotic Synthesis of α-Amino Acids Towards a Primitive Translation Apparatus for the Synthesis of Peptides. in *Prebiotic Chemistry* 69–122 (Springer-Verlag, Berlin/Heidelberg, 2005).

95. Frenkel-Pinter, M., Samanta, M., Ashkenasy, G. & Leman, L. J. Prebiotic peptides: Molecular hubs in the origin of life. *Chem Rev* **120**, 4707–4765 (2020).

96. Ashe, K. *et al.* Selective prebiotic synthesis of phosphoroaminonitriles and aminothioamides in neutral water. *Commun Chem* **2**, 23 (2019).

97. Metzler, D. E., Olivard, J. & Snell, E. E. Transamination of pyridoxamine and amino acids with glyoxylic acid. *J Am Chem Soc* **76**, 644–648 (1954).

98. Metzler, D. E., Snell, E. E. & Snell, E. E. Some transamination reactions involving vitamin B6. *J Am Chem Soc* **74**, 979–983 (1952).

99. Metzler, D. E., Longenecker, J. B. & Snell, E. E. The reversible catalytic cleavage of hydroxyamino acids by pyridoxal and metal Salts. *J Am Chem Soc* **76**, 639–644 (1954).

100. Muchowska, K. B., Varma, S. J. & Moran, J. Synthesis and breakdown of universal metabolic precursors promoted by iron. *Nature* **569**, 104–107 (2019).

101. Mayer, R. J. & Moran, J. Quantifying reductive amination in nonenzymatic amino acid synthesis. *Angewandte Chemie - International Edition* **61**, e202212237 (2022).

102. Springsteen, G., Yerabolu, J. R., Nelson, J., Rhea, C. J. & Krishnamurthy, R. Linked cycles of oxidative decarboxylation of glyoxylate as protometabolic analogs of the citric acid cycle. *Nat Commun* **9**, 91 (2018).

103. Islam, S. & Powner, M. W. Prebiotic systems chemistry: Complexity overcoming clutter. *Chem* **2**, 470–501 (2017).

104. Butlerow, A. Bildung einer zuckerartigen substanz durch synthese. *Liebigs. Ann. Chem.* **120**, 295–298 (1861).

105. Kirschning, A. Coenzymes and their role in the evolution of Life. *Angewandte Chemie International Edition* **60**, 6242–6269 (2020).

106. Decker, P. & Schweer, H. Gas — liquid chromatography on OV-225 of tetroses and aldopentoses as their O-methoxime and O-n-butanxime pertrifluoroacetyl derivatives and of C3-C6 alditol pertrifluoroacetates. *J Chromatogr A* **236**, 369–373 (1982).

107. Ricardo, A., Carrigan, M. A., Olcott, A. N. & Benner, S. A. Borate Minerals Stabilize Ribose. *Science* **303**, 196 (2004).

108. Ricardo, A. Borate Minerals Stabilize Ribose. *Science* **303**, 196–196 (2004).

109. Ritson, D. & Sutherland, J. D. Prebiotic synthesis of simple sugars by photoredox systems chemistry. *Nat Chem* **4**, 895–899 (2012).

110. Ritson, D. J. & Sutherland, J. D. Thiophosphate photochemistry enables prebiotic access to sugars and terpenoid precursors. *Nat Chem* **15**, 1470–1477 (2023).

111. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).

112. Oró, J. Synthesis of adenine from ammonium cyanide. *Biochem Biophys Res Commun* **2**, 407–412 (1960).

113. Bray, D. & Robbins, P. Mechanism of ε15 conversion studied with bacteriophage mutants. *J Mol Biol* **30**, 457–475 (1967).

114. BRAY, D. & ROBBINS, P. Mechanism of ε15 conversion studied with bacteriophage mutants. *J Mol Biol* **30**, 457–475 (1967).

115. Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Studies in prebiotic synthesis. *J Mol Biol* **38**, 121–128 (1968).

116. Ferris, J. P., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis. *J Mol Biol* **33**, 693–704 (1968).

117. Robertson, M. P. & Miller, S. L. An efficient prebiotic synthesis of cytosine and uracil. *Nature* **375**, 772–774 (1995).

118. Fuller, W. D., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis. *J Mol Biol* **67**, 25–33 (1972).

119. Krishnamurthy, R., Guntha, S. & Eschenmoser, A. Regioselective α-phosphorylation of aldoses in aqueous solution. *Angewandte Chemie - International Edition* **39**, 2281–2285 (2000).

120. Kim, H. J. & Benner, S. A. Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. *Proceedings National Academy of Sciences* **114**, 11315–11320 (2017).

121. Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis. *J Mol Biol* **47**, 531–543 (1970).

122. Tapiero, C. M. & Nagyvary, J. Prebiotic formation of cytidine nucleotides. *Nature* **231**, 42–43 (1971).

123. Powner, M. W. & Sutherland, J. D. Potentially prebiotic synthesis of pyrimidine β-D-ribonucleotides by photoanomerization/hydrolysis of α-D-cytidine-2′-phosphate. *ChemBioChem* **9**, 2386–2387 (2008).

124. Powner, M. W. *et al.* On the prebiotic synthesis of ribonucleotides: Photoanomerisation of cytosine nucleosides and nucleotides revisited. *ChemBioChem* **8**, 1170–1179 (2007).

125. Anastasi, C., Crowe, M. A., Powner, M. W. & Sutherland, J. D. Direct assembly of nucleoside precursors from two- and three-carbon Units. *Angewandte Chem. International Edition* **45**, 6176–6179 (2006).

126. Islam, S., Bučar, D.-K. & Powner, M. W. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nat Chem* **9**, 584–589 (2017).

127. Powner, M. W. & Sutherland, J. D. Phosphate-mediated Interconversion of ribo- and arabino-configured prebiotic nucleotide intermediates. *Angewandte Chemie International Edition* **49**, 4641–4643 (2010).

128. Powner, M. W., Sutherland, J. D. & Szostak, J. W. Chemoselective multicomponent one-pot assembly of purine precursors in water. *J Am Chem Soc* **132**, 16677–16688 (2010).

129. Becker, S. *et al.* Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* **366**, (2019).

130. Becker, S. *et al.* Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. *Nat Commun* **9**, 163 (2018).

131. Becker, S. *et al.* Origin of life: A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833–836 (2016).

132. Stairs, S. *et al.* Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. *Nat Commun* **8**, 15270 (2017).

133. Roberts, S. J. *et al.* Selective prebiotic conversion of pyrimidine and purine anhydronucleosides into Watson-Crick base-pairing arabino-furanosyl nucleosides in water. *Nat Commun* **9**, 4073 (2018).

134. Coggins, A. J. & Powner, M. W. Prebiotic synthesis of phosphoenol pyruvate by α-phosphorylation-controlled triose glycolysis. *Nat Chem* **9**, 310–317 (2017).

135. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).

136. Jaradat, D. M. M. Thirteen decades of peptide synthesis: key developments in solid phase peptide synthesis and amide bond formation utilized in peptide ligation. *Amino Acids* **50**, 39–68 (2018).

137. Rode, B. M. Peptides and the origin of life1. *Peptides (N.Y.)* **20**, 773–786 (1999).

138. Pederson, T. RNA: Life's Indispensable Molecule, by James E. Darnell. 2011. Cold Spring Harbor Laboratory Press. *RNA* **17**, (2011).

139. Jaradat, D. M. M. Thirteen decades of peptide synthesis: key developments in solid phase peptide synthesis and amide bond formation utilized in peptide ligation. *Amino Acids* **50**, 39–68 (2018).

140. Caetano-Anollés, G., Wang, M., Caetano-Anollés, D. & Mittenthal, J. E. The origin, evolution and structure of the protein world. *Biochemical Journal* **417**, 621–637 (2009).

141. Plankensteiner, K., Reiner, H. & Rode, B. Prebiotic chemistry: The amino acid and peptide world. *Curr Org Chem* **9**, 1107–1114 (2005).

142. Ikehara, K. Evolutionary steps in the emergence of life deduced from the bottom-up approach and GADV hypothesis (Top-down approach). *Life* **6**, 6 (2016).

143. Ikehara, K. Possible steps to the emergence of life: The [GADV]-protein world hypothesis. *Chemical Record* **5**, 107–118 (2005).

144. Ikehara, K. [GADV]-Protein world hypothesis on the origin of life. *Origins of Life and Evolution of Biospheres* **44**, 299–302 (2014).

145. Bryan, M. C. *et al.* Key green chemistry research areas from a pharmaceutical manufacturers' perspective revisited. *Green Chemistry* **20**, 5082–5103 (2018).

146. Constable, D. J. C. *et al.* Key green chemistry research areas — a perspective from pharmaceutical manufacturers. *Green Chem.* **9**, 411–420 (2007).

147. Ratti, R. Industrial applications of green chemistry: Status, Challenges and Prospects. *SN Appl Sci* **2**, 263 (2020).

148. Pattabiraman, V. R. & Bode, J. W. Rethinking amide bond synthesis. *Nature* **480**, 471–479 (2011).

149. Curtius, T. Ueber einige neue der hippursäure analog constituirte, synthetisch dargestellte amidosäuren. *Journal für Praktische Chemie* **26**, 145–208 (1882).

150. Fischer, E. & Fourneau, E. Ueber einige derivate des glykocolls. *Berichte der deutschen chemischen Gesellschaft* **34**, 2868–2877 (1901).

151. Fischer, E. Synthese von polypeptiden. XVII. *Berichte der deutschen chemischen Gesellschaft* **40**, 1754–1767 (1907).

152. Fischer, E. Synthese von polypeptiden. IX. chloride der aminosäuren und ihrer acylderivate. *Berichte der deutschen chemischen Gesellschaft* **38**, 605–619 (1905).

153. McKay, F. C. & Albertson, N. F. New amine-masking groups for peptide synthesis. *J Am Chem Soc* **79**, 4686–4690 (1957).

154. Martin, R. B. Free energies and equilibria of peptide bond hydrolysis and formation. *Biopolymers* **45**, 351–353 (1998).

155. Schreiner, E., Nair, N. N. & Marx, D. Peptide synthesis in aqueous environments: The role of extreme conditions on peptide bond formation and peptide hydrolysis. *J Am Chem Soc* **131**, 13668–13675 (2009).

156. Flegmann, A. W. & Tattersall, R. Energetics of peptide bond formation at elevated temperatures. *J Mol Evol* **12**, 349–355 (1979).

157. Flegmann, A. W. & Tattersall, R. Energetics of peptide bond formation at elevated temperatures. *J Mol Evol* **12**, 349–355 (1979).

158. Shock, E. L. Stability of peptides in high-temperature aqueous solutions. *Geochim Cosmochim Acta* **56**, 3481–3491 (1992).

159. Perrin, D. D. Dissociation contants of inorganic acids and bases in aqueous solution. *Pure and Applied Chemistry* **20**, 133–236 (1969).

160. Merrifield, R. B. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J Am Chem Soc* **85**, 2149–2154 (1963).

161. Merrifield, R. B. Solid-phase peptide synthesis. III. An improved synthesis of Bradykinin. *Biochemistry* **3**, 1385–1390 (1964).

162. Merrifield, R. B., Gisin, B. F. & Bach, A. N. The limits of reaction of radioactive dicyclohexylcarbodiimide with amino groups during solid-phase peptide synthesis. *Journal of Organic Chemistry* **42**, 1291–1295 (1977).

163. Hirschmann, R. F. Controlled synthesis of peptides in aqueous medium. III. Use of Leuch's anhydrides in the synthesis of di-peptides. Mechanism and control of side reactions. *J Org Chem* **32**, 3415–3425 (1967).

164. Hirschmann, R. *et al.* Synthesis of peptides in aqueous medium. VII. Preparation and use of 2,5-thiazolidinediones in peptide synthesis. *J Org Chem* **36**, 49–59 (1971).

165. Hirschmann, R. *et al.* Controlled synthesis of peptides in aqueous medium. VIII. Preparation and use of novel. alpha. -amino acid N-carboxyanhydrides. *J Am Chem Soc* **93**, 2746–2754 (1971).

166. Leman, L. Carbonyl sulfide-mediated prebiotic formation of peptides. *Science* **306**, 283–286 (2004).

167. Leman, L. J., Orgel, L. E. & Ghadiri, M. R. Amino acid dependent formation of phosphate anhydrides in water mediated by carbonyl sulfide. *J Am Chem Soc* **128**, 20–21 (2006).

168. Leman, L. & Ghadiri, M. Potentially prebiotic synthesis of α-amino thioacids in water. *Synlett* **28**, 68–72 (2016).

169. Pascal, R., Boiteau, L. & Commeyras, A. From the Prebiotic Synthesis of α-Amino Acids Towards a Primitive Translation Apparatus for the Synthesis of Peptides. in *Prebiotic Chemistry* 69–122 (Springer-Verlag, Berlin/Heidelberg, 2005).

170. Forsythe, J. G. *et al.* Ester-mediated amide bond formation driven by wet-dry cycles: A possible path to polypeptides on the prebiotic Earth. *Angewandte Chemie International Edition* **54**, 9871–9875 (2015).

171. Frenkel-Pinter, M. *et al.* Thioesters provide a plausible prebiotic path to proto-peptides. *Nat Commun* **13**, 2569 (2022).

172. Duve, C. The beginnings of life on Earth. *Am Sci* **83**, 428–437 (1995).

173. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).

174. Canavelli, P., Islam, S. & Powner, M. W. Peptide ligation by chemoselective aminonitrile coupling in water. *Nature* **571**, 546–549 (2019).

175. Maurel, M. C. & Orgel, L. E. Oligomerization of α-thioglutamic acid. *Origins of Life and Evolution of the Biosphere* **30**, 423–430 (2000).

176. Liu, R. & Orgel, L. E. Oxidative acylation using thioacids. *Nature* **389**, 52–54 (1997).

177. Thoma, B. & Powner, M. W. Selective synthesis of lysine peptides and the prebiotically plausible synthesis of catalytically active diaminopropionic acid peptide nitriles in water. *J Am Chem Soc* **145**, 3121–3130 (2023).

178. Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* **409** 387–390 (2001).

179. Schwarz, P. S. *et al.* Parasitic behavior in competing chemically fueled reaction cycles. *Chem Sci* **12**, 7554–7560 (2021).

180. Altay, M., Altay, Y. & Otto, S. Parasitic behavior of self-replicating molecules. *Angewandte Chemie International Edition* **57**, 10564–10568 (2018).

181. Ruiz-Mirazo, K., Briones, C. & de la Escosura, A. Prebiotic systems chemistry: New perspectives for the origins of life. *Chem Rev* **114**, 285–366 (2014).

182. Brack, A. *The Molecular Origins of Life. The Molecular Origins of Life* (Cambridge University Press, 1998).

183. Ourisson, G. & Nakatani, Y. Addendum: Origins of cellular life: Molecular foundations and new approaches. *Tetrahedron* **55**, 3183–3190 (1999).

184. Thomas, J. A. & Rana, F. R. The influence of environmental conditions, lipid composition, and phase behavior on the origin of cell membranes. *Origins of Life and Evolution of Biospheres* **37**, 267–285 (2007).

185. Davis, B. H. & Occelli, M. L. *Advances in Fischer-Tropsch Synthesis, Catalysts, and Catalysis. Advances in Fischer-Tropsch Synthesis, Catalysts, and Catalysis* (CRC Press, 2009).

186. van de Loosdrecht, J. *et al.* Fischer–Tropsch Synthesis: Catalysts and Chemistry. in *Comprehensive Inorganic Chemistry II* vol. 7 525–557 (Elsevier, 2013).

187. Podolsky, K. A. & Devaraj, N. K. Synthesis of lipid membranes for artificial cells. *Nat Rev Chem* **5**, 676–694 (2021).

188. Mccollom, T. M., Ritter, G. & Simoneit, B. R. T. Lipid synthesis under hydrothermal conditions by Fischer-Tropsch-type reactions. *Origins of Life and Evolution of the Biosphere* **29**, 153–166 (1999).

189. Fernández-García, C. & Powner, M. Selective Acylation of Nucleosides, Nucleotides, and Glycerol-3-phosphocholine in Water. *Synlett* **28**, 78–83 (2016).

190. Gurr, M. I., Harwood, J. L. & Frayn, K. N. *Lipid Biochemistry. Lipid Biochemistry* (Wiley, 2002).

191. Vance, J. E. & Vance, D. E. *Biochemistry of Lipids, Lipoproteins and Membranes. Biochemistry of Lipids, Lipoproteins and Membranes* (Elsevier, 2008).

192. Fuchs, G. Alternative pathways of carbon dioxide fixation: Insights into the early evolution of life? *Annu Rev Microbiol* **65**, 631–658 (2011).

193. Wächtershäuser, G. Groundworks for an evolutionary biochemistry: The iron-sulphur world. *Prog Biophys Mol Biol* **58**, 85–201 (1992).

194. Poole, L. B. The basics of thiols and cysteines in redox biology and chemistry. *Free Radic Biol Med* **80**, 148–157 (2015).

195. Goldford, J. E., Hartman, H., Smith, T. F. & Segrè, D. Remnants of an ancient metabolism without phosphate. *Cell* **168**, 1126-1134.e9 (2017).

196. Eck, R. V. & Dayhoff, M. O. Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. *Science* **152**, 363–366 (1966).

197. Bonfio, C. *et al.* UV-light-driven prebiotic synthesis of iron–sulfur clusters. *Nat Chem* **9**, 1229–1234 (2017).

198. Foden, C. S., Islam, S., Fern, C., Maugeri, L. & Sheppard, T. D. Prebiotic synthesis of cysteine peptides that catalyse peptide ligation in neutral water. *Pre-print https://chemrxiv.org/articles/preprint/Prebiotic_Synthesis_of_Cysteine_Peptid es_That_Catalyze_Peptide_Ligation_in_Neutral_Water/12578699/1* 1–27.

199. Foden, C. S. *et al.* Prebiotic synthesis of cysteine peptides that catalyze peptide ligation in neutral water. *Science* **370**, 865–869 (2020).

200. Fujishima, K. *et al.* Reconstruction of cysteine biosynthesis using engineered cysteine-free enzymes. *Sci Rep* **8**, 1776 (2018).

201. Ksander, G. *et al.* Chemie der α-aminonitrile 1. Mitteilung einleitung und wege zu uroporphyrinogen-octanitrilen. *Helv Chim Acta* **70**, 1115–1172 (1987).

202. Leonardi, R., Zhang, Y., Rock, C. & Jackowski, S. Coenzyme A: Back in action. *Prog Lipid Res* **44**, 125–153 (2005).

203. Daugherty, M. *et al.* Complete reconstitution of the human coenzyme A biosynthetic pathway via comparative genomics. *Journal of Biological Chemistry* **277**, 21431–21439 (2002).

204. Walsh, C. T., Tu, B. P. & Tang, Y. Eight kinetically stable but thermodynamically activated molecules that power cell metabolism. *Chem Rev* **118**, 1460–1494 (2018).

205. Shi, L. & Tu, B. P. Acetyl-CoA and the regulation of metabolism: Mechanisms and consequences. *Curr Opin Cell Biol* **33**, 125–131 (2015).

206. Strieker, M., Tanović, A. & Marahiel, M. A. Nonribosomal peptide synthetases: structures and dynamics. *Curr Opin Struct Biol* **20**, 234–240 (2010).

207. Keefe, A. D., Newton, G. L. & Miller, S. L. A possible prebiotic synthesis of pantetheine, a precursor to coenzyme A. *Nature* **373**, 683–685 (1995).

208. Dawson, P., Muir, T., Clark-Lewis, I. & Kent, S. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).

209. Dawson, P., Muir, T., Clark-Lewis, I. & Kent, S. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).

210. Jin, K. & Li, X. Advances in native chemical ligation-desulfurization: A powerful strategy for peptide and protein Synthesis. *Chemistry European Journal* **24**, 17397–17404 (2018).

211. Singh, J. *et al.* Prebiotic catalytic peptide ligation yields proteinogenic peptides by intramolecular amide catalyzed hydrolysis facilitating regioselective lysine ligation in neutral water. *J Am Chem Soc* **144**, 10151–10155 (2022).

212. Jadhav, V. R. & Yarus, M. Coenzymes as coribozymes. *Biochimie* **84**, 877–888 (2002).

213. Lipmann, F. Attempts to map a process evolution of peptide biosynthesis. *Science* **173**, 875–884 (1971).

214. Eschenmoser, A. & Loewenthal, E. Chemistry of potentially prebiological natural products. *Chem Soc Rev* **21**, 1 (1992).

215. Wu, L.-F. & Sutherland, J. D. Provisioning the origin and early evolution of life. *Emerg Top Life Sci* **3**, 459–468 (2019).

216. White, S. W., Zheng, J., Zhang, Y.-M. & Rock, C. O. The structural biology of type II fatty acid biosynthesis. *Annu Rev Biochem* **74**, 791–831 (2005).

217. Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chem Rev* **106**, 3468–3496 (2006).

218. Huang, F., Bugg, C. W. & Yarus, M. RNA-catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN. *Biochemistry* **39**, 15548–15555 (2000).

219. Begley, T. P., Kinsland, C. & Strauss, E. The biosynthesis of coenzyme a in bacteria. in *Vitamins and Hormones* **61**, 157–171 (2001).

220. Miller, StanleyL. & Schlesinger, G. Prebiotic syntheses of vitamin coenzymes: II. Pantoic acid, pantothenic acid, and the composition of coenzyme A. *J Mol Evol* **36**, (1993).

221. Choughuley, A. S. U. & Lemmon, R. M. Production of cysteic acid, taurine and cystamine under primitive Earth conditions. *Nature* **210**, 628–629 (1966).

222. Raulin, F. & Toupance, G. The role of sulphur in chemical evolution. *J Mol Evol* **9**, 329–338 (1977).

223. Miller, StanleyL. & Schlesinger, G. Prebiotic syntheses of vitamin coenzymes: I. Cysteamine and 2-mercaptoethanesulfonic acid (coenzyme M). *J Mol Evol* **36**, 302–307 (1993).

224. Parker, E. T. *et al.* Prebiotic synthesis of methionine and other sulfur-containing organic compounds on the primitive Earth: A contemporary reassessment based on an unpublished 1958 Stanley Miller experiment. *Origins of Life and Evolution of Biospheres* **41**, 201–212 (2011).

225. Sampedro, A., Rodriguez Granger, J., Ceballos, J. & Aliaga, L. Pantothenic acid: An overview focused on medical aspects. *Eur Sci J* **11**, (2016).

226. Williams, R. J., Lyman, C. M., Goodyear, G. H., Truesdail, J. H. & Holaday, D. "Pantothenic acid," a growth determinant of universal biological occurrence. *J Am Chem Soc* **55**, 2912–2927 (1933).

227. Kresge, N., Simoni, R. D. & Hill, R. L. Fritz Lipmann and the discovery of coenzyme A. *Journal of Biological Chemistry* **280**, 164–166 (2005).

228. Baddiley, J., Thain, E. M., Novelli, G. D. & Lipmann, F. Structure of coenzyme A. *Nature* **171**, 76–76 (1953).

229. Lipmann, F. Development of the acetylation problem, a personal account. *Science* **120**, 855–865 (1954).

230. Lipmann, F. A phosphorylated oxidation product of pyruvic acid. *Journal of Biological Chemistry* **134**, 463–464 (1940).

231. Lipmann, F. Acetylation of sulfanilamide by liver homogenates and extracts. *Journal of Biological Chemistry* **160**, 173–190 (1945).

232. Lipmann, Fritz., Kaplan, N. O., Novelli, G. D., Tuttle, L. C. & Guirard, B. M. Isolation of coenzyme A. *Journal of Biological Chemistry* **186**, 235–243 (1950).

233. Lipmann, F., Kaplan, N. O., Novelli, G. D., Tuttle, L. C. & Guirard, B. M. Coenzyme for acetylation, a pantothenic acid derivative. *Journal of Biological Chemistry* **167**, 869–870 (1947).

234. Brown, G. M. Biosynthesis of pantothenic acid and coenzyme A. in *Comprehensive Biochemistry* **21**, 73–80 (1970).

235. Leonardi, R. & Jackowski, S. Biosynthesis of pantothenic acid and coenzyme A. *EcoSal Plus* **2**, (2007).

236. Knowles, J. R. Enzyme catalysis: not different, just better. *Nature* **350**, 121–124 (1991).

237. Kowtoniuk, W. E., Shen, Y., Heemstra, J. M., Agarwal, I. & Liu, D. R. A chemical screen for biological small molecule–RNA conjugates reveals CoA-linked RNA. *Proceedings of the National Academy of Sciences* **106**, 7768–7773 (2009).

238. Bird, J. G. *et al.* The mechanism of RNA 5′ capping with NAD+, NADH and desphospho-coA. *Nature* **535**, 444–447 (2016).

239. Breaker, R. R. & Joyce, G. F. Self-Incorporation of coenzymes by ribozymes. *J Mol Evol* **40**, 551–558 (1995).

240. Pietrocola, F., Galluzzi, L., Bravo-San Pedro, J. M., Madeo, F. & Kroemer, G. Acetyl coenzyme A: A central metabolite and second messenger. *Cell Metab* **21**, 805–821 (2015).

241. Martin, W. F. Older than genes: The acetyl coA pathway and origins. *Front Microbiol* **11**, (2020).

242. Harrison, S. A. & Lane, N. Life as a guide to prebiotic nucleotide synthesis. *Nat Commun* **9**, 5176 (2018).

243. Muchowska, K. B., Varma, S. J. & Moran, J. Nonenzymatic metabolic reactions and life's origins. *Chem Rev* **120**, 7708–7744 (2020).

244. Islam, S., Bučar, D.-K. & Powner, M. W. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nat Chem* **9**, 584–589 (2017).

245. Parker, E. T. *et al.* Primordial synthesis of amines and amino acids in a 1958 Miller H_2S-rich spark discharge experiment. *Proceedings of the National Academy of Sciences* **108**, 5526–5531 (2011).

246. Edward, J. T., Farrell, P. G., Job, J. L. & Poh, B.-L. Re-examination of the Kirkwood–Westheimer theory of electrostatic effects. II. Possible conformations of α,ω-amino-acids in aqueous solutions, as deduced from dissociation constants. *Can J Chem* **56**, 1122–1129 (1978).

247. Ferris, J. P. Life at the margins. *Nature* **373**, 659–659 (1995).

248. Hawker, J. R. & Oró, J. Cyanamide mediated syntheses of peptides containing histidine and hydrophobic amino acids. *J Mol Evol* **17**, 285–294 (1981).

249. Ponnamperuma, C. & Peterson, E. Peptide Synthesis from Amino Acids in Aqueous Solution. *Science* **147**, 1572–1574 (1965).

250. Nooner, D. W. & Oró, J. Direct synthesis of polypeptides. *J Mol Evol* **3**, 79–88 (1974).

251. Danger, G. *et al.* 5(4 *H*)-Oxazolones as Intermediates in the Carbodiimide- and Cyanamide-Promoted Peptide Activations in Aqueous Solution. *Angewandte Chemie International Edition* **52**, 611–614 (2013).

252. Liu, R. & Orgel, L. E. Polymerization of β-amino acids in aqueous solution. *Origins of Life and Evolution of the Biosphere* **28**, (1998).

253. Danger, G., Plasson, R. & Pascal, R. Pathways for the formation and evolution of peptides in prebiotic environments. *Chem Soc Rev* **41**, 5416 (2012).

254. Danger, G., d'Hendecourt, L. L. S. & Pascal, R. On the conditions for mimicking natural selection in chemical systems. *Nat Rev Chem* **4**, 102–109 (2020).

255. Whicher, A., Camprubi, E., Pinna, S., Herschy, B. & Lane, N. Acetyl phosphate as a primordial energy currency at the origin of life. *Origins of Life and Evolution of Biospheres* **48**, 159–179 (2018).

256. Griesser, H., Bechthold, M., Tremmel, P., Kervio, E. & Richert, C. Amino acid-specific, ribonucleotide-promoted peptide formation in the absence of enzymes. *Angewandte Chemie International Edition* **56**, 1224–1228 (2017).

257. Schimpl, A., Lemmon, R. M. & Calvin, M. Cyanamide Formation under Primitive Earth Conditions. *Science* **147**, 149–150 (1965).

258. Pinna, S. *et al.* A prebiotic basis for ATP as the universal energy currency. *PLoS Biol* **20**, e3001437 (2022).

259. Shimizu, M., Ohta, G., Nagase, O., Okada, S. & Hosokawa, Y. Investigations on pantothenic acid and its related compounds. I. Chemical studies. (1). A novel synthesis of pantethine. *Chem Pharm Bull (Tokyo)* **13**, 180–188 (1965).

260. Capon, P. K., Avery, T. D., Purdey, M. S. & Abell, A. D. An improved synthesis of 4-aminobutanenitrile from 4-azidobutanenitrile and comments on room temperature stability. *Synth Commun* **51**, 428–436 (2021).

261. Keillor, J. W. & Brown, R. S. Attack of zwitterionic ammonium thiolates on a distorted anilide as a model for the acylation of papain by amides. A simple demonstration of a bell-shaped pH/rate profile. *J Am Chem Soc* **114**, 7983–7989 (1992).

262. Stiller, E. T., Harris, S. A., Finkelstein, J., Keresztesy, J. C. & Folkers, K. Pantothenic acid. VIII. The total synthesis of pure pantothenic acid. *J Am Chem Soc* **62**, 1785–1790 (1940).

263. Williams, R. J., Mitchell, H. K., Weinstock, H. H. & Snell, E. E. Pantothenic acid. VII. Partial and total synthesis studies. *J Am Chem Soc* **62**, 1784–1785 (1940).

264. Wessely, L. Uber die einwirkung von kali auf 2-dimethyl-3-oxypropionaldehyd. *Monatshefte für Chemie* **22**, 66–68 (1901).

265. Magalhães, Á. F. & Powner, M. W. Prebiotic triose glycolysis promoted by co-catalytic proline and phosphate in neutral water. *Chemical Communications* **58**, 13519–13522 (2022).

266. Markert, M., Scheffler, U. & Mahrwald, R. Asymmetric histidine-catalyzed cross-aldol reactions of enolizable aldehydes: Access to defined configured quaternary stereogenic centers. *J Am Chem Soc* **131**, 16642–16643 (2009).

267. Kuhn, R. & Weiser, D. Aminozucker-synthesen VII. α-amino-β,β-dimethyl-γ-hydroxy-butyraldehyd. *Justus Liebigs Ann Chem* **602**, 208–217 (1957).

268. Marshall, M. How the first life on Earth survived its biggest threat — water. *Nature* **588**, 210–213 (2020).

269. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat Chem Biol* **5**, 593–599 (2009).

270. Hagan, W. J. Uracil-catalyzed synthesis of acetyl phosphate: A photochemical driver for protometabolism. *ChemBioChem* **11**, 383–387 (2010).

271. Liu, H. *et al.* Potassium thioacids mediated selective amide and peptide constructions enabled by visible light photoredox catalysis. *ACS Catal* **6**, 1732–1736 (2016).

272. Bonfio, C. *et al.* UV-light-driven prebiotic synthesis of iron–sulfur clusters. *Nat Chem* **9**, 1229–1234 (2017).

273. Wu, L.-F., Liu, Z. & Sutherland, J. D. pH-Dependent peptide bond formation by the selective coupling of α-amino acids in water. *Chemical Communications* **57**, 73–76 (2021).

274. Zhang, S. J., Duzdevich, D. & Szostak, J. W. Potentially prebiotic activation chemistry compatible with nonenzymatic RNA copying. *J Am Chem Soc* **142**, 14810–14813 (2020).

275. Okamoto, R. *et al.* Regioselective α-peptide bond formation through the oxidation of amino thioacids. *Biochemistry* **58**, 1672–1678 (2019).

6. Experimental

General Experimental and Safety

Reagents and solvents were obtained from commercial sources and used without further purification, unless stated otherwise. Deionised water was obtained from an *Elga Option 3* purification system and degassing by three freeze-vacuum-thaw cycles. Temperatures of 0 °C was achieved by ice-water bath. The cryogenic temperatures were obtained using dry ice in acetone (-78 °C) and dry ice in 30% methanol/water (-20 °C). Thin layer chromatography (TLC) was carried out on *Merck* aluminium-backed DC 60 F_{254} 0.2 mm pre-coated plates. TLC plates were visualised under ultraviolet light and stained with permanganate solution when required. Photochemical reactions were carried out in quartz screw-cap cuvettes, using a Rayonet RPR-200 photochemical chamber reactor, acquired from The Southern New England Ultraviolet Company, with RPR2547A or RPR3000A lamps. Solution pH values were measured using a *Mettler Toledo Seven Compact* pH meter with a *Mettler Toledo InLab* semi-micro pH probe. Flash column chromatography was performed on a *Biotage Isolera One* using *Kinesis TELOS* cartridges.^{1}H and ^{13}C NMR spectra were recorded on *Bruker* NMR spectrometers *AVANCE Neo 700*, equipped with a *Bruker* room temperature 5 mm multinuclear gradient probe, a 5 mm DCH cryoprobe (700 MHz) or on a *AVANCE III 600 equipped* with a 5 mm DCH cryoprobe (600 MHz) and a *AVANCE Neo 500*, equipped with a *Bruker* room temperature 5 mm multinuclear gradient probe (500 MHz). [13]C spectra were proton decoupled. Chemical shifts (δ) are reported in parts per million (ppm) relative to residual solvent peaks, and ^{1}H chemical shifts relative to TMS were calibrated using the residual solvent peak or an internal NMR reference ((methanesulfonyl)methane (MSM) or potassium hydrogen phthalate (KHP)). Nuclear assignments were made using 2D NMR homo- and heteronuclear correlation spectroscopy (^{1}H-^{1}H COSY; ^{1}H-^{13}C HSQC; ^{1}H-^{13}C HMBC). Solvent suppression pulse sequence with pre-saturation and spoil gradients were used to obtain ^{1}H NMR spectra (noesygppr1d, *Bruker*) and ^{1}H-^{13}C HMBC NMR spectra (hmbcgplpndprqf, *Bruker*). Coupling constants are reported in hertz (Hz). Spin multiplicities are indicated by symbols: s (singlet); d (doublet); t (triplet); q (quartet); qn (quintet); spt (septet); oct (octet), m (multiplet); obs. (obscured/coincidental signals); app. (apparent), or a combination of these. Diastereotopic geminal spin systems are reported as AB. Diastereotopic geminal (AB) spin systems coupled to one or two additional nuclei are reported as ABX and ABXY, respectively. NMR data are reported as follows: chemical shift (number of protons, multiplicity, coupling constants (J), nuclear assignment). NMR spectra were recorded at 298 K. Infrared spectra (IR) were recorded on a *Bruker Alpha FT-IR* with a platinum-ATR (attenuated total reflection) attachment as a solid or neat oil/liquid. Absorption maxima are reported in wavenumber (cm^{-1}) and the spectral range was between 400 cm^{-1} and 4000 cm^{-1} with a resolution of 0.01 cm^{-1}. Mass spectra and accurate mass measurements were recorded on a *Waters LCT Premier QTOF* connected to a *Waters Autosampler Manager 2777C*, *Thermo Finnigan MAT900*, and an *Agilent LC* connected to an *Agilent 6510 QTOF* mass spectrometer. Melting points were determined using an *Electrothermal* standard digital apparatus for all solids and are quoted to the nearest ºC and uncorrected.

WARNING: Hydrogen cyanide, hydrogen sulfide, and their salts are highly toxic poisons by contact, inhalation, and ingestion. The salts generate poisonous hydrogen cyanide (pK_a = 9.2) and hydrogen sulfide (pK_a = 7.1) gas at neutral or acidic pH. Any solutions that contain (hydro)sulfide or cyanide, or compounds that may generate these must be handled in a well-ventilated fume hood with appropriate chemical quenches at hand (such as sodium hypochlorite (bleach) or iron (II) sulphate solution). Material safety data sheets (MSDS) and instructions for personnel handling, exposure, and disposal information must be read and followed. Local safety personnel should be consulted for regulations concerning proper and safe disposal according to local environmental guidelines.

Appendix 1

Experimental Data for the Prebiotic Synthesis of Pantetheine

1. Attempted syntheses of pantothenic acid 6 and pantetheine 1 from pantolactone 2, β-alanine 3, and cysteamine 4

Acid-catalysed cyclisation of pantoic acid 5 to pantolactone 2
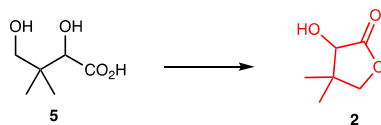
A solution of pantoate 5 (Na$^+$-salt; 50 mM) in phosphate buffer (pH 7; 500 mM) was adjusted to the required pH using HCl (conc.) and incubated at 20 ºC. The reaction was periodically monitored by NMR spectroscopy. Yield of pantolactone 2 formation is given in Supplementary Table 1.
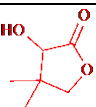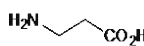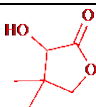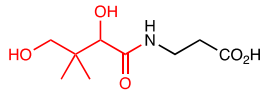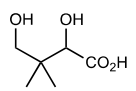
pH	Time (days)	Yield (%) 2
6.1	6	4
5.0	13	22
3.7	10	46
3.0	10	88
2.0	2	97
1.5	1	>95

Supplementary Table 1. Quantification of pantolactone 2 formation from pantoic acid 5 (50 mM) in phosphate buffer (500 mM) at the specified pH and 20 ºC.

Reaction of pantolactone 2 and β-alanine 3 in water
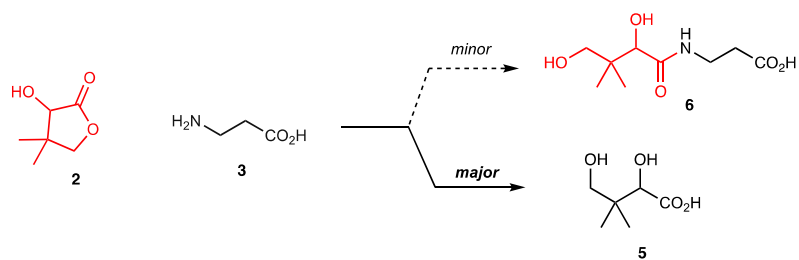
Pantolactone **2** (0.05–0.5 M) and β-alanine **3** (0.1–1 M) were dissolved in phosphate buffer (0.5 M) at pH 7.0 at 20 ºC. The solution was adjusted to the specified pH with 1-4 M NaOH. The reaction was periodically analysed by ^{1}H NMR spectroscopy and its pH continually monitored. Yields of pantothenic acid **6** and pantoic acid **5** are reported in Supplementary Table 2. ^{1}H NMR spectra are given in Supplementary Figure 1 and Supplementary Figure 2.

Concentration (mM)		pH	Time (days)	Yield (%)		
![2] 2	![3] 3			![2] 2	![6] 6	![5] 5
0.05	0.1	7.5	8	30	0	70
0.05	0.1	8.5	8	12	0	88
0.05	0.1	9.0	8	6	0	94
0.10	0.2	9.0	7	11	1	88
0.25	0.5	9.0	7	6	7	87
0.50	1.0	9.0	7	14	12	74

Supplementary Table 2. Quantification of species observed from the reaction of pantolactone **2** with β-alanine **3** at specified pH in phosphate buffer (500 mM) and at 20 ºC.
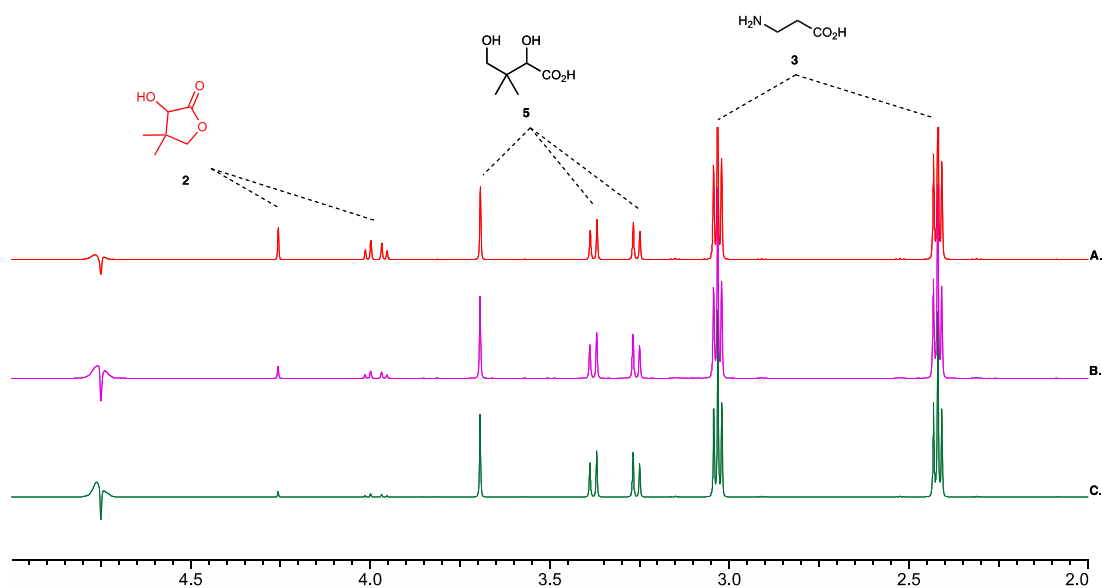
Supplementary Figure 1. ^{1}H NMR (700 MHz, H_2O, noesygppr1d, 2.0–5.0 ppm) spectra to show the reaction of pantolactone **2** (50 mM) with β-alanine **3** (2 equiv.) in phosphate buffer (500 mM) at 20 °C, and at (A.) pH 7.5 after 8 days; (B.) pH 8.5 after 8 days; (C.) pH 9.0 after 8 days.
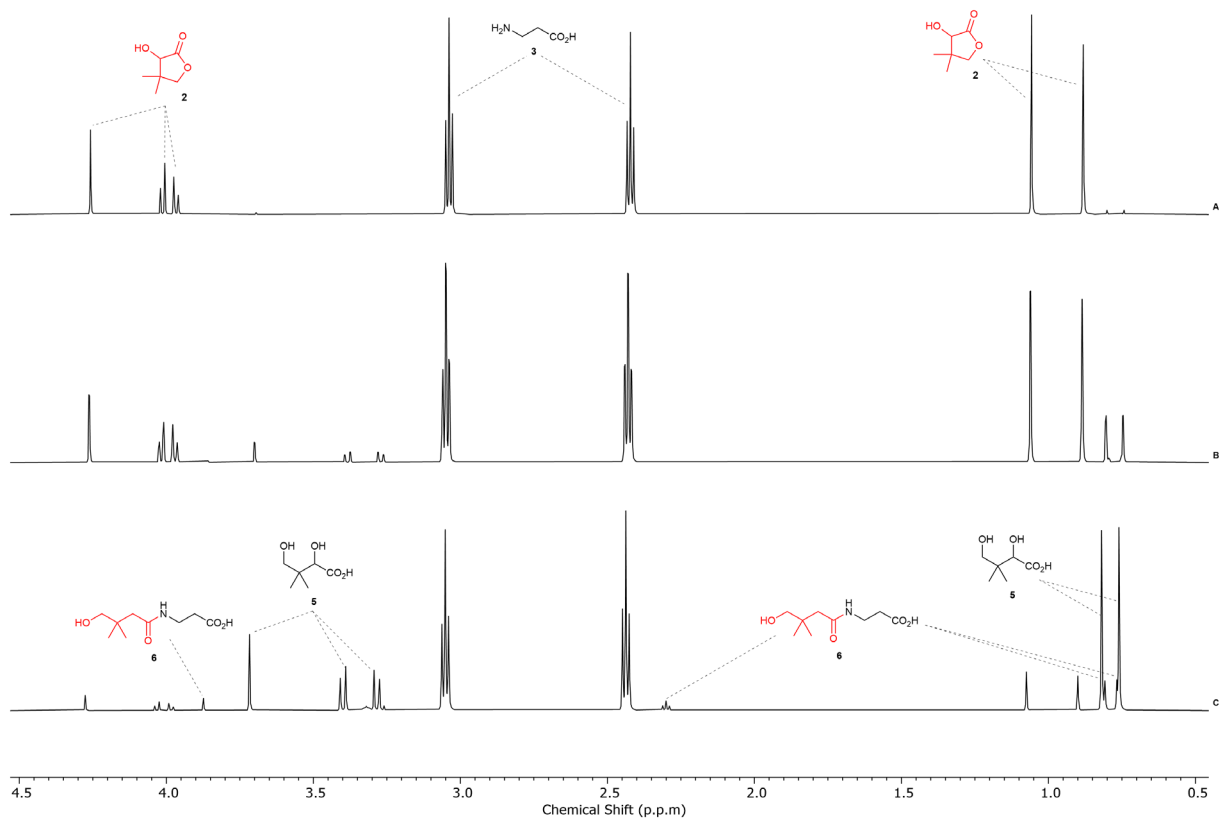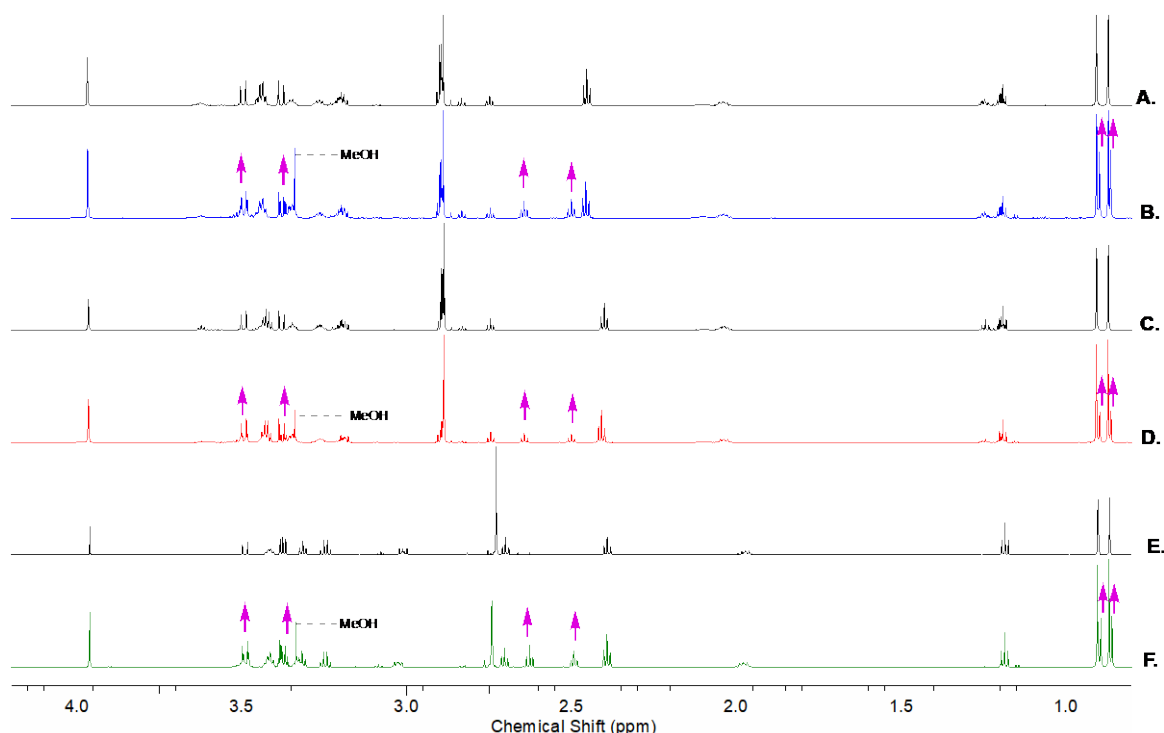

Supplementary Figure 2. ^{1}H NMR (600 MHz, H_2O, noesygppr1d, 0.5–4.5 ppm) spectra to show the reaction of pantolactone **2** (500 mM) and β-alanine **3** (2 equiv.) in phosphate buffer (pH 9.0; 500 mM) and 20 °C at: **A.** <10 mins; **B.** 3 days; **C.** 7 days.

*Competitive reaction of pantolactone **2** with β-alanine **3** and glycine **Gly** in water*
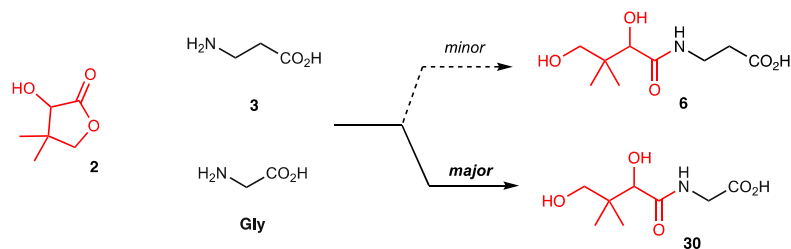
Pantolactone **2** (0.5 M), β-alanine **3** (1 M), and glycine **Gly** (1 M) were dissolved in phosphate buffer at pH 7.0 (0.5 M) at 20 ºC. The solution pH was adjusted to pH 9 with 1-4 M NaOH. The resulting mixture was periodically analysed by ^{1}H NMR spectroscopy and its pH monitored. Selective formation of pantoylglycine **30** (29%) over pantothenic acid **6** (11%) was observed, alongside competing hydrolysis of pantolactone **2** to pantoic acid **5** (59%) after 7 days. ^{1}H NMR spectra are shown in Supplementary Figure 3.

Supplementary Figure 3. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–4.5 ppm) spectra to show the competitive reaction of pantolactone **2** (500 mM) with β-alanine **3** (2 equiv.) and glycine **Gly** (2 equiv.) in phosphate buffer (pH 9.0; 500 mM) at 20 ºC and pH 9 after: (A.) 10 mins; (B.) 1 day (C.) 7 days. Inset: Expanded ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 3.6–4.4 ppm) spectral region.
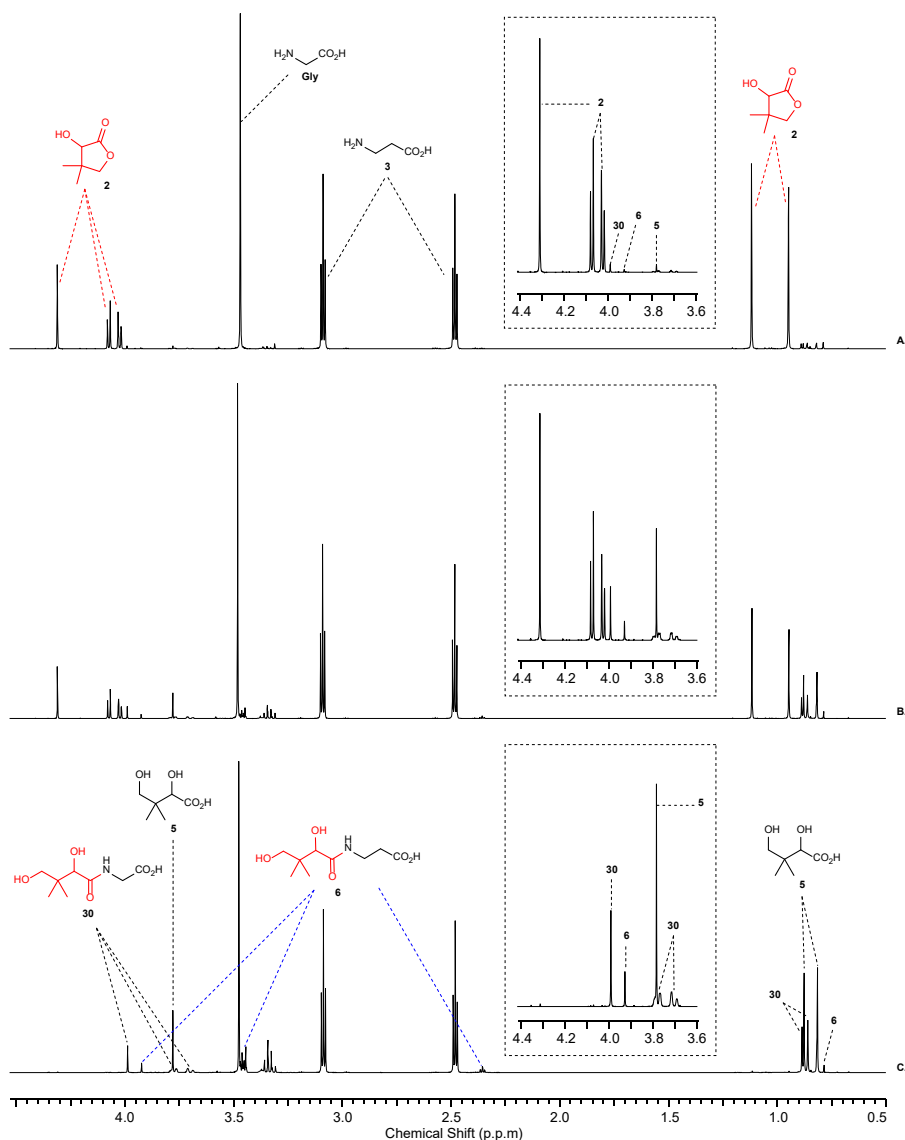
*Failed syntheses of pantothenic acid **6** from pantolactone **2** and β-alanine **3** using a wet-dry cycle*
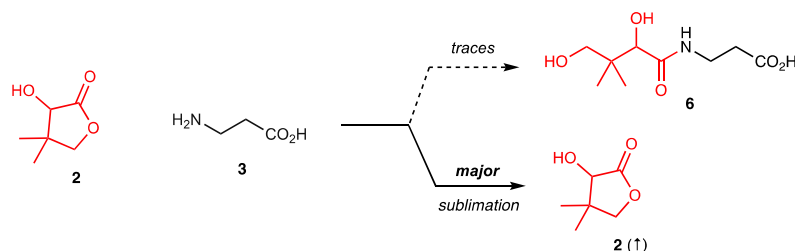
A solution of pantolactone **2** (500 mM) and β-alanine **3** (500 mM) in water was adjusted to pH 7.0 with 8 M HCl at 20 ºC. This solution was analysed by ^{1}H NMR spectroscopy; only pantolactone **2** and β-alanine **3** were observed (Supplementary Figure 4A).

*Reduced pressure-promoted sublimation of pantolactone **2** at room temperature preventing pantothenic acid **6** synthesis*

Pantolactone **2** (500 mM) and β-alanine **3** (500 mM) in water (1 mL) at pH 7.0 were concentrated *in vacuo* (0.04–0.08 mBar) at 20 ºC for 2 hours or 1 day. The resulting residue was dissolved in water and analysed by NMR spectroscopy. Substantial loss (92%) of pantolactone **2** was observed after 2 hours (Supplementary Figure 4B). Loss of pantolactone **2** was complete after 1 day. Pantothenic acid **6** was not observed.

*Heat-promoted sublimation of pantolactone **2** at atmospheric pressure preventing pantothenic acid **6** synthesis*

Experiment A
A stream of air was passed over pantolactone **2** (500 mM) and β-alanine **3** (500 mM) in water (1 mL) at pH 7.0 for 2.5 days at 20 ºC in an open vial. The resulting residue was then heated at 100 °C for 24 hours in an open vial. The reaction was cooled to 20 ºC, the resulting residue was dissolved in water, and analysed by NMR spectroscopy to show complete loss (>99%) of pantolactone **2**. Only β-alanine **3** was detected by NMR spectroscopy. (Supplementary Figure 4C).

Experiment B
Pantolactone **2** (500 mM) and β-alanine **3** (500 mM) in water (1 mL) at pH 7.0 were heated at 100 °C for 24 hours in an open vial. The reaction was then cooled to 20 ºC. The resulting residue was dissolved in water (2 mL) and analysed by NMR spectroscopy. Substantial depletion of pantolactone **2** was observed; 13% residual **2** was observed. Multiple trace (<1%) resonances in the characteristic pantoylamide ^{1}H NMR spectral region [(CO)CHOH (3.80–4.05 ppm)] were observed, the combined yield of these pantoyl-products was 5% (based on initial **2**). The β-alanyl resonances were more complex, and 10% of the initial β-alanine **3** had been converted into β-alanyl-products. This result indicated that multiple pantoyl-β-alanyl species and oligo-β-alanyl species were formed from oligomerisation of β-alanine **3**. Spiking with authentic standard of

pantothenic acid **6** allowed identification of **6** as a minor trace product, amongst the multitude of low levels of pantoyl-amide species (see Supplementary Figure 4D and Supplementary Figure 5). No further attempt was made to quantify the trace yield of **6**, or isolated **6** from the multiple (related) observed by-products.
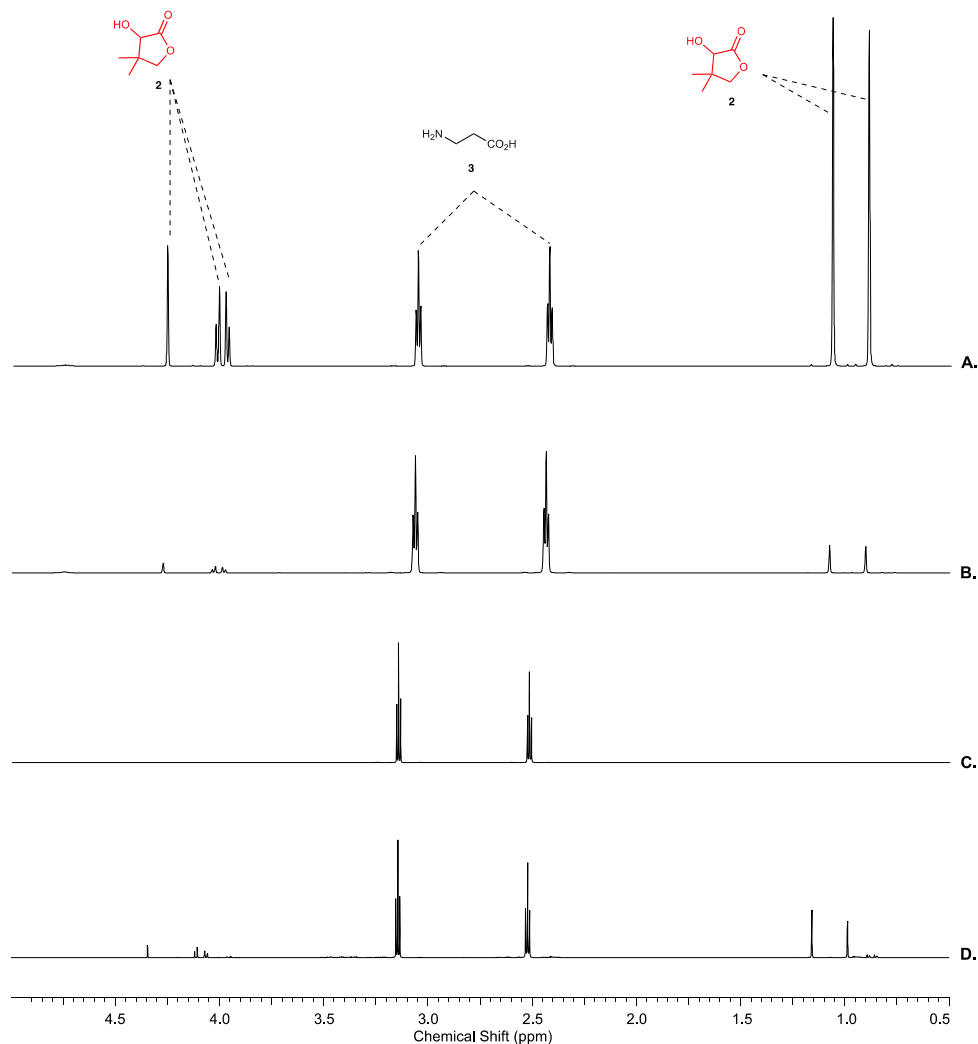
Supplementary Figure 4. ^{1}H NMR (600 MHz, H$_2$O, noessygppr1d, 0.5–5.0 ppm) spectra to show (A.) pantolactone **2** (500 mM) and β-alanine **3** (1 equiv.) at pH 7.0; (B.) after leaving pantolactone **2** (500 mM) with β-alanine **3** (1 equiv.) at pH 7.0 *in vacuo* (0.04–0.08 mbar) at 20 ºC for 2 hours; (C.) after leaving a sample of pantolactone **2** (500 mM) and β-alanine **3** (1 equiv.) at pH 7.0 at atmospheric pressure under a slow stream of air for 60 hours at 20 ºC, followed by heating of the resulting residue at 100 °C for 24 hours at atmospheric pressure; (D.) after heating a sample of pantolactone **2** (500 mM) with β-alanine **3** (1 equiv.) at pH 7.0 at 100 °C for 24 hours at atmospheric pressure – see Supplementary Figure 5 for ^{1}H NMR spectra after acute signal magnification.
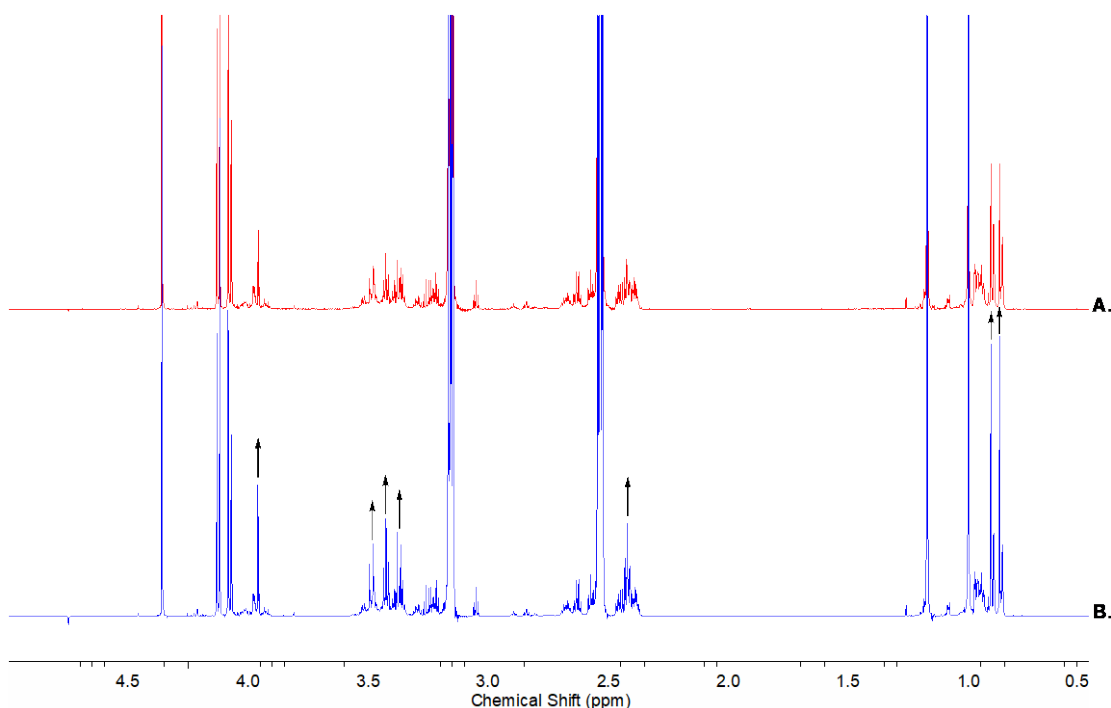
Supplementary Figure 5. ^{1}H NMR (600 MHz, H_2O, noessygppr1d, 0.5–5.0 ppm) spectra after acute signal amplification. (A.) Heating a sample of pantolactone **2** (500 mM) with β-alanine **3** (1 equiv.) at pH 7.0 and 100 °C for 24 hours at atmospheric pressure – see Supplementary Figure 4D for an attenuated ^{1}H NMR spectrum. Multiple resonances in the characteristic pantoylamide ^{1}H NMR spectral region [(CO)CHOH (3.80–4.05 ppm)] were observed (<5% total amide formation). However, a greater depletion of β-alanine **3** resonances was accompanied by greater signal complexity derived from β-alanine **3** (10%), suggesting that multiple pantoyl-β-alanyl (including pantothenic acid **6**) and homopoly-β-alanyl species were formed from uncontrolled polymerisation reactions of β-alanine **3**. Spiking with a commercial sample of pantothenic acid **6** confirmed its presence amongst the multitude of low levels of pantoyl-derived species that are not pantolactone **2** or pantoic acid **5**. (B.) Spiking of the crude reaction sample shown in spectrum A. with commercial pantothenic acid **6**.

Sublimation of pantolactone 2

Solid pantolactone **2** (D-, L- or *rac*-**2**) was heated at 100 °C in an open vessel. Needle-like crystals were observed to collect at the neck of the vessel within 2 hours. After 24 hours the bulk of the pantolactone **2** solid had sublimed and collected at the neck of the open vessel (see Supplementary Figure 6 for representative photographs for the sublimation of D-**2**).
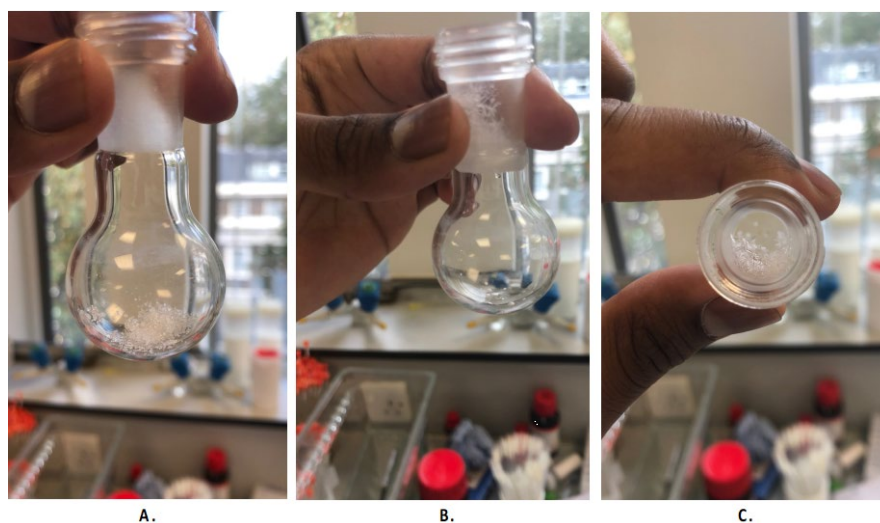
Supplementary Figure 6. Photographs of (A.) D-pantolactone D-**2** (133 mg) in a 10 mL volume borosilicate glass with short jointed B19/23 neck before heating at 100 °C. (B.) After heating D-pantolactone D-**2** at 100 °C for 24 hours, where the vast majority of the compound had migrated from the bottom of the flask, and (C.) the sublimate of D-pantolactone D-**2** that had collected at the neck of the flask as needle-like crystals.

Decomposition of pantothenic acid 6 during a wet-dry cycle
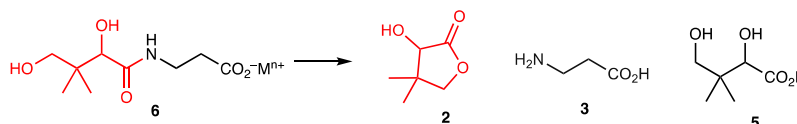
Pantothenic acid **6** ((hemi)-Ca^{2+} or Na$^+$ salt; 500 mM) was dissolved in water (1 mL) and the solution was adjusted to pH 7.0 with 1-4 M HCl. The solution was heated in an open vessel at 100 ºC for 24 hours. The resulting residue was allowed to cool to 20 ºC, and then dissolved in D$_2$O (2 mL). The resultant solution was analysed by NMR spectroscopy (Supplementary Figure 7). Yield of pantothenic acid **6** recovery and hydrolysis products observed are reported in Supplementary Table 3 .

Concentration (mM)		Yield (%)			
6	M^{n+}	**6**	**2**	**3**	**5**
500	Ca^{2+}	46	3	51	51
500	Na$^+$	75	3	25	17

Supplementary Table 3. Quantification of hydrolysis observed after heating a solution of pantothenic acid **6** (500 mM) at pH 7 and heating at 100 ºC for 24 hours.
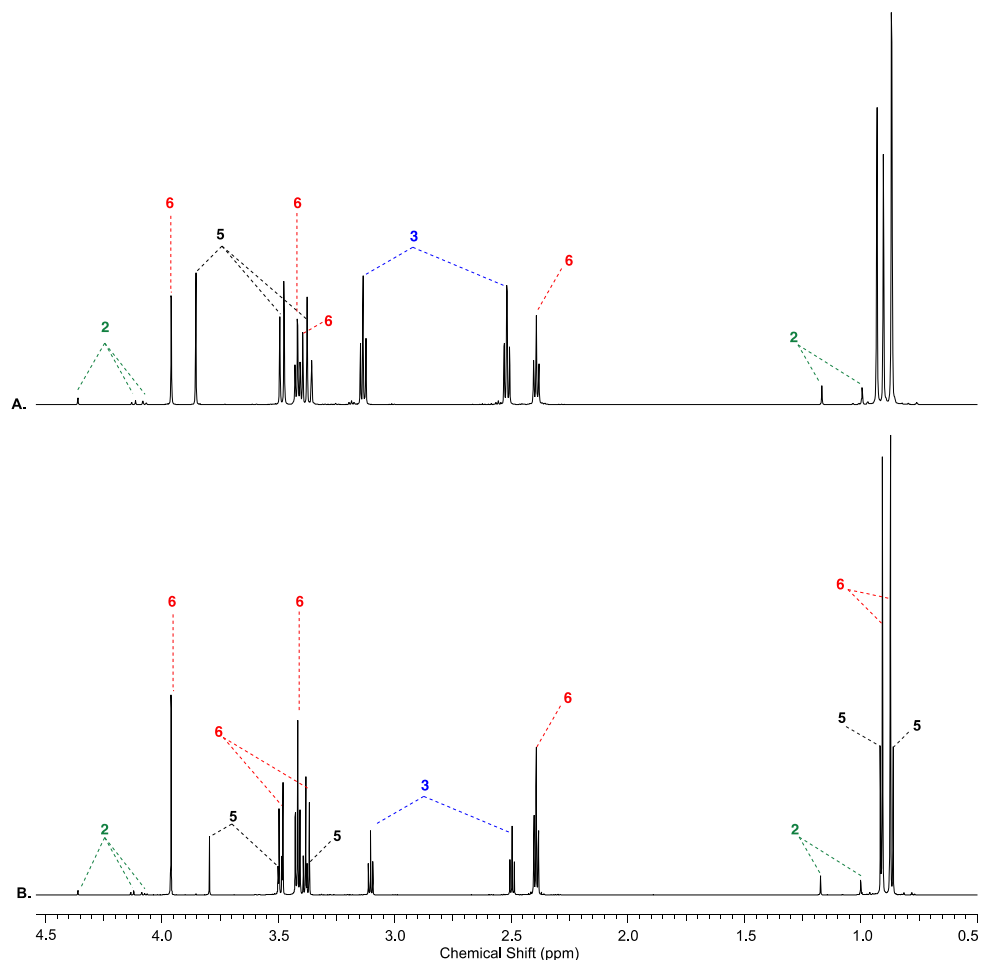
Supplementary Figure 7. ^{1}H NMR (700 MHz, D$_2$O, noessygppr1d, 0.5–4.5 ppm) spectra to show the hydrolysis of pantothenic acid **6** after heating a solution of pantothenic acid **6** (500 mM) at pH 7 in an open vessel at 100 ºC for 24 hours. (A.) pantothenic acid **6**-(hemi) calcium salt. (B.) pantothenic acid **6**-sodium salt.

Attempted syntheses of pantetheine 1 from pantolactone 2, β-alanine 3 and cysteamine 4 using a wet-dry cycle

not detectable to traces
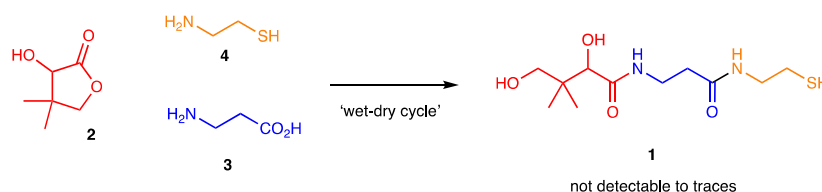
A solution of pantolactone **2** (500 mM), β-alanine **3** (500 mM), and cysteamine **4** (500 mM) in water was adjusted to pH 7.0 with 8 M HCl at 20 ºC. This solution was analysed by ^{1}H NMR spectroscopy; only pantolactone **2**, β-alanine **3**, cysteamine **4** and cysteamine disulfide [**4**]$_2$ (5%) were observed (Supplementary Figure 8A).

Incubation in water

Pantolactone **2** (500 mM), β-alanine **3** (500 mM) and cysteamine **4** (500 mM) in water were incubated at pH 7.0 and 20 ºC for 4 days. Trace (<1%) of pantoic acid **5** from the hydrolysis of pantolactone **2** was observed. No other pantoyl-derived species were observed by NMR spectroscopy (Supplementary Figure 8B).

Wet-dry cycle and desiccation at room temperature

Pantolactone **2** (500 mM), β-alanine **3** (500 mM) and cysteamine **4** (500 mM) in water (1 mL) at pH 7.0 were concentrated *in vacuo* at 20 ºC (over P_2O_5 as a desiccant) under reduced pressure (220 mbar – a sufficiently high pressure was maintained to remove water but to also suppress the sublimation of pantolactone **2**). The residue was redissolved in water after 1 day, and analysed by NMR spectroscopy. Pantolactone **2** remained unchanged. Only partial (10%) oxidation of cysteamine **4** to cysteamine disulfide [**4**]$_2$ was observed (Supplementary Figure 9).

Wet-dry cycle and desiccation at room temperature, followed by heating

Pantolactone **2** (500 mM), β-alanine **3** (500 mM) and cysteamine **4** (500 mM) in water (1 mL) at pH 7.0 were concentrated *in vacuo* at 20 ºC over P_2O_5 (as a desiccant) under (moderately) reduced pressure (220 mbar) for 1 day. The residue was then heated for 1 d: (a.) in an open vessel, or (b.) in a seal and argon filled vessel. After cooling to 20 ºC, resultant residues were dissolved in water and analysed by NMR spectroscopy.

   a. *Open vessel* – substantial loss of pantolactone **2** (78%) was caused by heat-promoted sublimation at atmospheric pressure. Very little (<5%) pantoyl-amide resonances ([(CO)C<u>H</u>OH (3.80–4.00 ppm)]) were observed, and 13% pantolactone **2** remaining Sequential spiking with pantetheine **1** and pantethine [**1**]$_2$ (a.k.a. 'pantetheine disulfide') could not confirm their presence within the detection limits of ^{1}H NMR spectroscopy (Supplementary Figure 10). A substantial increase in complexity in the 2.3–3.6 ppm ^{1}H NMR spectral region was indicative of multiple side reactions of: (i) cysteamine **4**, including oxidation of **4** to cysteamine disulfide [**4**]$_2$, and the formation of β-alanyl-cysteamine amides derived therefrom, and (ii) homo-oligomers of β-alanine **3**.

b. *Argon-filled, sealed vessel* – loss of pantolactone **2** (25%) was caused by heat-promoted sublimation despite being a closed system[*]. However, the loss of pantolactone **2** was substantially reduced compared with the analogous reaction in an open system (see above; *Experiment A.*), and so no further efforts were made to prevent sublimation. Spiking with an authentic sample of pantetheine **1** and pantethine [**1**]$_2$ ('pantetheine disulfide') indicated that a combined trace (<1%) yield of **1** and [**1**]$_2$ was formed (Supplementary Figure 11), which is consistent with the previously reported low-yielding synthesis of pantetheine **1** under comparable conditions[2]. Complexity in the 2.3-3.6 ppm ^{1}H NMR spectral region indicated the formation multiple side reactions and by-products of (i) cysteamine **4**, including oxidation of **4** to cysteamine disulfide [**4**]$_2$ and the formation of β-alanyl-cysteamine amides derived therefrom, and (ii) homo-oligomers of β-alanine **3**. No further attempt to characterise these by-products was undertaken.

Supplementary Figure 8. ^{1}H NMR (600 MHz, H$_2$O, noessygppr1d, 0.5–5.0 ppm) spectra to show (A.) pantolactone **2** (500 mM), β-alanine **3** (1 equiv.), and cysteamine **4** (1 equiv.) at pH 7.0 immediately after dissolution in water at 20 ºC and (B.) after incubation at 20 ºC and pH 7.0 for 4 days.

Supplementary Figure 9. ^{1}H NMR (600 MHz, H_2O, noessygppr1d, 0.5–5.0 ppm) spectra to show pantolactone **2** (500 mM), β-alanine **3** (1 equiv.) and cysteamine **4** (1 equiv.) in water (1 mL) at pH 7.0 after concentrating *in vacuo* at 20 ºC over P_2O_5 as a desiccant under moderately reduced pressure (220 mbar) after 1 day at 20 ºC.

Supplementary Figure 10. ^{1}H NMR (600 MHz, H$_2$O, noessygppr1d, 2.3–4.5 ppm) spectra to show pantolactone **2** (500 mM), β-alanine **3** (1 equiv.) and cysteamine **4** (1 equiv.) in water (1 mL) at pH 7.0 after concentrating *in vacuo* at 20 ºC over P$_2$O$_5$ as a desiccant under moderately reduced pressure (220 mbar) for 1 day, followed by (A.) heating at 100 °C for 1 day in an open vessel; (B.) spiking with an authentic sample of pantetheine **1** did not confirm its presence; (C.) spiking with an authentic sample of pantethine [**1**]$_2$ also did not confirm its presence.
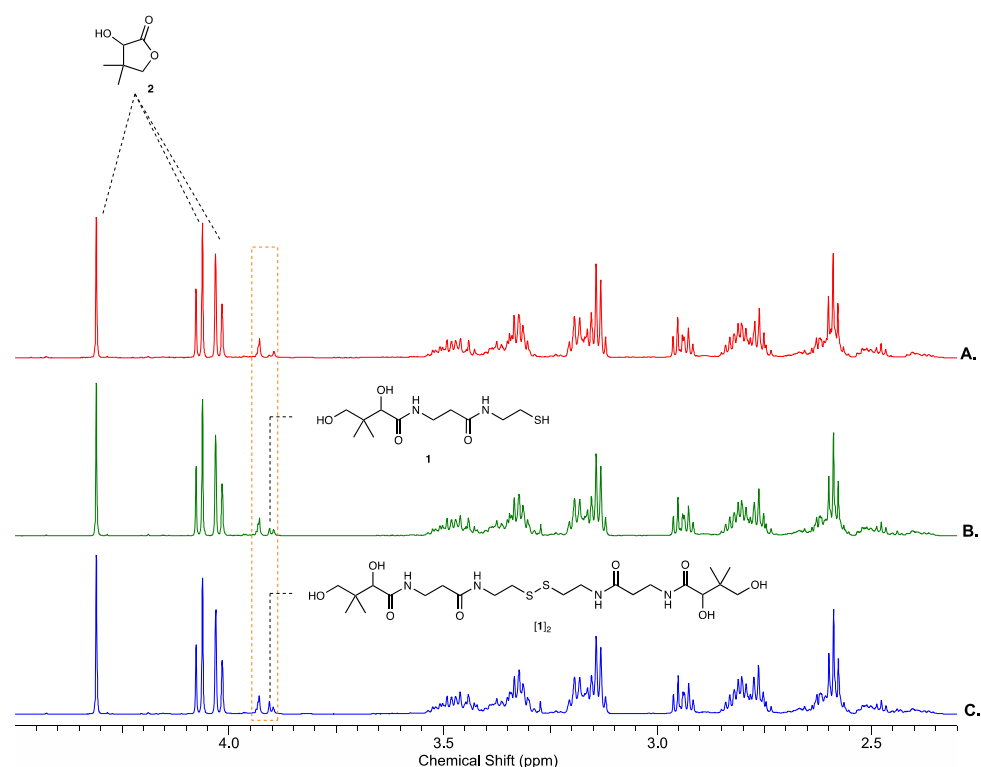
Supplementary Figure 11. ^{1}H NMR (600 MHz, H$_2$O, noessygppr1d, 2.3–5.0 ppm) spectra to show pantolactone **2** (500 mM), β-alanine **3** (1 equiv.) and cysteamine **4** (1 equiv.) in water (1 mL) at pH 7.0 after concentrating *in vacuo* at 20 ºC over P$_2$O$_5$ as a desiccant under moderately reduced pressure (220 mbar) for 1 day, followed by (A.) heating at 100°C for 1 day in an closed argon-flushed vessel; (B.) spiking with an authentic sample of pantetheine **1**, and (C.) pantethine [**1**]$_2$ confirming that a combined trace (<1%) yield of **1** and [**1**]$_2$ was formed.

Attempted pantetheine 1 syntheses with electrophilic activating agents in water

Attempted synthesis of pantetheine 1 from pantothenic acid 6 and cysteamine 4 by acetyl phosphate-mediated activation
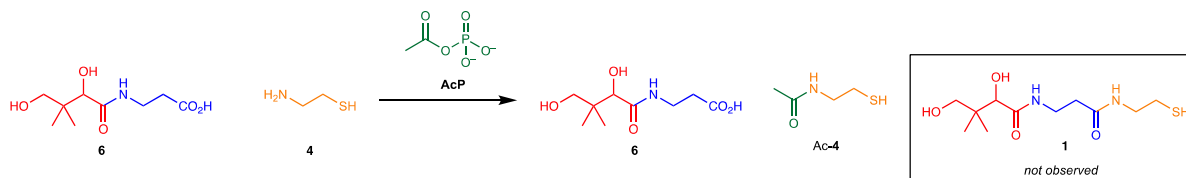
Pantothenic acid **6** (100 mM) and cysteamine **4** (100 mM) were dissolved phosphate buffer (pH 7; 500 mM). To the solution was added acetyl phosphate **AcP** (1.1 equiv.). The reaction was kept at 20 ºC and analysed over 24 hours by ^{1}H NMR spectroscopy. The acetylation of cysteamine **4** to yield *N*-acetylcysteamine Ac-**4** (31%) was observed. Pantetheine **1** was not detected. The hydrolysis of acetyl phosphate **AcP** to acetic acid (40%) was also observed.

Data for N-acetylcysteamine Ac-4
^{1}H NMR (600 MHz, H$_2$O/D$_2$O 9:1, noesygppr1d) δ_H 3.30 (t, 2H, *J* = 6.5 Hz, NHC$\underline{H}_2$), 2.61 (t, *J* = 6.5 Hz, 2H, C$\underline{H}_2$SH, 1.95 (s, 3H, (COC$\underline{H}_3$). ^{13}C NMR (151 MHz, H$_2$O/D$_2$O 9:1) δ_C 174.7 ($\underline{C}$OCH$_3$), 38.6 (NH$\underline{C}$H$_2$), 24.3 ($\underline{C}$H$_2$SH), 22.6 (COC$\underline{H}_3$).
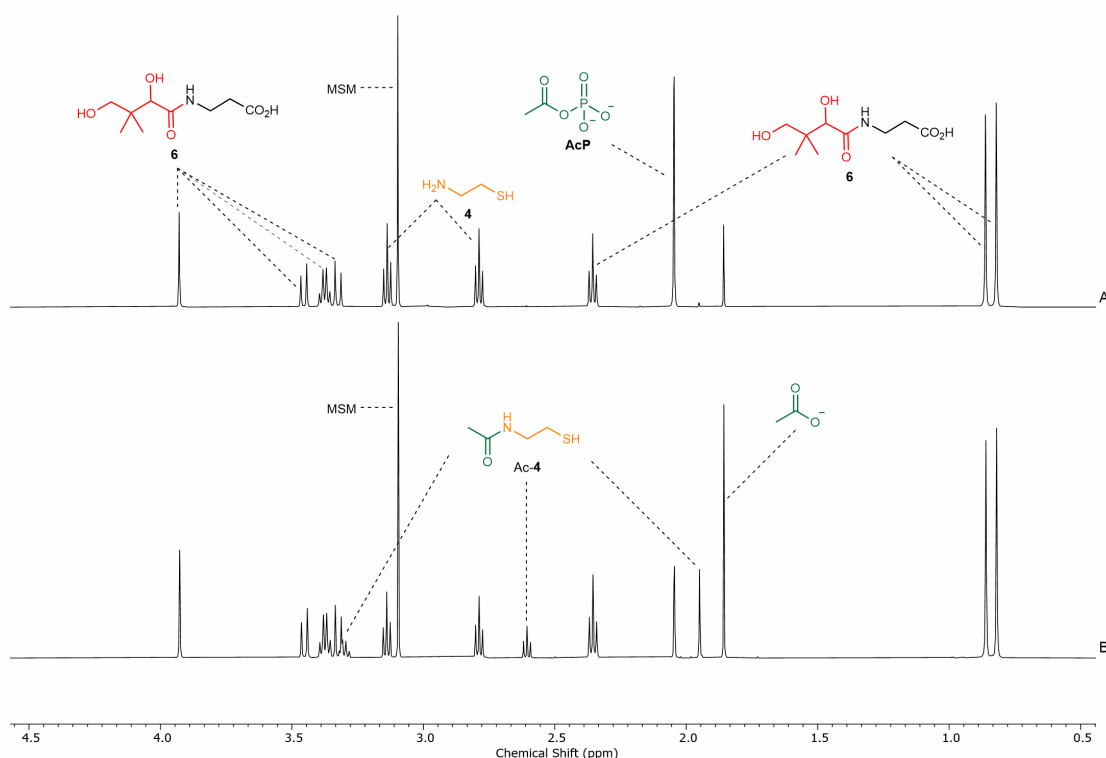
Supplementary Figure 12. ^{1}H NMR (500 MHz, H$_2$O/D$_2$O 9:1, noessygppr1d, 0.5–4.5 ppm) spectra to show the reaction of acetyl-phosphate **AcP** with pantothenic acid **6** (100 mM) and cysteamine **4** (100 mM) in phosphate buffer (pH 7, 500 mM) after: (A.) within 10 mins; and (B.) 24 hours, showing the formation of *N*-acetylcysteamine Ac-**4** (31%) from the reaction of cysteamine **4** and acetyl-phosphate **AcP**. MSM = (methanesulfonyl)methane (internal standard).

Carbodiimide-mediated cyclisation of pantoic acid 5 to pantolactone 2
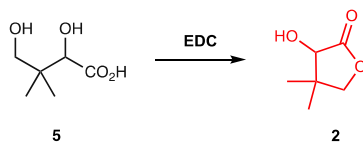
To a solution of pantoic acid **5** (100 mM) in H_2O/D_2O (99:1) at a specified pH (5.5, 7.0 or 9.0) was added *N*-(3-dimethylaminopropyl)-*N′*-ethylcarbodiimide hydrochloride (EDC·HCl; 1.2 equiv.). The solution was maintained at the required pH using HCl and NaOH (1–4 M) until no further pH changes were observed. The reactions were analysed by NMR spectroscopy (Supplementary Figure 13). After 2 days at (A.) pH 9.0, pantolactone **2** (5%) and recovery of pantoic acid **2** (95%) was observed; (B.) pH 7.0, EDC-promoted cyclisation of pantoic acid **5** to pantolactone **2** (32%) was observed; (C.) pH 5.5, EDC-promoted cyclisation of pantoic acid **5** to pantolactone **2** (90%) was observed.
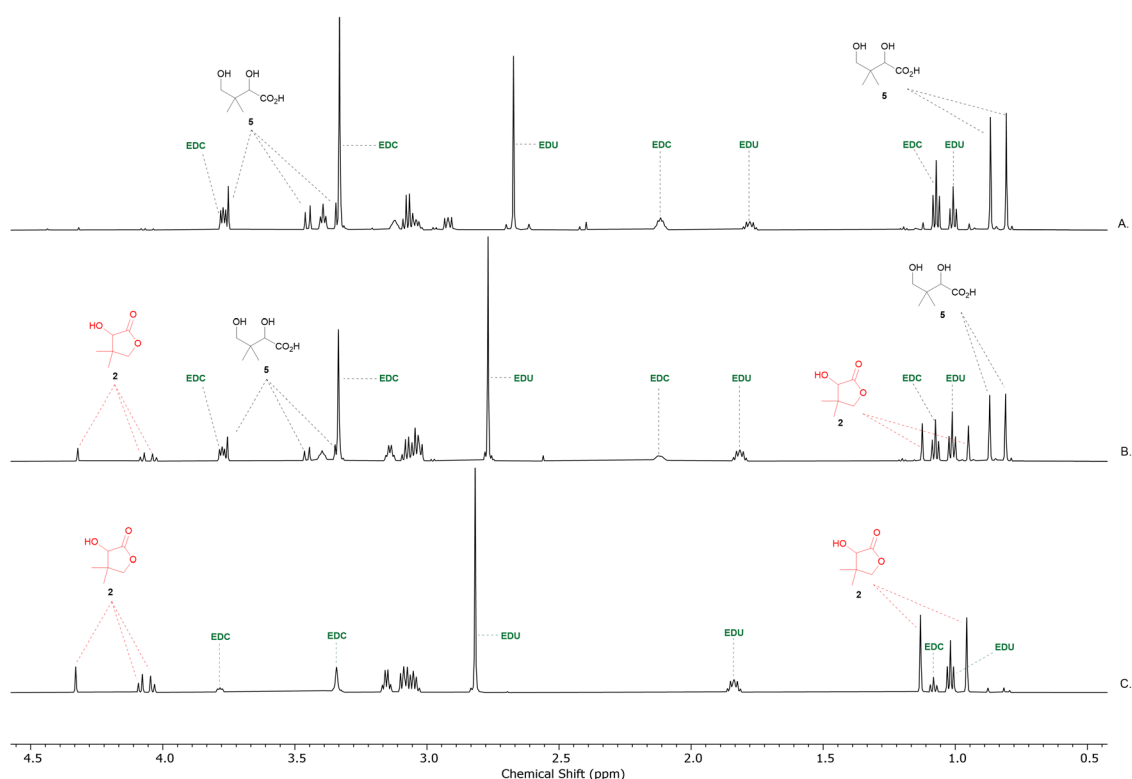
Supplementary Figure 13. ¹H NMR (600 MHz, H_2O/D_2O 99:1, noessygppr1d, 0.5–4.5 ppm) spectra to show the reaction of pantoic acid **5** (100 mM) and EDC·HCl (120 mM) after 2 days at 20 °C and: (A.) pH 9, (B.) pH 7, and (C.) pH 5.5.

*Failed synthesis of pantetheine **1** by reaction of pantothenic acid **6** with cysteamine **4** using carbodiimide activation*
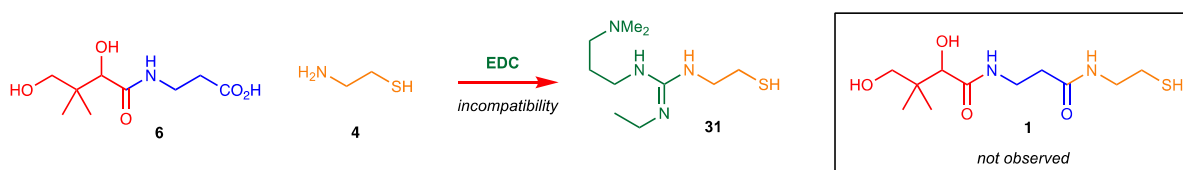
A stirred solution of pantothenic acid **6** (100 mM) and cysteamine **4** (100 mM) in water was adjusted to the specified pH (5.5, 7.0 or 9.0) using HCl and NaOH (1–4 M) at 20 ºC. *N*-(3-Dimethylaminopropyl)-*N′*-ethylcarbodiimide hydrochloride (EDC·HCl; 1 equiv.) was added and the solution was maintained at the required pH (5.5, 7.0 or 9.0) using HCl and NaOH (1–4 M) until no further pH changes were observed. The reactions were monitored by NMR spectroscopy for 24 hours. Pantothenic acid **6** remained unchanged, and the EDC was consumed by transformation into guanidine **31**. The reactions were spiked with commercial pantetheine **1**, resulting in the appearance of new resonances (Supplementary Figure 14). This confirmed that the condensation of pantothenic acid **6** with cysteamine **4** had not occurred. EDC underwent preferential reaction with cysteamine **4** to form the guanidine **31** (see Supplementary Figure 15 and Supplementary Figure 16). This demonstrates the incompatibility of carbodiimide-mediated activation of pantothenic acid **6** with a 1,2-aminothiols such as cysteamine **4**.

Supplementary Figure 14. ^{1}H NMR (700 MHz, H_2O, noessygppr1d, 0.8–4.2 ppm) spectra to show the reaction of pantothenic acid **6** (100 mM) with cysteamine **4** (100 mM) and EDC·HCl (100 mM) at pH 5.5, 7.0, and 9.0 and their subsequent spiking with authentic pantetheine **1** resulting in new resonances appearing in the NMR spectra (magenta arrows). (A.) pantothenic acid **5** (100 mM) with cysteamine **4** (100 mM) and EDC·HCl (100 mM) at pH 5.5 at 20 ºC and (B.) spiking of the crude mixture in spectrum (A.) with authentic pantetheine **1** (magenta arrows). (C.) pantothenic acid **5** (100 mM) with cysteamine **4** (100 mM) and EDC·HCl (100 mM) at pH 7.0 at 20 ºC and (D.) spiking of the crude mixture in spectrum (C.) with authentic pantetheine **1** (magenta arrows). (E.) pantothenic acid **5** (100 mM) with cysteamine **4** (100 mM) and EDC·HCl (100 mM) at pH 9.0 at 20 ºC and (F.) spiking of the crude mixture shown in spectrum (E.) with authentic pantetheine **1** (magenta arrows).

N-Guanidylation of cysteamine **4** by carbodiimide
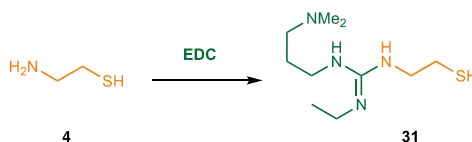
A stirred solution of cysteamine **4** (100 mM) in water was adjusted to the specified pH (7.0 or 9.0) using HCl and NaOH (1–4 M) at 20 ºC. *N*-(3-Dimethylaminopropyl)-*N′*-ethylcarbodiimide hydrochloride (EDC·HCl; ~1 equiv.) was added and the solution was maintained at the required pH (7.0 or 9.0) using HCl and NaOH (1–4 M) until no further pH changes were observed (<15 min). The reactions were analysed by NMR spectroscopy to observe quantitative yield of guanidine **31** (Supplementary Figure 15 and Supplementary Figure 16).

*Data for guanidine **31** (pH 7.0)*

^{1}H NMR (600 MHz, H_2O, pH 7.0, noesygppr1d) δ_H 3.36 (t, *J* = 6.7 Hz, 2H, NC$\underline{H}_2$CH$_2$SH), 3.27 (t, *J* = 6.8 Hz, 2H, (CH$_3$)$_2$NCH$_2$CH$_2$C$\underline{H}_2$), 3.19 (q, *J* = 7.3 Hz, 2H, CH$_3$C$\underline{H}_2$N), 3.13 – 3.08 (m, 2H, (CH$_3$)$_2$NC$\underline{H}_2$CH$_2$CH$_2$), 2.80 (s, 6H, (C$\underline{H}_3$)$_2$NCH$_2$CH$_2$CH$_2$), 2.67 (t, *J* = 6.7 Hz, 2H, NCH$_2$C$\underline{H}_2$SH), 2.00 - 1.91 (m, 2H, (CH$_3$)$_2$NCH$_2$C$\underline{H}_2$CH$_2$)), 1.12 (t, *J* = 7.3 Hz, 3H, C$\underline{H}_3$CH$_2$N). ^{13}C NMR (151 MHz, H_2O, pH 7.0, partial assignment) δ_C 155.2 (C=N), 55.8 ((CH$_3$)$_2$N$\underline{C}$H$_2$CH$_2$CH$_2$), 44.6 (($\underline{C}$H$_3$)$_2$NCH$_2$CH$_2$CH$_2$ × 2), 43.6 (N$\underline{C}$H$_2$CH$_2$SH), 38.8 ((CH$_3$)$_2$NCH$_2$CH$_2$$\underline{C}H_2$), 37.2 (CH$_3$$\underline{C}H_2$N), 24.4 ((CH$_3$)$_2NCH_2$$\underline{C}H_2CH_2$), 23.9 (NCH$_2$$\underline{C}H_2$SH), 14.1 ($\underline{C}H_3CH_2$N). HRMS-ESI [M+H]$^+$ calculated for $C_{10}H_{25}N_4S$ 233.1794; observed 233.1795.

*Data for guanidine **31** (pH 9.0)*

^{1}H NMR (600 MHz, H_2O, pH 9.0, noesygppr1d) δ_H 3.26 (t, *J* = 6.7 Hz, 2H, NC$\underline{H}_2$CH$_2$SH), 3.22 (t, *J* = 6.8 Hz, 2H, (CH$_3$)$_2$NCH$_2$CH$_2$C$\underline{H}_2$), 3.16 (q, *J* = 7.3 Hz, 2H, CH$_3$C$\underline{H}_2$N), 2.78 - 2.73 (m, 2H, (CH$_3$)$_2$NC$\underline{H}_2$CH$_2$CH$_2$), 2.60 (t, *J* = 6.7 Hz, 2H, NCH$_2$C$\underline{H}_2$SH), 2.50 (s, 6H, (C$\underline{H}_3$)$_2$NCH$_2$CH$_2$CH$_2$), 1.87 - 1.81 (m, 2H, (CH$_3$)$_2$NCH$_2$C$\underline{H}_2$CH$_2$)), 1.11 (t, *J* = 7.3 Hz, 3H, C$\underline{H}_3$CH$_2$N). ^{13}C NMR (151 MHz, H_2O, pH 9.0) δ_C 155.3 (C=N), 55.9 ((CH$_3$)$_2$N$\underline{C}$H$_2$CH$_2$CH$_2$), 46.4 (N$\underline{C}$H$_2$CH$_2$SH), 43.9 (($\underline{C}$H$_3$)$_2$NCH$_2$CH$_2$CH$_2$ × 2), 39.3 ((CH$_3$)$_2$NCH$_2$CH$_2$$\underline{C}H_2$), 37.1 (CH$_3$$\underline{C}H_2$N), 25.3 ((CH$_3$)$_2NCH_2$$\underline{C}H_2CH_2$), 24.6 (NCH$_2C\underline{H}_2$SH), 14.2 ($\underline{C}H_3CH_2$N).
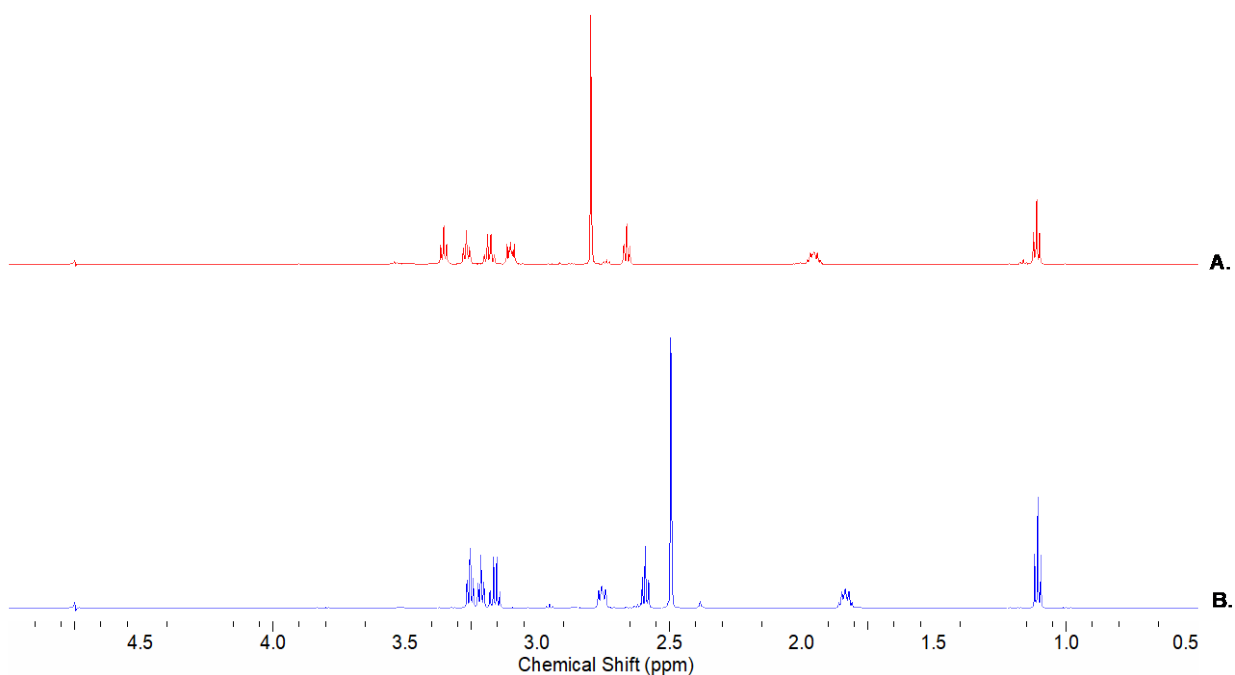
Supplementary Figure 15. ^{1}H NMR (700 MHz, H$_2$O, noessygppr1d, 0.5–5.0 ppm) spectra to show the reaction of cysteamine **4** (100 mM) and EDC·HCl (100 mM) at 20 ºC and: (A.) pH 7.0; (B.) pH 9.0.
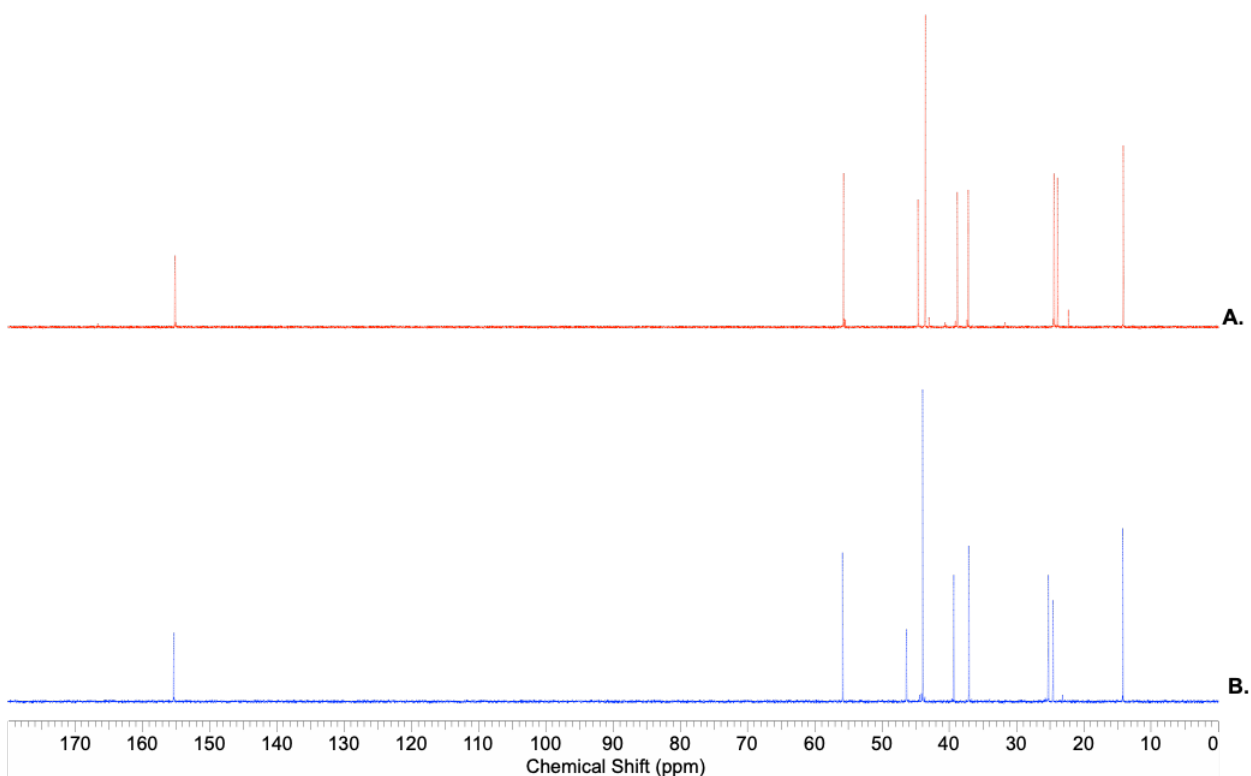
Supplementary Figure 16. ^{13}C NMR (176 MHz, H$_2$O, noessygppr1d, 0–180 ppm) spectra to show the reaction of cysteamine **4** (100 mM) and EDC·HCl (100 mM) at 20 ºC and: (A.) pH 7.0; (B.) pH 9.0.

*Fragmentation of pantothenic acid **6** by carbodiimide activation*

*Pantothenic acid nitrile **9** is resistant to carbodiimide-mediate fragmentation*
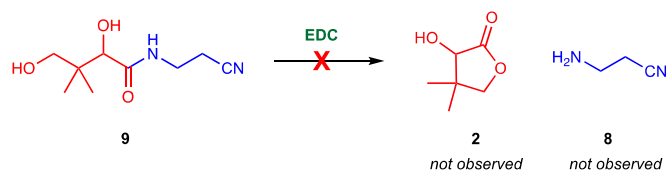
Pantothenic acid nitrile **9** (100 mM) and *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC·HCl; 1 equiv.) were dissolved in water. The solution pH was adjusted to pH 5.5 using HCl/NaOH (1–4 M). The reaction was monitored at 20 ºC by NMR spectroscopy for 12 hours. Quantitative recovery (>95%) of pantothenic acid nitrile **9** was observed. Pantolactone **2** and β-alanine nitrile **8** were not detected (Supplementary Figure 17)
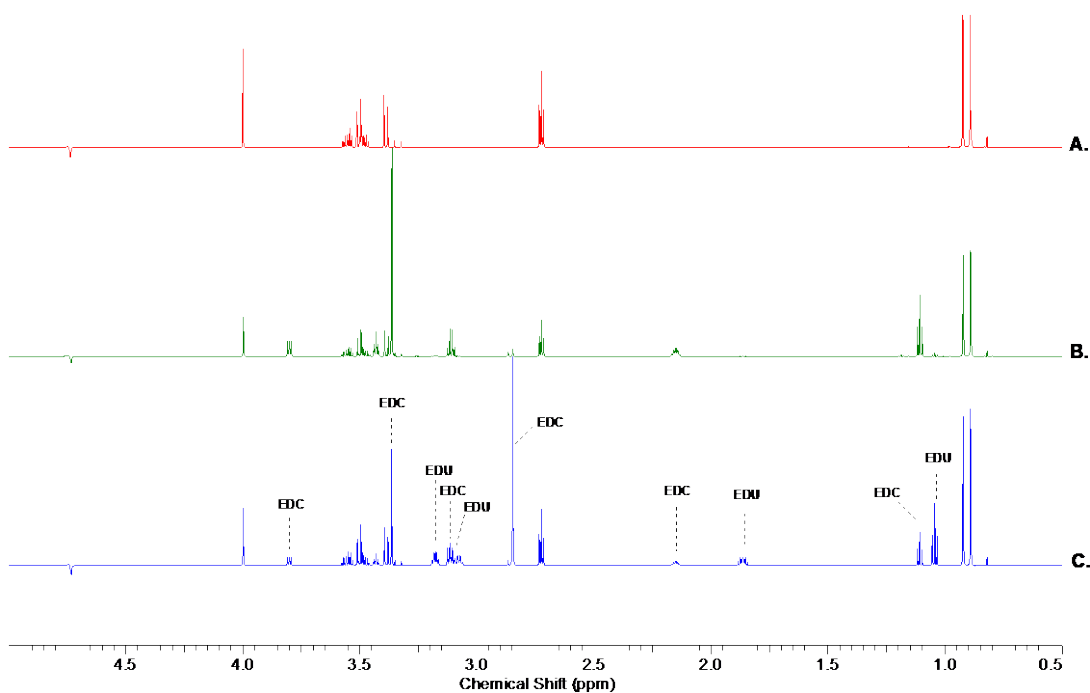
Supplementary Figure 17. ^{1}H NMR (700 MHz, H$_2$O, noessygppr1d, 0.5–5.0 ppm) spectra to show the reaction of pantothenic acid nitrile **9** (100 mM) and EDC·HCl (100 mM) at 20 ºC and pH 5.5. (A.) authentic pantothenic acid nitrile **9**, (B.) the reaction of pantothenic acid nitrile **9** (100 mM) and EDC·HCl (100 mM) within 20 mins of EDC addition, and (C.) after 12 hours. Pantothenic acid nitrile **9** remained unchanged.

Carbodiimide-mediated decomposition of pantothenic acid 6
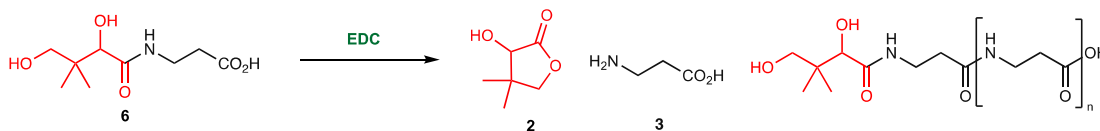
Pantothenic acid **6** (100 mM) and *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC·HCl; 1 equiv.) were dissolved in water. The solution was adjusted to pH 5.5, 7.0 or 9.0 using HCl/NaOH (1–4 M), and the reactions were monitored and analysed at 20 ºC by NMR spectroscopy. Quantitative recovery (>95%) of pantothenic acid **6** at pH 7.0 and 9.0 was observed alongside EDC hydrolysis to 1-(3-(dimethylamino)propyl)-3-ethylurea (EDU) after 24 hours. At pH 5.5 the fragmentation of pantothenic acid **6** to pantolactone **2** (31%) was observed, alongside β-alanine **3** (24%) and residual pantothenic acid **6** (49%) after 4 hours. New triplets at 2.44 - 2.50 and 2.34 - 2.37 ppm in the ¹H NMR spectrum (26% of total (C2)–H_2 of expected β-alanyl-derived species) indicated that 20% of pantothenic acid **6** underwent EDC-mediated side reactions with fragmentation product β-alanine **3** to form oligomeric products. These oligomeric byproducts were not further characterised (see Supplementary Figure 18). The proposed mechanism of **6** fragmentation to **2** and **3** is shown in Supplementary Figure 19.
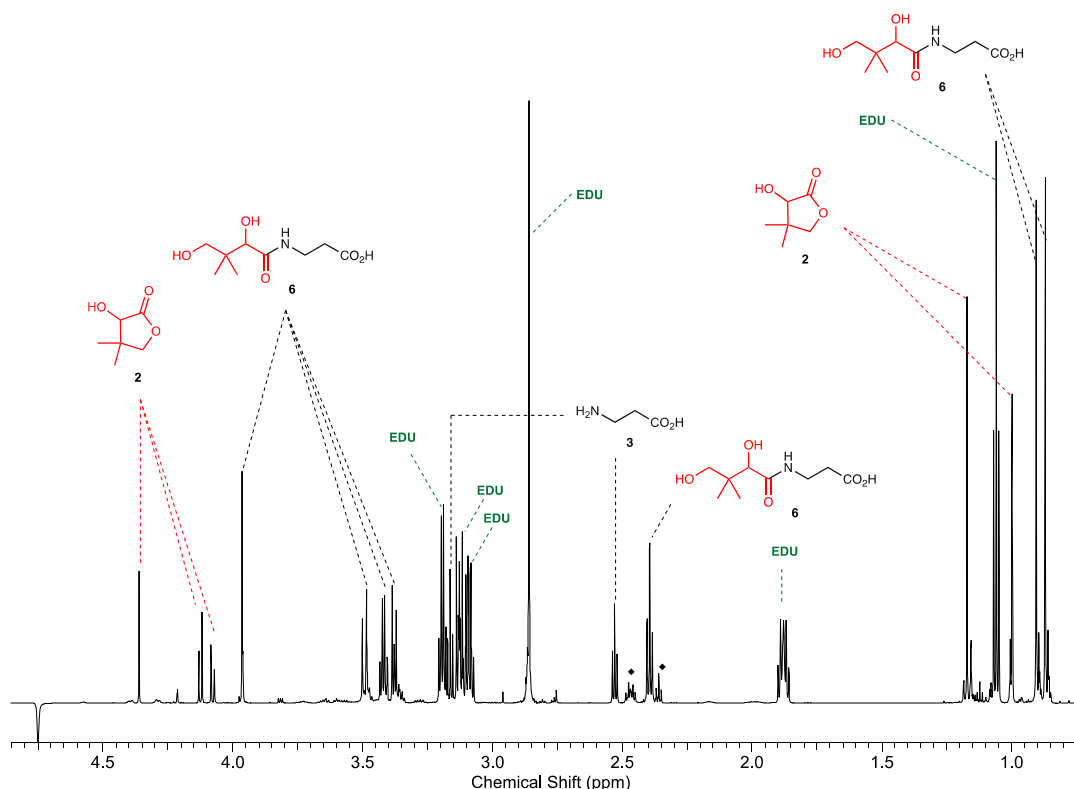
Supplementary Figure 18. ¹H NMR (700 MHz, H_2O, noessygppr1d, 0.5–5.0 ppm) spectrum to show the reaction of pantothenic acid **5** (100 mM) and EDC·HCl (100 mM) at 20 ºC and pH 5.5 after 4 hours. ♦ = overlapping triplet (C2)–H_2 resonances of β-alanyl byproducts.
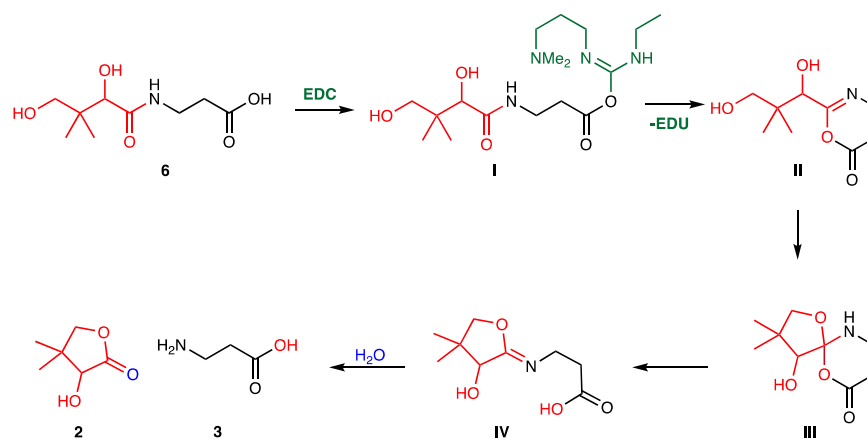
Supplementary Figure 19. Proposed mechanism for the observed EDC-mediated fragmentation of pantothenic acid **6** to pantolactone **2** and β-alanine **3**.

Selective carbodiimide-mediated fragmentation of pantothenic acid 6 in the presence of pantothenic acid nitrile 9
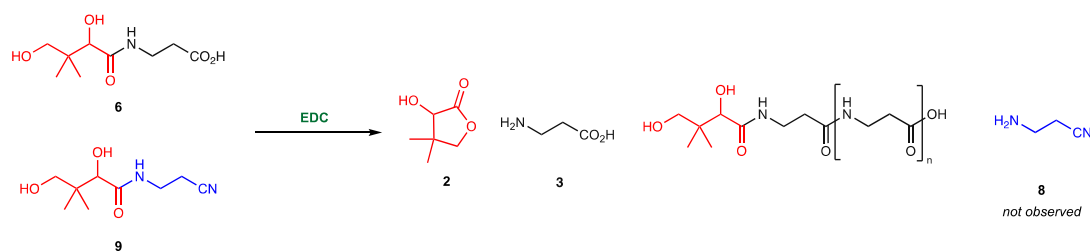
Pantothenic acid **6** (100 mM) and pantothenic acid nitrile **9** (70 mM) were dissolved in water and the solution was adjusted to pH 5.5 using 1–4 M HCl. *N*-(3-Dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC·HCl; 100 mM) was added and the solution was maintained at pH 5.5 using dilute HCl (<10 min). The reaction was kept at 20 ºC and analysed by NMR spectroscopy. Pantothenic acid nitrile **9** remained intact and stable to EDC-mediated fragmentation, whereas the formation of pantolactone **2** (38%) at the expense of pantothenic acid **6** was observed after 12 hours alongside new triplet resonances at 2.39 and 2.47 ppm, suggesting that the β-alanine **3** generated from pantothenic acid **6** fragmentation underwent further EDC-mediated side reactions with fragmentation product β-alanine **3** to form oligomeric products. These oligomeric byproducts were not further characterised. The EDC by-products were also observed in the reaction of pantothenic acid **6** and EDC (Supplementary Figure 18).

Supplementary Figure 20. ^{1}H NMR (700 MHz, H$_2$O, noessygppr1d, 0.5–5.0 ppm) spectra to show the selective EDC-mediated (100 mM) fragmentation of pantothenic acid **6** (100 mM) in the presence of pantothenic acid nitrile **9** (70 mM) at pH 5.5 and 20 ºC. (A.) before the addition of EDC, and (B.) 12 hours after the addition of EDC, showing substantial depletion in the characteristic resonances corresponding to pantothenic acid **6**. ♦ = triplet (C2)–H$_2$ resonances of β-alanyl byproducts.

*Attempted synthesis of pantothenic acid **6** by reaction of pantoic acid **5** and β-alanine **3** using carbodiimide activation*
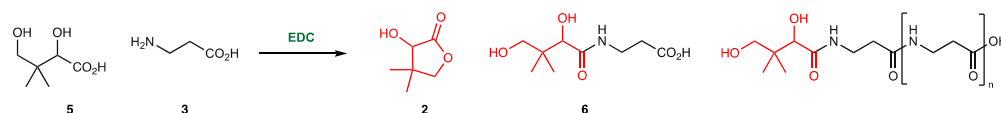
A solution of pantoic acid **5** (100 mM) and β-alanine **3** (1.1 equiv.) in water was adjusted to the specified pH (5.5, 7.0 or 9.0) using HCl and NaOH (1-4 M). *N*-(3-Dimethylaminopropyl)-*N′*-ethylcarbodiimide hydrochloride (EDC·HCl; ~1.1 equiv.) was added and the solution pH was maintained at the required pH (5.5, 7.0 or 9.0) using HCl and NaOH (1-4 M) until no further pH changes were observed. The reactions were monitored by NMR spectroscopy for 24 hours. At pH 9 recovery of pantoic acid **5** (94%) and β-alanine **3** (82%) was observed, but pantothenic acid **6** was not detected. At pH 7.0, β-alanine **3** (90%) was recovered, alongside EDC-mediated cyclisation of pantoic acid **5** to pantolactone **2** (32%), and the EDC-mediated reaction of pantoic acid **5** and β-alanine **3** to pantothenic acid **6** (4%). At pH 5.5, β-alanine **3** (75%) was recovered, alongside EDC-mediated cyclisation of pantoic acid **5** to pantolactone **2** (78%), and the EDC mediated coupling of pantoic acid **5** and β-alanine **3** to pantothenic acid **6** (7%). At pH 5.5 additional new triplet resonances (~13%; between 2.34 and 2.59 ppm), suggesting that β-alanine **3** underwent further EDC-mediated side reactions, including reaction with pantothenic acid **6** to form oligomeric products. These oligomeric byproducts were not further characterised.
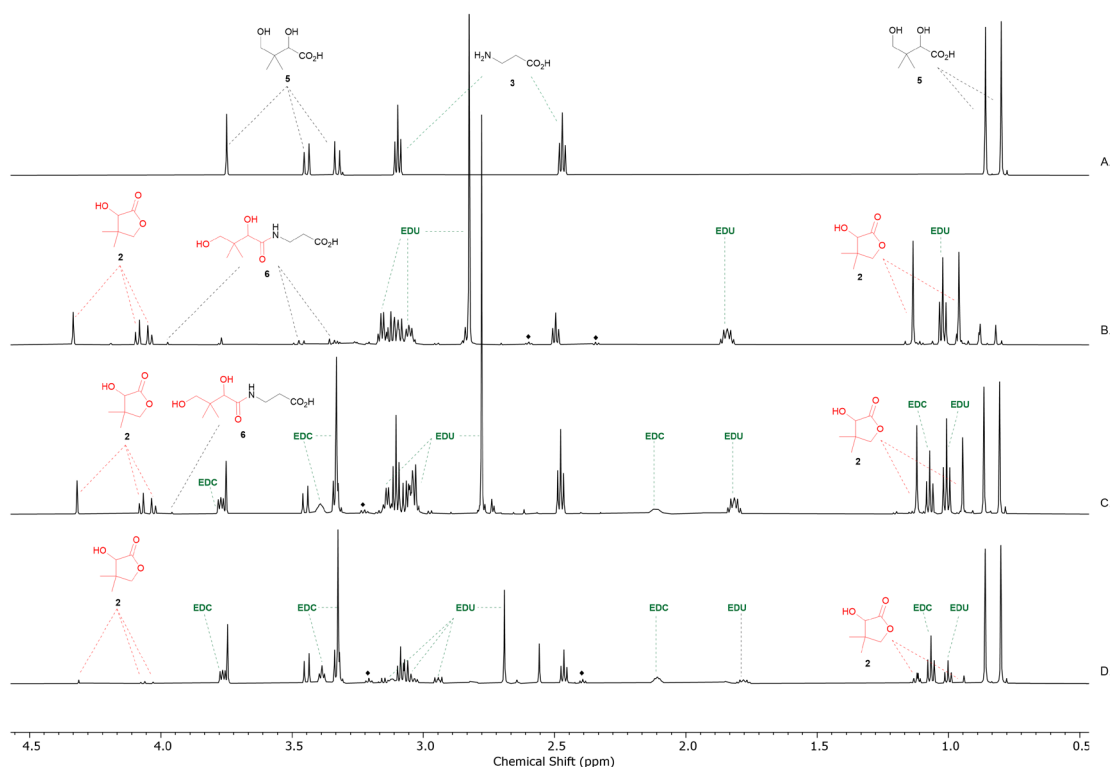
Supplementary Figure 21. ^{1}H NMR (600 MHz, H$_2$O/D$_2$O 99:1, noessygppr1d, 0.5–4.5 ppm) spectrum to show the reaction of pantoic acid **5** (100 mM), β-alanine **3** (100 mM), and EDC·HCl (110 mM) at pH 5.5, 7, and 9 and 20 ºC. (A.) The starting reagents at pH 7, before addition of EDC·HCl. (B.) The reaction at pH 5.5 after 24 hours. (C.) The reaction at pH 7 after 24 hours. (D.) The reaction at pH 9 after 24 hours. ♦ = triplet (C2)–H$_2$ resonances of β-alanyl byproducts.

*Failed one-pot synthesis of pantetheine **1** through the reaction of pantoic acid **5**, β-alanine **3** and cysteamine **4** using carbodiimide activation*
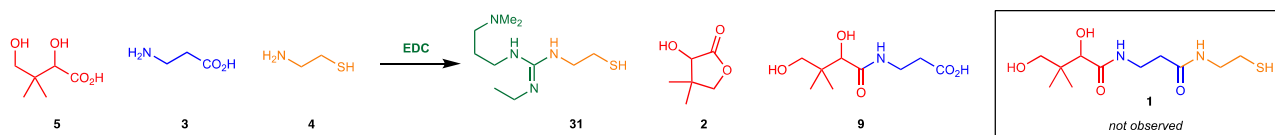
To the solution of pantoic acid **5** (100 mM) at pH 7 was added β-alanine **3** (1.0 equiv.) and cysteamine **4** (1.0 equiv.). The solution was adjusted to the specified pH (5.5, 7.0 or 9.0) using HCl and NaOH (1-4 M). *N*-(3-Dimethylaminopropyl)-*N*′-ethylcarbodiimide hydrochloride (EDC·HCl; ~1.1 equiv.) was added and the solution was maintained at the required pH (5.5, 7.0 or 9.0) using HCl and NaOH (1-4 M) until no further pH changes were observed. The reaction was kept at 20 ºC and analysed by ¹H NMR spectroscopy. At pH 5.5, cysteamine **4** underwent *N*-guanidylation by EDC to give **31** (87%), and quantitative recovery of pantoic acid **5** and β-alanine **3** was observed after 24 hours. A second portion of EDC·HCl (~1.1 equiv., ~2.2 equiv. total) to the reaction mixture led to quantitative (>95%) conversion of cysteamine **4** to guanidine **31**, and the formation of pantolactone **2** (35%) from pantoic acid **5**. Pantothenic acid **6** (<6%) was observed, but pantetheine **1** was not detected after a further 24 hours (Supplementary Figure 22). At pH 7.0, after two additions of EDC·HCl (1.1 equiv. × 2), cysteamine **4** was quantitatively converted to guanidine **31** (>95%), pantoic acid **5** cyclised to pantolactone **2** (14%), but pantothenic acid **6** and pantetheine **1** were not detected. At pH 9.0, after two additions of EDC·HCl (1.1 equiv × 2), quantitative conversion of cysteamine **4** to guanidine **31** (>95%) was observed, and pantoic acid **5** and β-alanine **3** were quantitatively recovered.

These reactions demonstrate that the condensation of pantoic acid **5,** β-alanine **3** and cysteamine **4** to pantetheine **1** had not occurred through EDC-mediated reaction. EDC underwent preferential reaction with cysteamine **4** to form the guanidine **31** at pH 5.5, 7.0 and 9.0 (see Supplementary Figure 15–Supplementary Figure 16). Furthermore, EDC-mediated cyclisation of pantoic acid **5** to pantolactone **2** is preferential over the condensation of pantoic acid **5** with β-alanine **3**. These experiments confirm the incompatibility of carbodiimide-mediated activation of pantothenic acid **6**, pantoic acid **5** and β-alanine **3** in the presence of a 1,2-aminothiol such as cysteamine **4** to form pantetheine **1** in one pot.
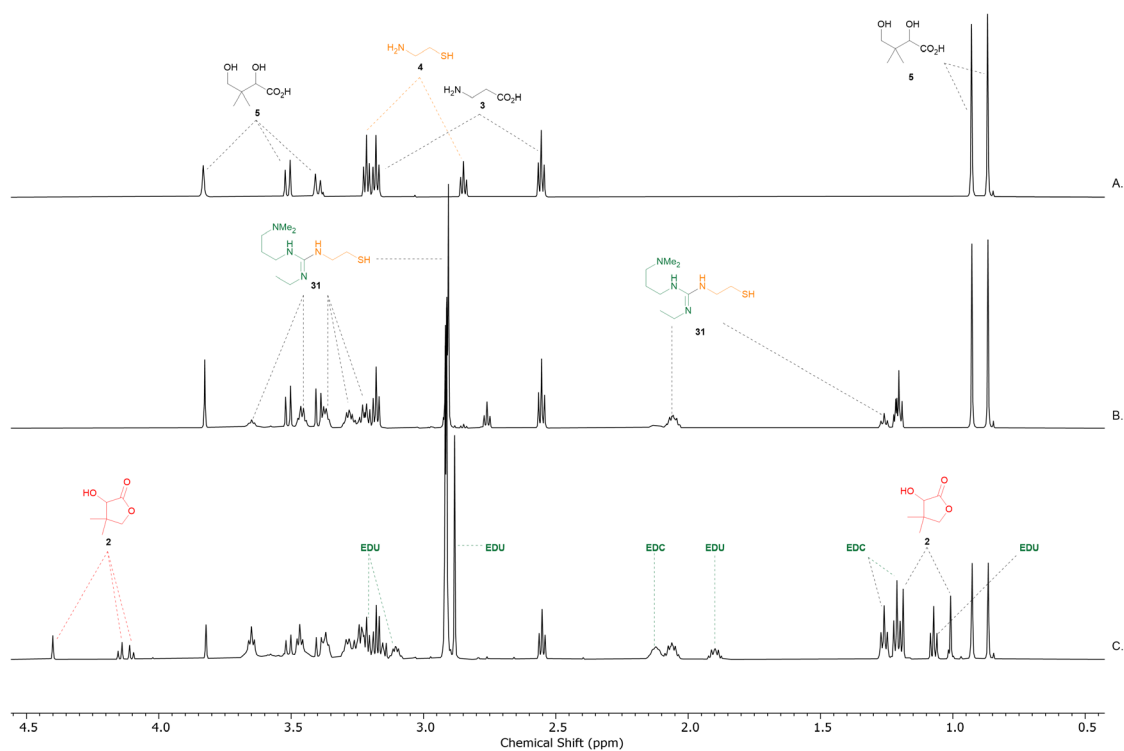
Supplementary Figure 22. ^{1}H NMR (700 MHz, H$_2$O/D$_2$O 99:1, noessygppr1d, 0.5–4.5 ppm) spectra to show the attempted one-pot condensation of pantoic acid **5** (100 mM), β-alanine **3** (100 mM) and cysteamine **4** (100 mM) at pH 5.5 and 20 ºC. (A.) before the addition of EDC, (B.) 24 hours after the addition of EDC (110 mM) and (C.) 24 hours after the addition of a second portion of EDC (110 mM, total 220 mM).

*Carbodiimide-mediated synthesis of pantothenic acid nitrile **9** by amine exchange of pantothenic acid **6***
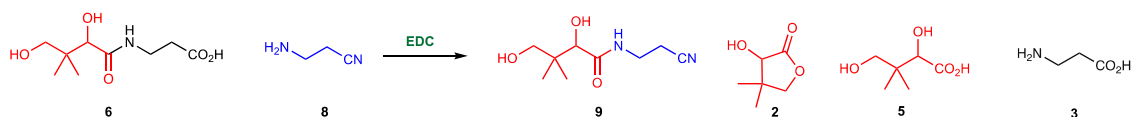
Pantothenic acid **6** (100 mM) and β-alanine nitrile **8** (100 mM) were dissolved in H_2O/D_2O (99:1) and the solution was adjusted to the specified pH (pH 5.5, 7.0 or 9.0) using HCl/NaOH (1-4 M). *N*-(3-Dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC·HCl; ~1 equiv.) was added, and the solution was maintained at the required pH (5.5, 7.0 or 9.0) using HCl and NaOH (1-4 M) until no further pH changes were observed. The reaction was kept at 20 ºC and analysed by NMR spectroscopy. At pH 5.5, the formation of pantothenic acid nitrile **9** (39%) was observed by transamidation of **6** by β-alanine nitrile **8** after 30 minutes. The liberated β-alanine **3** (35%) was also observed alongside new triplet resonance (~31%; at 2.40, 2.50, 3.76 and 3.83 ppm), indicating that the β-alanine **3** generated from pantothenic acid **6** fragmentation underwent further EDC-mediated side reactions. These β-alanyl byproducts were also observed in the decomposition of pantothenic acid **6** by EDC (see Supplementary Figure 18). At pH 7.0, 85% of pantothenic acid **6** was recovered, alongside the formation of pantothenic acid nitrile **9** (15%) and β-alanine **3** (15%), but pantolactone **2** was not detected. At pH 9.0, quantitative recovery of pantothenic acid **6** (>95%) was observed.

Supplementary Figure 23. ^1H NMR (700 MHz, H_2O/D_2O 99:1, noessygppr1d, 0.5–5.0 ppm) spectra to show the EDC-mediated (100 mM) fragmentation of pantothenic acid **6** (100 mM) in the presence of β-alanine nitrile **8** (90 mM) at pH 5.5 and 20 ºC. (A.) before the addition of EDC, and (B.) 30 min after the addition of EDC, showing depletion in the characteristic resonances corresponding to pantothenic acid **6** and the formation of pantothenic acid nitrile **9**. ♦ = triplet (C2)–H_2 resonances of β-alanyl byproducts. * = transient intermediate observed upon activation of **6** with EDC.

Carbodiimide-mediated fragmentation of pantothenic acid 6 and oligomerisation side reactions in the presence of β-alanine 3
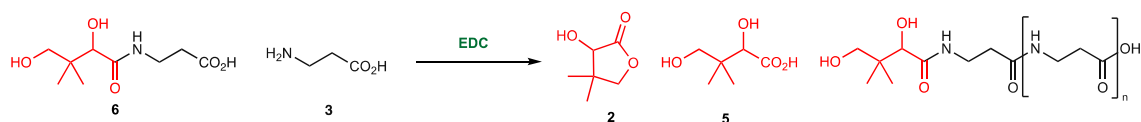
Pantothenic acid **6** (100 mM) and β-alanine **3** (100 mM) were dissolved in 99:1 H_2O/D_2O and the pH was adjusted to the specified pH (5.5, 7.0 or 9.0) using HCl and NaOH (1-4 M). *N*-(3-dimethylaminopropyl)-*N′*-ethylcarbodiimide hydrochloride (EDC·HCl; 1.1 equiv.) was added and the solution pH was maintained at the required pH (5.5, 7.0 or 9.0) using HCl and NaOH (1-4 M) until no further pH changes were observed. The reaction was kept at 20 °C and immediately analysed by NMR spectroscopy. At pH 5.5 pantothenic acid **6** (44%, recovered) underwent EDC-mediated fragmentation resulting in the formation of pantolactone **2** (12%) at the expense of pantothenic acid **6** after 30 minutes (Supplementary Figure 24). The low recovery of pantothenic acid **6** (expected to be 88%) was due to additional reactions of pantothenic acid **6** with β-alanine **3**, which was confirmed by the depletion of β-alanine **3** (72% recovery; expected 100–156 mol%). The marked depletion of β-alanine **3** and the observation of new unassigned triplet resonances (~34%) at 2.31, 2.44, 2.53 ppm, suggested that the β-alanine **3** generated from pantothenic acid **6** fragmentation underwent further EDC-mediated side reactions. At pH 7.0, the majority of pantothenic acid **6** (93%, recovered) remained unreacted, and we observed quantitative recovery of β-alanine **3**. (1 equiv.). The characteristic peaks of pantolactone **2** and pantoic acid **5** were not detected. At pH 9.0, quantitative recovery of pantothenic acid **6** (>95%) was observed.

Supplementary Figure 24. ^{1}H NMR (700 MHz, H_2O/D_2O 99:1, noessygppr1d, 0.5–5.0 ppm) spectra to show the EDC-mediated fragmentation of pantothenic acid **6** (100 mM) in the presence of β-alanine **3** (100 mM) at pH 5.5 and 20 °C. (A.) before the addition of EDC, and (B.) 30 minutes after the addition of EDC, showing depletion in the characteristic resonances corresponding to pantothenic acid **6** and the formation of pantoic acid **5**. ♦ = triplet (C2)–H$_2$ resonances of β-alanyl by-products. * = Transient intermediate observed upon activation of **6** with EDC.

## 2. Syntheses of pantetheine 1 from pantolactone 2 using nitrile activation

*Reactions of pantolactone **2** and β-alanine nitrile **3***
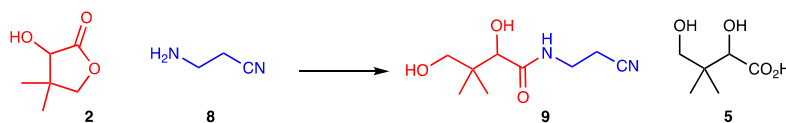
Pantolactone **2** (3.1–500 mM) and β-alanine nitrile **8** (1–5 equiv.) were dissolved in phosphate buffer at pH 7.0 (1–10 equiv.) and 20 ºC. The solution pH was adjusted to the specified pH with 1-4 M NaOH. The resulting mixture was periodically analysed by NMR spectroscopy and its pH monitored. Yield of pantothenic acid nitrile **9** are reported in Supplementary Table 4. Representative ^{1}H NMR spectra are shown in Supplementary Figure 25–Supplementary Figure 27.

Concentration (mM)		P_i (equiv.)	pH	Time (days)	Yield (%)		
2	**8**				**2**	**9**	**5**
3.1	6.3	10	9.0	6	30	2	68
50	100	10	7.0	6	60	2	38
50	100	10	8.0	12	15	13	72
50	100	10	9.0	5	1	37	62
100	300	5	7.0	7	66	9	25
100	500	5	7.0	7	57	19	24
100	100	5	9.0	7	20	22	58
100	200	5	9.0	7	0	50	50
100	300	5	9.0	5	1	63	36
100	500	5	9.0	2	0	77	23
500	500	1	7.0	10	76	2	22
500	2500	1	7.0	10	59	12	28
500	500	2	9.0	5	16	48	36
500	500	1	9.0	7	16	54	30
500	1000	1	9.0	3	0	83	17
500	1500	1	9.0	1	0	91	9
500	2500	1	9.0	0.5	0	94	6

Supplementary Table 4. Quantification of species observed from the reaction of pantolactone **2** (3.1–500 mM) with β-alanine nitrile **8** (1–5 equiv.) at the specified pH in phosphate buffer (1–10 equiv.) at 20 ºC.
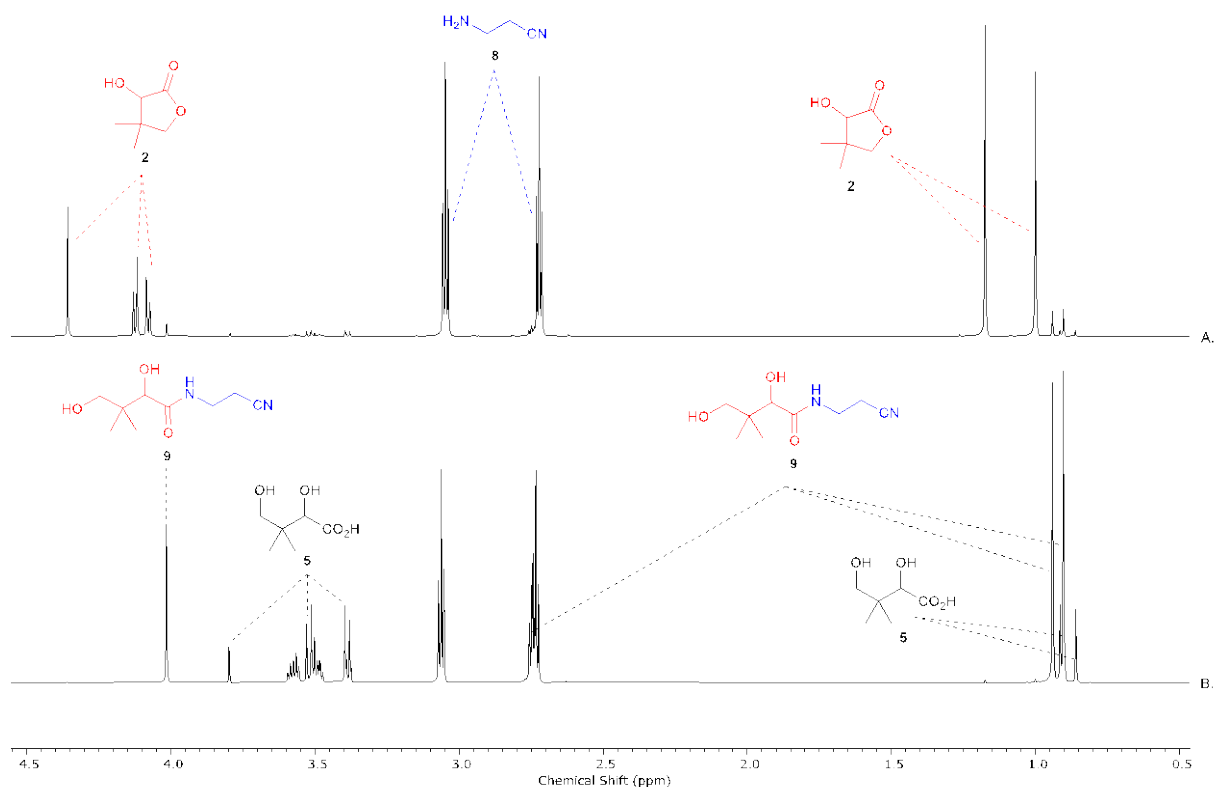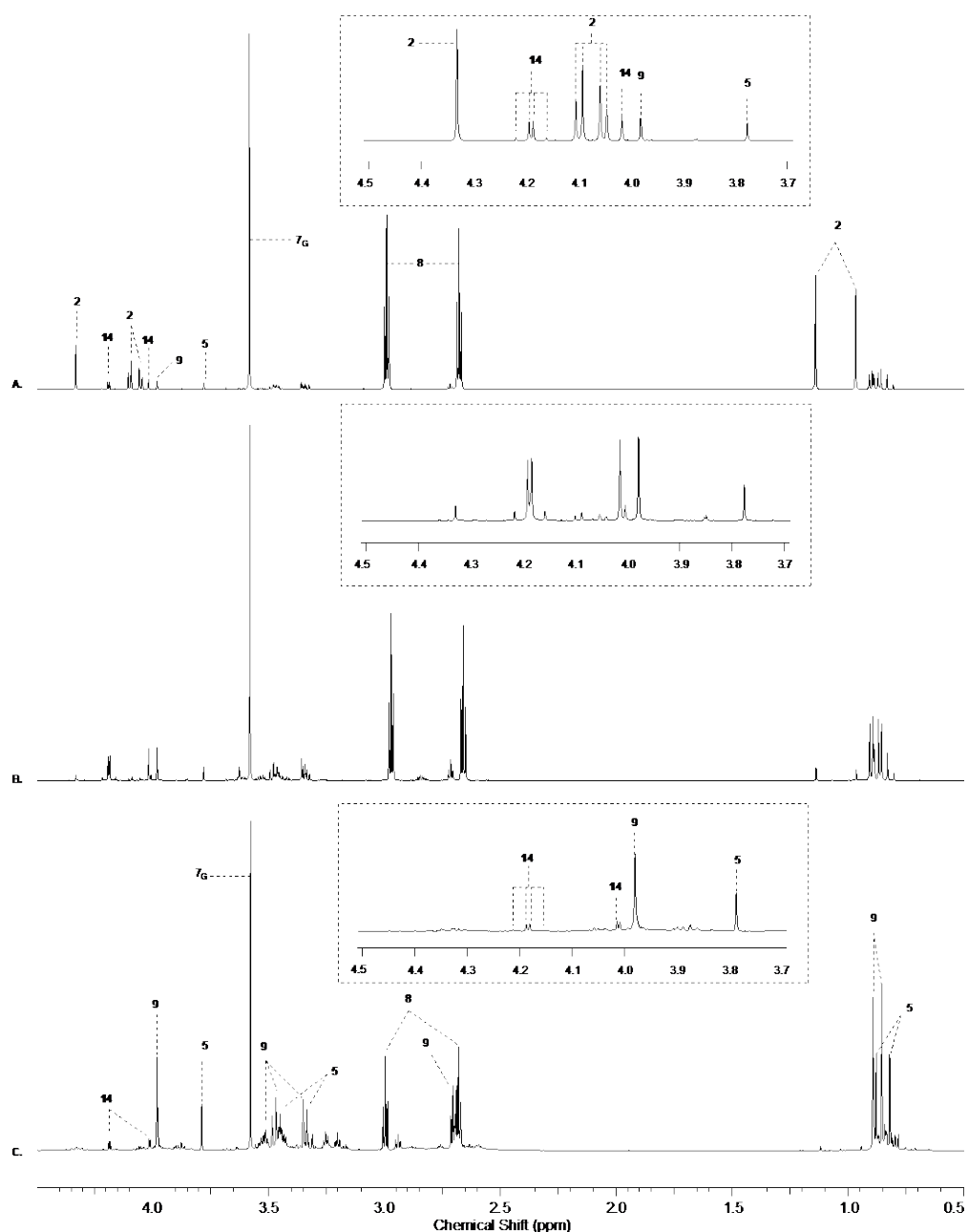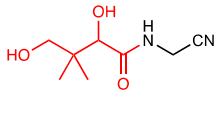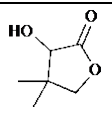
Supplementary Figure 25. ¹H NMR (600 MHz, 9:1 H_2O/D_2O, noesygppr1d, 0.5–4.5 ppm) spectra to show the reaction of pantolactone **2** (500 mM) with β-alanine nitrile **8** (2 equiv.) in phosphate buffer (pH 9.0; 500 mM) at 20 ºC after; (A.) 10 min; (B.) 3 days, showing the formation of pantothenic acid nitrile **9** (83%) and pantoic acid **5** (17%).
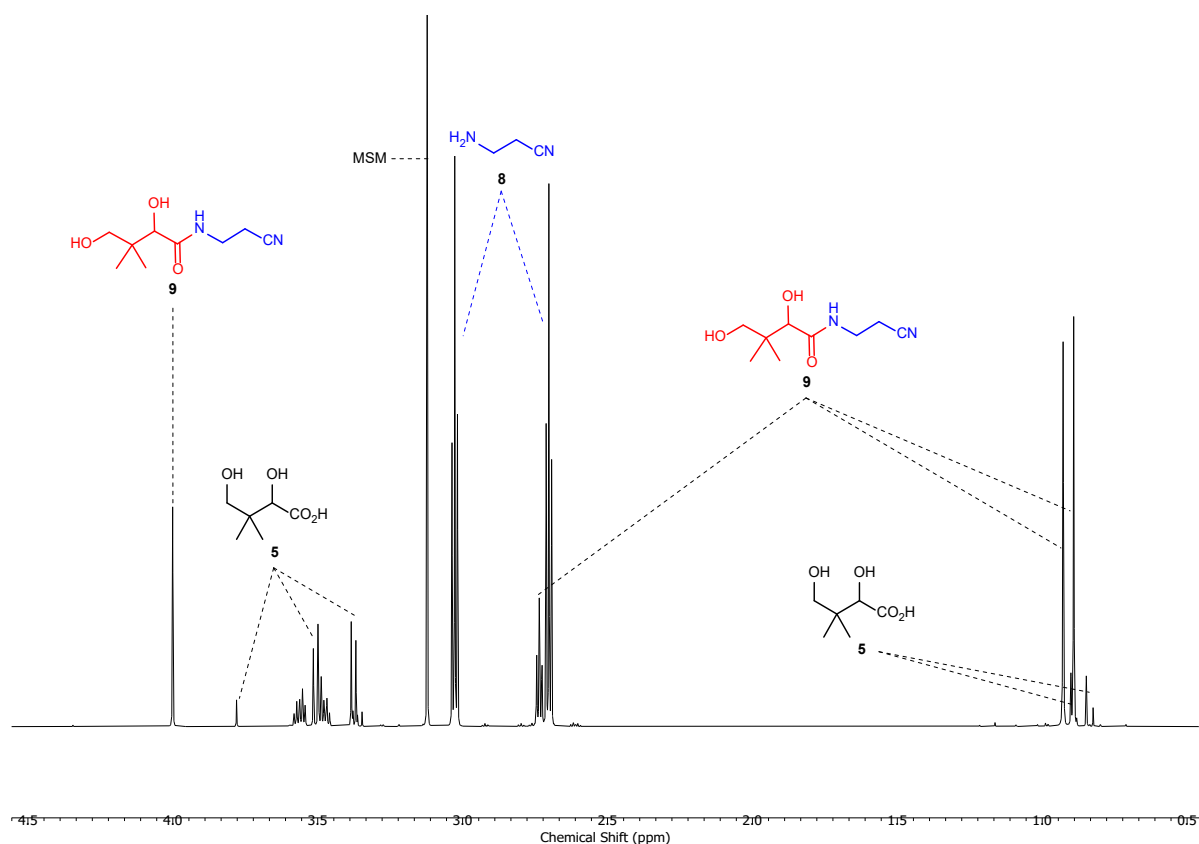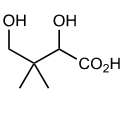

Supplementary Figure 26. ¹H NMR (600 MHz, 9:1 H_2O/D_2O, noesygppr1d, (0.5–4.5 ppm) spectrum to show the reaction of pantolactone **2** (500 mM) with β-alanine nitrile **8** (3 equiv.) in phosphate buffer (pH 9.0; 500 mM) at 20 ºC after 1 day, showing the formation of pantothenic acid nitrile **9** (91%) and pantoic acid **5** (9%). MSM = (methanesulfonyl)methane (internal standard)).
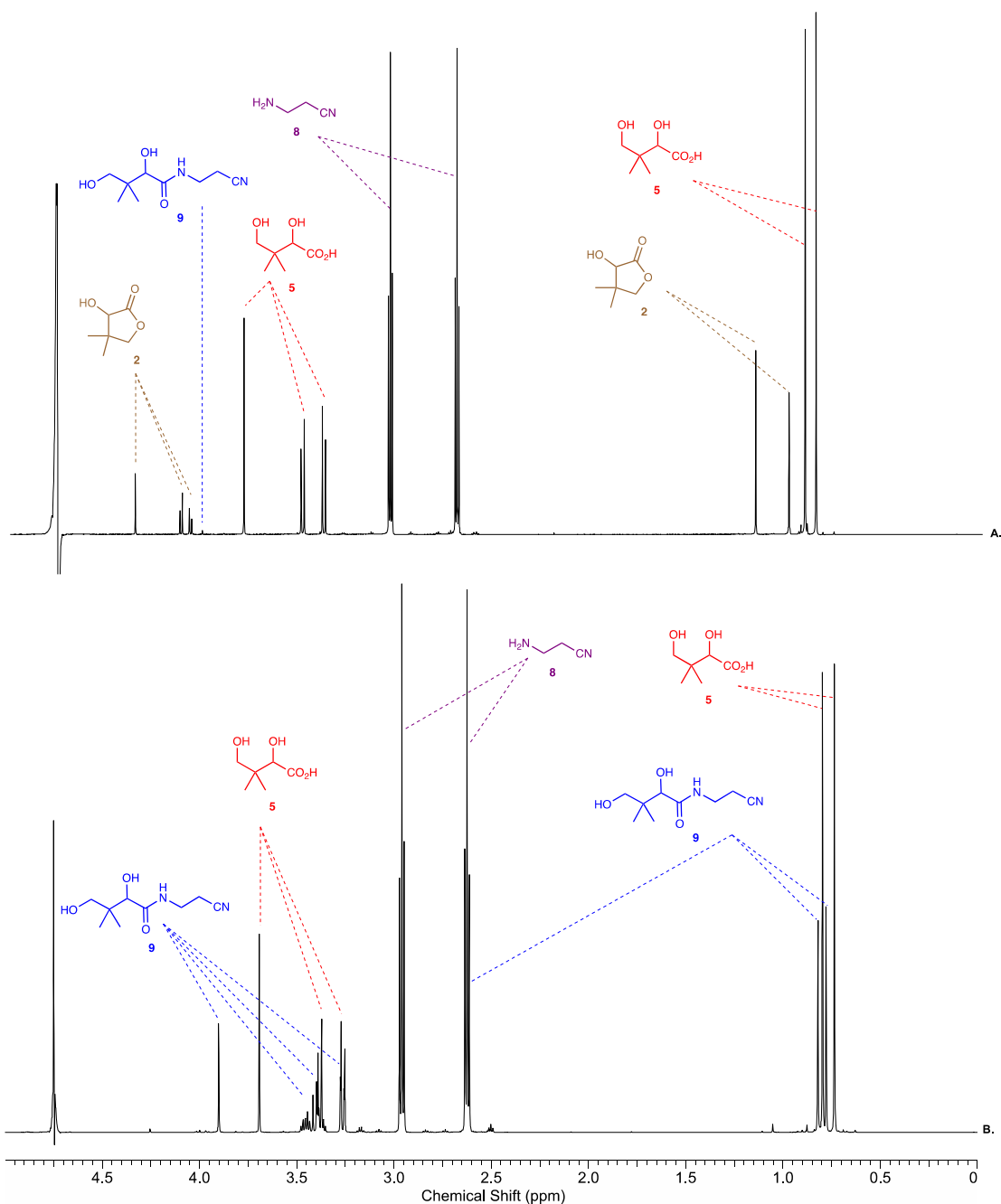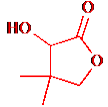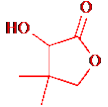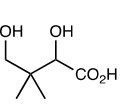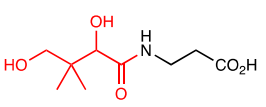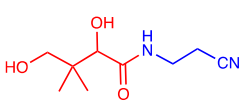
Supplementary Figure 27. ^{1}H NMR (H$_2$O, noesygppr1d, 0–5 ppm) spectra to show the reaction of (A., 700 MHz) pantolactone **2** (3.13 mM) with β-alanine nitrile **8** (6.25 mM) in phosphate buffer (pH 9.0; 31.3 mM) at 20 ºC after 6 days, and (B. 600 MHz) pantolactone **2** (50 mM) with β-alanine nitrile **8** (2 equiv.) in phosphate buffer (pH 9.0; 500 mM) at 20 ºC after 5 days.

*Competitive reaction of pantolactone **2** with β-alanine **3** and β-alanine nitrile **9***
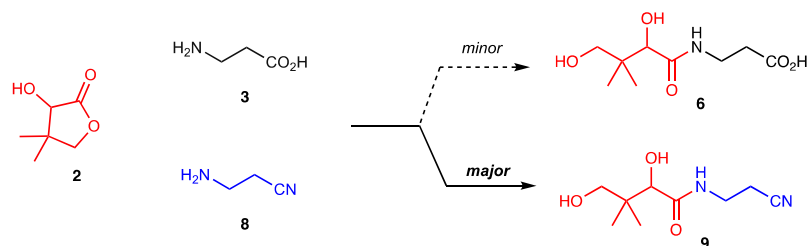
Pantolactone **2** (100 or 500 mM), β-alanine **3** (2 equiv.) and β-alanine nitrile **8** (2 equiv.) were dissolved in phosphate buffer (pH 7.0; 500 mM) at 20 ºC. The solution pH was adjusted to the required pH with 1-4 M NaOH. The resulting mixture was periodically analysed by NMR spectroscopy and its pH monitored. Selective formation of pantothenic acid nitrile **9** over pantothenic acid **6** was always observed. Yields are reported in Supplementary Table 5. ^{1}H NMR spectra are shown in Supplementary Figure 28 and Supplementary Figure 29

Concentration (mM)	pH	Time (days)	Yield (%)			
2			**2**	**5**	**6**	**9**
100	7.0	3	83	17	0	0
100	8.0	4	50	45	0	5
100	9.0	4	5	49	1	46
500	9.0	2	tr.	11	3	84

Supplementary Table 5. Quantification of species observed from the competitive reaction of pantolactone **2** (100–500 mM) with β-alanine **3** (2 equiv.) and β-alanine nitrile **8** (2 equiv.) at the specified pH in phosphate buffer (500 mM) at 20 ºC. tr. = trace.
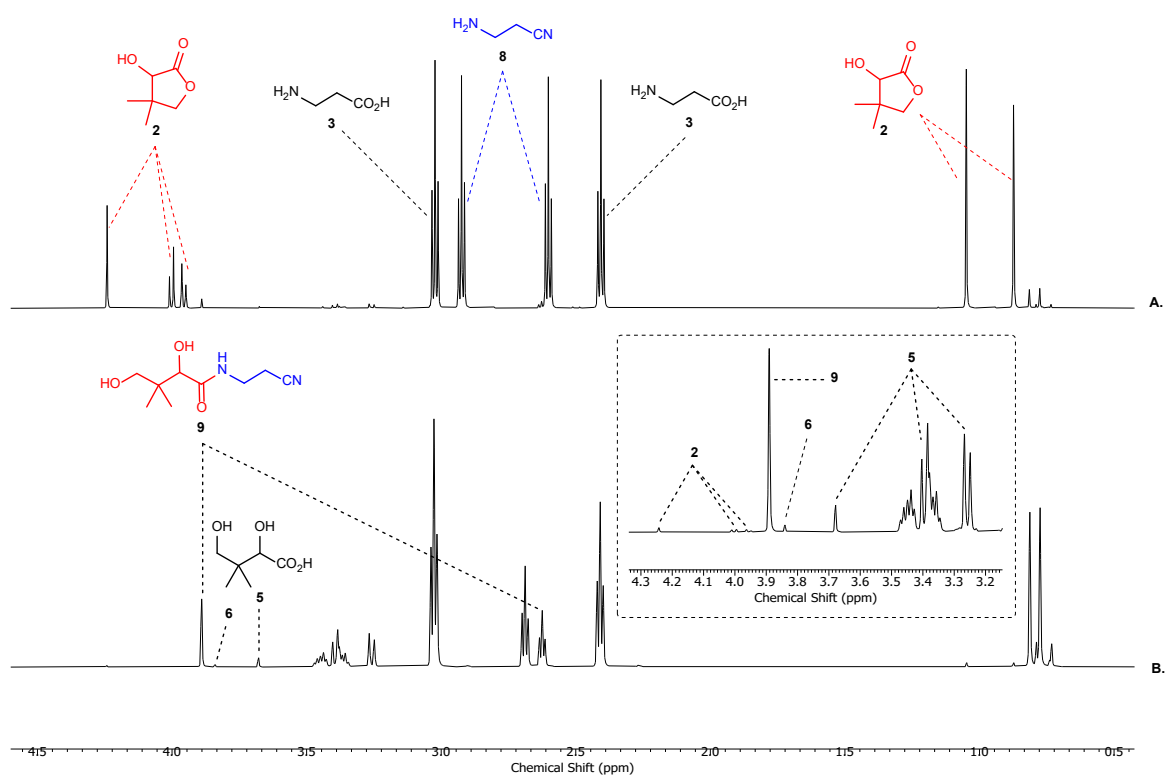
Supplementary Figure 28. ^{1}H NMR (600 MHz, H$_2$O, noesygppr1d, 0.5–4.5 ppm) spectra to show the competitive reaction of pantolactone **2** (500 mM) with β-alanine **3** (2 equiv.) and β-alanine nitrile **8** (2 equiv.) in phosphate buffer (pH 9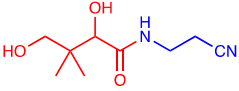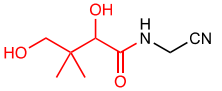.0; 500 mM) at 20 °C after: (A.) 10 mins; (B.) 2 days. Inset: Expanded ^{1}H NMR spectral region between 3.2–4.3 ppm after 2 days.
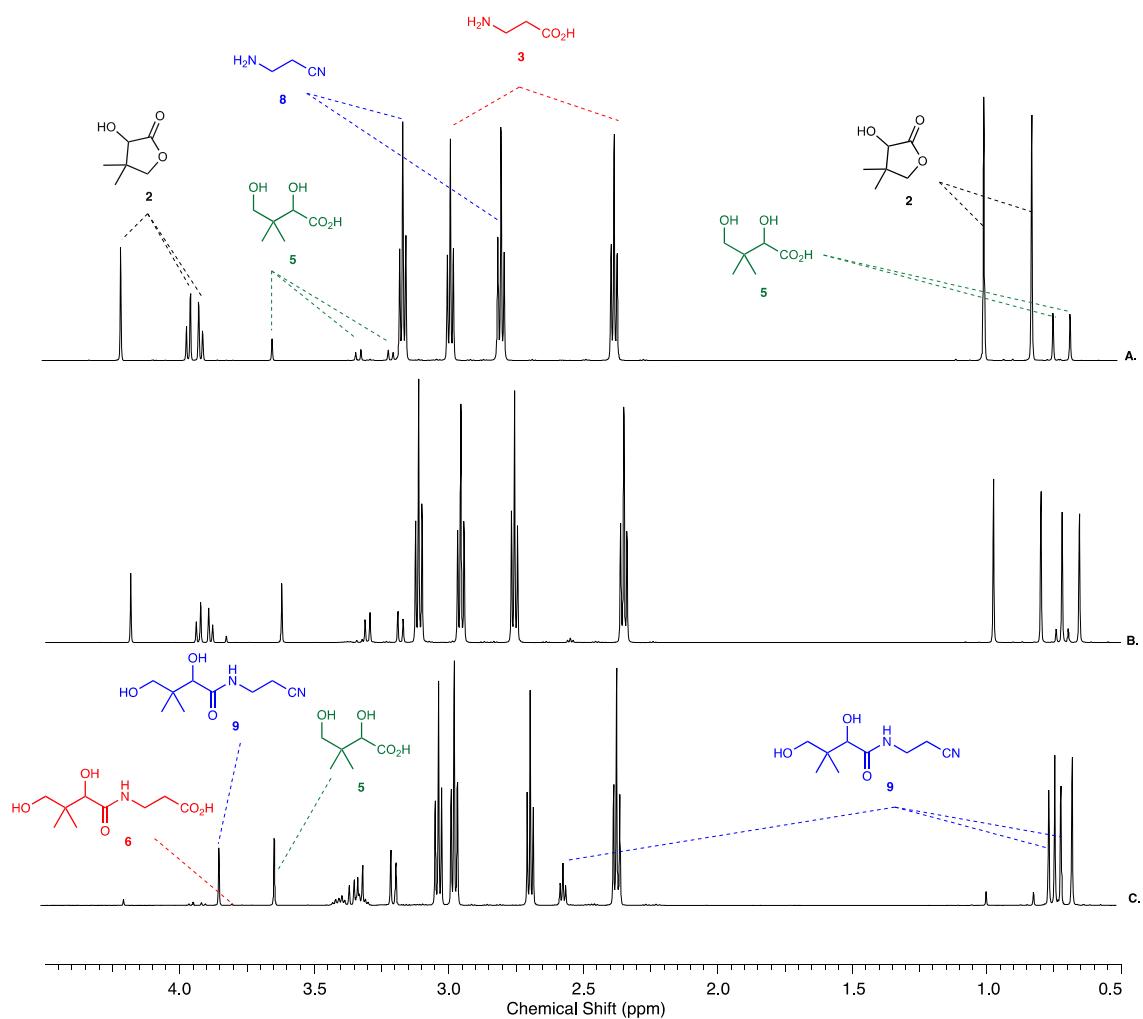
Supplementary Figure 29. ^{1}H NMR (700 MHz, H$_2$O, noessygppr1d) spectra to show the reactions of pantolactone **2** (100 mM) with β-alanine **3** (2 equiv.) and β-alanine nitrile **8** (2 equiv.) in phosphate buffer (500 mM) at 20 ºC at (A.) pH 7.0 after 3 days; (B.) pH 8.0 after 4 days; (C.) pH 9.0 after 4 days.

*Thiazoline **10** and pantetheine **1** formation by reaction of pantothenic acid nitrile **9** with cysteamine **4***
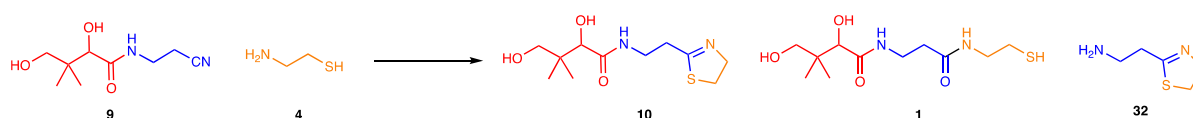
Pantothenic acid nitrile **9** (3.1 – 500 mM) and cysteamine **4** (1.1 – 2 equiv.) were dissolved in water or phosphate buffer (1 – 5 equiv.). The solution was adjusted to the specified pH with 1–4 M HCl/NaOH and incubated at 20 ºC. The reaction was then monitored by NMR spectroscopy. Yields are reported Supplementary Table 6.

Concentration (mM)		pH	Time (days)	Yield (%)			
9	**4**			**10**	**32**	**5**	**1**
3.1	4.7	9.0[a]	8	15	tr.	tr.	tr.
12.5	18.8	9.0[a]	8	26	tr.	tr.	tr.
25	37.5	9.0[a]	8	45	2	2	3
50	75	9.0[a]	8	64	3	3	4
100	110	9.0[b]	7	63	8	6	9
100	200	9.0[b]	7	83	7	7	5
500	550	9.0[b]	5	79	5	9	n.d
500	1000	9.0[b]	3	84	4	1	n.d
500	1000	9.0[c]	2	80	8	5	n.d
500	550	7.0[b]	16	65	2	1	4
500	1000	7.0[b]	9	75	2	1	4
500	1000	7.0[c]	7	38	n.d	n.d	15
			14	41	n.d	n.d	35
			21	34	tr.	tr.	52
			30	28	tr.	tr.	62
			36	18	tr.	tr.	77
			60	4	n.d	1	93

Supplementary Table 6. Quantification of species observed from the reaction of pantothenic acid nitrile **9** (3.1–500 mM) with cysteamine **4** (1.1–2 equiv.) at pH 7 and 9 at 20 ºC, n.d = not detected. tr. = trace. [a] Reactions carried out with phosphate buffer (5 equiv.); [b] no buffer. [c] phosphate buffer (1 equiv.).
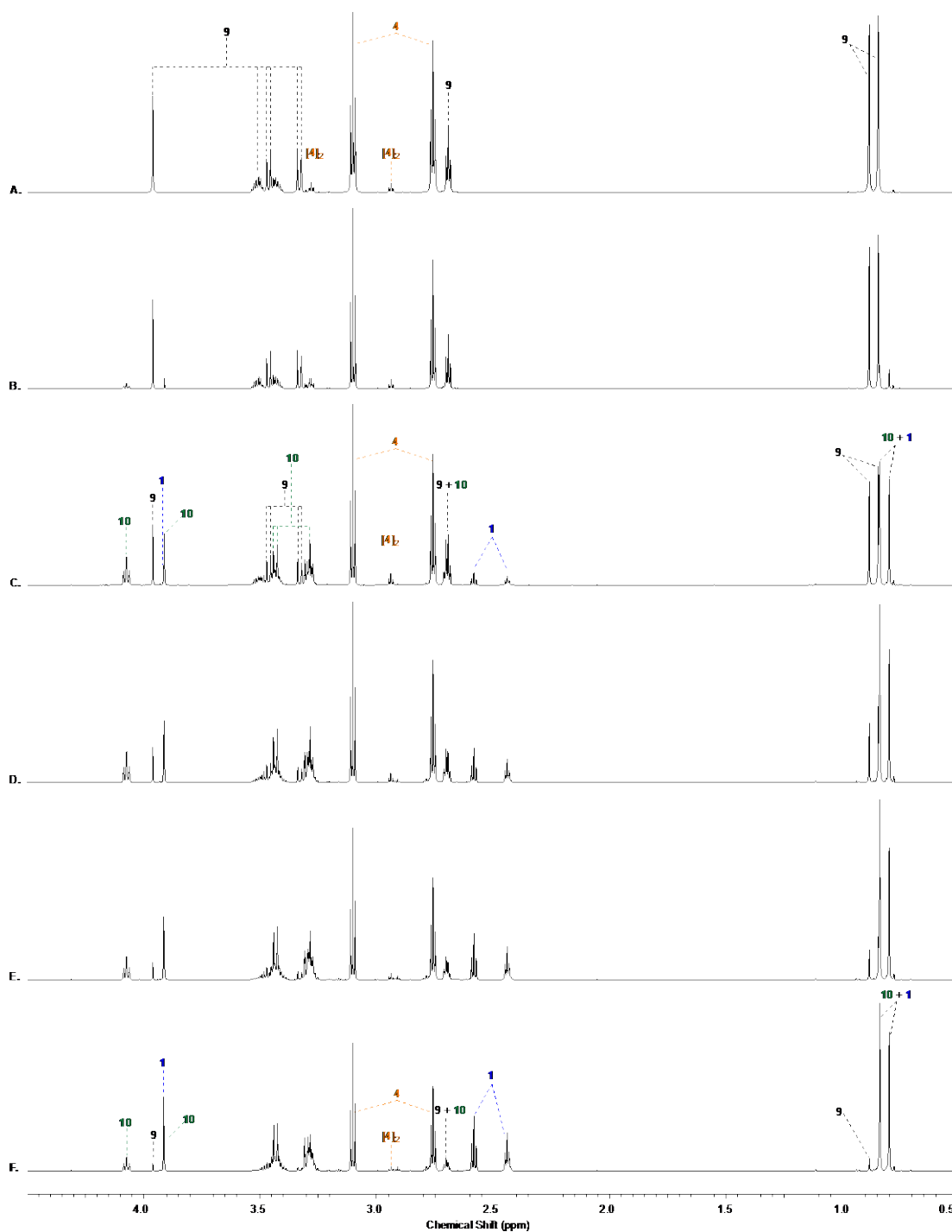
Supplementary Figure 30. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, (0.5–4.5 ppm) spectra to show the reaction of pantothenic acid nitrile **9** (500 mM) with cysteamine **4** (2 equiv.) at pH 7.0 and 20 ºC. after: (**A.**) 10 min; (**B.**) 1 day; (**C.**) 7 days; (**D.**) 14 days; (**E.**) 21 days; (**F**.) 30 days, where pantetheine **1** (62% and thiazoline **10** (28%) were the major species. See Supplementary Table 6, and see Supplementary Figure 31 for the NMR spectra for the continuation of the reaction from 36 days (**1** [77%]; **10** [18%]) to 60 days (**1** [93%]; **10** [4%]).
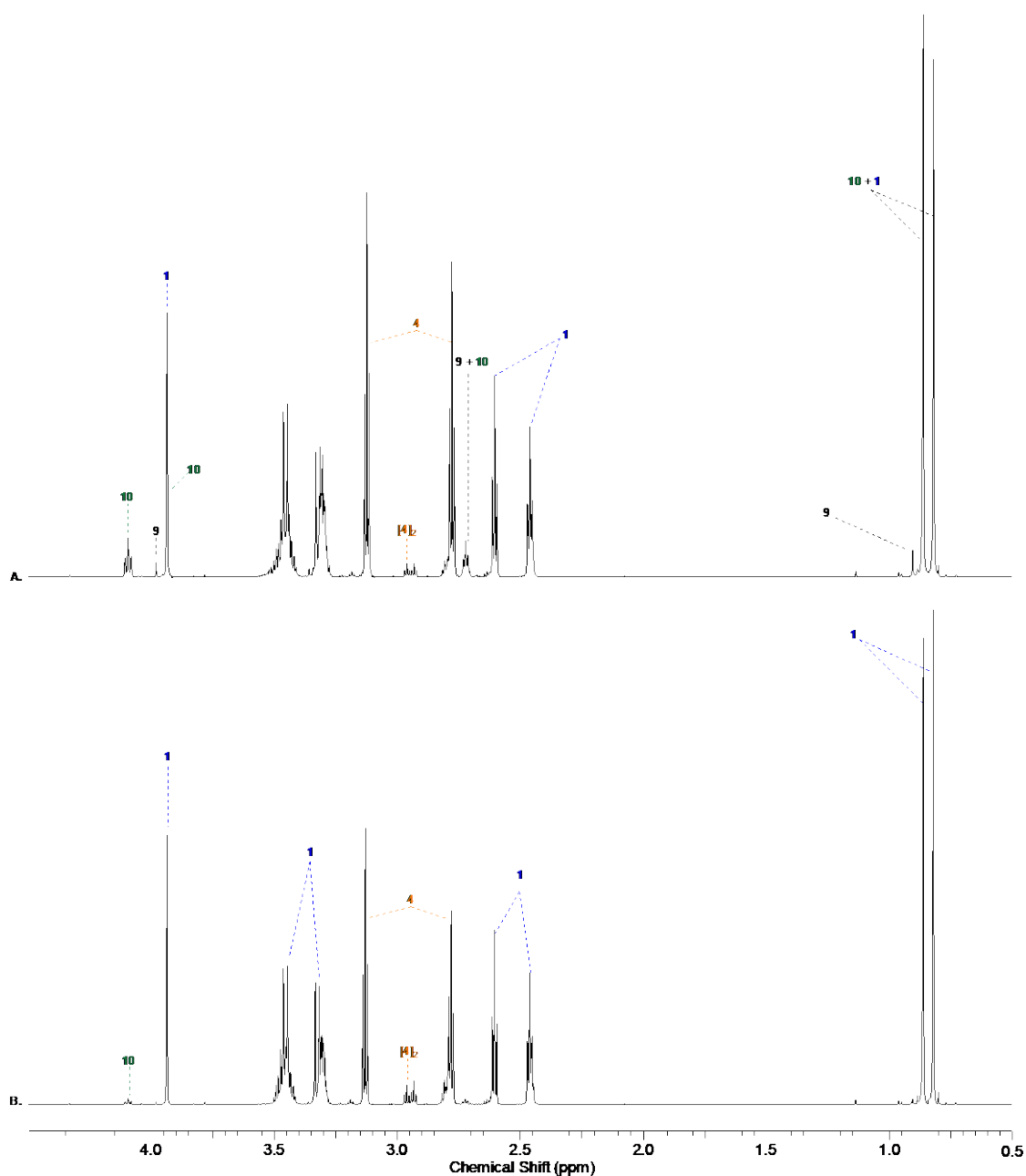
Supplementary Figure 31. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, (0.5–4.5 ppm) spectra to show the reaction of pantothenic acid nitrile **9** (500 mM) with cysteamine **4** (2 equiv.) at pH 7.0 and 20 °C. after: (**A.**) 36 days, where pantetheine **1 (**77%) and thiazoline **10** (18%) were observed, and (**B**.) after 60 days to give pantetheine **1** (93%) and thiazoline **10** (4%). See Supplementary Table 6.
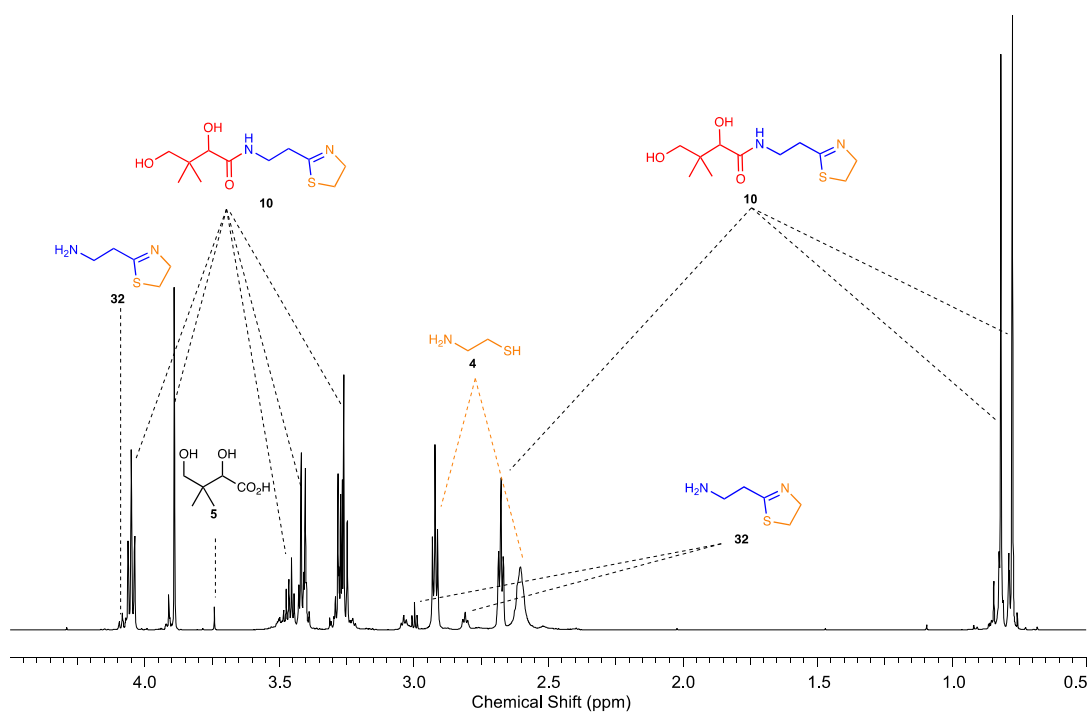
Supplementary Figure 32. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–4.5 ppm) spectrum to show the reaction of pantothenic acid nitrile **9** (500 mM) with cysteamine **4** (2 equiv.) at pH 9.0 and 20 ºC after 3 days.
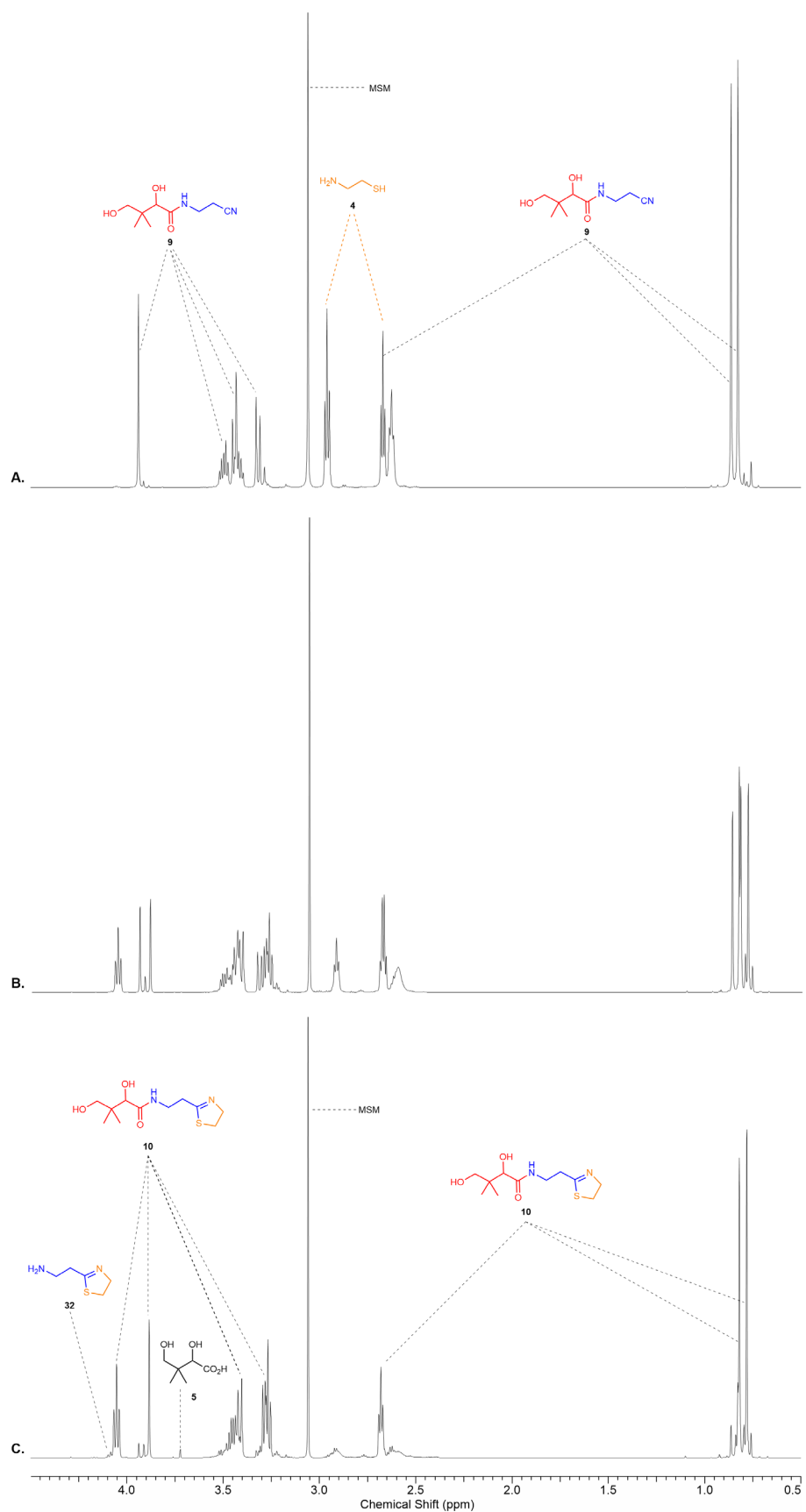
Supplementary Figure 33. ^{1}H NMR (600 MHz, H$_2$O, noesygppr1d, (0.5–4.5 ppm) spectra to show the reaction of pantothenic acid nitrile **9** (500 mM) with cysteamine **4** (1.1 equiv.) at pH 9.0 after: (A.) 10 min; (B.) 1 day; and (C.) 3 days. MSM = (methanesulfonyl)methane (internal standard)).

Hydrolysis of thiazoline 10 to pantetheine 1
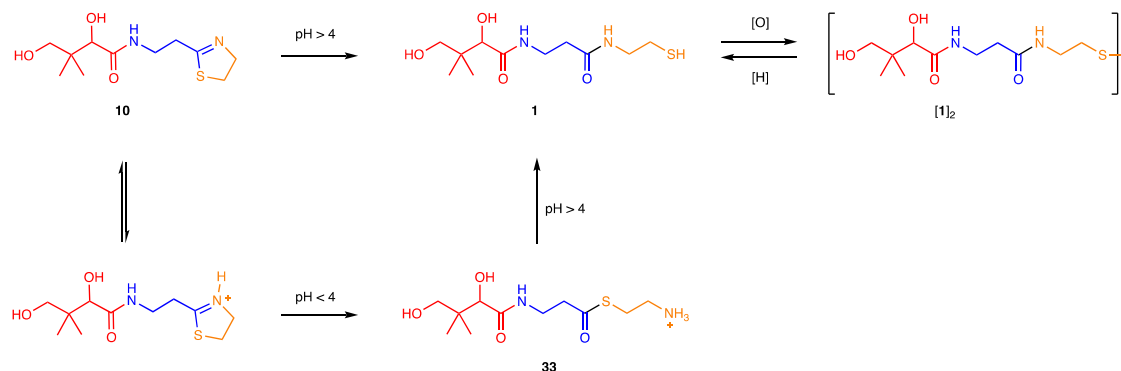
Thiazoline **10** (100 – 500 mM) and MSM (50 mM, internal standard) were dissolved in degassed acetate or phosphate buffer (500 mM; H_2O/D_2O (9:1)) at the specified pH and then incubated at 20 ºC or 60 ºC. The solution was analysed periodically by NMR spectroscopy (Supplementary Figure 34–Supplementary Figure 38). The yields of pantetheine **1** and pantethine **[1]$_2$** are given in Supplementary Table 7. In acidic solution (pH <4) thioester **33** was also observed. Quantitative conversion of thioester **33** to pantetheine **1** was observed upon addition of NaOH (pH 3 → pH 7).

Concentration **10** (mM)	Temperature (°C)	pH	Time (days)	Yield (%)			
				10	**1**	**[1]$_2$**	**33**
500	20	7.0[a]	43	1	95	4	-
500	20	6.0[a]	14	3	95	2	-
500	20	4.0[b]	3	2	86	3	9
100	20	9.0[a]	7	89	6	0	-
100	20	7.0[a]	43	2	92	8	-
100	20	6.0[a]	14	1	95	4	-
100	20	5.0[b]	5	2	90	8	-
100	20	4.0[b]	2	4	93	3	-
100	20	3.0[b]	2	5	57	-	38
100	20	2.0	2	2	46	-	42
100	20	1.0	2	5	32	-	48
100	60	7.0[a]	5	3	86	11	-
100	60	6.0[a]	5	2	81	17	-

Supplementary Table 7. Quantification of species observed upon hydrolysis of thiazoline **10**. [a] Phosphate buffer (500 mM). [b] Acetate buffer (500 mM).

*Data for pantetheine **1** (Supplementary Figure 34)*
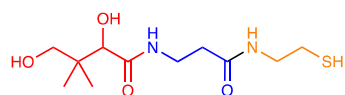
^{1}H NMR (700 MHz, H_2O/D_2O 9:1, noesygppr1d, pH 7.0) δ_H 3.94 (s, 1H, (C2)–H), 3.53-3.42 (m, 2H, NH-C$\underline{H}_2$CH$_2$CO), 3.47 (AB, J = 11.3 Hz, 1H, (C4)–H), 3.34 (AB, J = 11.3 Hz, 1H, (C4)–H′), 3.36-3.30 (m, 2H, NH-C$\underline{H}_2$CH$_2$SH), 2.56 (t, J = 6.6 Hz, 2H, NHCH$_2$C$\underline{H}_2$SH), 2.48 (t, J = 6.5 Hz, 2H, NHCH$_2$C$\underline{H}_2$CO), 0.82 (s, 3H, (C3)-CH$_3$), 0.78 (s, 3H, (C3)-CH$_3$′). ^{13}C NMR (176 MHz, H_2O/D_2O 9:1) δ_C 175.7 (C1), 174.5 (NHCH$_2$CH$_2$$\underline{C}$O), 76.4 (C2), 69.1 (C4), 43.0 (NH$\underline{C}$H$_2$CH$_2$SH), 39.2 (C3), 36.2 (NH$\underline{C}$H$_2$CH$_2$CO), 36.0 (NHCH$_2$$\underline{C}H_2$CO), 23.9 (CH$_2$SH), 21.2 ((C3)-$\underline{C}H_3$), 19.8 ((C3)-$\underline{C}H_3$′). The NMR spectral data for pantetheine **1** synthesised by hydrolysis of thiazoline **10** matched an authentic sample of D-**1** synthesised by disulfide reduction of commercially available pantethine D-[**1**]$_2$. See Supplementary Page S**Error! Bookmark not defined.** for synthesis and characterisation of D-**1** from D-[**1**]$_2$.

*Data for pantethine [**1**]$_2$ (Supplementary Figure 34)*
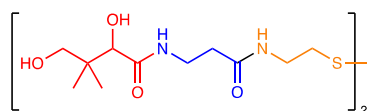
^{1}H NMR (700 MHz, H_2O/D_2O 9:1, noesygppr1d, pH 7.0, partial assignment) δ_H 3.94 (s, 2H, (C2)–H), 2.80 (t, J = 6.5 Hz, 4H, NHCH$_2$C$\underline{H}_2$SS).

*Data for thioester **33** (Supplementary Figure 37 and Supplementary Figure 38)*
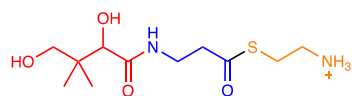
^{1}H NMR (700 MHz, H_2O/D_2O 9:1, noesygppr1d, pH 3.0, partial assignment) δ_H 3.96 (s, 1H, (C1)-H), 3.58-3.42 (m, 2H, NHC$\underline{H}_2$CH$_2$), 3.47 (AB, J = 11.0 Hz, 1H, (C4)–H), 3.35 (AB, J = 11.0 Hz, 1H, (C4)–H′), 3.23-3.14 (br. m, 2H, NHCH$_2$C$\underline{H}_2$COS & COSC$\underline{H}_2$CH$_2$NH$_2$), 2.90 (t, J = 6.3 Hz, 2H, (COSC$\underline{H}_2$CH$_2$NH$_2$), 0.87 (s, 3H, CH$_3$), 0.83 (s, 3H, CH$_3$′). ^{13}C NMR (176 MHz, H_2O/D_2O 9:1) δ_C 201.1 (COS), 175.9 (C1), 76.5 (C2), 69.1 (C4), 43.5 (COS$\underline{C}$H$_2$CH$_2$NH$_2$), 40.0 (NH$_2$CH$_2$$\underline{C}H_2$COS), 39.3 (C3), 35.8 (NH$\underline{C}$H$_2$CH$_2$CO), 26.6 (CH$_2$$\underline{C}H_2NH_2$), 21.1 ((C3)-$\underline{C}H_3$), 19.8 ((C3)-$\underline{C}H_3$′).
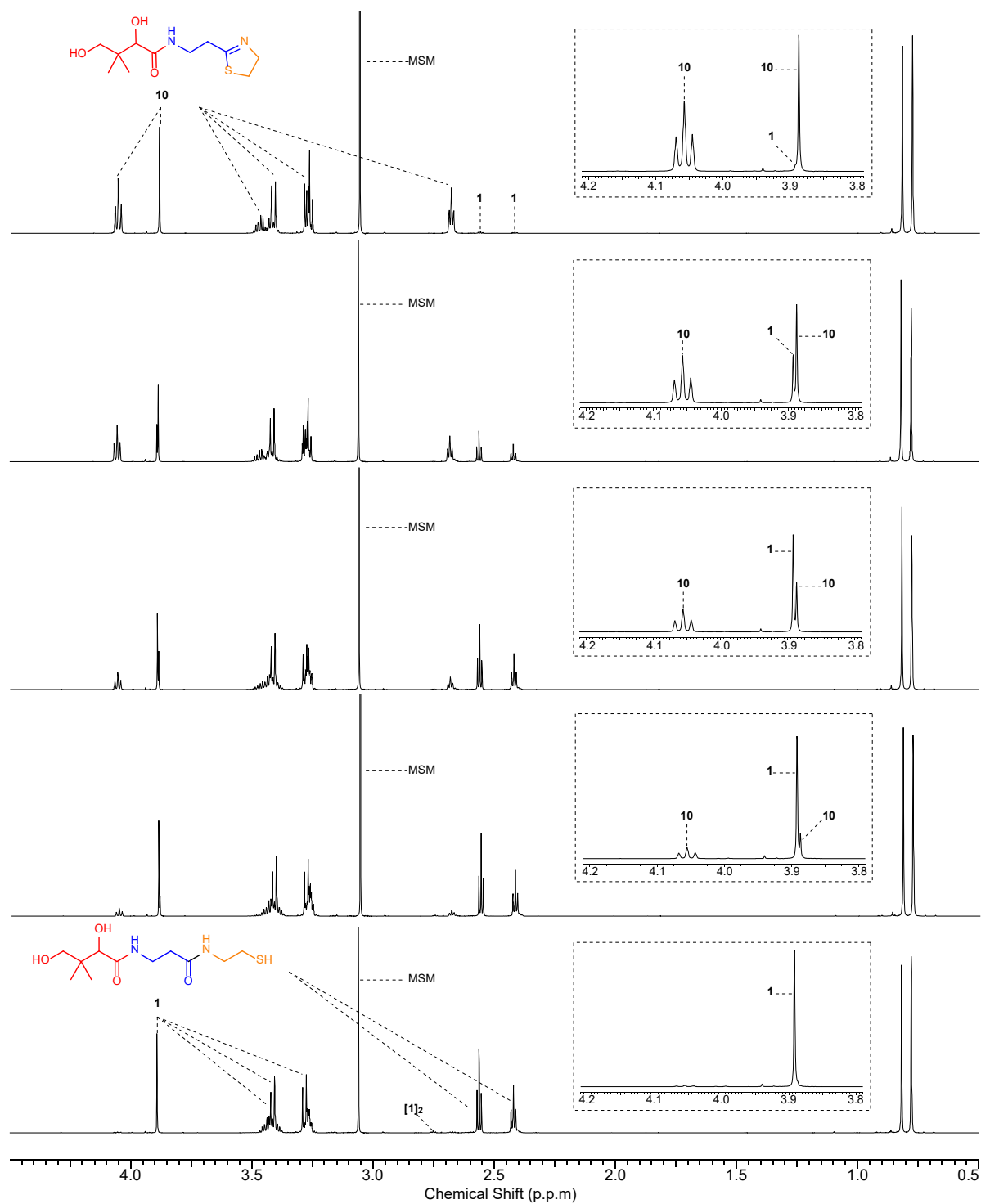
Supplementary Figure 34. ¹H NMR (700 MHz, H₂O/D₂O 9:1, noesygppr1d, 0.5–4.5 p.p.m) spectra to show the hydrolysis of thiazoline **10** (500 mM) in phosphate buffer (pH 7; 500 mM) at 20 °C to yield pantetheine **1**: (A.) 10 mins; (B.) 4 days; (C.) 12 days; (D.) 21 days; (E.) 43 days, yielding pantetheine **1** (95%), alongside pantethine [**1**]₂ (4%).
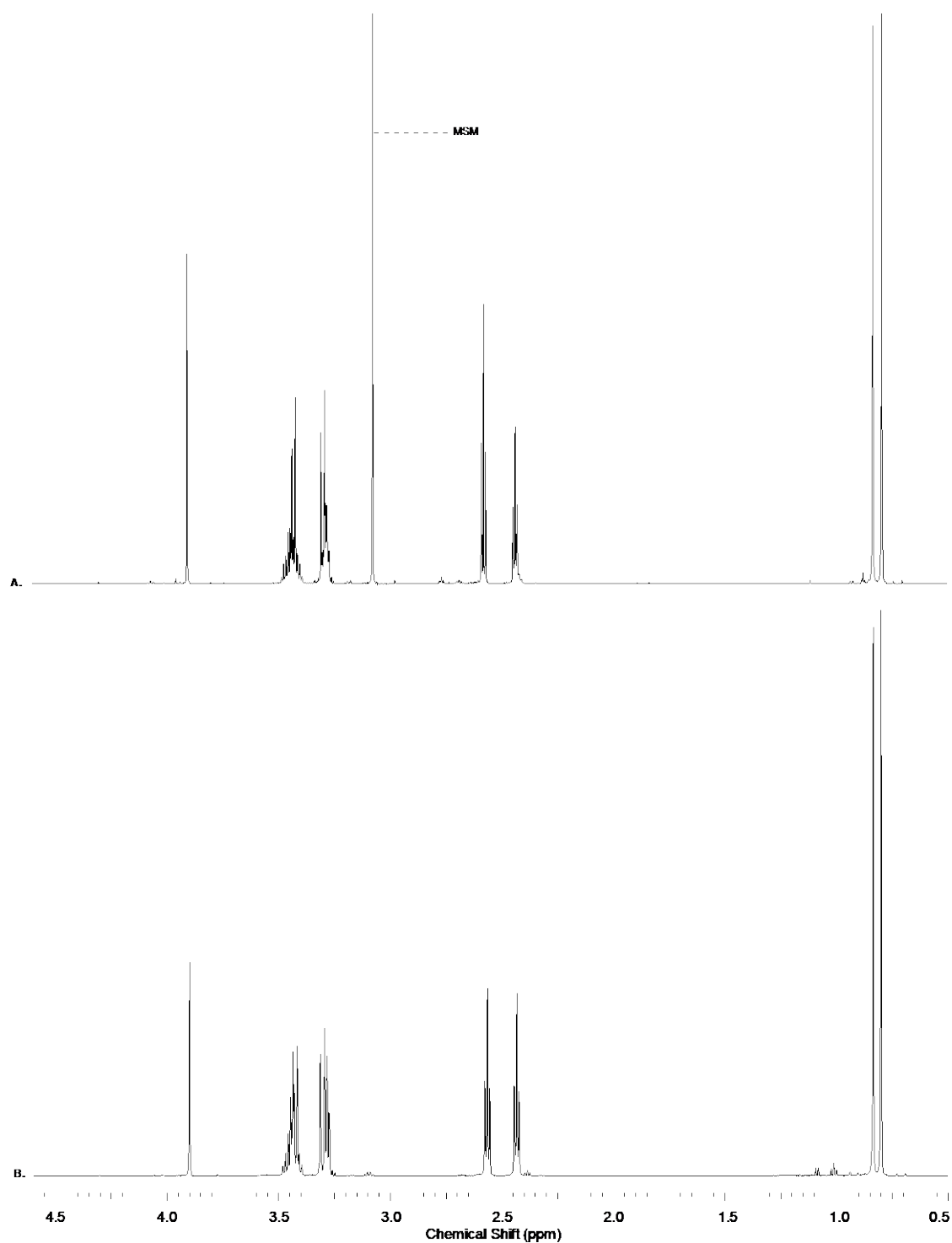
Supplementary Figure 35. Comparative ^{1}H NMR spectra to show (A.) pantetheine **1** generated from the hydrolysis of thiazoline **10** (500 mM) in phosphate buffer (pH 7; 500 mM) at 20 °C after 43 days, and (B.) authentic D-pantetheine D-**1**. The ^{1}H NMR spectral data for pantetheine **1** synthesised by hydrolysis of thiazoline **10** matched an authentic sample of D-**1** synthesised by disulfide reduction of commercially available pantethine D-[**1**]$_2$. See Supplementary Page S**Error! Bookmark not defined.** for synthesis and characterisation of D-**1** from D-[**1**]$_2$.
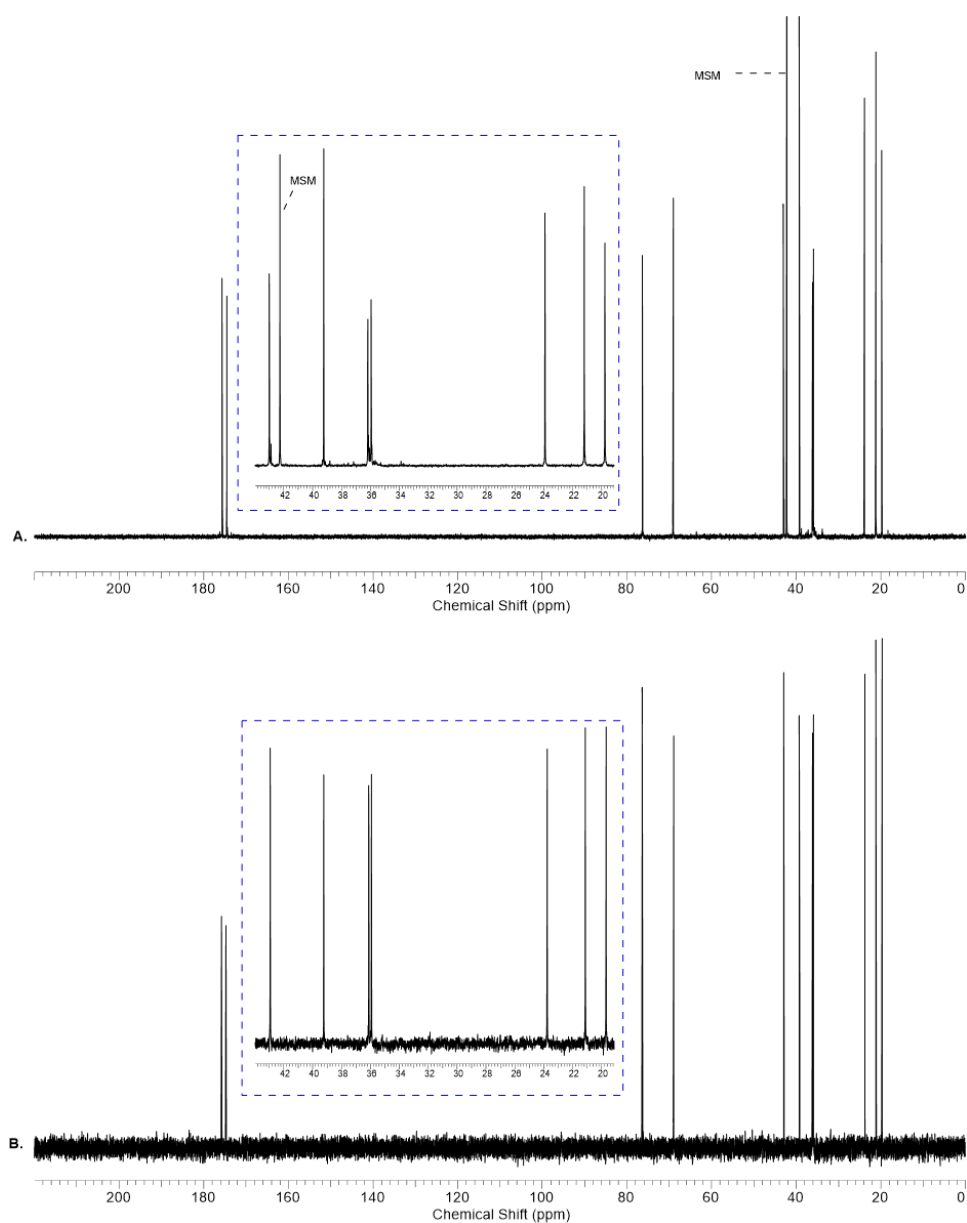
Supplementary Figure 36. Comparative ^{13}C NMR spectra to show (A.) pantetheine **1** generated from the hydrolysis of thiazoline **10** (500 mM) in phosphate buffer (pH 7; 500 mM) at 20 °C after 43 days, and (B.) authentic D-pantetheine D-**1**. The ^{13}C NMR spectral data for pantetheine **1** synthesised by hydrolysis of thiazoline **10** matched an authentic sample of D-**1** synthesised by disulfide reduction of commercially available pantethine [**1**]$_2$. See Supplementary Page S**Error! Bookmark not defined.** for synthesis and characterisation of D-**1** from D-[**1**]$_2$.
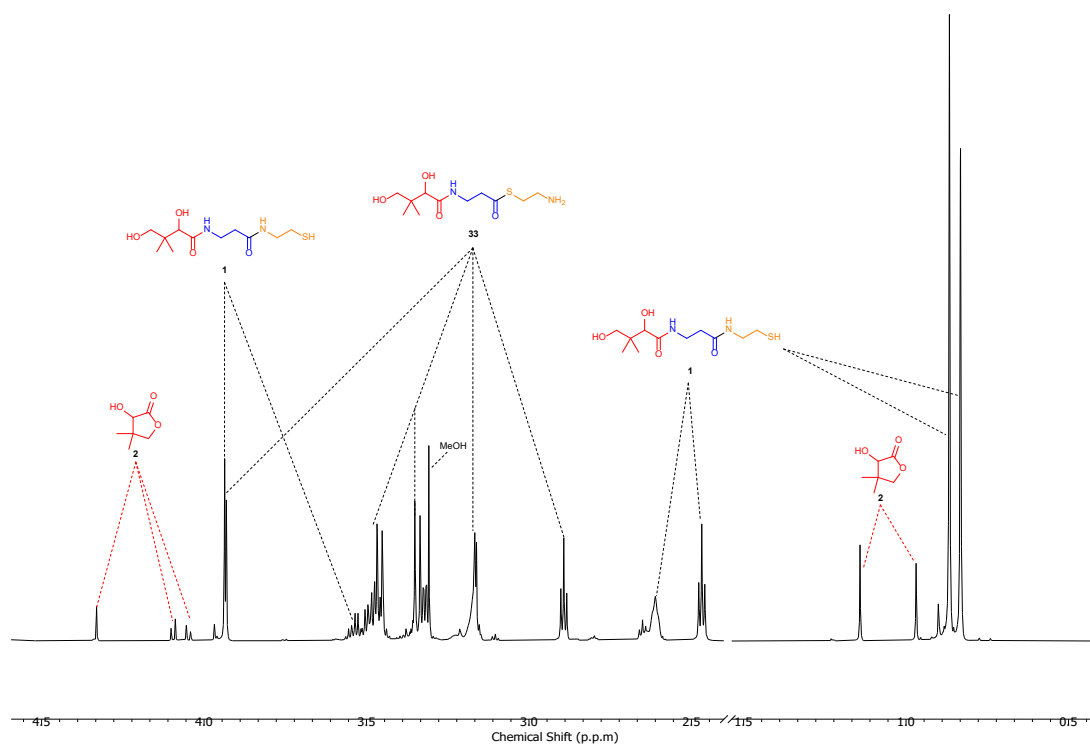
Supplementary Figure 37. ^{1}H NMR (700 MHz, H$_2$O/D$_2$O 9:1, noesygppr1d, 0.5–1.5 & 2.4–4.5 p.p.m) spectra to show the formation of thioester **33** from thiazoline **10** (100 mM) at pH 2.0 after 5 days at 20 °C.
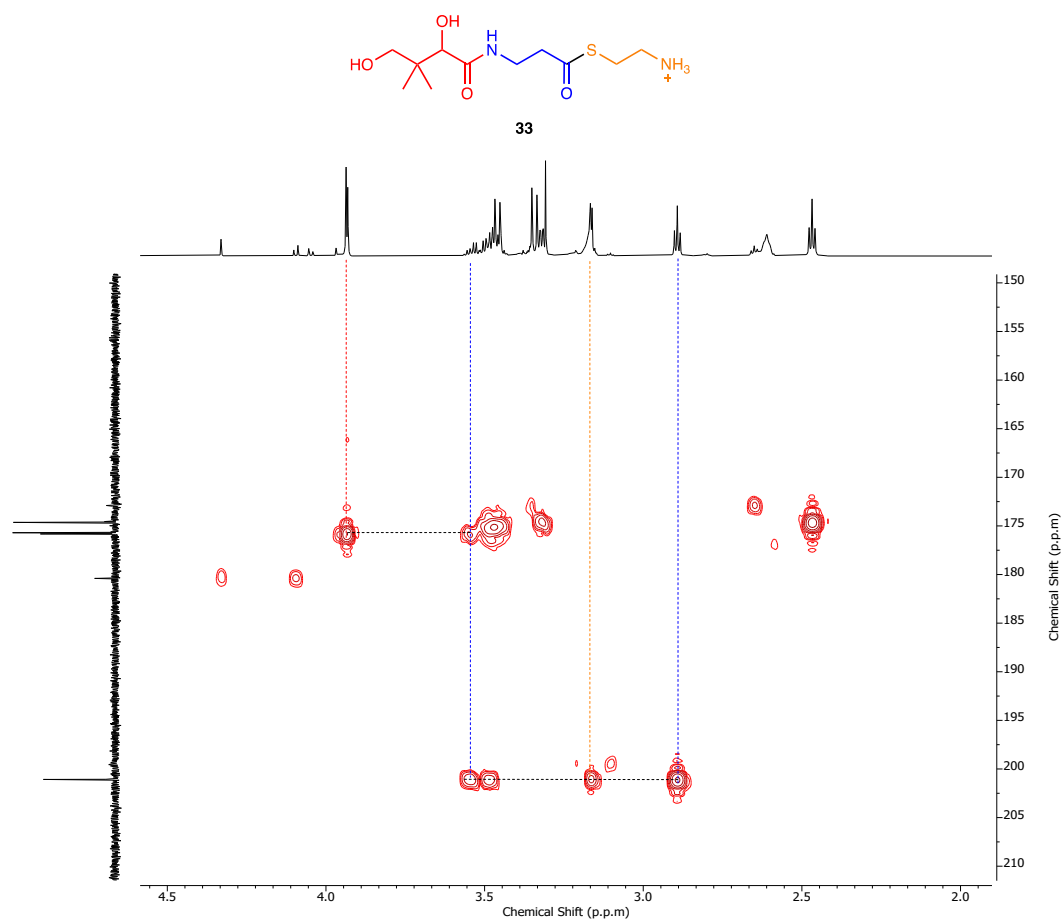
Supplementary Figure 38. ^{1}H–^{13}C HMBC (^{1}H: 700 MHz [2.0–4.5 p.p.m], ^{13}C: 176 MHz [150–210 p.p.m], H$_2$O/D$_2$O 9:1) spectrum showing the diagnostic $^2J_{CH}$ and $^3J_{CH}$ of thioester **33** ^{1}H resonances with the <u>C</u>OS ^{13}C resonance at 201.1 p.p.m, which is characteristic of a thioester. See Supplementary Figure 37 for labelled ^{1}H NMR spectrum.

*Determination of thiazoline **10** pK$_{aH}$*

10
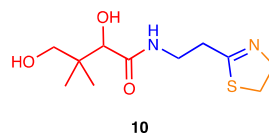
Thiazoline **10** (100 mM) and MSM (50 mM, internal standard) were dissolved in H$_2$O/D$_2$O (9:1). The solution was adjusted to the specified pH using HCl/NaOH (1–4 M) and analysed by ^{1}H NMR spectroscopy. ^{1}H NMR spectra were reference to the MSM methyl resonance (δ_H = 3.10 p.p.m) and the chemical shift of the α-methylene ($\delta_{\alpha\text{-CH2}}$) resonance of thiazoline **10** was measured (Supplementary Figure 39).
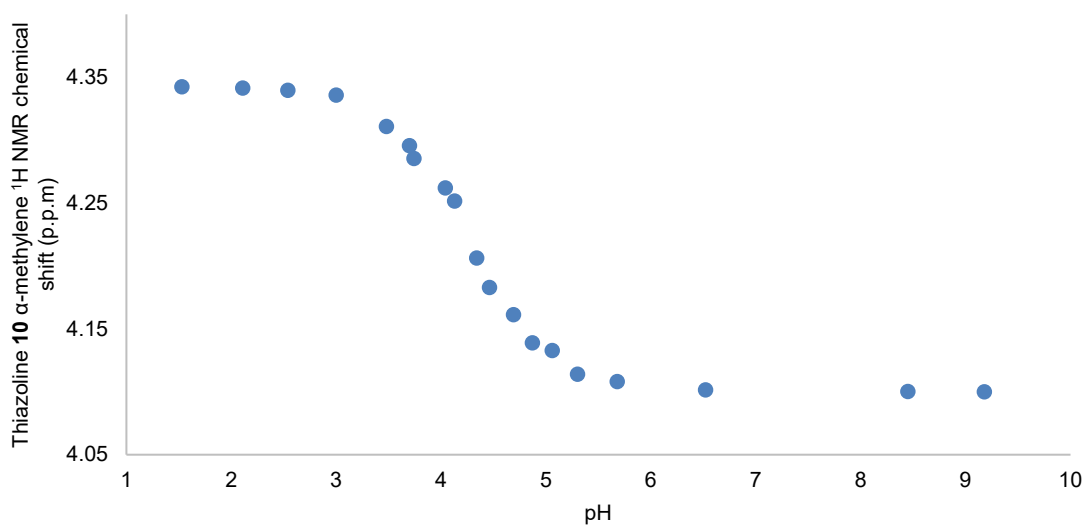
Supplementary Figure 39. Thiazoline **10** α-methylene ^{1}H NMR resonance chemical shift change ($\delta_{\alpha\text{-CH2}}$) observed upon pH titration of thiazoline **10** (α-CH$_2$NH$^+$ → α-CH$_2$N).

Reaction of pantolactone 2 and glycine nitrile 7$_G$
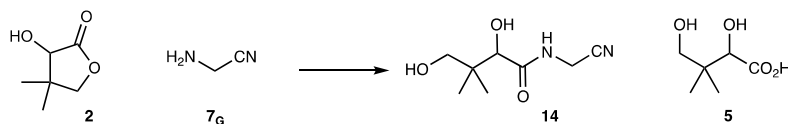
Pantolactone **2** (100 mM) and glycine nitrile **7**$_G$ (1–2 equiv.) were dissolved in phosphate buffer at pH 7.0 (500 mM) at 20 °C. The solution pH was adjusted to the required pH with 1-4 M NaOH. The resulting solution was periodically analysed by NMR spectroscopy and its pH monitored. Yields of pantoylglycine nitrile **14** and pantoic acid **5** are reported in Supplementary Table 8.

Concentration (mM)		pH	Time (days)	Yield (%)		
2	**7**$_G$			**2**	**14**	**5**
100	100	7.0	7	60	13	27
100	200	7.0	7	53	18	29
100	100	9.0	5	33	25	42
100	200	9.0	5	25	39	36

Supplementary Table 8. Quantification of species observed from the reaction of pantolactone **2** with glycine nitrile **7**$_G$ in phosphate buffer (500 mM; 10 equiv.) and at 20 °C.
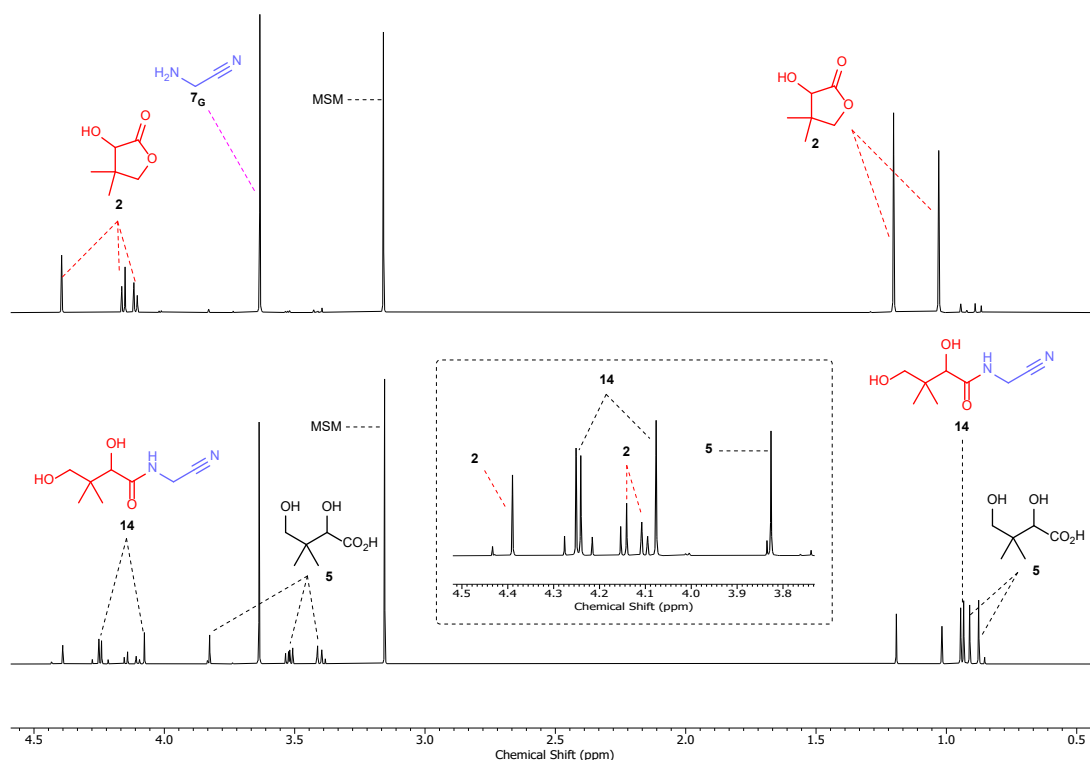
Supplementary Figure 40. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–4.5 ppm) spectra to show the reaction of pantolactone **2** (100 mM) with glycine nitrile **7**$_G$ (2 equiv.) in phosphate buffer (pH 9.0; 0.5 M) at 20 °C after; (A.) 10 mins; (B.) 5 day; Inset: Expansion of the ^{1}H NMR spectral region between 3.8–4.5 ppm. MSM = (methanesulfonyl)methane (internal standard).

*Competitive reaction of pantolactone **2** with glycine nitrile **7**$_G$ and β-alanine nitrile **8***

Pantolactone **2** (500 mM), β-alanine nitrile **8** (2 equiv.) and glycine nitrile **7**$_G$ (2 equiv.) were dissolved in phosphate buffer (pH 7.0; 500 mM) at 20 ºC. The solution pH was adjusted to pH 9 with 1–4 M NaOH. The resultant solution was maintained at pH 9 and periodically analysed by NMR spectroscopy. The near-equivalent formation of pantothenic acid nitrile **9** and pantoylglycine nitrile **14** (**9**/**14**; 1.1:1) was observed after 6 hours (Supplementary Figure 41B). However, nitrile **9** was the major pantoylamide product after 6 days (**9**/**14**; >5:1) (Supplementary Figure 41C).
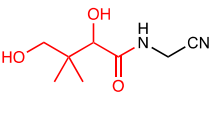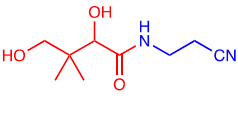
Supplementary Figure 41. ^{1}H NMR (700 MHz, H_2O, noesygppr1d, 0.5–4.5 ppm) spectra to show the competitive reaction of pantolactone **2** (500 mM) with β-alanine nitrile **8** (2 equiv.) and glycine nitrile **8** (2 equiv.) in phosphate buffer (pH 9; 500 mM) at 20 ºC after: (A.) 10 mins; (B.) 6 h, and (C.) 6 days. Inset: Expanded ^{1}H NMR spectral region between 3.7–4.5 ppm.

*Synthesis of pantothenic acid nitrile **9** by transamidation of pantoylglycine nitrile **14** with β-alanine nitrile **8***
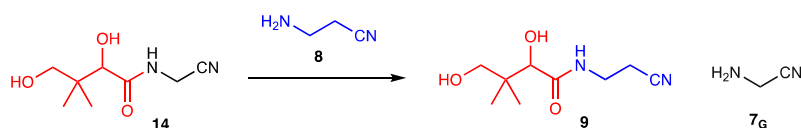
Pantoylglycine nitrile **14** (500 mM), β-alanine nitrile **8** (1.25-5 equiv.) and methanol (250 mM; internal standard) were dissolved in phosphate buffer (pH 7.0; 500 mM) at 20 ºC. The solution was adjusted to pH 9.0 with 1-4 M NaOH. The resulting mixture was periodically analysed by ^{1}H NMR spectroscopy. The yield of pantothenic acid nitrile **9** is reported in Supplementary Table 2. Representative ^{1}H NMR spectra are given in Supplementary Figure 42 and Supplementary Figure 43.

Concentration (M)		Time (days)	Yield (%)			
14	**8**		**9**	**14**	**2**	**5**
0.5	0.63	10	57	21	2	20
0.5	1	10	82	7	0	11
0.5	1.5	19	87	5	0	8
0.5	2.5	19	93	0	0	6

Supplementary Table 9. Quantification of pantoyl species observed from the reaction of pantoylglycine nitrile **14** with β-alanine nitrile **8** at pH 9 in phosphate buffer (500 mM) and at 20 ºC.
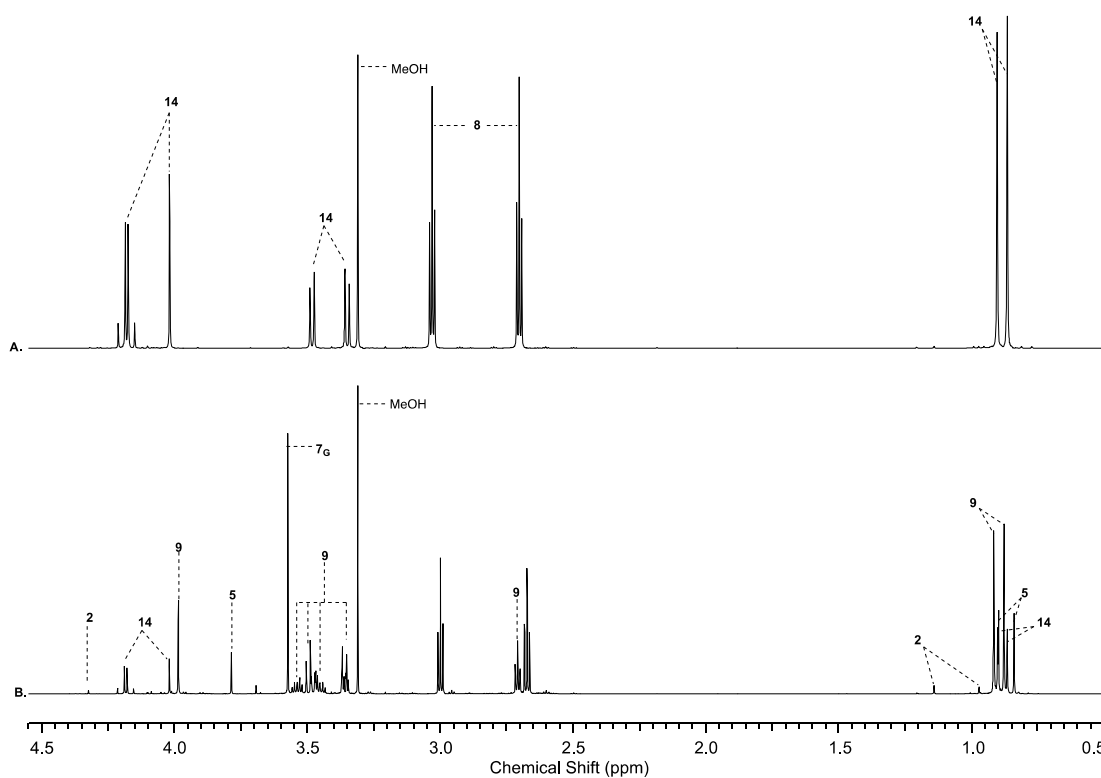
Supplementary Figure 42. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–4.5 ppm) spectra to show the reaction of pantoylglycine nitrile **14** (500 mM) with β-alanine nitrile **8** (1.25 equiv.) in phosphate buffer (pH 9.0; 500 mM) and at 20 ºC after: (A.) 10 mins; and (B.) 10 days. MeOH (internal standard).
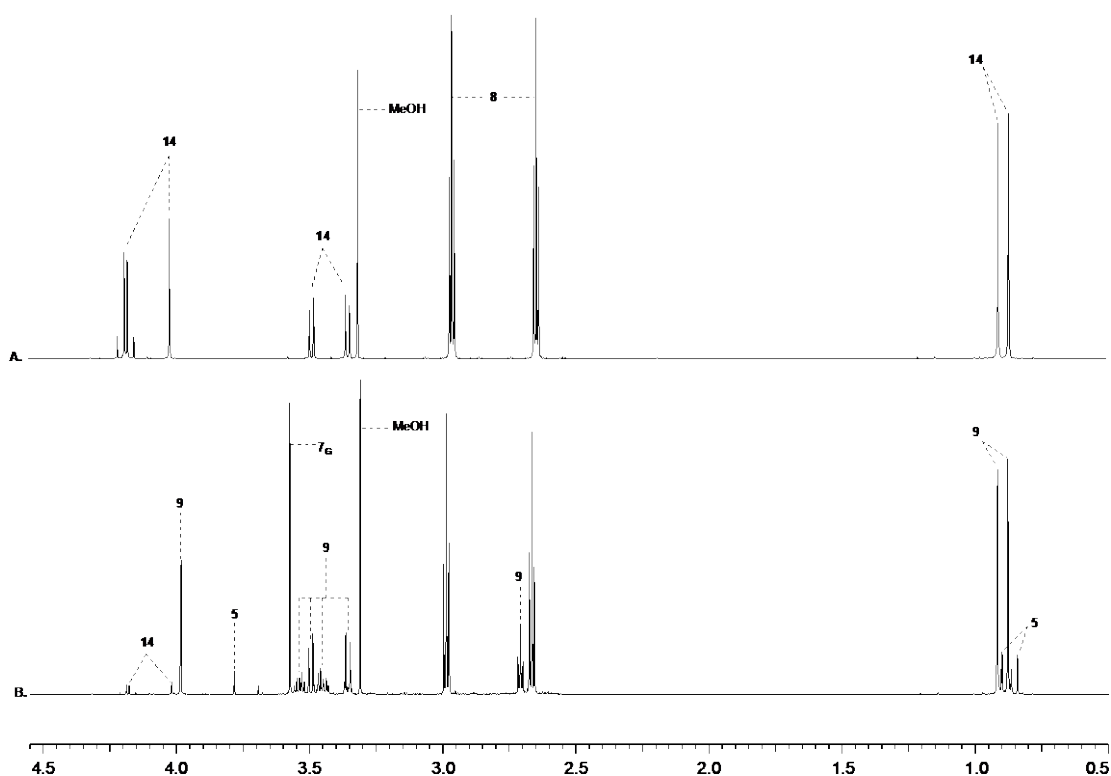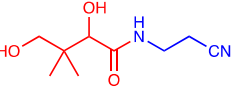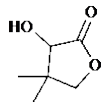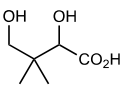
Supplementary Figure 43. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–4.5 ppm) spectra to show the reaction of pantoylglycine nitrile **14** (500 mM) with β-alanine nitrile **8** (2 equiv.) in phosphate buffer (pH 9.0; 500 mM) at 20 ºC after: (A.) 10 mins; and (B.) 10 days. MeOH (internal standard).

*Attempted synthesis of pantoylglycine nitrile **14** by transamidation of pantothenic acid nitrile **9** with glycine nitrile **7G***
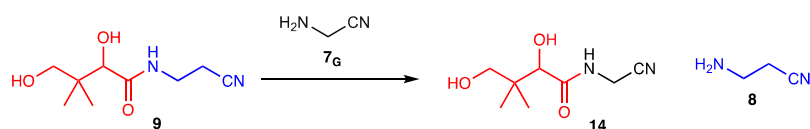
Pantothenic acid nitrile **9** (500 mM) and glycine nitrile **7G** (1–5 equiv.) were dissolved in phosphate buffer (pH 7.0; 500 mM) at 20 ºC. The solution was adjusted to pH 9.0 with 1-4 M NaOH. The resulting solution was periodically analysed by ^{1}H NMR spectroscopy. The yield of pantoylglycine nitrile **14** is reported in Supplementary Table 10. ^{1}H NMR spectra are given in Supplementary Figure 44 and Supplementary Figure 45.

Concentration (M)		Time (days)	Amounts (%)			
9	**7G**		**14**	**9**	**2**	**5**
0.5	0.5	10	3	93	1	3
0.5	1	10	5	90	0	4
0.5	1.5	10	7	86	0	6
0.5	2.5	10	7	83	0	9

Supplementary Table 10. Quantification of pantoyl species observed from the reaction of pantothenic acid nitrile **9** with glycine nitrile **7G** in phosphate buffer (pH 9.0; 500 mM) at 20 ºC.
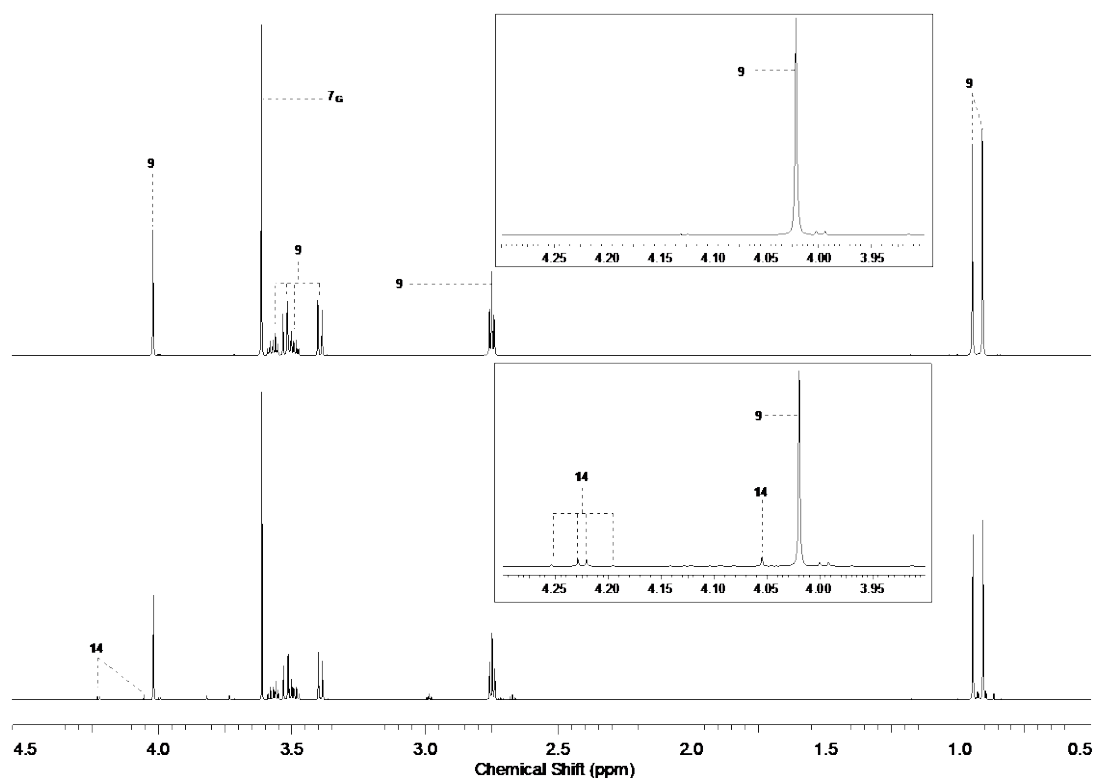
Supplementary Figure 44. ^{1}H NMR (700 MHz, H_2O, noesygppr1d, 0.5–4.5 ppm) spectra to show the attempted synthesis of pantoylglycine nitrile **14** from the reaction of pantothenic acid nitrile **9** (500 mM) with glycine nitrile **7$_G$** (1 equiv.) in phosphate buffer (pH 9; 500 mM) at 20 ºC after: (A.) 10 mins; and (B.) 10 days.
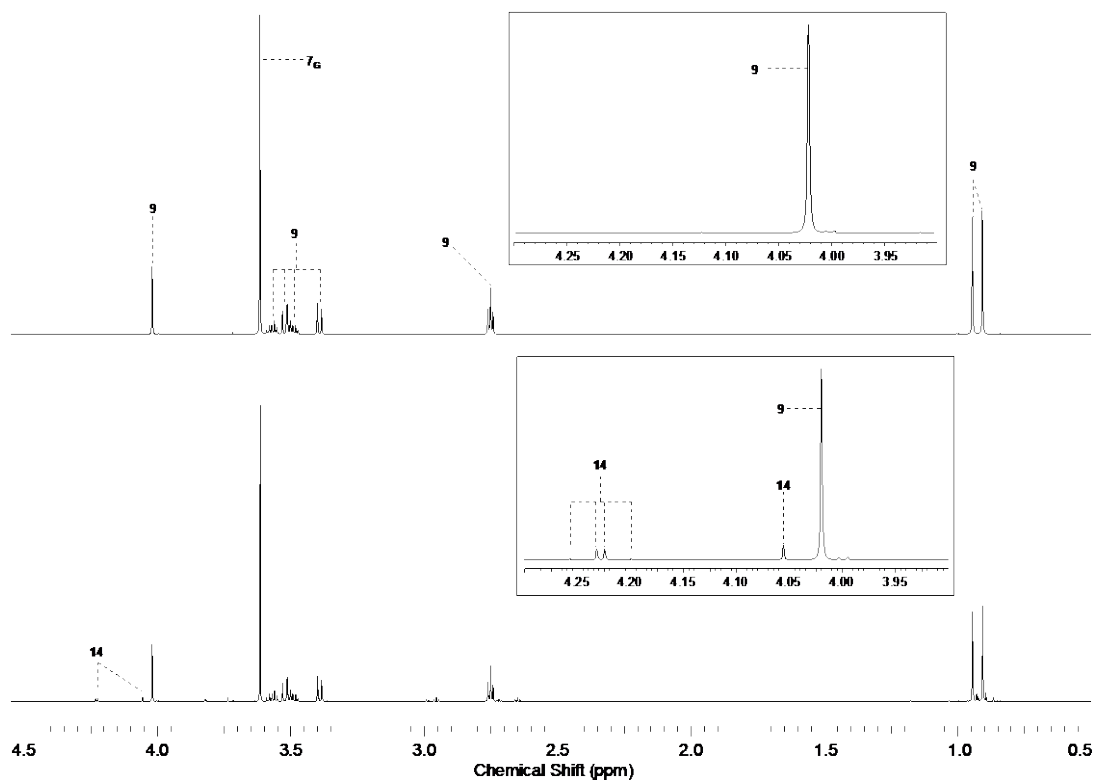
Supplementary Figure 45. ^{1}H NMR (700 MHz, H_2O, noesygppr1d, 0.5–4.5 ppm) spectra to show the attempted synthesis of pantoylglycine nitrile **14** from the reaction of pantothenic acid nitrile **9** (500 mM) with glycine nitrile **7$_G$** (2 equiv.) in phosphate buffer (pH 9; 500 mM) at 20 ºC after: (A.) 10 mins; and (B.) 10 days.

*Multicomponent one-pot synthesis of pantetheine **1** by reaction of pantolactone **2**, β-alanine nitrile **8**, and cysteamine **4***
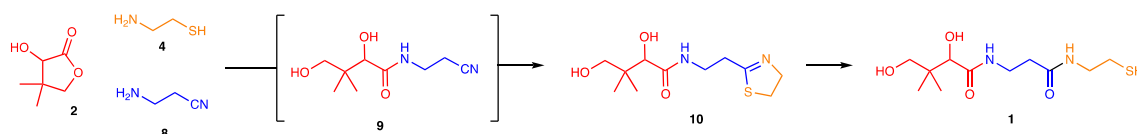
Pantolactone **2** (500 mM), cysteamine **4** (2 equiv.), and β-alanine nitrile **8** (2 equiv.) were dissolved in phosphate buffer (pH 7.0; 500 mM) at 20 ºC, and the solution was adjusted to pH 9 with 1–4 M NaOH. The reaction was periodically analysed by NMR spectroscopy. After 4 days, **2** and **8** were consumed and the pantothenic acid nitrile **9** (formed transiently) had been converted in situ to thiazoline **10** by reaction with cysteamine **4** (Supplementary Figure 46). The solution was adjusted to pH 4 with 4 M HCl and incubated at 20 ºC for 24 hours, then adjusted back to pH 9 with 1-4 M NaOH and analysed by NMR spectroscopy. Quantitative hydrolysis of **10** gave pantetheine **1** (57% based on **2**), alongside pantoic acid **5** (19%) (Supplementary Figure 47).
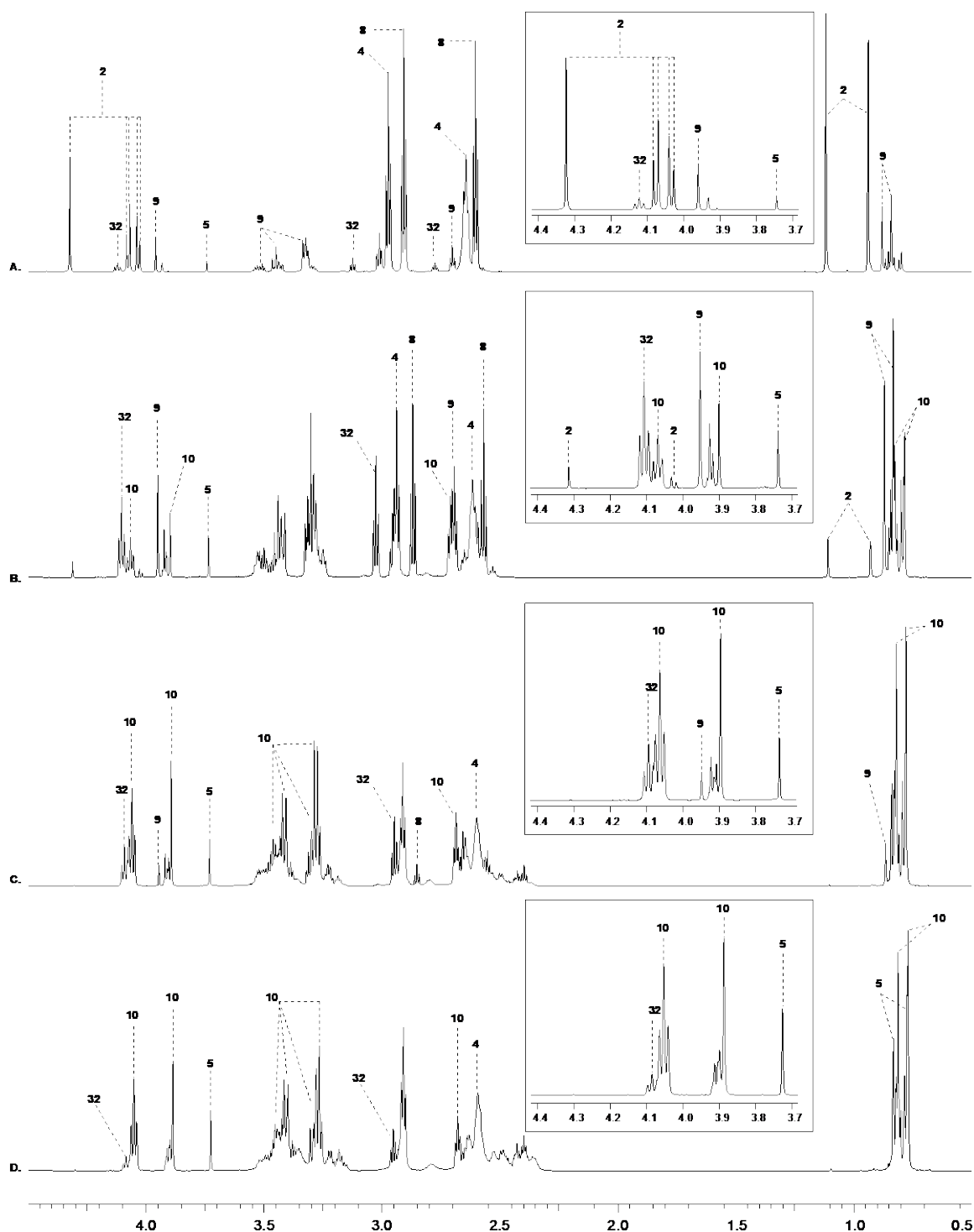
Supplementary Figure 46. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–4.5 ppm) spectra to show the multicomponent one-pot synthesis of thiazoline **10** by reaction of pantolactone **2** (500 mM), cysteamine **4** (2 equiv.) and β-alanine nitrile **8** (2 equiv.) in phosphate buffer (pH 9; 500 mM) at 20 ºC after: (A.) 10 mins; (B.) 3 hours; (C.) 1 day; (D.) 4 days. Inset: Expanded ^{1}H NMR spectral region between 3.7–4.4 ppm. See Supplementary Figure 47 for the quantitative conversion of thiazoline **10** to pantetheine **1** by acid-catalysed hydrolysis.
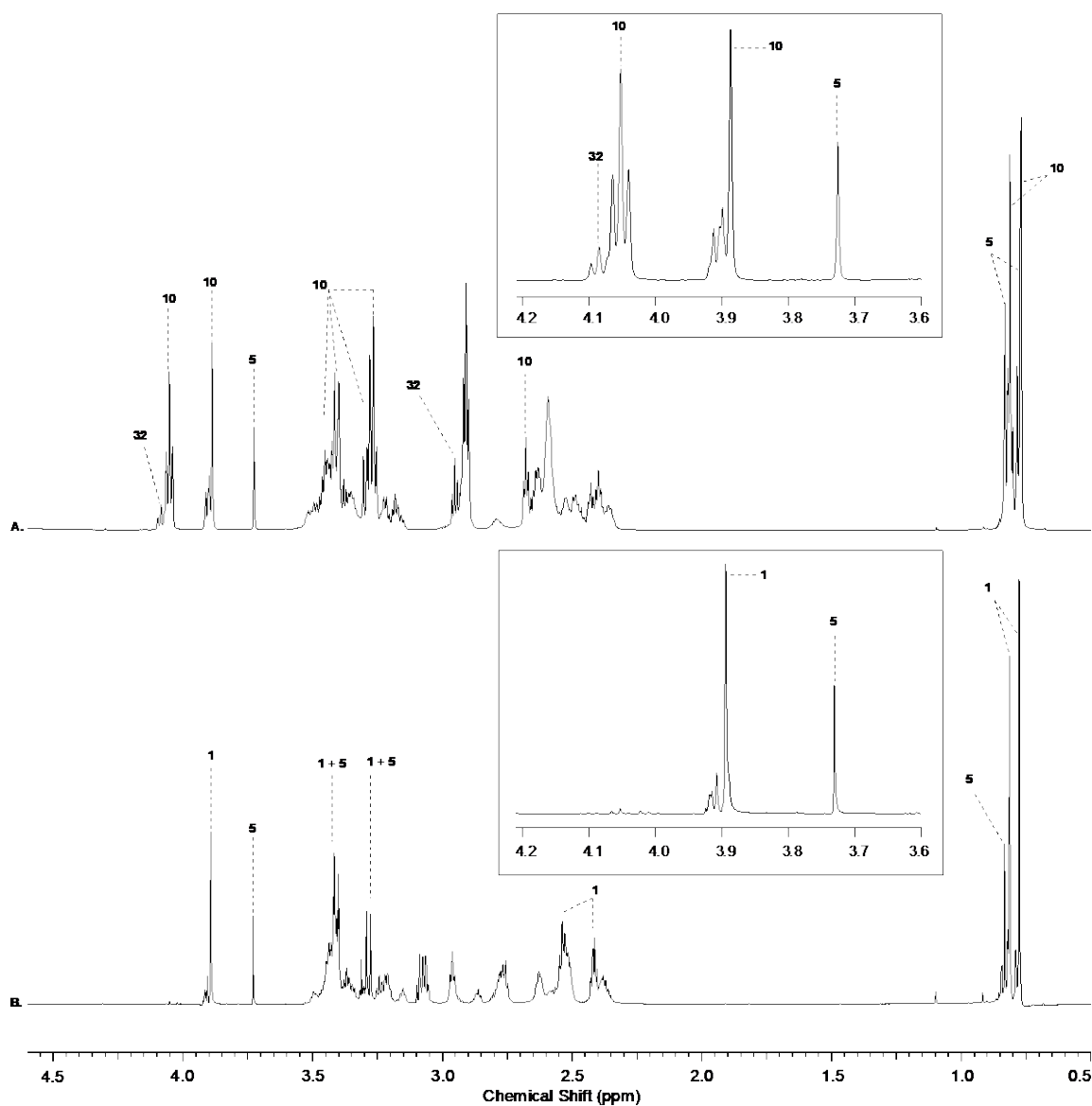
Supplementary Figure 47. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–4.5 ppm) spectra to show the acid-catalysed hydrolysis of thiazoline **10** to pantetheine **1**. (A.) The multicomponent one-pot synthesis of **10** by reaction of pantolactone **2** (500 mM), cysteamine **4** (2 equiv.) and β-alanine nitrile **8** (2 equiv.) in phosphate buffer (pH 9, 500 mM) at 20 °C after 4 days (as shown in Supplementary Figure 46D); (B.) after incubating the crude reaction mixture at pH 4 and 20 °C for 1 day, followed by NMR analysis at pH 9. Inset: Expanded ^{1}H NMR spectral region between 3.7–4.4 ppm.

Competitive multicomponent one-pot synthesis of thiazoline 10 by reaction of pantolactone 2, β-alanine nitrile 8, and cysteamine 4 in the presence of glycine nitrile 7$_G$
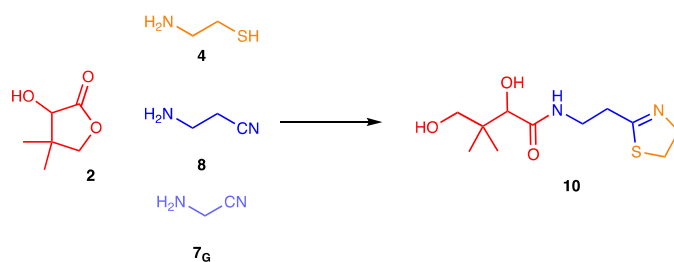
Pantolactone **2** (500 mM), cysteamine **4** (2 equiv.), β-alanine nitrile **8** (2 equiv.), and glycine nitrile **7$_G$** (2 equiv.) were dissolved in phosphate buffer (pH 7.0; 500 mM) at 20 °C, and the solution was adjusted to pH 9 with 1–4 M NaOH. The reaction was periodically analysed by NMR spectroscopy. After 5 days, thiazoline **10** (33%) and pantoic acid **5** (35%) were observed as the major pantoyl species (See Supplementary Figure 48).
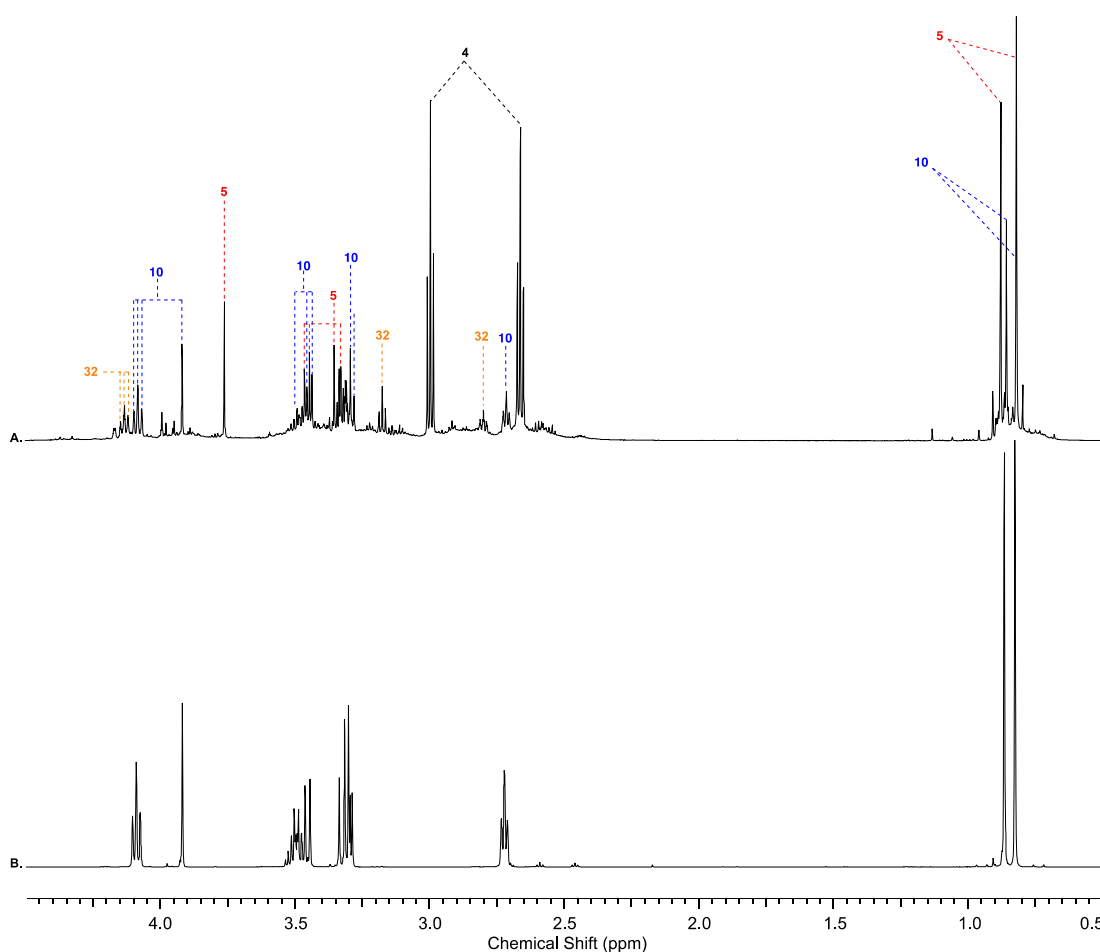
Supplementary Figure 48. ¹H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–4.5 ppm) spectra to show the multicomponent one-pot synthesis of thiazoline **10** by reaction of pantolactone **2** (500 mM), cysteamine **4** (2 equiv.), β-alanine nitrile **8** (2 equiv.) and glycine nitrile **7$_G$** (2 equiv.) in phosphate buffer (pH 9; 500 mM) at 20 °C after: (A.) 5 days; and (B.) authentic thiazoline **10**. See Supplementary Figure 49 for ¹H NMR time course spectra.
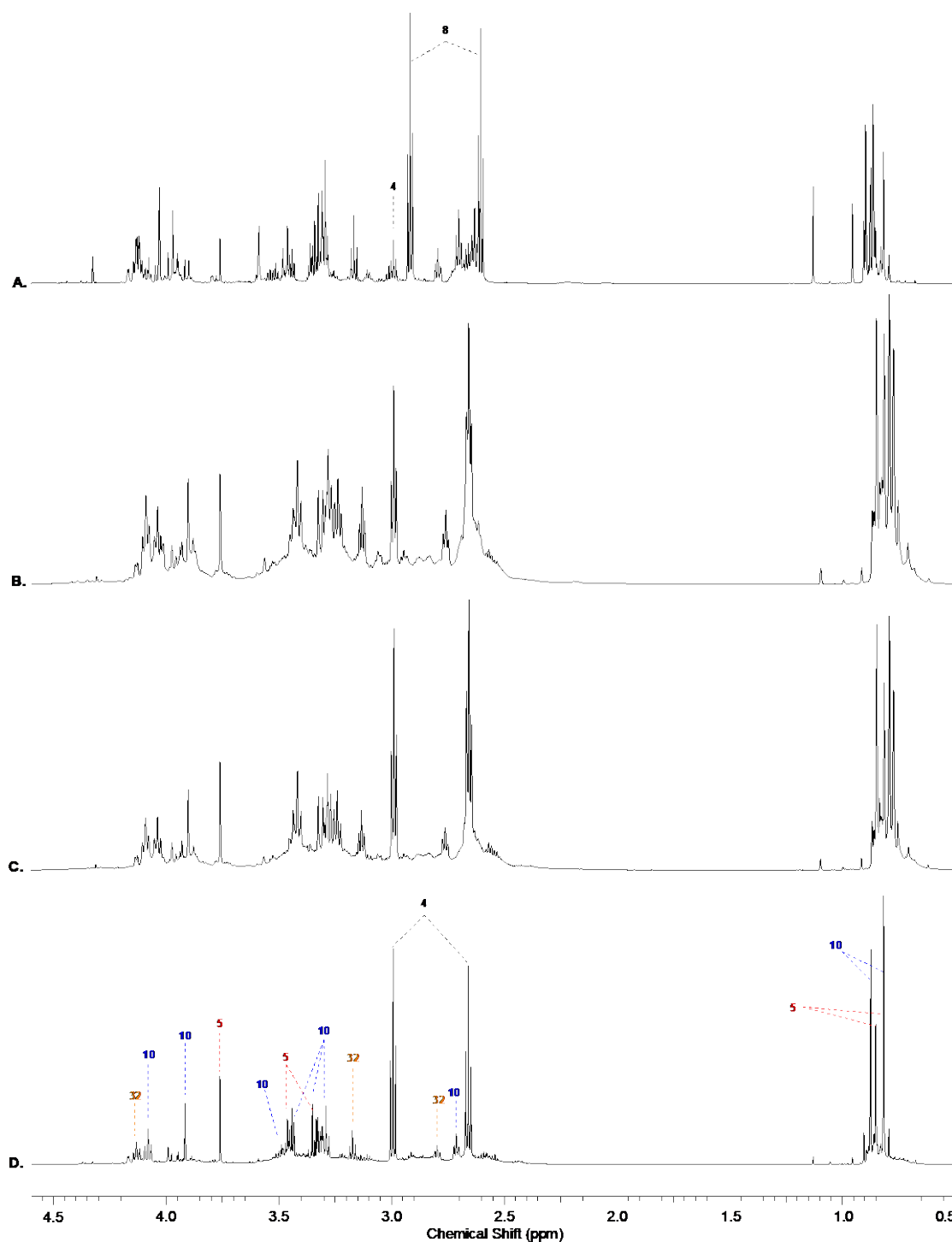
Supplementary Figure 49. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–4.5 ppm) spectra to show the multicomponent one-pot synthesis of thiazoline **10** by reaction of pantolactone **2** (500 mM), cysteamine **4** (2 equiv.), β-alanine nitrile **8** (2 equiv.) and glycine nitrile **7$_G$** (2 equiv.) in phosphate buffer (pH 9; 500 mM) at 20 ºC after: (A.) 1 hour; (B.) 1 day; (C.) 2 days; (D.) 5 days. See Supplementary Figure 48A for reproduction of ^{1}H NMR spectrum (D.) and comparison with authentic thiazoline **10** in Supplementary Figure 48B

Recovery of cysteamine 4 in a multicomponent reaction with glycine nitrile 7$_G$ and β-alanine nitrile 8
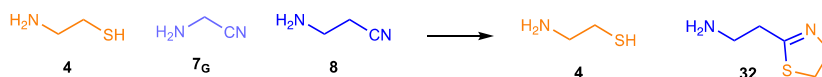
To a solution of glycine nitrile **7$_G$** (500 mM), β-alanine nitrile **8** (500 mM) and potassium hydrogen phthalate (250 mM; internal standard) in phosphate buffer (500 mM) at pH 9.0 and 20 ºC was added cysteamine **4** (1 M). The reaction was monitored periodically by NMR spectroscopy for 4 days. Cysteamine **4** was observed to initially deplete over 6 hours (1 hour [41%]; 6 hours [36%]), but its concentration increased thereafter (12 hours [48%]; 1 day [54%]; 2 days [65%]; 4 days [73%]; see Supplementary Figure 50 [**A.** black, left]). 2-Aminoethylthiazoline **32** (20%) was also detected after 4 days, and an orange suspension was also observed (see below: *Hydrolysis of crude mixture generated from the reaction of cysteamine 4 in a multicomponent reaction with glycine nitrile 7$_G$ and β-alanine nitrile 8*).

The recovery of **4** in this multicomponent reaction with aminonitrile **7$_G$** and **8** starkly contrasts the observed reaction of **4** with pantothenic acid nitrile **9**, where the formation of thiazoline **10** (73%) is observed – **10** is a key intermediate in the nitrile mediated synthesis of pantetheine **1** (see Supplementary Figure 50 [**A.** blue, right]). See Supplementary Page S42–52 for full experimental details for reactions of pantothenic acid nitrile **9** and cysteamine **4** to yield thiazoline **10**, and the conversion of **10** to pantetheine **1**.

Hydrolysis of crude mixture generated from the reaction of cysteamine 4 in a multicomponent reaction with glycine nitrile 7$_G$ and β-alanine nitrile 8
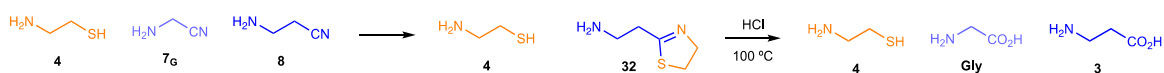
The crude reaction mixture generated from incubation of cysteamine **4** (1 M) in a multicomponent reaction with glycine nitrile **7$_G$** (500 mM) and β-alanine nitrile **8** (500 mM) in phosphate buffer (500 mM) at pH 9.0 and 20 ºC after 4 days (see Supplementary Figure 50A) was incubated in 6 M HCl at 100 ºC for 3 day. After cooling to 20 ºC, the solution pH was adjusted to pH 9.0 with 6 M NaOH and analysed by NMR spectroscopy. Glycine **Gly** (>90%), β-alanine **3** (>90%) and cysteamine **4** (>90%) were recovered (Supplementary Figure 51).
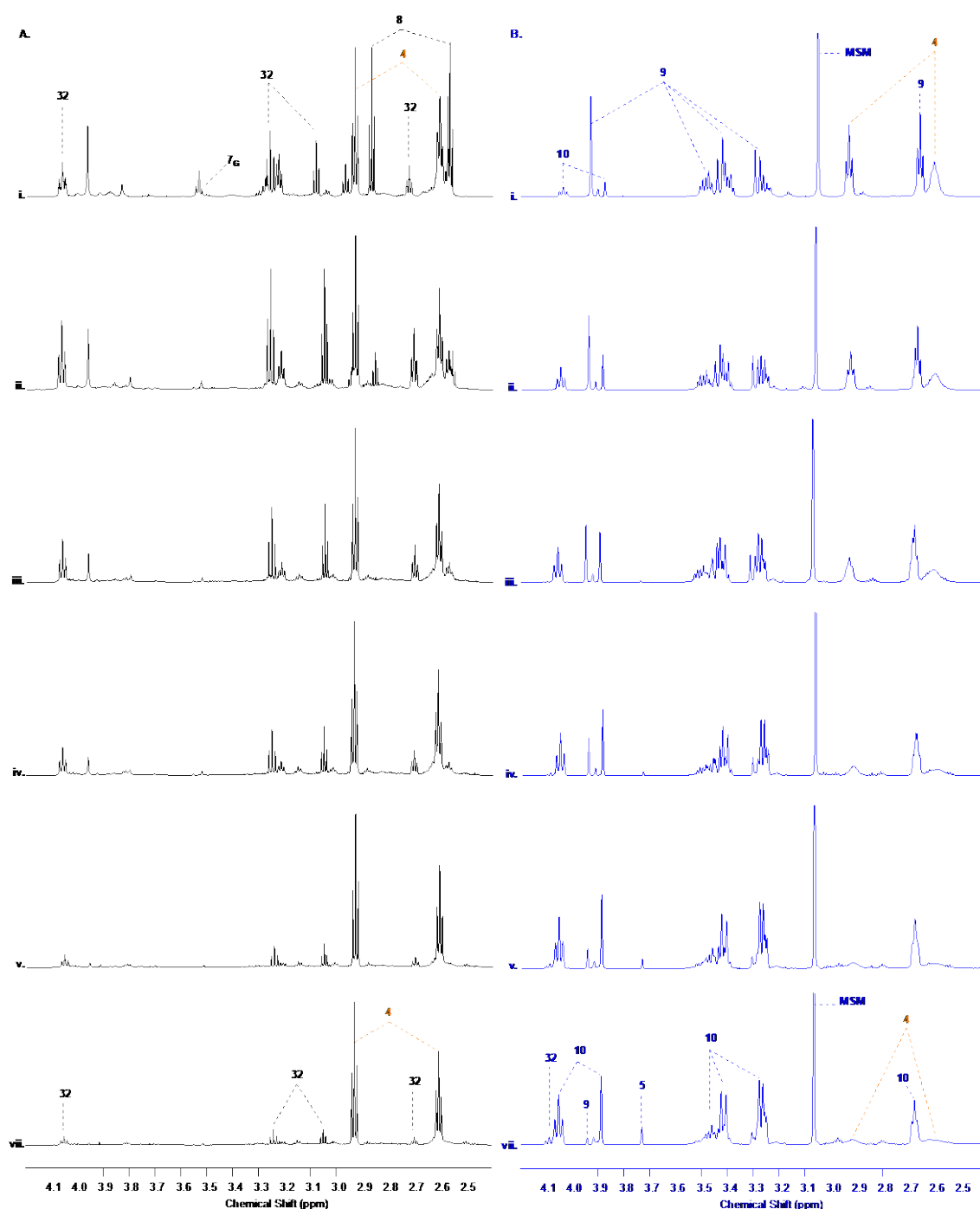
Supplementary Figure 50. ^{1}H NMR (H$_2$O, noesygppr1d, 2.4–4.2 ppm) spectra to show the contrasting reactions of cysteamine **4** with aminonitriles (i.e., glycine nitrile **7$_G$** and β-alanine nitrile **8**; spectrum **A.** [700 MHz, black])) and pantothenic acid nitrile **9** (spectrum **B.** [600 MHz, blue]). Incubation of **7$_G$** (500 mM; 1 equiv.), **8** (1 equiv.) with **4** (2 equiv.) in phosphate buffer (500 mM) at pH 9.0 and 20 ºC led to quantitative depletion of **7$_G$** and >90% depletion of **8** after 4 days. **4** also initially depleted, with only 36% of the initial **4** remaining after 6 hours. **4** was then observed to return and increase to 73% after 4 days. After 4 days, **4** was the major detectable species. However, the recovery of **4** was not observed during its stoichiometric reaction with amidonitrile **9** (500 mM) (spectrum **B.**). The reaction of **4** and **9** yielded thiazoline **10** (73%) – **10** is a key intermediate in pantetheine **1** synthesis (see text for details) – and pantoic acid **5** (18%). (**A.**) The reaction of **7$_G$** (500 mM) and **8** (500 mM) with **4** (1 M) in phosphate buffer (500 mM) at pH 9.0 and 20 ºC after: (i) 1 hour; (ii) 6 hours; (iii) 12 hours; (iv) 1 day; (v) 2 days; and (vi) 4 days. (**B.**) The reaction of **9** (500 mM) and **4** (550 mM) in phosphate buffer (500 mM) at pH 9.0 and 20 ºC after: (i) 1 hour; (ii) 6 hours; (iii) 12 hours; (iv) 1 day; (v) 2 days; and (vi) 4 days. See Supplementary Page S42–52 for full experimental details.
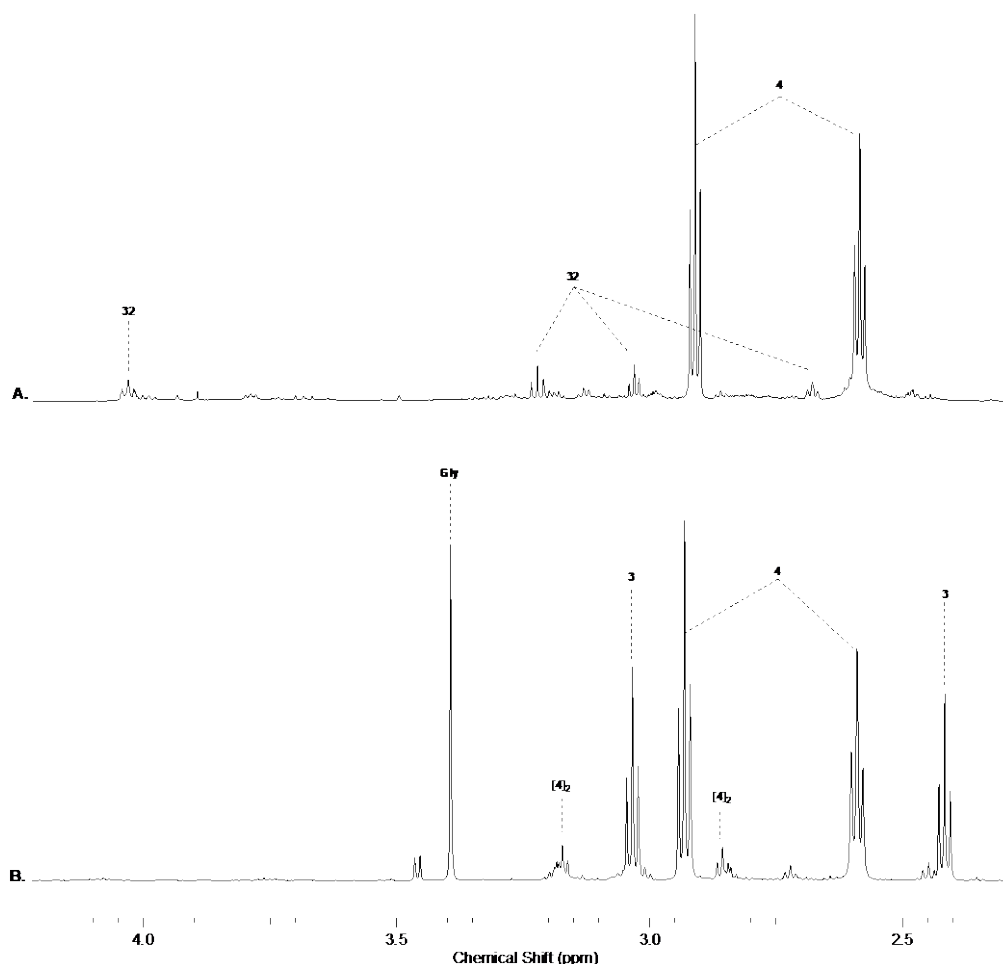
Supplementary Figure 51. ^{1}H NMR (600 MHz, H$_2$O, noesygppr1d, 2.3–4.2 ppm) spectra to show the (**A.**) the crude mixture after the reaction of glycine nitrile of **7$_G$** (500 mM), β-alanine nitrile **8** (500 mM) and cysteamine **4** (1 M) in phosphate buffer (500 mM) at pH 9.0 and 20 ºC and 4 days, followed by (**B.**) incubation of the crude reaction mixture (shown in spectrum A., and Supplementary Figure 50A) in 6 M HCl at 100 ºC for 24 hours, followed by cooling to 20 ºC, and adjustment of the solution pH to 9.0. Substantial recovery of glycine **Gly** (>90%) and β-alanine **3** (>90%) was observed in addition to cysteamine **4** and cysteamine disulfide **[4]$_2$** (**4** + **[4]$_2$**; >90%).

*Data for glycine **Gly** (Supplementary Figure 51)*

^{1}H NMR (600 MHz, H$_2$O) δ$_H$ 3.39 (s, 2H, (C2)–H$_2$). ^{13}C NMR (151 MHz, H$_2$O) δ$_C$ 174.7 (C1), 43.0 (C2).

*Data for β-alanine **3** (Supplementary Figure 51)*

^{1}H NMR (600 MHz, H$_2$O) δ$_H$ 3.03 (t, *J* = 6.7 Hz, 2H, (C3)–H$_2$), 2.42 (t, *J* = 6.7 Hz, 2H, (C2)–H$_2$). ^{13}C NMR (151 MHz, H$_2$O) δ$_C$ 179.5 (C1), 37.6 (C3), 34.7 (C2).

*Data for cysteamine **4** (Supplementary Figure 51)*

^{1}H NMR (600 MHz, H$_2$O) δ$_H$ 2.93 (t, *J* = 6.9 Hz, 2H, NH$_2$CH$_2$CH$_2$SH), 2.59 (t, *J* = 6.9 Hz, 2H, NH$_2$CH$_2$CH$_2$SH). ^{13}C NMR (151 MHz, H$_2$O) δ$_C$ 44.5 (NH$_2$CH$_2$CH$_2$SH), 23.0 (NH$_2$CH$_2$CH$_2$SH).

*Data for cysteamine disulfide **[4]$_2$** (Supplementary Figure 51)*

^{1}H NMR (600 MHz, H$_2$O) δ$_H$ 3.17 (t, *J* = 6.5 Hz, 4H, (NH$_2$CH$_2$CH$_2$S–S), 2.86 (t, *J* = 6.5 Hz, 4H, NH$_2$CH$_2$CH$_2$S–S).

Reaction of pantolactone 2 with γ-aminobutyronitrile 11
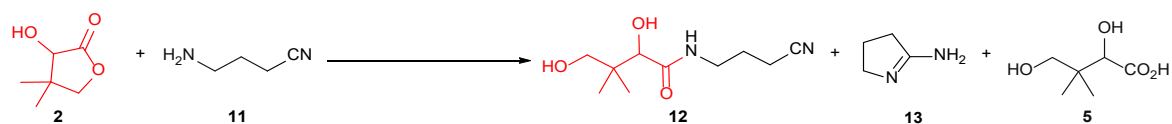
Pantolactone **2** (100 - 500 mM) and γ-aminobutyronitrile **11** (1 - 3 equiv.) were dissolved in phosphate buffer (pH 7.0; 500 mM) at 20 ºC, and the solution was adjusted to pH 9.0 with 1-4 M NaOH. The resulting solution was periodically analysed by NMR spectroscopy and its pH monitored. Yields of nitrile **12** and amidine **13** are reported in Supplementary Table 11. ^{1}H NMR spectra are shown in Supplementary Figure 52.

Concentration (mM)		Time (days)	Yield (%)			
2	**11**		**2**	**12**	**13**	**5**
100	100	9	34	1	7	64
100	200	14	21	4	15	75
100	300	14	18	6	21	71
500	500	8	12	6	6	76
500	1000	6	11	21	11	68

Supplementary Table 11. Quantification of species observed from the reaction of pantolactone **2** with γ-aminobutyronitrile **11** in phosphate buffer (pH 9.0; 500 mM) at 20 ºC.
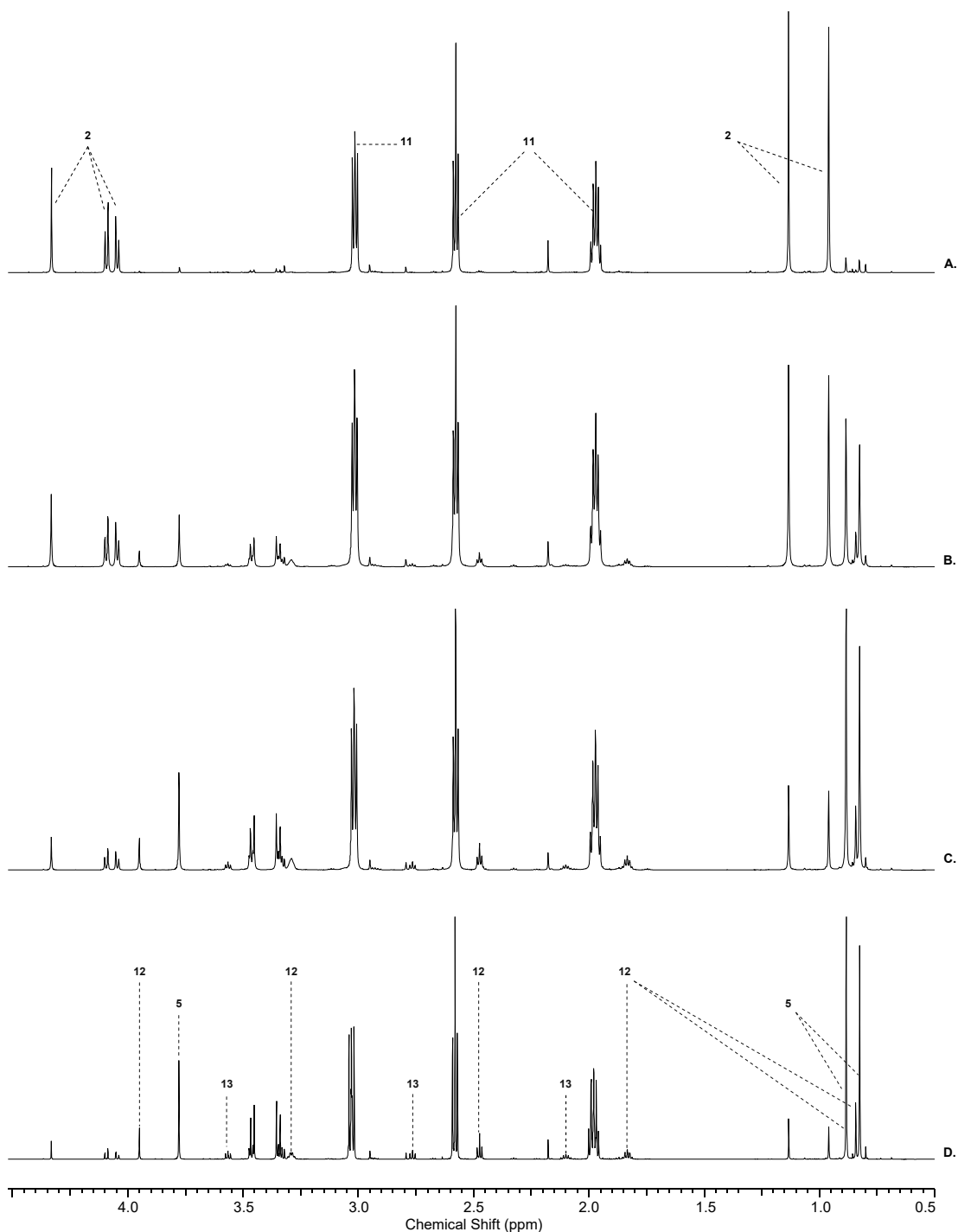
Supplementary Figure 52. ¹H NMR (700 MHz, H_2O/D_2O 9:1, noesygppr1d, 0.5–4.5 ppm) spectra to show the reaction of pantolactone **2** (500 mM) and γ-aminobutyronitrile **11** (2 equiv.) in phosphate buffer (pH 9.0; 500 mM) at 20 ºC after: (A.) 10 min; (B.) 1 day; (C.) 4 days; and (D.) 6 days.

Data for amidine **13**

¹H NMR (700 MHz, H_2O/D_2O 9:1, noesygppr1d, pH 9) δ_H 3.57 (t, 2H, *J* = 7.2 Hz, α-CH₂), 2.77 (t, 2H, *J* = 8.1 Hz, γ-CH₂), 2.13 (tt, *J* = 7.2, 8.1 Hz, 1H, β-CH₂). ¹³C NMR (125 MHz, H_2O/D_2O 9:1) δ_C 171.5 (H₂NC), 47.5 (α-CH₂), 30.3 (γ-CH₂), 20.3 (β-CH₂).

*Determination of γ-aminobutyronitrile **11** pK$_{aH}$*

11
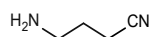
γ-Aminobutyronitrile **11** (100 mM) and (methanesulfonyl)methane (MSM; 50 mM, internal standard) were dissolved in H_2O/D_2O (99:1). The solution was adjusted to the specified pH using HCl/NaOH (1–4 M) and analysed by ^{1}H NMR spectroscopy. ^{1}H NMR spectra were reference to the MSM methyl resonance (δ_H = 3.10 ppm) and the chemical shift of the γ-aminobutyronitrile γ-methylene ($\delta_{γ\text{-}CH2}$) resonance measured. (Supplementary Figure 53)
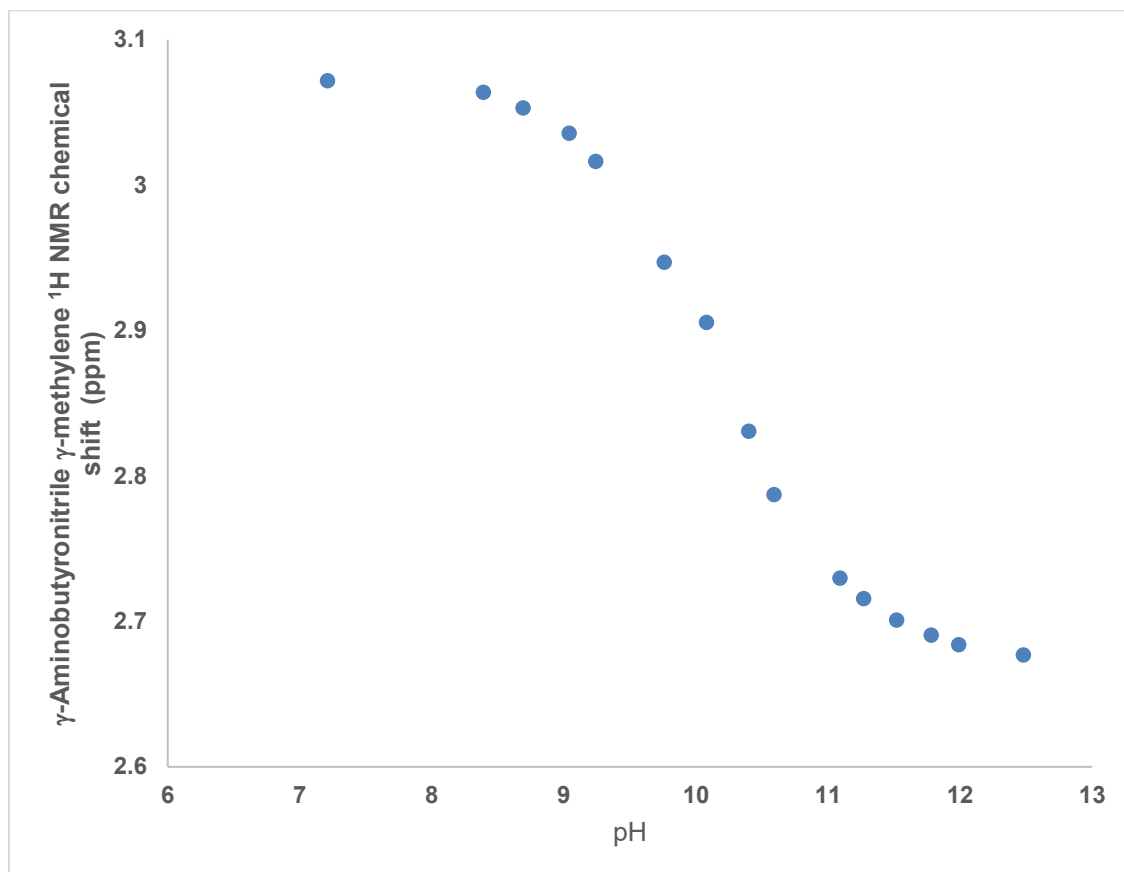
Supplementary Figure 53. γ-Aminobutyronitrile **11** γ-methylene ^{1}H NMR resonance chemical shift change ($\delta_{γ\text{-}CH2}$) observed upon pH titration of γ-aminobutyronitrile **11** (γ-CH$_2$NH$_3$+ → γ-CH$_2$NH$_2$).

Competitive reaction of pantolactone 2 with β-alanine nitrile 8 and γ-aminobutyronitrile 11
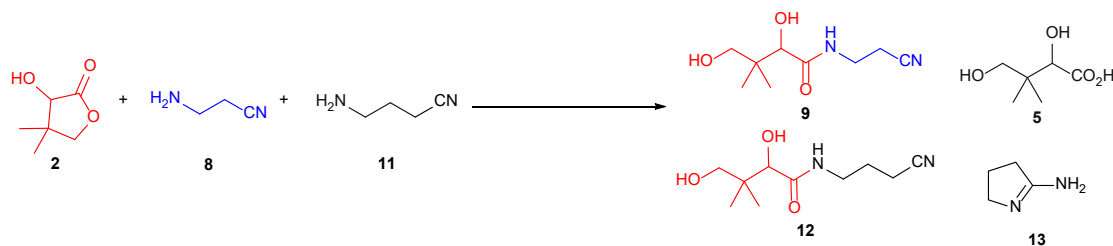
Pantolactone **2** (100 – 500 mM), β-alanine nitrile **8** (1 – 2 equiv.) and γ-aminobutyronitrile **11** (1 – 2 equiv.) were dissolved in phosphate buffer (pH 7.0; 500 mM) at 20 ºC. The solution was adjusted pH 9.0 with 1-4 M NaOH. The resulting solution was periodically analysed by NMR spectroscopy and its pH monitored. Yields of nitriles **8**, **12** and **13** are reported in Supplementary Table 12. ^{1}H NMR spectra are shown in Supplementary Figure 54.

Concentration (mM)			Time	Yield (%)				
2	**8**	**11**	(days)	**2**	**9**	**12**	**13**	**5**
100	200	200	7	1	59	7	21	34
500	500	500	4	1	60	10	8	38
500	1000	1000	2	0	71	10	11	19

Supplementary Table 12. Quantification of species observed from the reaction of pantolactone **2**, β-alanine nitrile **8** and γ-aminobutyronitrile **11** in phosphate buffer (pH 9.0; 500 mM) at 20 ºC.
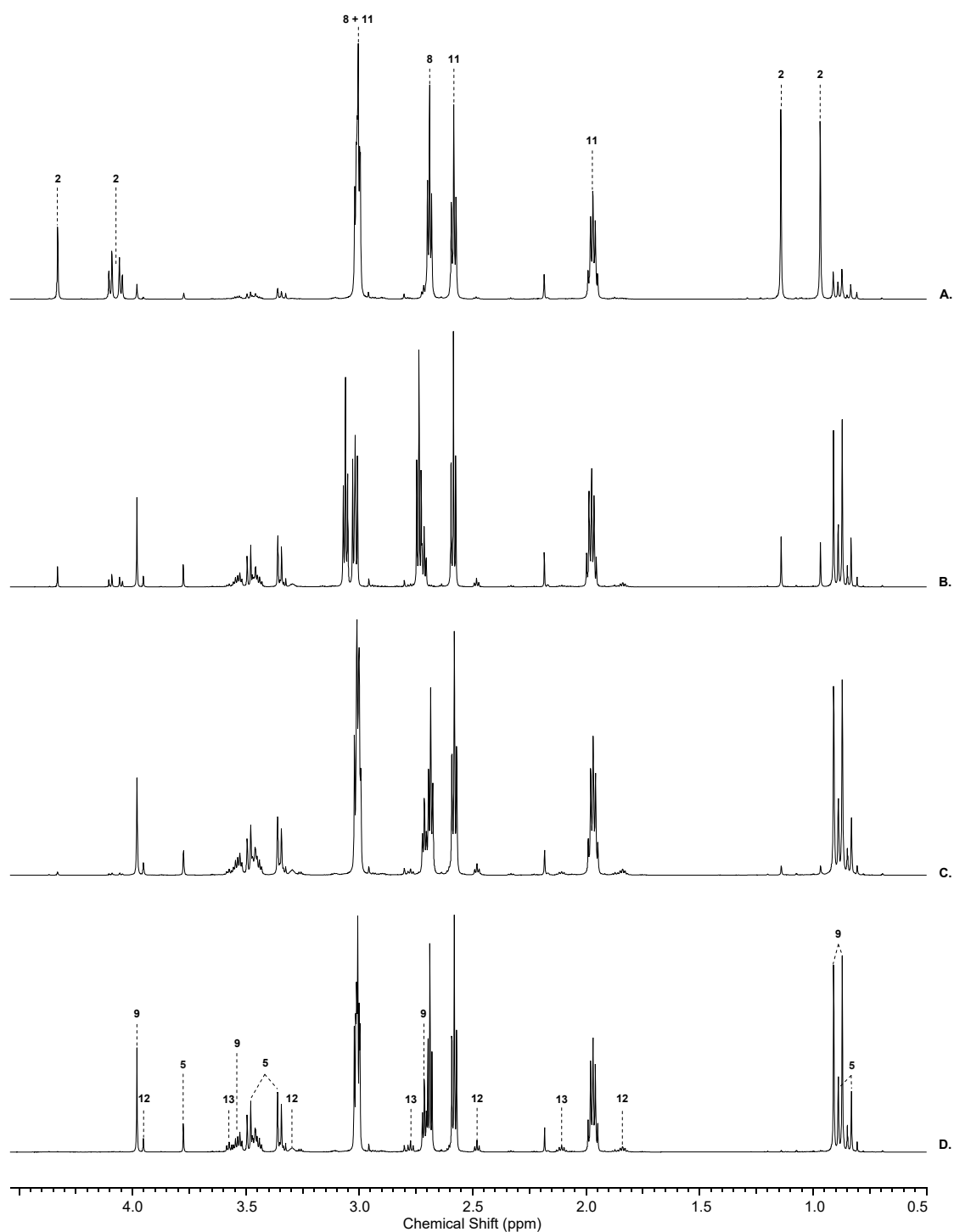
Supplementary Figure 54. ^{1}H NMR (700 MHz, H$_2$O/D$_2$O, noesygppr1d, 0.5–4.5 ppm) to show the reaction of pantolactone **2** (500 mM), β-alanine nitrile **8** (2 equiv.) and γ-aminobutyronitrile **11** (2 equiv.) in phosphate buffer (pH 9.0; 500 mM) at 20 °C after: (A.) 10 mins; (B.) 12 hours; (C.) 24 hours; (D.) 48 hours.

3. Chemoselective aldol synthesis of hydroxypivaldehyde 15 and pantolactone 2 at neutral pH

Incubation of pantoic acid nitrile 24 at varied pH

Reaction of pantoic acid nitrile 24 at pH 7.0 and below
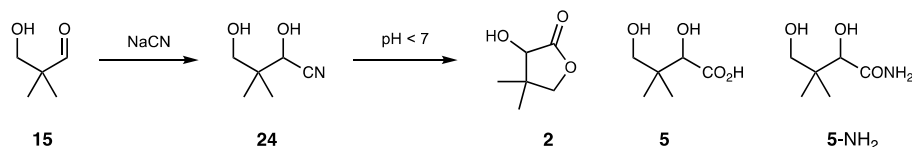
To a stirred solution of hydroxypivaldehyde **15** (50 mM) in phosphate buffer (500 mM) at pH 7.0 and 20 ºC was added sodium cyanide (1.5 equiv., 75 mM). The solution was adjusted from pH 7.5 to 7.0 with 8 M HCl and NMR spectra were acquired to show quantitative conversion of hydroxypivaldehyde **15** to pantoic acid nitrile **24**. The required solution pH was obtained using concentrated HCl solution (38%), and the reaction was monitored periodically NMR spectroscopy at the specified temperature. Yields are reported in Supplementary Table 13. Iminopantolactone **26** was not observed at pH 7.0 or below, but was observed at pH 7.5 and above (see **Error! Reference source not found.–Error! Reference source not found.**).

Reaction of pantoic acid nitrile 24 at pH 7.0 and above, and at room temperature.
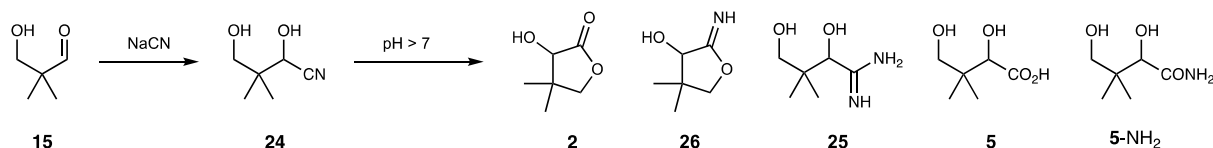
To a stirred solution of hydroxypivaldehyde **15** (50 mM) in phosphate buffer (pH 7.0; 500 mM) at 20 ºC was added sodium cyanide (1.5 equiv., 75 mM). The solution was adjusted from pH 7.5 to 7.0 using 8 M HCl and NMR spectra were acquired to show quantitative conversion of hydroxypivaldehyde **15** to pantoic acid nitrile **24**. The required solution pH was obtained using 1-8 M NaOH, and the reaction was monitored periodically NMR spectroscopy. Pantoic acid nitrile **24** underwent pH-dependent conversion to mixtures of iminopantolactone **26** (detected only above pH 7.5), pantolactone **2**, pantoamidine **25** (detected only above pH 8.0), pantoamide **5-NH₂**, and pantoic acid **5**. Yields are reported in Error! Reference source not found. to Error! Reference source not found.. Representative ¹H NMR spectra for the reaction of pantoic acid nitrile **24** at pH 9.0 are shown in Supplementary Figure 55 and Supplementary Figure 56. See Supplementary Figure 76 for the formation of pantolactone **2** (54%) at pH 7.5 from **15** in the presence of β-alanine nitrile **8**.

*Data for pantoic acid nitrile **24***
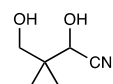
 ^{1}H NMR (700 MHz, H_2O, noesygppr1d, pH 9.0) δ_H 4.50 (s, 1H, (C2)–H), 3.46 (AB, *J* = 11.5 Hz, 1H, (C4)–H), 3.37 (AB, *J* = 11.5 Hz, 1H, (C4)–H′), 0.95 (s, 3H, CH$_3$), 0.94 (s, 3H, CH$_3$′).

*Data for pantoic acid **5***
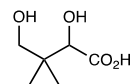
 ^{1}H NMR (700 MHz, H_2O, noesygppr1d, pH 9.0) δ_H 3.75 (s, 1H, (C2)–H), 3.44 (AB, *J* = 11.0 Hz, 1H, (C4)–H), 3.33 (AB, *J* = 11.0 Hz, 1H′), 0.86 (s, 3H, CH$_3$), 0.80 (s, 3H, CH$_3$′).

*Data for pantolactone **2***
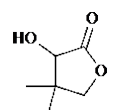
 ^{1}H NMR (700 MHz, H_2O, noesygppr1d, pH 9.0) δ_H 4.31 (s, 1H, (C2)–H), 4.07 (AB, *J* = 9.0 Hz, 1H, (C4)–H), 4.02 (AB, *J* = 9.0 Hz, 1H, (C4)–H′), 1.12 (s, 3H, CH$_3$), 0.94 (s, 3H, CH$_3$′).

*Data for iminopantolactone **26***
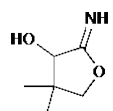
 ^{1}H NMR (700 MHz, H_2O, noesygppr1d, pH 9.0) δ_H 4.23 (s, 1H, (C2)–H), 3.96 (AB, *J* = 9.0 Hz, 1H, (C4)–H), 3.91 (AB, *J* = 9.0 Hz, 1H, (C4)–H′), 1.07 (s, 3H, CH$_3$), 0.89 (s, 3H, CH$_3$′).

*Data for pantoamide **5**-NH$_2$*
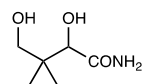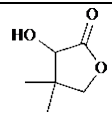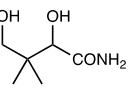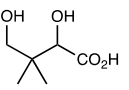
 ^{1}H NMR (700 MHz, H_2O, noesygppr1d, pH 9.0) δ_H 3.95 (s, 1H, (C2)–H), 3.47 (d, *J* = 11.2 Hz, 1H, (C4)–H), 3.33 (d, *J* = 11.2 Hz, 1H, (C4)–H′, overlapping with (C4)–H′ of pantoic acid **5**), 0.88 (s, 3H, CH$_3$), 0.84 (s, 3H, CH$_3$′).

*Data for pantoamidine **25***
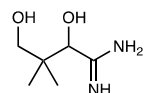
 ^{1}H NMR (700 MHz, H_2O, noesygppr1d, pH 9.0, partial assignment) δ_H 4.31 (s, 1H, (C2)–H).

pH	Temp (°C)	Time (days)	Yield (%)			
			24	2	5-NH$_2$	5
-0.6*	25	2	82	18	0	0
-0.6*	25	3	75	24	0	0
-0.6*	25	4	64	36	0	0
-0.6*	25	7	34	66	0	0
-0.6*	40	1	10	90	0	0
-0.6*	60	0.25	0	>95	0	0
1.5	60	1	95	4	0	0
3.0	60	1	>95	tr.	tr.	0
5.0	60	1	89	9	tr.	tr.
7.0	60	1	0	15	6	79

Supplementary Table 13. Conditions and quantification of species observed from the reaction of hydroxypivaldehyde **15** (50 mM) with sodium cyanide (NaCN; 75 mM) and phosphate buffer (500 mM). * 4 M HCl solution. tr. = trace.
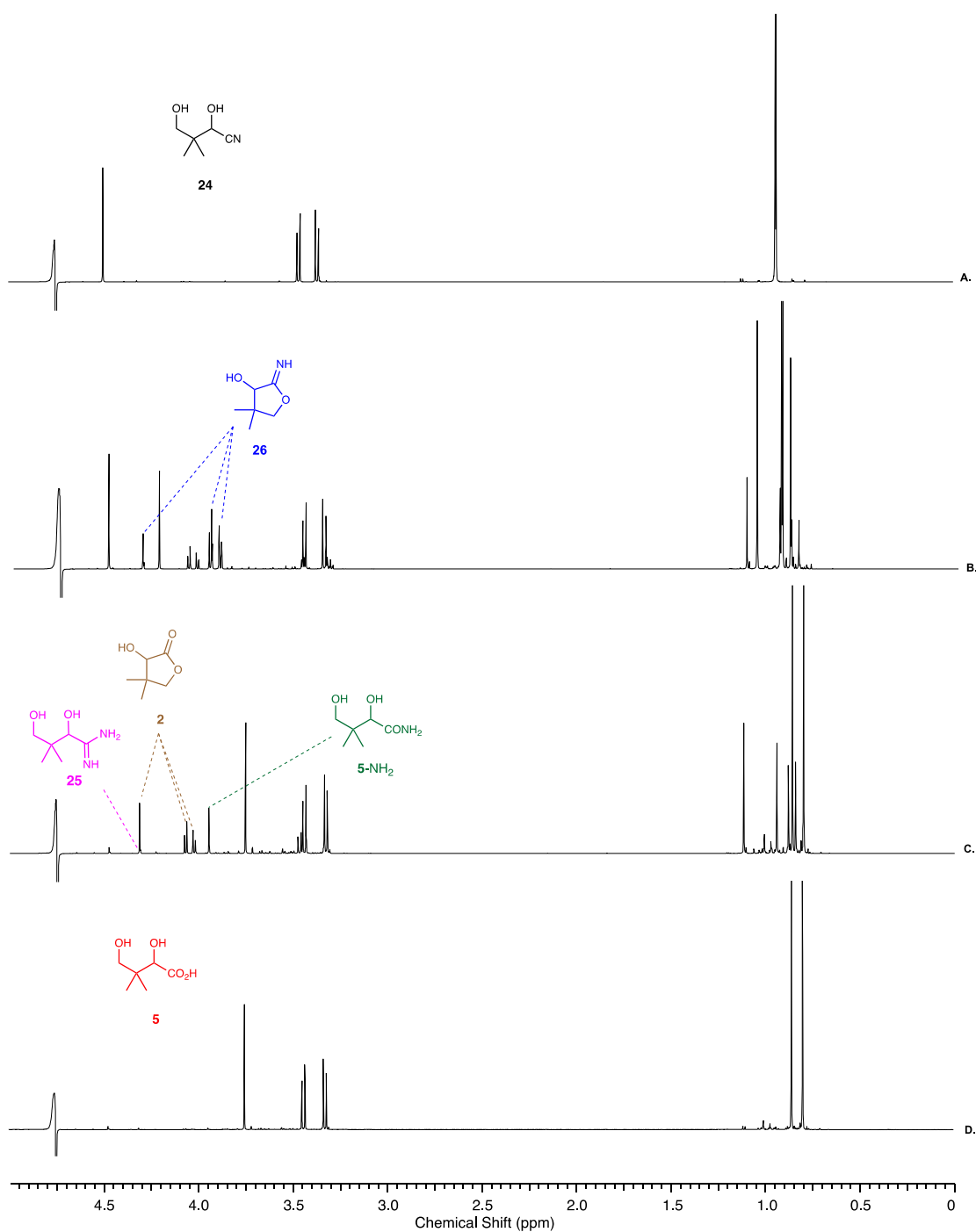
Supplementary Figure 55. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0–5 ppm) spectra to show the (A.) formation of pantoic acid nitrile **24** from hydroxypivaldehyde **15** (50 mM) and sodium cyanide (75 mM) in phosphate buffer (500 mM; pH 7.0) at 20 ºC, followed by an increase in the solution to pH 9.0 and further incubation at 20 ºC, after (B.) 2 hours; (C.) 24 hours; (D.) 8 days. See Supplementary Figure 56 for an expanded ^{1}H NMR spectrum between 3.85–4.85 ppm.
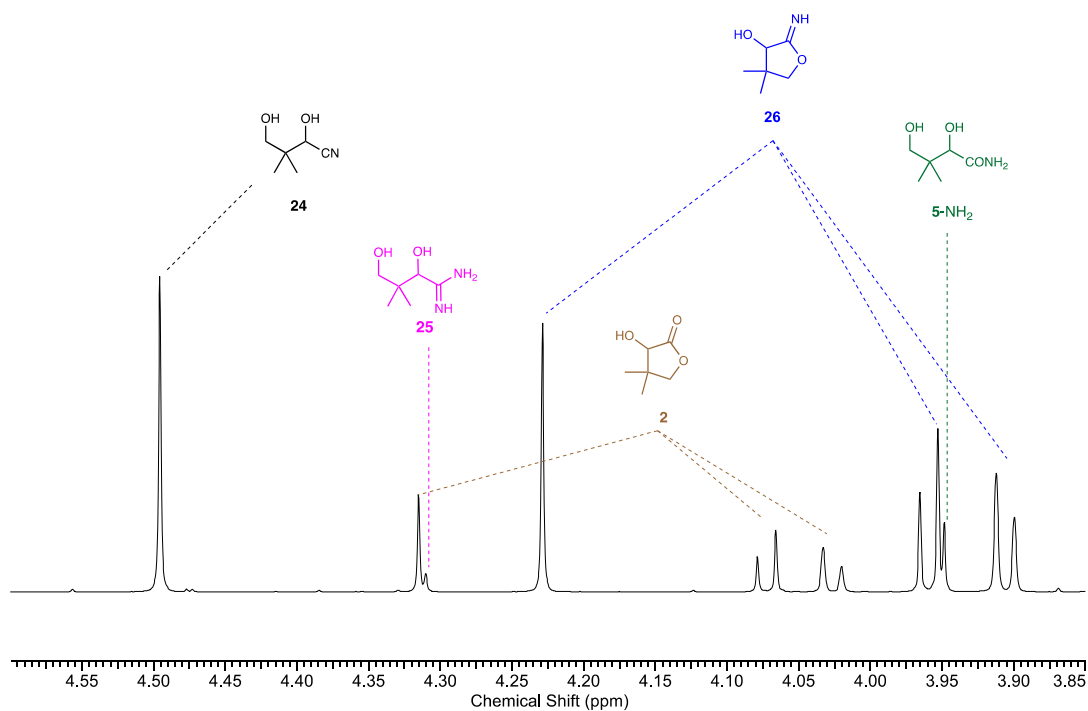
Supplementary Figure 56. Expanded ^{1}H NMR (700 MHz, H_2O, noesygppr1d, 3.85–4.85 ppm) spectrum after 2 hours of incubation of pantoic acid nitrile **24** in phosphate buffer (pH 9.0; 500 mM) at 20 ºC. A mixture of pantoic acid nitrile **24**, pantoamidine **25**, iminopantolactone **26**, pantolactone **2** and pantoamide **5**-NH$_2$ were detected. See Supplementary Figure 55C for ^{1}H NMR spectrum between 0–5 ppm.

Crossed-aldol reaction of isobutyraldehyde 17 with formaldehyde 16
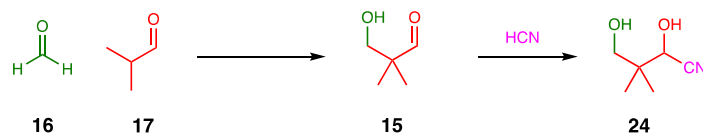
Isobutyraldehyde **17** (1.6–50 mM) and formaldehyde **16** (2.7–85 mM; 37% wt.% in water) were incubated in water or phosphate buffer (16–500 mM) at a specified temperature and pH, and analysed periodically by NMR spectroscopy. Hydroxypivaldehyde **15** formation was quantified by the disappearance of the ^{1}H NMR (C2)–H resonances of isobutyraldehyde **17** (and its hydrate **17**·H_2O) (1.50–2.50 ppm; Supplementary Figure 57), relative to the combined total integral of CH_3 region at 0.50–1.10 ppm.[*] Hydroxypivaldehyde **15** formation and quantification was further verified by its quantitative conversion to pantoic acid nitrile **24** (see below *'Quantification of hydroxypivaldehyde 15 by conversion to pantoic acid nitrile 24'*).[†]

Quantification of hydroxypivaldehyde 15 by conversion to pantoic acid nitrile 24
Sodium cyanide (NaCN; 3 equiv.; 150 mM, relative to starting isobutyraldehyde **17** concentration) was added to the crude aldol mixture at 20 ºC at pH 7.0. The resulting solution was adjusted from pH 7.2–7.5 to pH 7.0 with 1–4 M HCl and analysed by NMR spectroscopy to show the quantitative conversion of residual isobutyraldehyde **17** to 2-hydroxy-3-methylbutane nitrile **21**, formaldehyde **16** to glycolonitrile **20**, and hydroxypivaldehyde **15** to pantoic acid nitrile **24** (Supplementary Figure 57). Yields of pantoic acid nitrile **24** are given in Supplementary Table 14 (20 ºC), Supplementary Table 15 (40 ºC), and Supplementary Table 16 (60 ºC).

17 (mM)	16 (mM)	P_i (equiv.)	pH	Time (hours)	15 (%)
50	80	-	4.3	168	6
50	80	-	7	168	2
50	80	-	9	168	4
50	80	-	9.8	168	15
50	80	10	7	120	28
50	85	10	7	528	55
25	42.5	10	7	528	40
12.5	21.3	10	7	528	30
6.3	10.6	10	7	528	20
3.1	5.3	10	7	528	13
1.6	2.7	10	7	528	0

Supplementary Table 14. Yields and conditions for the formation of hydroxypivaldehyde **15** by crossed-aldolisation of isobutyraldehyde **17** and formaldehyde **16** at 20 ºC in water or phosphate buffer. Hydroxypivaldehyde **15** yields were further confirmed by subsequent quantitative conversion to pantoic acid nitrile **24** by the addition of sodium cyanide (NaCN) to the crude hydroxypivaldehyde **15** at pH 7.0 and 20 ºC.

---

[*] Hydroxypivaldehyde **15** exists as a concentration-dependent mixture of aldehyde, hydrates, and (hemi)acetals, and the rapid and quantitative conversion to pantoic acid nitrile **24** allowed unambiguous quantification of the preceding aldol condensation of isobutyraldehyde **17** and formaldehyde **16**.
[†] Residual isobutyraldehyde **17** and formaldehyde **16** were also quantitatively converted to their cyanohydrins (isobutyraldehyde **17** → 2-hydroxy-3-methylbutane nitrile **21**; formaldehyde **16** → glycolonitrile **20**).

17 (mM)	16 (mM)	Time (hours)	15 (%)
50	80	168	83
50	85	144	77
25	42.5	144	64
12.5	21.3	144	46
6.3	10.6	144	32
3.1	5.3	144	13
1.6	2.7	144	12

Supplementary Table 15. Yields and conditions for the formation of hydroxypivaldehyde **15** by crossed-aldolisation of isobutyraldehyde **17** and formaldehyde **16** at 40 ºC in phosphate buffer (10 equiv.) at pH 7. Hydroxypivaldehyde **15** yields were further confirmed by subsequent quantitative conversion to pantoic acid nitrile **24** by the addition of sodium cyanide (NaCN) to hydroxypivaldehyde **15** at pH 7.0 and 20 ºC.

17 (mM)	16 (mM)	P_i (equiv.)	pH	Temp (°C)	Time (hours)	15 (%)
50	80	-	4.3	60	168	15
50	80	-	7	60	168	21
50	80	-	9	60	168	19
50	80	10	7	60	48	94
50	85	10	7	60	48	92
25	42.5	10	7	60	72	91
12.5	21.3	10	7	60	72	83
6.3	10.6	10	7	60	120	80
3.1	5.3	10	7	60	48	46
1.6	2.7	10	7	60	672	78

Supplementary Table 16. Yields and conditions for the formation of hydroxypivaldehyde **15** by crossed-aldolisation of isobutyraldehyde **17** and formaldehyde **16** at 60 ºC in water or phosphate buffer (10 equiv.) at pH 7. Hydroxypivaldehyde **15** yields were further confirmed by subsequent quantitative conversion to pantoic acid nitrile **24** by the addition of sodium cyanide (NaCN) to hydroxypivaldehyde **15** at pH 7.0 and 20 ºC.
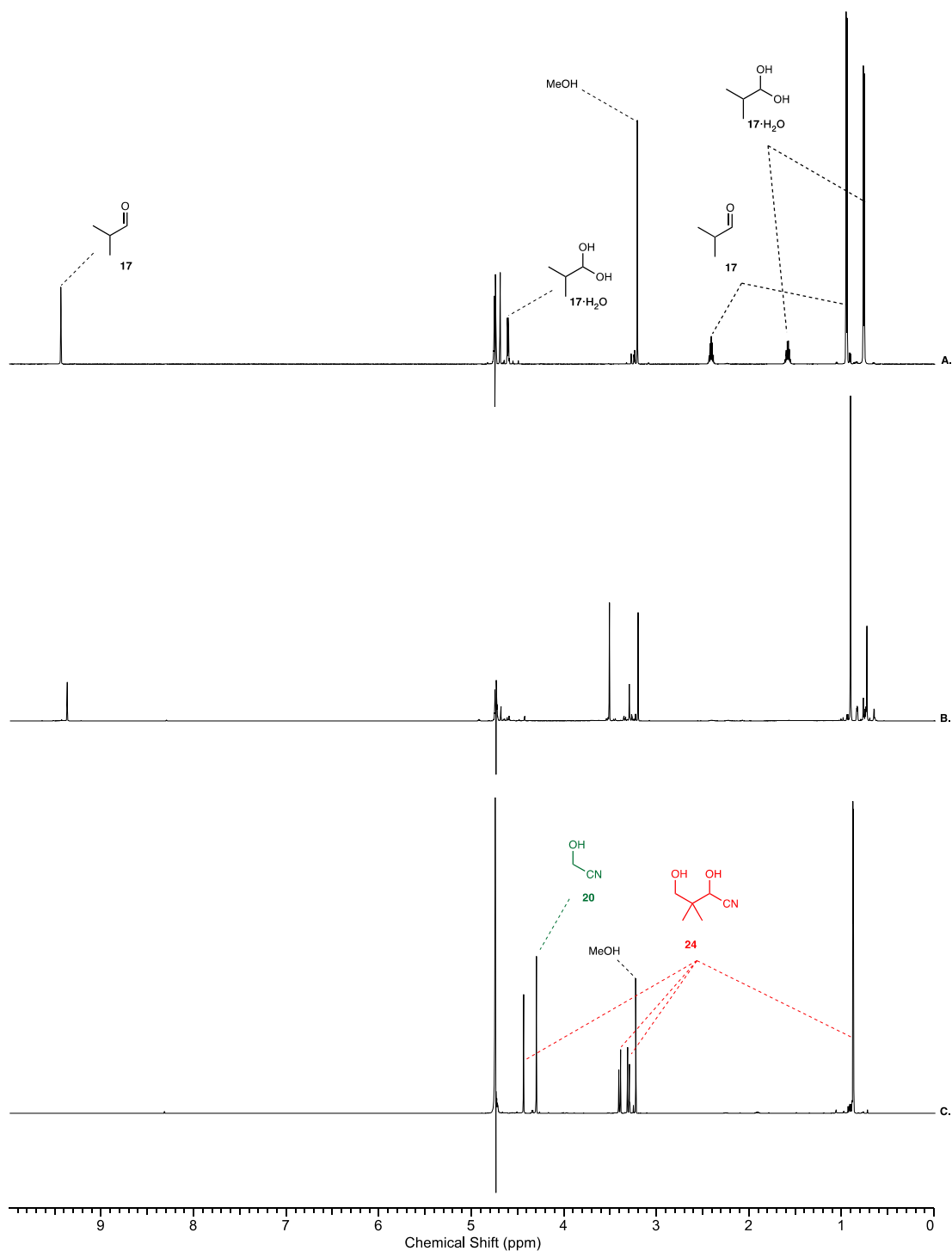
Supplementary Figure 57. ^{1}H NMR (600 MHz, H$_2$O, noesygppr1d, 0.0–10.0 ppm) spectra to show the reaction of isobutyraldehyde **17** (50 mM) and formaldehyde **16** (85 mM) in phosphate buffer (pH 7.0; 500 mM); **A.** before heating at 60 °C; **B**. after heating at 60 °C for 48 hours, showing the appearance of a new aldehyde species at the expense of isobutyraldehyde **17**; **C**. after addition of sodium cyanide (NaCN; 3 equiv.) at pH 7.0 and 20 ºC, showing the formation of pantoic acid nitrile **24** (94%) alongside the conversion of residual formaldehyde **16** to glycolonitrile **20**, and residual isobutyraldehyde **17** to 2-hydroxy-3-methylbutane nitrile **21**.
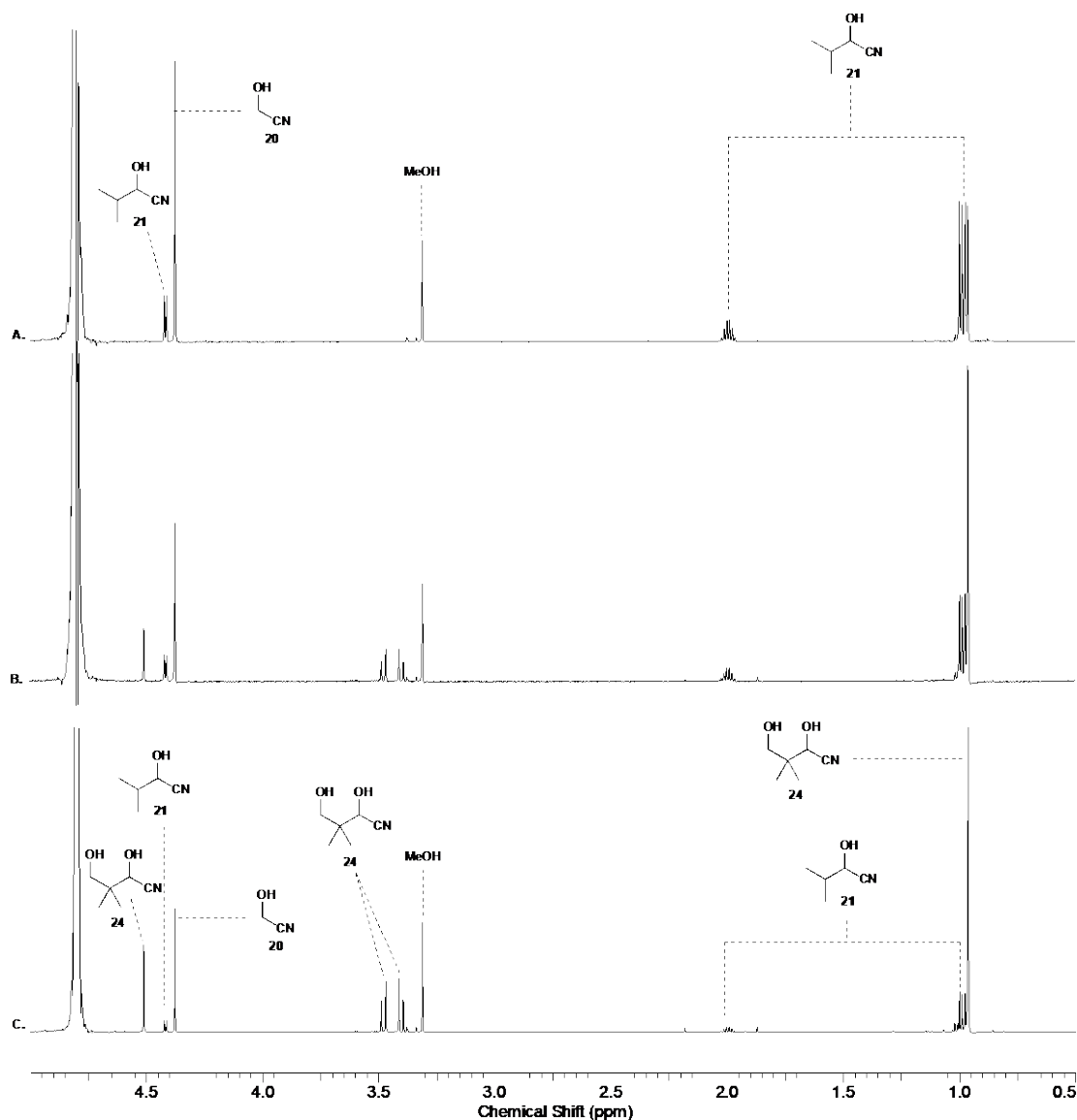
Supplementary Figure 58. ^{1}H NMR (600 MHz, H$_2$O, noesygppr1d, 0.5–5.0 ppm) spectra to show the cyanohydrin mixtures obtained after the addition of sodium cyanide (NaCN) at 20 ºC following the reactions of isobutyraldehyde **17** (1.6 mM) and formaldehyde **16** (2.8 mM) in phosphate buffer (pH 7.0; 16 mM) at 60 ºC. **A.** before heating at 60 °C, showing only glycolonitrile **20** (from formaldehyde **16**) and 2-hydroxy-3-methylbutane nitrile **21** (from isobutyraldehyde **17**); **B.** after heating at 60 °C for 8 days; and **C.** after heating at 60 °C for 28 days, resulting in the formation of hydroxypivaldehyde **15**, which is observed as pantoic acid nitrile **24** (78%). The conversion of residual formaldehyde **16** to glycolonitrile **20**, and residual isobutyraldehyde **17** to 2-hydroxy-3-methylbutane nitrile **21**, were observed.

*Attempted synthesis of 3-hydroxypropanal **HP** by aldol condensation of acetaldehyde **18** with formaldehyde **16***

Acetaldehyde **18** (32 mM) and formaldehyde **16** (50 mM; 37% wt.% in water) were incubated in phosphate buffer (pH 7.0, 500 mM) at 60 º C for 2 days, and analysed by ^{1}H NMR spectroscopy (Supplementary Figure 59). The lack of 3-hydroxypropanal **HP** formation was further verified by a competitive reaction of acetaldehyde **18** and isobutyraldehyde **17** with formaldehyde **16**, where isobutyraldehyde **17** was converted to hydroxypivaldehyde **15** in near-quantitative yield, but 3-hydroxypropanal **HP** was not observed. See Supplementary Page S83 for further details.
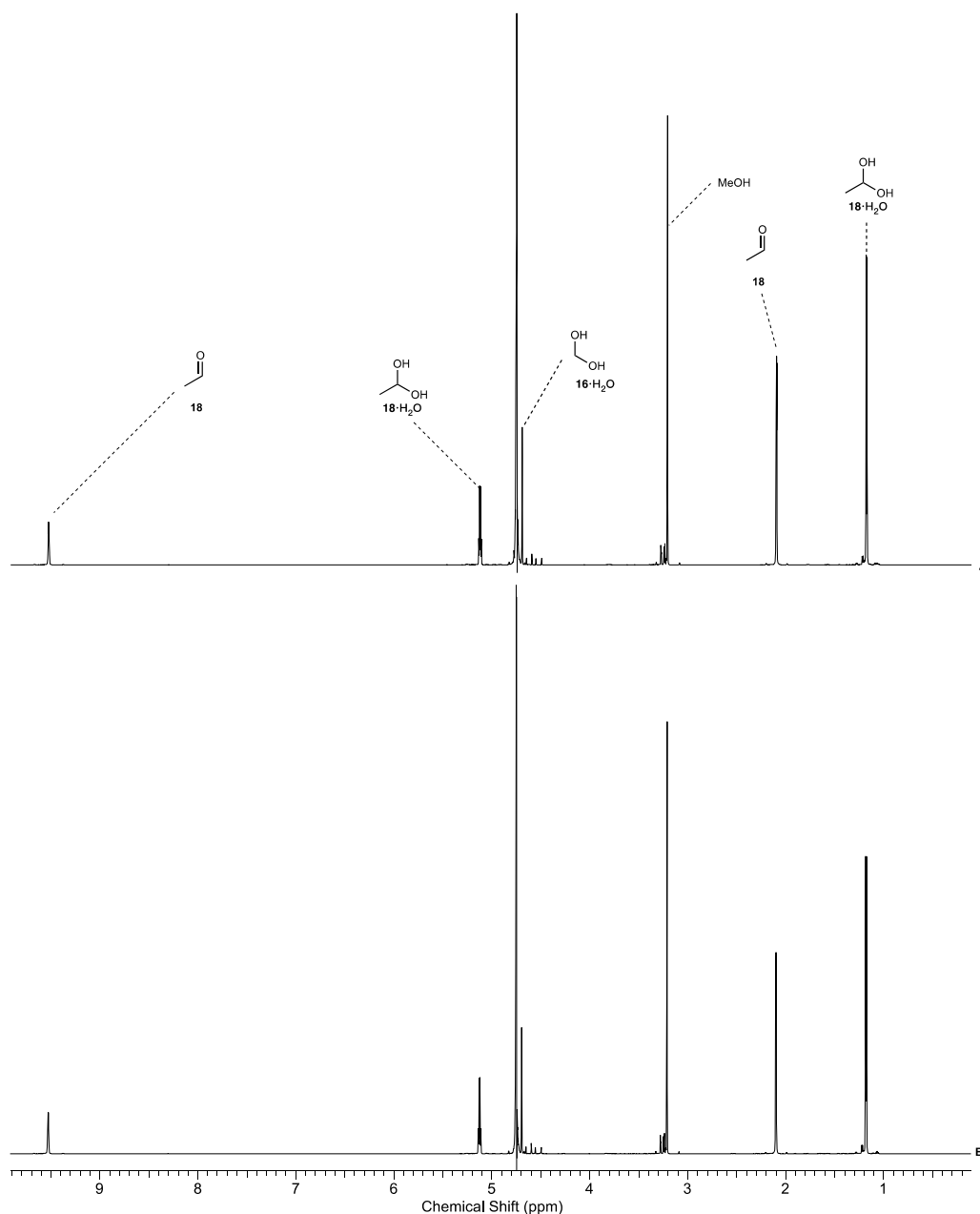
Supplementary Figure 59. ^{1}H NMR (600 MHz, H_2O, noesygppr1d, 0.5–10.0 ppm) spectra showing the reaction of acetaldehyde **18** (32 mM) and formaldehyde **16** (50 mM) in phosphate buffer (pH 7.0; 500 mM) after: (A.) 10 mins at room temperature; (B.) 2 days at 60 ºC.

Competitive aldol reaction of acetaldehyde 18 and isobutyraldehyde 17 with formaldehyde 16
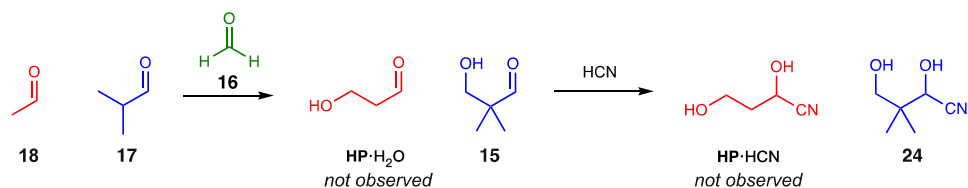
Acetaldehyde **18** (50 mM), isobutyraldehyde **17** (32 mM), formaldehyde **16** (100 mM; 37% wt.% in water) were incubated in phosphate buffer (pH 7.0; 500 mM) at 60 ºC for 48 hours. Sodium cyanide (NaCN; 200 mM) was added to the crude aldol mixture at 20 ºC and pH 7.0. The resulting solution was adjusted to pH 7.0 with 1–4 M HCl and analysed by NMR spectroscopy to show the quantitative conversion of residual isobutyraldehyde **17** to 2-hydroxy-3-methylbutane nitrile **21** (<5%), acetaldehyde **18** to lactonitrile **22** (>95%), formaldehyde **16** to glycolonitrile **20** (>95%), and hydroxypivaldehyde **15** to pantoic acid nitrile **24** (94%) (Supplementary Figure 60).
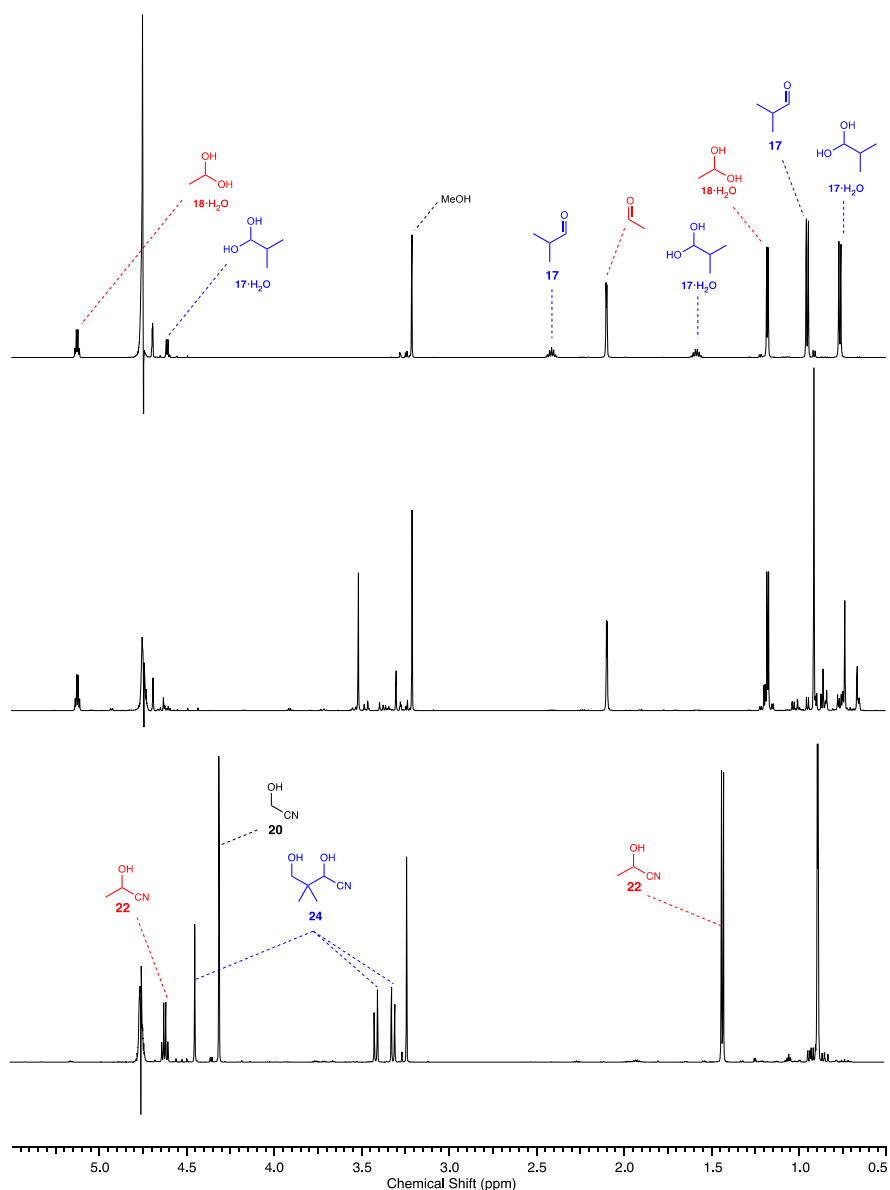
Supplementary Figure 60. ^{1}H NMR (600 MHz, H$_2$O, noesygppr1d, 0.50–5.50 ppm) spectra to show the reaction of isobutyraldehyde **17** (50 mM), acetaldehyde **18** (50 mM) and formaldehyde **16** (100 mM) in phosphate buffer (pH 7.0; 500 mM) after; (A.) 10 mins at room temperature; (B.) 48 hours at 60 °C; (C.) after addition of sodium cyanide (NaCN; 4 equiv.) at pH 7.0 and 20 ºC, showing the formation of pantoic acid nitrile **24** from hydroxypivaldehyde **15**, alongside the conversion of formaldehyde **16** to glycolonitrile **20**, residual isobutyraldehyde **17** to 2-hydroxy-3-methylbutane nitrile **21**, and acetaldehyde **18** to lactonitrile **22**.

*Conversion of hydroxypivaldehyde **15** to pantoic acid nitrile **24** at pH 7.0*
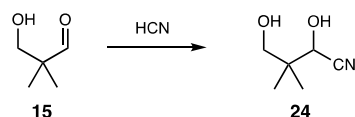
To a stirred solution of hydroxypivaldehyde **15** (50 mM) in phosphate buffer (pH 7.0; 500 mM) at 20 °C was added sodium cyanide (1.1 equiv., 55 mM). The solution was adjusted from pH 7.5 to pH 7.0 with 8 M HCl and NMR spectra were acquired to show quantitative conversion of hydroxypivaldehyde **15** to pantoic acid nitrile **24** (Supplementary Figure 61). 1H NMR (700 MHz, H_2O, noesygppr1d) δ 4.52 (s, 1H, (C2)–H), 3.49 (AB, *J* = 11.7 Hz, 1H, (C4)–H), 3.40 (d, *J* = 11.7 Hz, 1H, (C4)–H′), 0.97 (s, 3H, CH_3), 0.96 (s, 3H, CH_3′). ^{13}C NMR (176 MHz, H_2O, noesygppr1d) δ 120.8 (C1), 67.2 (C4), 66.9 (C2), 39.9 (C3), 20.4 (CH_3), 19.3 (CH_3′).
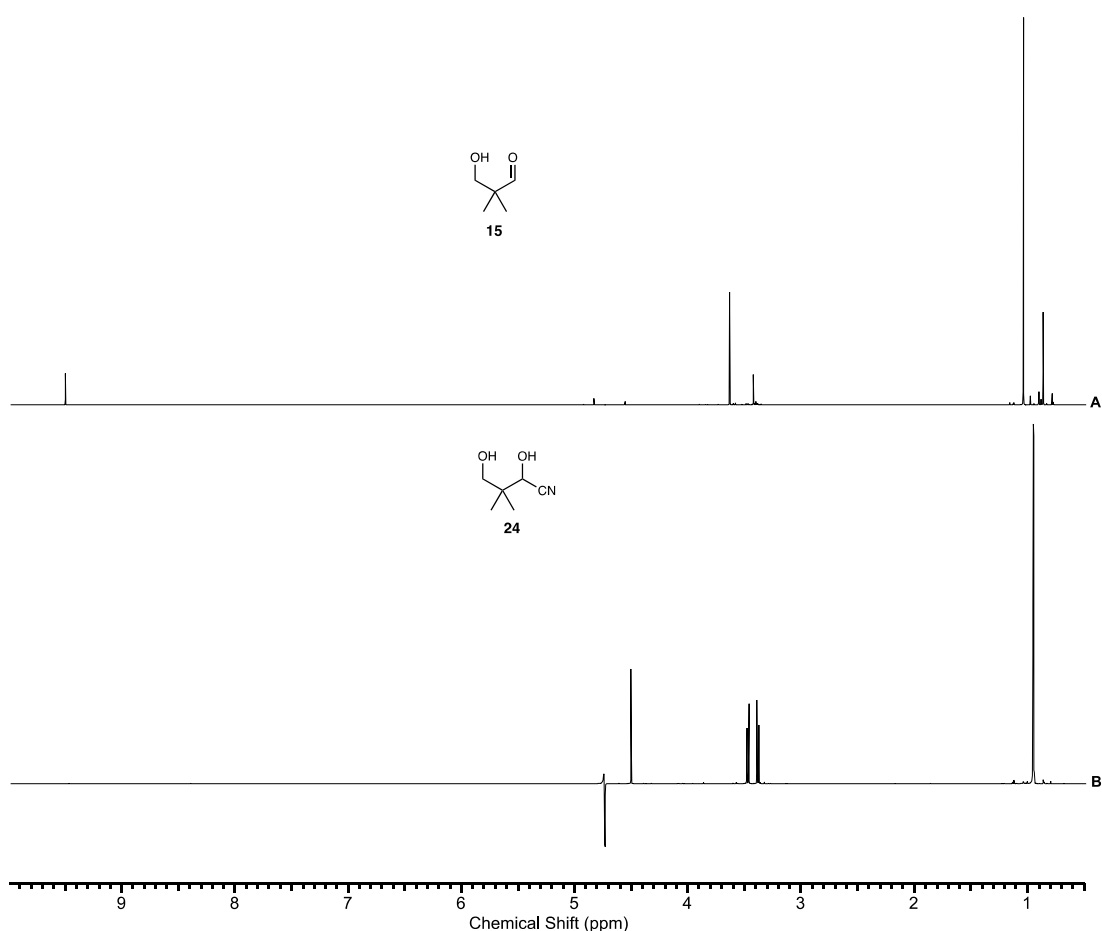
Supplementary Figure 61. ^1H NMR (700 MHz, H_2O, noesygppr1d) spectra of (A.) hydroxypivaldehyde **15** (50 mM) in phosphate buffer (pH 7.0; 500 mM) before the addition of sodium cyanide (NaCN), and (B.) pantoic acid nitrile **24** (>95%) formation after the addition of NaCN (55 mM) to hydroxypivaldehyde **15** (50 mM) in phosphate buffer (pH 7.0; 500 mM) at 20 °C, after 30 mins.
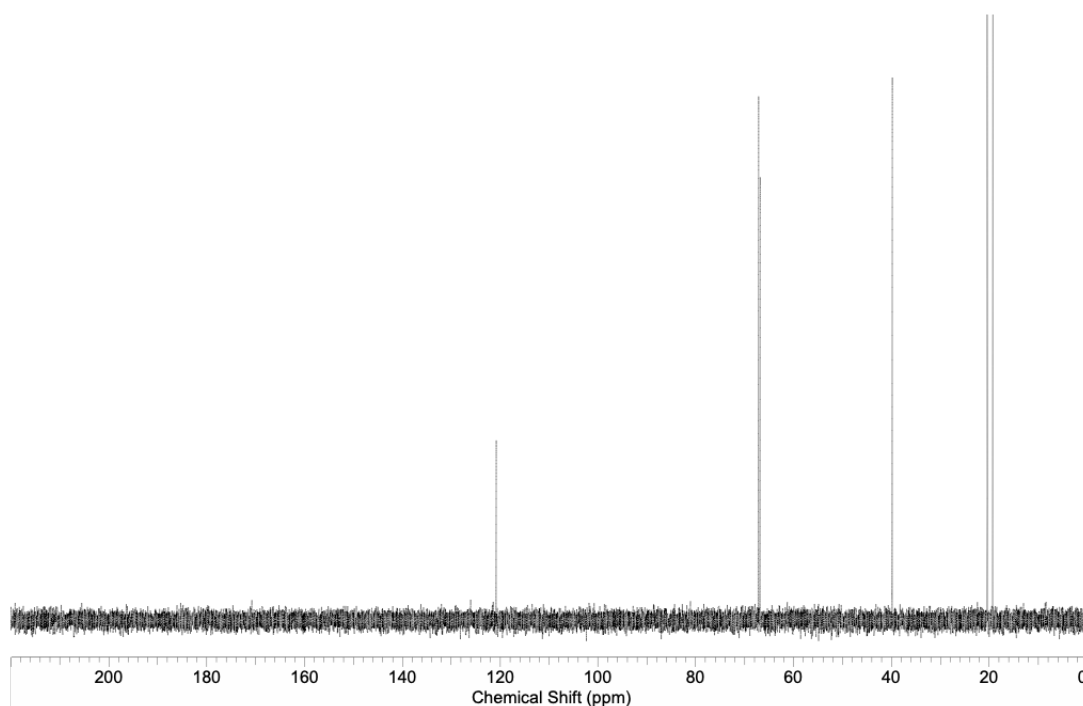
Supplementary Figure 62. ^{13}C NMR (176 MHz, H$_2$O, noesygppr1d, 0–220 ppm) spectrum to show pantoic acid nitrile **24** formation after the addition of sodium cyanide (NaCN; 55 mM) to hydroxypivaldehyde **15** (50 mM) in phosphate buffer (pH 7.0; 500 mM) at 20 ºC, after 30 mins.

Competitive aldol reaction of acetaldehyde 18, glycolaldehyde 19 and isobutyraldehyde 17 with formaldehyde 16 at pH 7.0 or 9.8
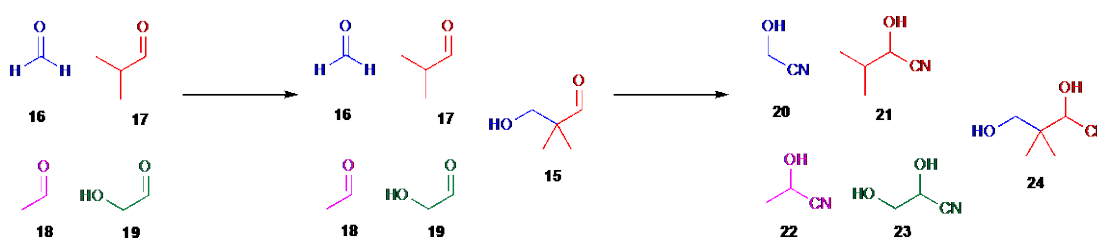
Synthesis of hydroxypivaldehyde 15 in an aldehyde mixture at pH 7

Acetaldehyde **18** (24–31 mM), isobutyraldehyde **17** (17 mM), formaldehyde **16** (20–156 mM; 37 wt.% in water), glycolaldehyde **19** (20–22 mM), acetone (0-1.8 mM) and MSM (20 mM) were dissolved in degassed phosphate buffer (pH 7.0; 500 mM)* and incubated under an argon atmosphere at 20, 40 or 60 ºC. An aliquot (1 mL) of the reaction mixture was periodically removed, cooled to room temperature and NMR spectra were acquired. Analysis confirmed that new ^{1}H NMR resonances that were characteristic of hydroxypivaldehyde **15** were detected (δ_H 9.40 (s), 3.55 (s), 0.94 (s) ppm; Supplementary Figure 66). Sodium cyanide (NaCN; 300–700 mM) was then added to the aliquot, and the resulting solution was adjusted from pH 7.5 to pH 7.0 with 1–4 M HCl. Hydroxypivaldehyde **15**, dihydroxyacetone (**DHA**) and residual aldehyde **16**-**19** were then quantified upon their quantitative conversion to pantoic acid nitrile **24** (**15**→**24**), formaldehyde **16** to glycolonitrile **20** (**16**→**20**), 2-hydroxy-3-methylbutane nitrile **21** (**17**→**21**), acetaldehyde **18** to lactonitrile **22** (**18**→**22**), glycolaldehyde **19** to glyceronitrile **23** (**19**→**23**), and dihydroxyacetone **DHA** (**DHA**→ **DHA**·HCN) upon the addition of sodium cyanide (NaCN) at pH 7.0 and 20 ºC (Supplementary Table 17).

Substantial loss of acetaldehyde 18, isobutyraldehyde 17, glycolaldehyde 19 and formaldehyde 16 at pH 9.8

A mixture of acetaldehyde **18** (31 mM), isobutyraldehyde **17** (17 mM), formaldehyde **16** (20 mM; 37% wt.% in water) and glycolaldehyde **19** (20 mM) were dissolved in phosphate buffer (pH 7.0; 500 mM) at 20 ºC. The solution was adjusted to pH 9.8 with 1-8 M NaOH and the reaction was monitored by NMR spectroscopy for 3 days at 20 ºC (Supplementary Figure 67). ^{1}H NMR resonances corresponding to the formation of hydroxypivaldehyde **15** were observed. However, a complex mixture and multitude of reaction products were observed that resulted in complete loss of glycolaldehyde **19**, and substantial loss of acetaldehyde **18**, formaldehyde **16** and isobutyraldehyde **17** that had not reacted to produce hydroxypivaldehyde **15**. This was confirmed upon reaction with cyanide where their corresponding cyanohydrins were not observed (see Supplementary Figure 68 and text above): the solution of the crude aldol mixture was lowered

---

* *Biotage*® microwave reaction vial (Part number: 351521) were used and sealed with a *Biotage*® aluminium cap with a *Reseal*™ septum (Part number: 352998) by a *Biotage*® mechanical crimper (Part number: 353671). Some vials were observed to contain residual acetone from cleaning. As acetone is an additional prebiotically plausible ketone we have provided this data. However, acetone remained unchanged during the reaction; when NaCN was added the expected equilibrium concentration of acetone cyanohydrin was formed. Replicate reactions without acetone were otherwise indistinguishable (e.g., see Supplementary Figure 63).

from pH 9.8 to pH 7.0 with 1-4 M HCl, followed by the addition of sodium cyanide (NaCN; 150 mM) to convert any aldehydes within the reaction mixture to their corresponding cyanohydrins. The resulting solution was adjusted from pH 7.5 to pH 7.0 with 1–4 M HCl and analysed by NMR spectroscopy to show the conversion of hydroxypivaldehyde **15** to pantoic acid nitrile **24** (41%), acetaldehyde **18** to lactonitrile **22** (28%), and formaldehyde **16** to glyceronitrile **20** (<5%) (Supplementary Table 17). However, 2-hydroxy-3-methylbutane nitrile **21** (from isobutyraldehyde **17**) and glyceronitrile **23** (from glycolaldehyde **19**) were not detected (Supplementary Figure 68) (Supplementary Table 17).

Starting aldehyde concentration (mM)				Conditions			Residual aldehyde (%)[*]				Yield (%)	
16	**17**	**18**	**19**	pH	Temp (ºC)	Time (days)	**16**	**17**	**18**	**19**	**15**	**DHA**
20	17	31	20	9.8	20	3	2	0	28	0	41	n.d
20	17	31	20	7.0	20	10	84	64	94	81	24	3
20	17	31	20	7.0	20	20	63	55	93	71	37	5
76	17	24	22	7.0	20	5	85	53	96	81	42	4
152	17	24	22	7.0	20	5	89	37	96	73	63	6
22	17	22	22	7.0	40	1	86	76	95	86	22	4
76	17	24	22	7.0	40	0.2	96	82	>99	94	18	2
76	17	24	22	7.0	40	1	85	47	92	79	50	16
76	17	24	22	7.0	40	2	78	23	91	59	74	14
22	17	22	22	7.0	60	0.5	55	56	87	61	45	10
22	17	22	22	7.0	60	1	44	38	79	33	57	15
76	17	24	22	7.0	60	0.2	84	52	96	77	48	11
76	17	24	22	7.0	60	1	64	<2	83	32	97	24

Supplementary Table 17. Yields and conditions for the formation of hydroxypivaldehyde **15** by crossed-aldolisation of isobutyraldehyde **17** and formaldehyde **16** in phosphate buffer (500 mM) in the presence of acetaldehyde **18** and glycolaldehyde **19**. n.d. = not detected.

---

[*] Hydroxypivaldehyde **15** and residual formaldehyde **16**, isobutyraldehyde **17**, acetaldehyde **18**, and glycolaldehyde **19** quantifications were confirmed by their subsequent quantitative conversion to pantoic acid nitrile **24** (**15**→**24**), formaldehyde **16** to glycolonitrile **20** (**16**→**20**), 2-hydroxy-3-methylbutane nitrile **21** (**17**→**21**), acetaldehyde **18** to lactonitrile **22** (**18**→**22**), glycolaldehyde **19** to glyceronitrile **23** (**19**→**23**), and dihydroxyacetone **DHA** (**DHA**→ **DHA**·HCN) upon the addition of sodium cyanide (NaCN) at pH 7.0 and 20 ºC. Commercial **16** and **17** contained methanol and isobutyric acid, respectively, which were unchanged during the reaction.
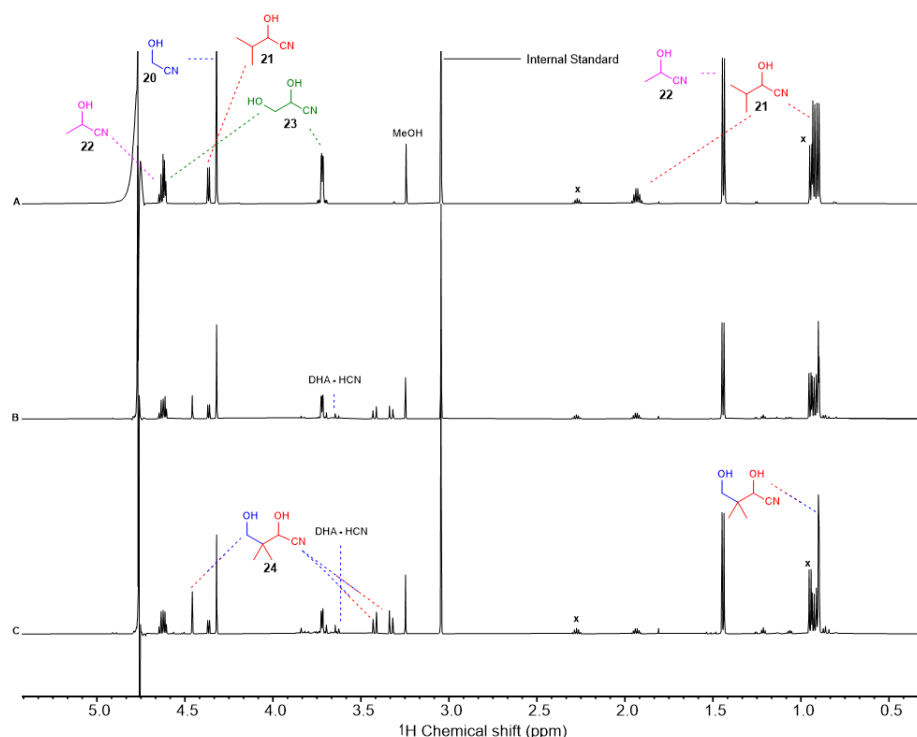
Supplementary Figure 63. ¹H NMR (600 MHz; noesygppr1d, 0.5–5.5 ppm) spectra to show the cyanohydrin mixture observed after incubation of formaldehyde **16** (22 mM), isobutyraldehyde **17** (17 mM), acetaldehyde **18** (22 mM), glycolaldehyde **19** (22 mM) and MSM (10 mM; internal standard) in phosphate buffer at pH 7.0 for: **(A.)** 20 min at 20 ºC followed by addition of NaCN (300 mM); **(B.)** 12 hours at 60 ºC followed by addition of NaCN (300 mM); and **(C.)** 1 day at 60 ºC followed by addition of NaCN (300 mM). In addition to glycolonitrile **20**, glyceronitrile **23**, 2-hydroxy-3-methylbutane nitrile **21** and lactonitrile **22** formation, the newly formed hydroxypivaldehyde **15** yielded pantoic acid nitrile **24**. Commercial **16** and **17** contained methanol and isobutyric acid (**x**), respectively, which were unchanged during the reaction.
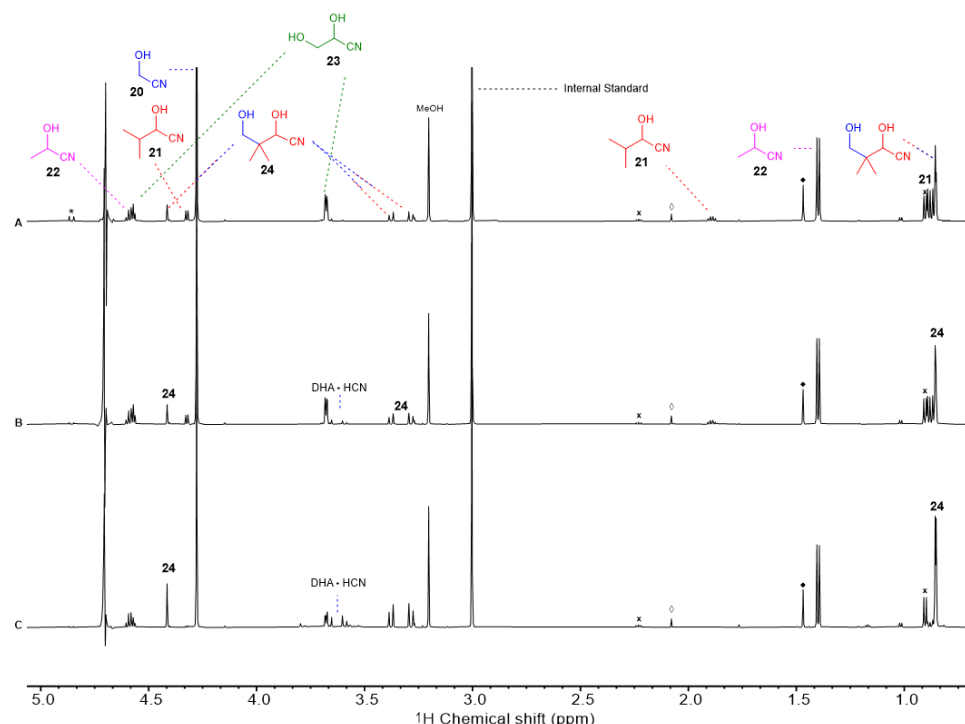
Supplementary Figure 64. ¹H NMR (600 MHz; noesygppr1d, 0.5–5.0 ppm) spectra to show the cyanohydrin mixture observed after incubating formaldehyde **16** (76 mM), isobutyraldehyde **17** (17 mM), acetaldehyde **18** (24 mM), glycolaldehyde **19** (22 mM), acetone (◊; 1.8 mM) and MSM (20 mM; internal standard) in phosphate buffer (pH 7; 500 mM) after: **(A.)** 5 days at 20 ºC, followed by addition of NaCN (500 mM) at pH 7; **(B.)** 1 day at 40 ºC, followed by the addition of NaCN (500 mM) at pH 7; and **(C.)** 1 day at 60 ºC, followed by the addition of NaCN (500 mM) at pH 7. In addition to glycolonitrile **20**, glyceronitrile **23**, 2-hydroxy-3-methylbutane nitrile **21**, lactonitrile **22**, acetone (◊) and acetone cyanohydrin (♦) formation, the newly formed hydroxypivaldehyde **15** yielded pantoic acid nitrile **24**. Recovery of aldehyde **16**–**19** is reported in Supplementary Table 17. Glycolic acid, acetic acid and tartaric acid dinitriles (*; formed via the addition of cyanide to glyoxal) from partial oxidation of **18** and **19** were observed in trace amounts (<5%). Commercial **16** and **17** contained methanol and isobutyric acid (**x**) respectively, which were unchanged during the reaction.
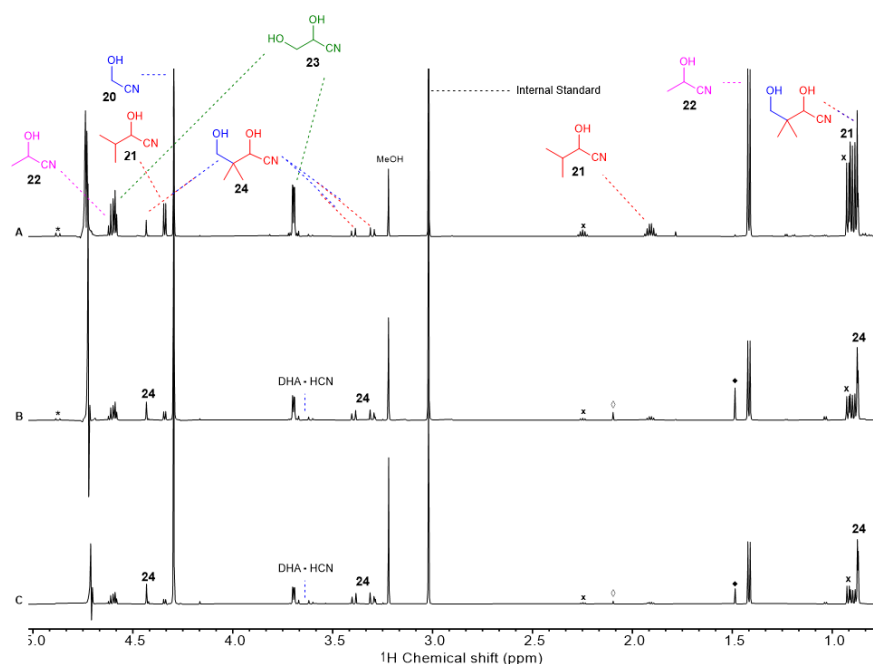
Supplementary Figure 65. ^{1}H NMR (600 MHz; noesygppr1d, 0.5–5.0 ppm) spectra to show the cyanohydrin mixture observed after incubating formaldehyde **16** (22–156 mM), isobutyraldehyde **17** (17 mM), acetaldehyde **18** (24 mM), glycolaldehyde **19** (22 mM), acetone (◊; 0-1.8 mM) and MSM (20 mM; internal standard) in phosphate buffer (pH 7.0; 500 mM) at 40 °C after 1 d, followed by the addition of NaCN (300-700 mM) at pH 7 and 20 °C: **(A.)** 22 mM **16**; **(B.)** 76 mM **16** and 1.8 mM acetone (◊); and **(C.)** 152 mM **16** and 1.8 mM acetone (◊). In addition to glycolonitrile **20**, glyceronitrile **23**, 2-hydroxy-3-methylbutane nitrile **21**, lactonitrile **22**, acetone (◊) and acetone cyanohydrin (♦) formation, the newly formed hydroxypivaldehyde **15** yielded pantoic acid nitrile **24**. Recovery of aldehyde **16**–**19** is reported in Supplementary Table 17. Glycolic acid, acetic acid, and tartaric acid dinitriles (* ; formed via addition of cyanide to glyoxal, from partial oxidation of **18** and **19**), were observed in trace amounts (<5%). Commercial **16** and **17** contained methanol and isobutyric acid (**x**) respectively, which were unchanged during the reaction.
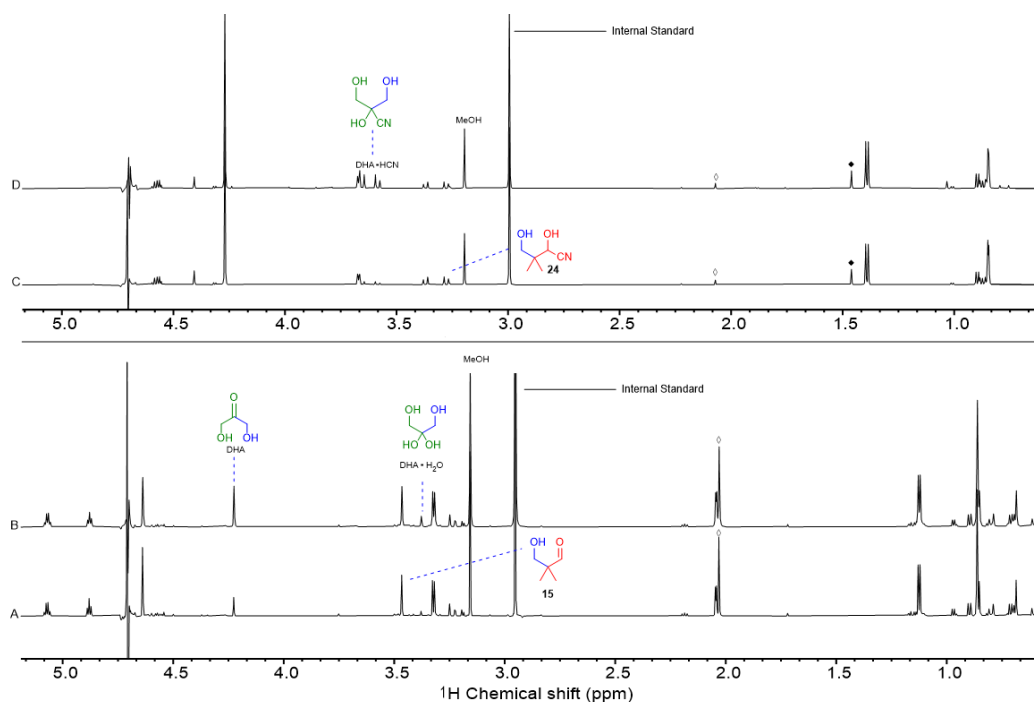
Supplementary Figure 66. ^{1}H NMR (600 MHz; noesygppr1d, 0.5–5.0 ppm) spectra to show the product mixture observed following incubation of formaldehyde **16** (76 mM), isobutyraldehyde **17** (17 mM), acetaldehyde **18** (24 mM), glycolaldehyde **19** (22 mM), acetone (◊; 1.8 mM) and MSM (20 mM; internal standard) incubated in phosphate buffer (pH 7.0; 500 mM) at 40 °C: **(A.)** after 2 d; and **(B.)** after the sample shown in spectrum A was subsequently spiked with authentic dihydroxyacetone (**DHA**) leading to enhancement of observed ^{1}H NMR resonances at 4.22 (s) and 3.38 (s) ppm. ^{1}H NMR (600 MHz; noesygppr1d, 0.5–5.0 ppm) spectra to show the product mixture observed following incubation of formaldehyde **16** (76 mM), isobutyraldehyde **17** (17 mM), acetaldehyde **18** (24 mM), glycolaldehyde **19** (22 mM), acetone (◊; 1.8 mM) and MSM (20 mM) in phosphate buffer (pH 7; 500 mM) at 40 °C; **(C.)** after 2 d and then the addition of NaCN (500 mM) at pH 7 and 20 °C; and **(D.)** after sample shown in spectrum C was subsequently spiked with authentic dihydroxyacetone-cyanohydrin (**DHA·HCN**) leading to enhancement of observed ^{1}H NMR resonances at 3.57 (AB, *J* =11.9 Hz) and 3.67 (AB, *J* = 11.9 Hz) ppm.
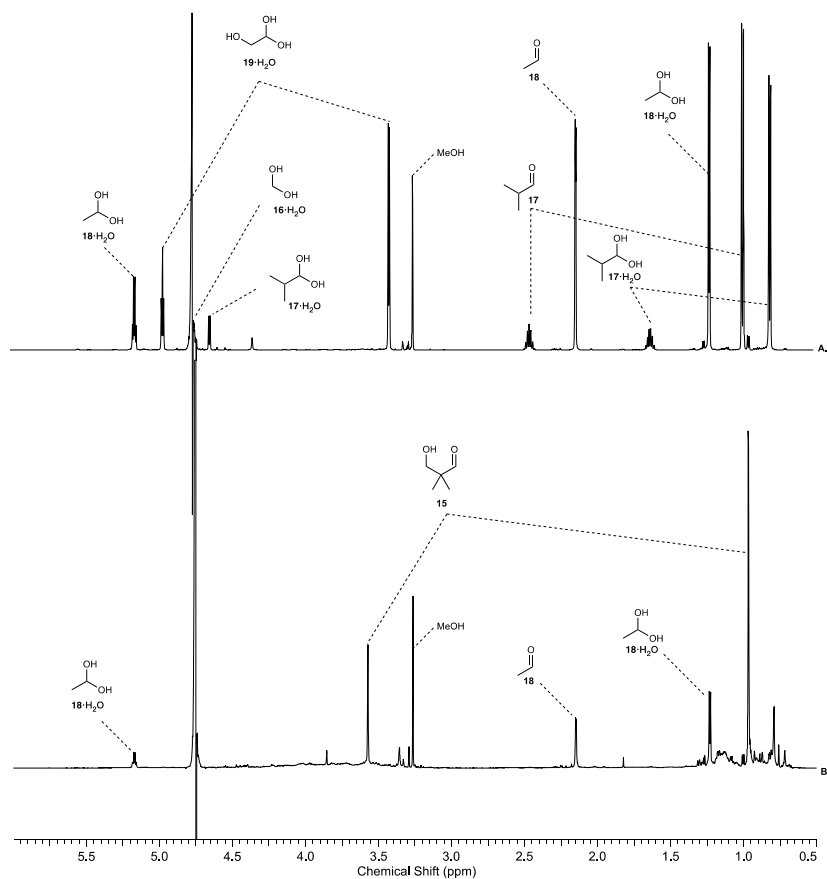
Supplementary Figure 67. ^{1}H NMR (600 MHz, noesygppr1d, 0.5–6.0 ppm) spectra showing a mixture of formaldehyde **16**, glycolaldehyde **19**, isobutyraldehyde **17** and acetaldehyde **18** in phosphate buffer (500 mM) after: (A.) 20 mins at pH 7.0, and (B.) 3 days at pH 9.8, showing the formation of hydroxypivaldehyde **15**. The loss of glycolaldehyde **19**, acetaldehyde **18**, formaldehyde **16**, and isobutyraldehyde **17** into a multitude of reaction products was confirmed upon conversion of aldehydes to cyanohydrins (see Supplementary Figure 68).
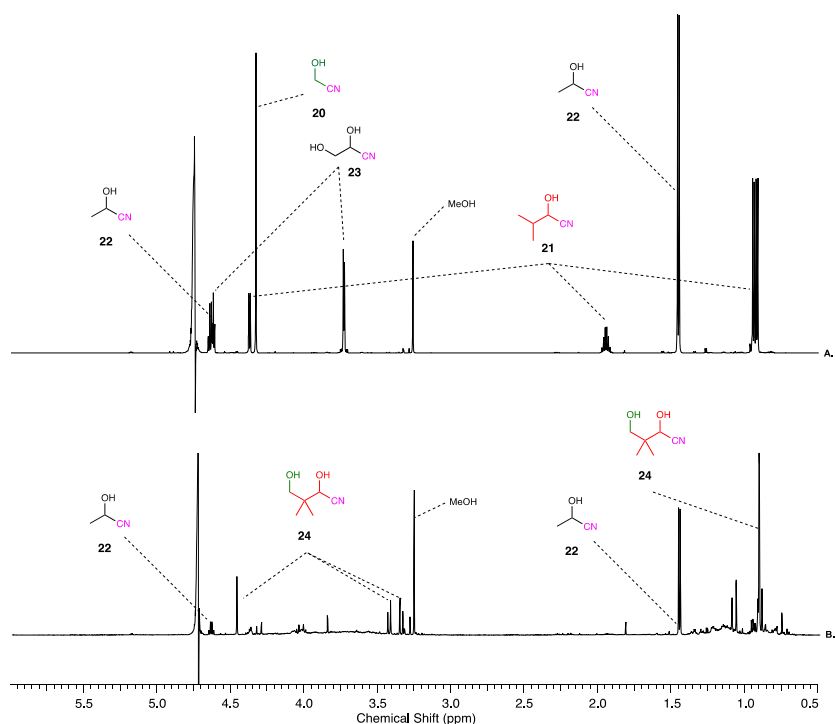
Supplementary Figure 68. ^{1}H NMR (600 MHz] noesygppr1d, 0.5–6.0 ppm) spectra showing (A.) glycolonitrile **20**, glyceronitrile **23**, 2-hydroxy-3-methylbutane nitrile **21** and lactonitrile **22** formed from the addition of sodium cyanide in phosphate buffer at pH 7.0 upon immediate mixing of formaldehyde **16**, glycolaldehyde **19**, isobutyraldehyde **17** and acetaldehyde **18** (see Supplementary Figure 67A for aldehyde mixture), and (B.) the resulting cyanohydrin mixture after incubation of formaldehyde **16**, glycolaldehyde **19**, isobutyraldehyde **17** and acetaldehyde **18** in phosphate buffer at pH 9.8 and 20 ºC for 3 days (See Supplementary Figure 67B for crude aldehyde mixture at pH 7.0 before the pH was increased to pH 9.8), followed by the addition of sodium cyanide. The newly formed hydroxypivaldehyde **15** (41%) underwent quantitative conversion to pantoic acid nitrile **24**. However, only partial recovery of acetaldehyde **18** (as lactonitrile **22** (28%)) and formaldehyde **16** (as glycolonitrile **20** (<5%)) was observed, with no detectable glycolaldehyde **19** (as glyceronitrile **23**) or isobutyraldehyde **17** (as 2-hydroxy-3-methylbutane nitrile **21**). See Supplementary Table 17.

*Data for acetaldehyde **18** (Supplementary Figure 66)*

^{1}H NMR (700 MHz, H$_2$O) Aldehyde δ_H 9.56 (q, J = 2.9 Hz, 1H, (C1)–H), 2.13 (d, J = 2.9 Hz, 3H, (C2)–H$_3$). Hydrate δ 5.15 (q, J = 5.3 Hz, 1H, (C1)–H), 1.21 (d, J = 5.4 Hz, 3H, (C2)–H$_3$).
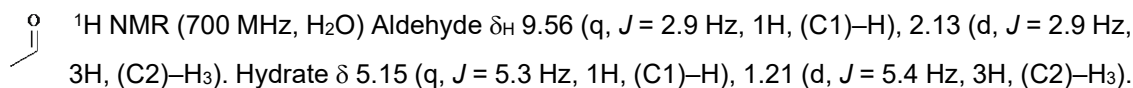
*Data for isobutyraldehyde **17** (Supplementary Figure 66)*

^{1}H NMR (700 MHz, H$_2$O) Aldehyde δ_H 9.47 (d, J = 1.3 Hz, 1H, (C1)–H), 2.48 - 2.41 (m, 1H, (C2)–H), 0.99 (s, 3H, (CH$_3$)), 0.98 (s, 3H, (CH$_3'$)). Hydrate (600 MHz, H$_2$O) δ_H 4.63 (d, J = 5.6 Hz, 1H, (C1)–H), 1.66 - 1.57 (m, 1H, (C2)–H), 0.80 (s, 3H, (CH$_3$)), 0.79 (s, 3H, (CH$_3'$)).
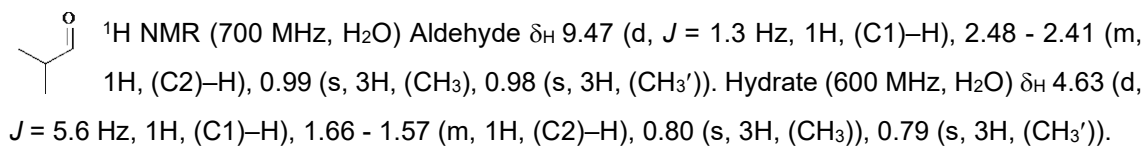
*Data for glycolaldehyde **20** (Supplementary Figure 66)*

^{1}H NMR (700 MHz, H$_2$O) Hydrate δ_H 4.96 (t, J = 5.2 Hz, 1H, (C1)–H), 3.41 (d, J = 5.2 Hz, 2H, (C2)–H$_2$).
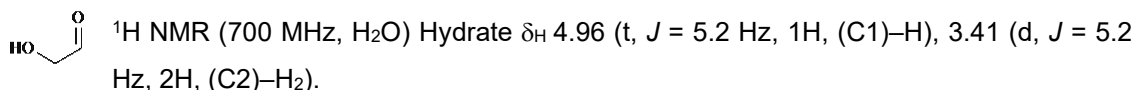
*Data for lactonitrile **22** (Supplementary Figure 63)*

^{1}H NMR (600 MHz, H$_2$O) δ_H 4.54 (q, J = 6.9 Hz, 1H, (C2)–H), 1.38 (d, J = 6.9 Hz, 3H, (C3)–H$_3$).
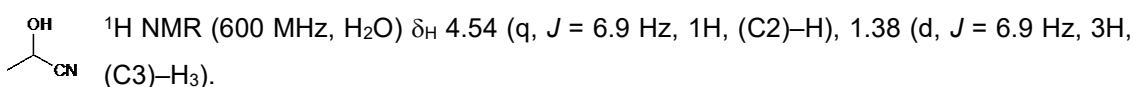
*Data for glycolonitrile **20** (Supplementary Figure 63)*

^{1}H NMR (600 MHz, H$_2$O) δ_H 4.32 (s, 2H, (C2)–H$_2$).

*Data for 2-hydroxy-3-methylbutane nitrile **21** (Supplementary Figure 63)*
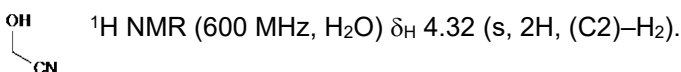
^{1}H NMR (600 MHz, H$_2$O) δ_H 4.36 (d, J = 6.2 Hz, 1H, (C2)–H), 1.89 -2.00 (m, 1H, (C3)–H), 0.92 (d, J = 6.7 Hz, 3H, (CH$_3$)), 0.90 (d, J = 6.7 Hz, 3H, (CH$_3'$)).
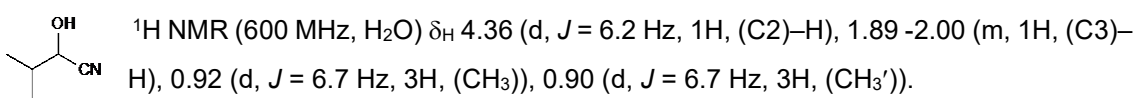
*Data for glyceronitrile **23** (Supplementary Figure 63)*

^{1}H NMR (600 MHz, H$_2$O) δ_H 4.62 (t, J = 4.3 Hz, 1H, (C2)–H), 3.71 (d, J = 4.3 Hz, 2H, (C3)–H$_2$).

*Data for pantoic acid nitrile **24** (Supplementary Figure 63)*
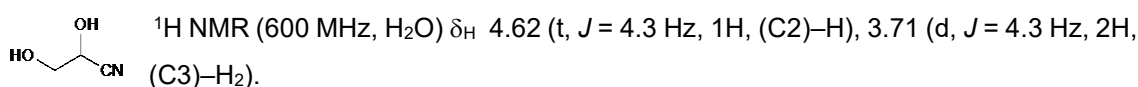
^{1}H NMR (600 MHz, H$_2$O) δ_H 4.46 (s, 1H, (C2)–H), 3.28 (AB, J = 11.4 Hz, 1H, (C4)–H), 3.37 (d, J = 11.4 Hz, 1H, (C4)–H$_2$), 0.85 (s, 6H, 2 × CH$_3$).
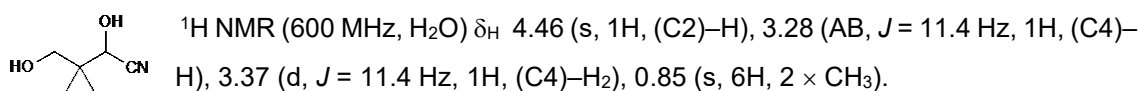
Data for tartaric acid dinitrile (Supplementary Figure 63)

^{1}H NMR (600 MHz, H$_2$O, major:minor 55:45) δ_H^{major} 4.86 (s, 2H, (C2)–H). δ_H^{minor} 4.83 (s, 2H, (C2)–H).
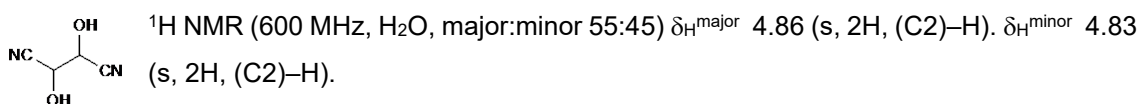
Data for dihydroxyacetone-cyanohydrin

^{1}H NMR (600 MHz, H$_2$O) δ_H 3.66 (AB, J = 11.9 Hz, 2H, (C3)–H$_2$), 3.59 (AB, J = 11.9 Hz, 2H, (C3')–H$_2$).
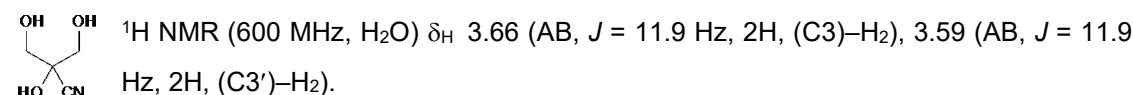
4. Reactions of hydroxypivaldehyde 15 with amines

Selective synthesis of proteinogenic α-aminonitriles 7 within a mixture of glycolaldehyde 19, formaldehyde 16, isobutyraldehyde 17, acetaldehyde 18, and hydroxypivaldehyde 15
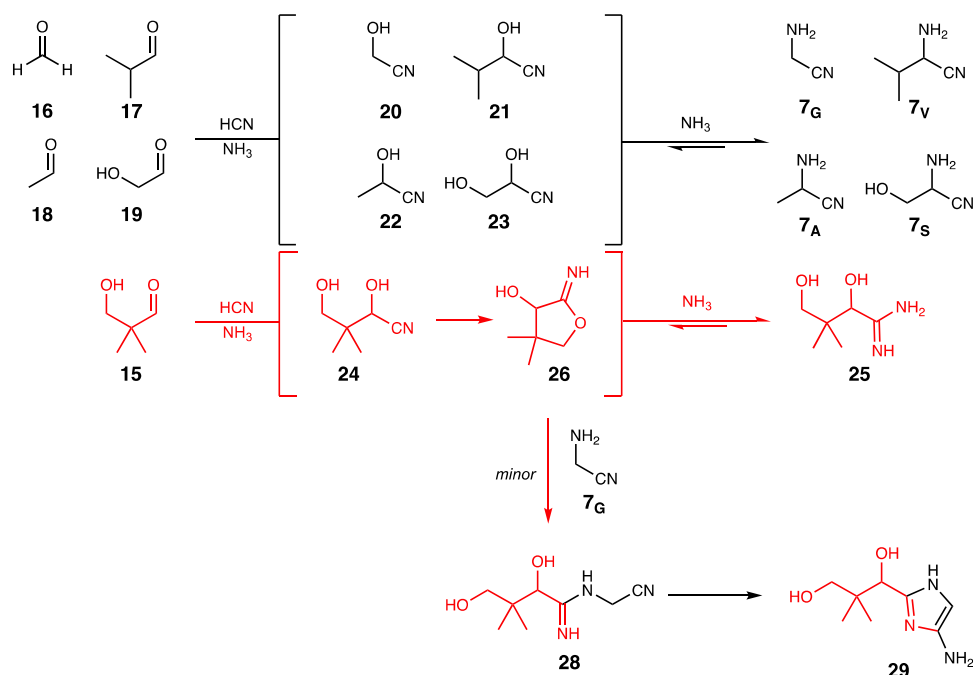
To a stirred solution of acetaldehyde **18** (31.3 mM), isobutyraldehyde **17** (17.2 mM), formaldehyde **16** (20 mM; 37% wt.% in water), hydroxypivaldehyde **15** (20 mM) and glycolaldehyde **19** (20 mM) in phosphate buffer (pH 7.0; 500 mM) at 20 ºC was added sodium cyanide (NaCN; 150 mM) and ammonium chloride (NH₄Cl; 500 mM). The resulting solution (pH 9.2) was adjusted to pH 9.5 using 1–8 M HCl/NaOH and monitored periodically by NMR spectroscopy. All aldehydes **16-19** underwent rapid and quantitative conversion to the corresponding cyanohydrins (isobutyraldehyde **17**→2-hydroxy-3-methylbutane nitrile **21**; acetaldehyde **18**→lactonitrile **22**; formaldehyde **16**→glycolonitrile **20**; glycolaldehyde **19**→glyceronitrile **23**; hydroxypivaldehyde **15**→pantoic acid nitrile **24**), followed by their conversions to the corresponding α-aminonitriles (2-hydroxy-3-methylbutane nitrile **21**→valine nitrile **7ᵥ** [>95%]; lactonitrile **22**→alanine nitrile **7ₐ** [80%]; glycolonitrile **20**→glycine nitrile **7₆** [60%]; glyceronitrile **23**→serine nitrile **7ₛ** [82%]) after 3 days at 20 ºC – with the notable exception of pantoic acid nitrile **24**, which underwent rapid formation of pantoamidine **25** (>95% after 4 hours).

Pantoamidine **25** was formed by rapid trapping of the incipient iminopantolactone **26** by ammonia, which then underwent subsequent reactions, including a slower hydrolysis to pantoic acid **5** (<10%) via pantoamide **5**-NH₂ and pantolactone **2** (detected in trace yield; <1%). Pantoamidine **25** underwent subsequent detectable side reactions with glycine nitrile **7₆** to generate aminoimidazole **29** (7% after 3 days; yield based on hydroxypivaldehyde **15**), which was further confirmed by the reaction of glycine nitrile **7₆** with pantoic acid nitrile **24** at pH 9.0 (See Supplementary Page 107). The detection of aminoimidazole **29** led to our discovery of a chemoselective synthesis of pantothenic acid nitrile **9** over pantoylglycine nitrile **14** via iminolactone **26** and pantoamidine **25** en route to pantetheine **1**. See associated paper for details.

Data for glyceronitrile 23
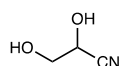
^{1}H NMR (700 MHz, H$_2$O, partial data) δ_H 3.80 (dd, J = 1.0, 4.8 Hz, 1H, (C3)–H$_2$)

Data for serine nitrile 7$_s$
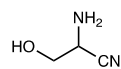
^{1}H NMR (700 MHz, H$_2$O) δ_H 3.95 (ABX, J = 4.7, 5.8 Hz, 1H, (C2)–H), 3.77 (ABX, J = 4.7, 11.4 Hz, 1H, (C3)–H), 3.70 (ABX, J = 5.8, 11.4 Hz, 1H, (C3)–H′).

Data for pantoic acid nitrile 24
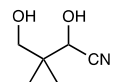
^{1}H NMR (700 MHz, H$_2$O, partial data) δ_H 4.51 (s, 1H, (C2)–H), 3.50 (AB, J = 11.4 Hz, 1H, (C4)–H), 3.42 (AB, J = 11.4 Hz, 1H, (C4)–H$_2$).

Data for pantoamidine 25
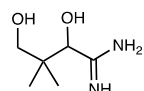
^{1}H NMR (700 MHz, H$_2$O) δ_H 4.35 (s, 1H, (C2)–H), 3.56 (AB, J = 11.4 Hz, 1H, (C4)–H), 3.37 (AB, J = 11.4 Hz, 1H, (C4)–H′), 0.94 (s, 3H, CH$_3$), 0.91 (s, 3H, CH$_3$′).

Data for iminopantolactone 26
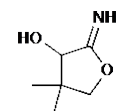
^{1}H NMR (700 MHz, H$_2$O, partial data) δ_H 4.28 (s, 1H, (C3)–H), 4.01 (AB, J = 8.8 Hz, (C5)–H).

Data for pantolactone 2
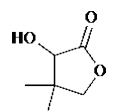
^{1}H NMR (700 MHz, H$_2$O, partial data) δ_H 4.42 (s, 1H, (C3)–H), 4.12 (AB, J = 8.9 Hz, 1H, (C5)–H), 4.08 (AB, J = 8.9 Hz, 1H, (C5)–H′).

Data for pantoic acid 5
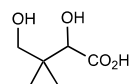
^{1}H NMR (700 MHz, H$_2$O, partial data) δ_H 3.78 (s, 1H, (C2)–H).

Data for glycolonitrile 20

^{1}H NMR (700 MHz, H$_2$O) δ_H 4.41 (s, 2H, (C2)–H$_2$).

Data for glycine nitrile 7$_G$

^{1}H NMR (700 MHz, H$_2$O) δ_H 3.59 (s, 2H, (C2)–H$_2$).

Data for 2-hydroxy-3-methylbutane nitrile 21
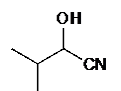
^{1}H NMR (700 MHz, H$_2$O) δ_H 4.45 (d, J = 6.1 Hz, 1H, (C2)–H), 2.05 - 1.98 (m, 1H, (C3)–H), 1.01 (d, J = 7.0 Hz, 3H, (CH$_3$)), 0.99 (d, J = 7.0 Hz, 3H, (CH$_3$′)).

Data for valine nitrile 7$_v$
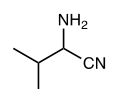
^{1}H NMR (700 MHz, H$_2$O) δ_H 3.73 (d, J = 5.6 Hz, 1H, (C2)–H), 1.97 - 1.89 (m, 1H, (C3)–H), 0.99 (d, J = 5.4 Hz, 3H, (CH$_3$)), 0.98 (d, J = 5.4 Hz, 3H, (CH$_3$′)).

Data for lactonitrile 22

¹H NMR (700 MHz, H₂O, partial data) δ$_H$ 1.53 (d, *J* = 7.0 Hz, 3H, (C3)–H₃).

Data for alanine nitrile 7$_A$

¹H NMR (700 MHz, H₂O) δ$_H$ 3.90 (q, *J* = 7.1 Hz, 1H, (C2)–H), 1.41 (d, *J* = 7.2 Hz, 3H, (CH₃)).
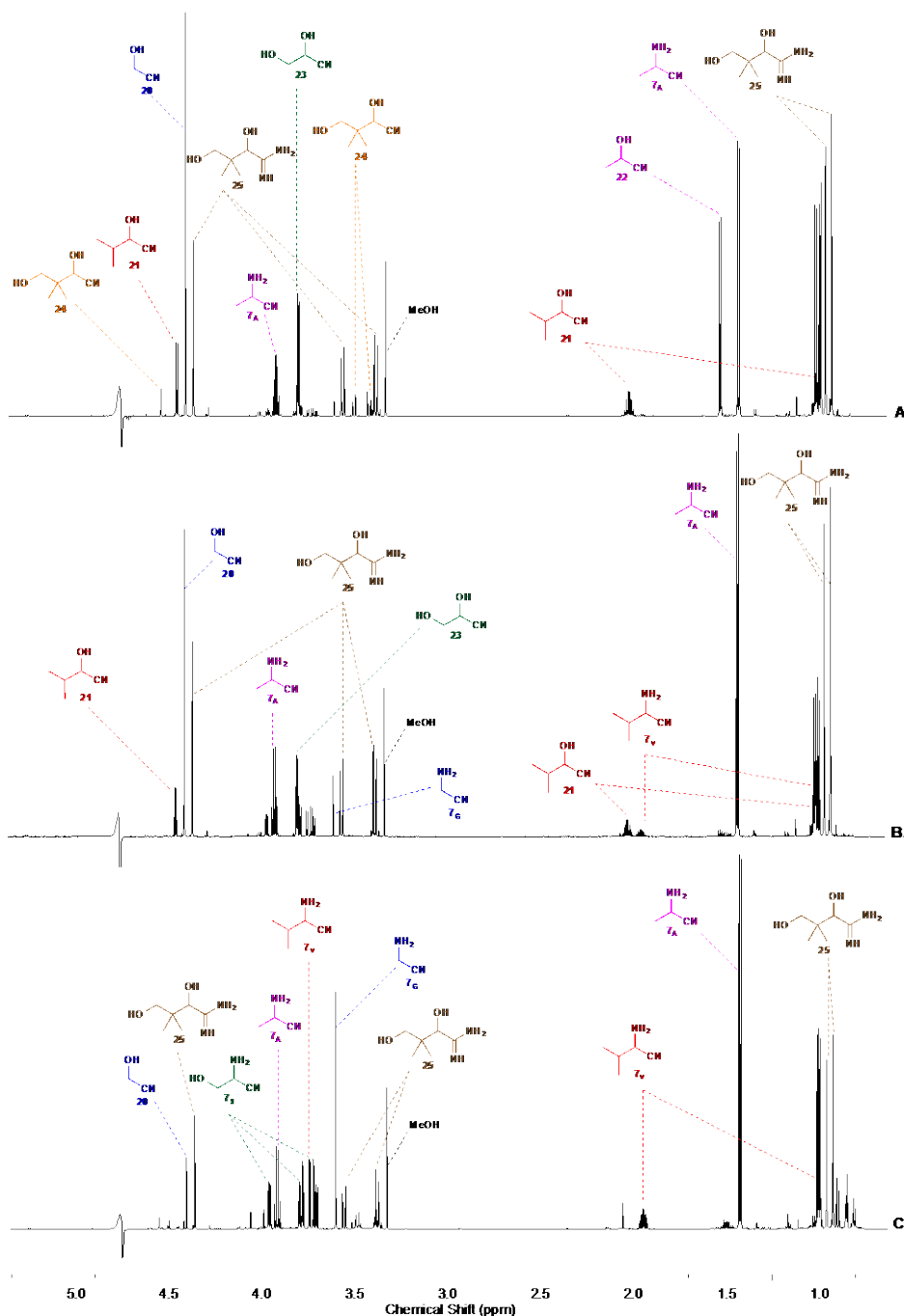
Supplementary Figure 69. ¹H NMR (700 MHz, H₂O, noesygppr1d, 0.5–5.0 ppm) spectra to show the selective synthesis of proteinogenic α-aminonitriles **7** from a mixture of formaldehyde **16** (20 mM), acetaldehyde **18** (31.3 mM), isobutyraldehyde **17** (17.2 mM), glycolaldehyde **19** (20 mM), hydroxypivaldehyde **15** (20 mM), sodium cyanide (NaCN, 150 mM) and ammonium chloride (NH₄Cl; 500 mM) in phosphate buffer (pH 9.5; 500 mM) at 20 ºC after (A.) 30 min; (B.) 4 hours; (C.) 3 days. Selective conversion of proteinogenic aldehydes **16-19** to α-aminonitriles **7** (via the corresponding cyanohydrins **20-23**) was observed (isobutyraldehyde **17**→2-hydroxy-3-methylbutane nitrile **21**→valine nitrile **7$_V$** [>95%]; acetaldehyde **18**→lactonitrile **22**→alanine nitrile **7$_A$** [80%]; formaldehyde **16**→glycolonitrile **20**→glycine nitrile **7$_G$** [60%]; glycolaldehyde **19**→glyceronitrile **23**→serine nitrile **7$_S$** [82%]). However, hydroxypivaldehyde **15** underwent rapid pantoamidine **25** formation (via pantoic acid nitrile **24**) followed by slower side reactions. See text for details.
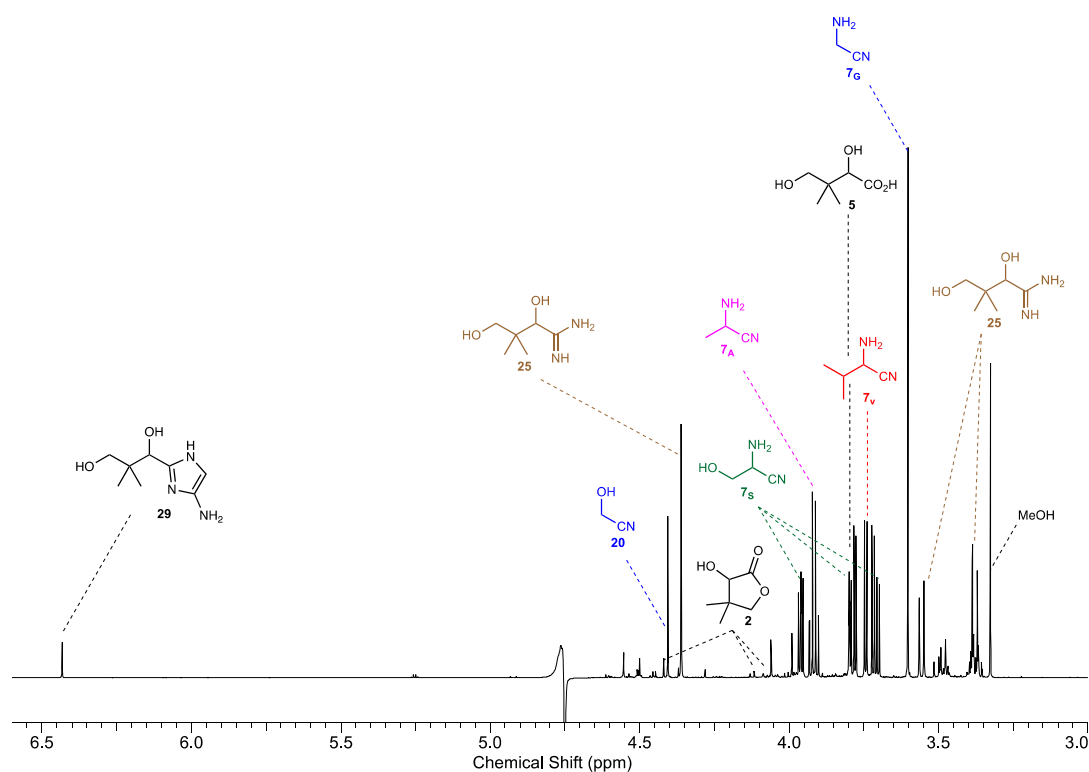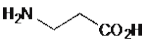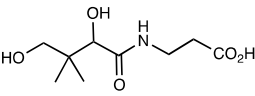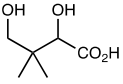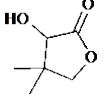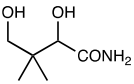
Supplementary Figure 70. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 3.0–6.6 ppm) spectrum to show the detection of the aminoimidazole **29** derived from the addition of glycine nitrile **7$_G$** to iminopantolactone **26** alongside the selective synthesis of proteinogenic α-aminonitriles **7** from a mixture of formaldehyde **16** (20 mM), acetaldehyde **18** (31.3 mM), isobutyraldehyde **17** (17.2 mM), glycolaldehyde **19** (20 mM), hydroxypivaldehyde **15** (20 mM), sodium cyanide (NaCN, 150 mM) and ammonium chloride (NH$_4$Cl; 500 mM) in phosphate buffer (pH 9.5; 500 mM) at 20 ºC after 3 days (the same spectrum including the region below 3.0 ppm is shown in Supplementary Figure 69C. Selective conversion of proteinogenic aldehydes **16**-**19** to α-aminonitriles **7** (via the corresponding cyanohydrins **20**-**23**) was observed (isobutyraldehyde **17**→2-hydroxy-3-methylbutane nitrile **21**→valine nitrile **7$_v$** [>95%]; acetaldehyde **18**→lactonitrile **22**→alanine nitrile **7$_A$** [80%]; formaldehyde **16**→glycolonitrile **20**→glycine nitrile **7$_G$** [60%]; glycolaldehyde **19**→glyceronitrile **23**→serine nitrile **7$_S$** [82%]).See text for details.

*Synthesis of pantothenic acid **6** by reaction of hydroxypivaldehyde **15** with β-alanine **3** and cyanide*
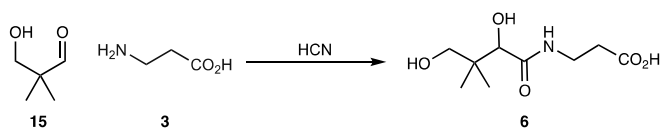
To a stirred solution of hydroxypivaldehyde **15** (1.6–50 mM), β-alanine **3** (3.1–100 mM) in phosphate buffer (pH 7.0; 10 equiv.) at 20 ºC was added sodium cyanide (2.4–75 mM). The solution was adjusted to the required pH by the addition of 1-8 M NaOH and the reaction was monitored periodically by NMR spectroscopy. Yields of products are reported in Supplementary Table 18. Representative ^{1}H NMR spectra are shown in Supplementary Figure 71 and Supplementary Figure 72.

Concentration (mM)		pH	Time (day)	Yield (%)			
15	**3**			**6**	**5**	**2**	**5-NH$_2$**
50[a]	100	7.5	13	21	30	12	18
50[a]	100	8.0	13	31	34	4	14
50[b]	100	9.0	5	43	54	0	tr.
25[b]	50	9.0	6	43	49	2	tr.
12.5[b]	25	9.0	7	51	44	tr.	tr.
6.3[b]	12.5	9.0	7	46	48	3	tr.
3.1[b]	6.3	9.0	7	36	55	tr.	tr.
1.6[b]	3.1	9.0	8	18	60	9	12

Supplementary Table 18. Quantification of species observed from the reaction of hydroxypivaldehyde **15** with sodium cyanide (NaCN; 1.5 equiv.) in phosphate buffer (10 equiv.) at the specified pH and concentrations and at 20 ºC. tr. = trace. [a] Reactions carried out with crude hydroxypivaldehyde **15** (88% purity) that was synthesised from the aldol condensation of isobutyraldehyde **17** and formaldehyde **16**, and contained residual formaldehyde **16** using potassium carbonate (K$_2$CO$_3$) according to Supplementary Reference [22]. [b] Reactions carried out with hydroxypivaldehyde **15** of >95% purity, isolated after aldol condensation of isobutyraldehyde **17** and formaldehyde **16** in phosphate buffer.
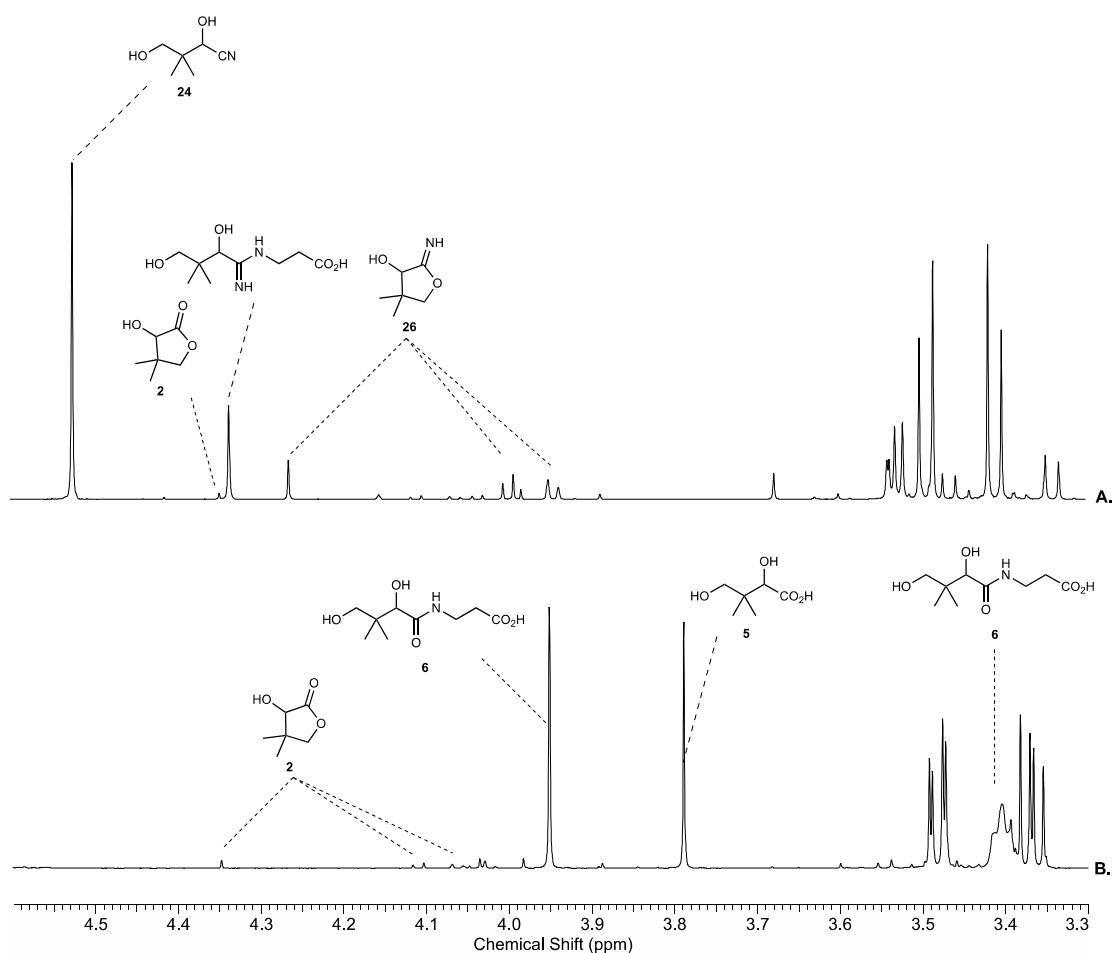
Supplementary Figure 71. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 3.3–4.6 ppm) spectra of the reaction of pantoic acid nitrile (generated *in situ* from hydroxypivaldehyde **15** (12.5 mM) and sodium cyanide (NaCN; 18.8 mM)) with β-alanine **3** (25 mM) in phosphate buffer (pH 9.0; 125 mM) at 20 ºC after (A.) 30 mins; and (B.) 7 days.
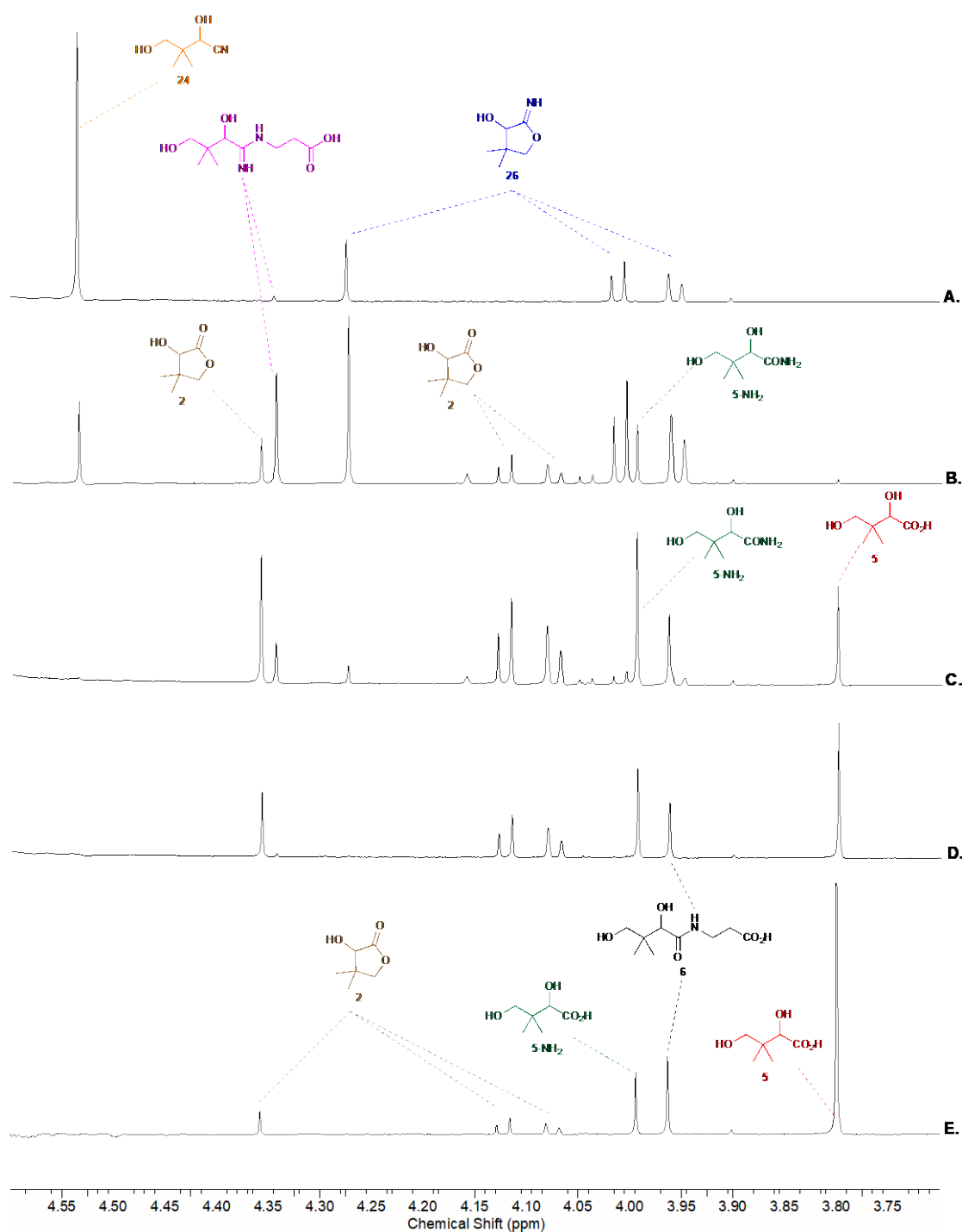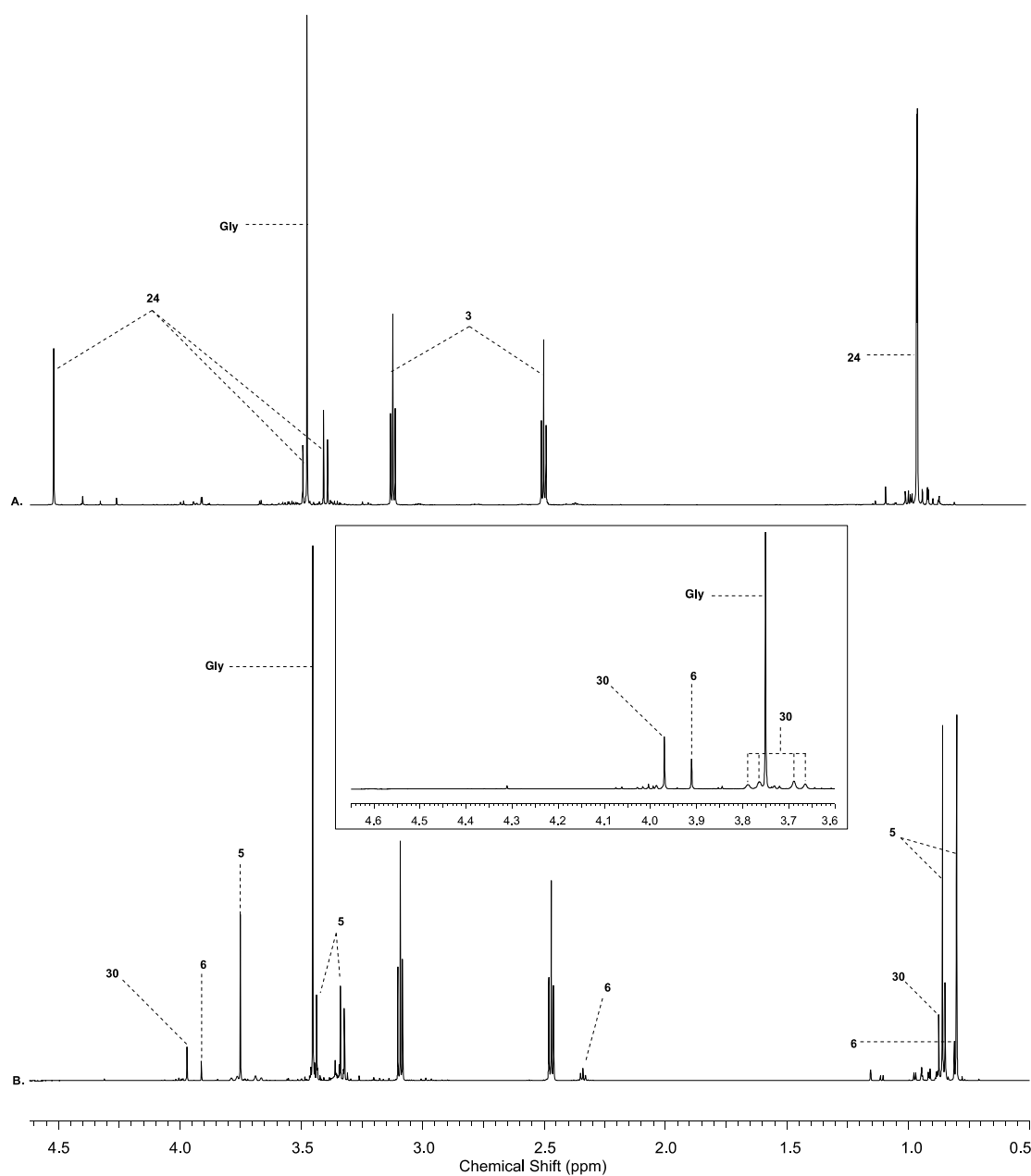
Supplementary Figure 72. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 3.7–4.6 ppm) spectra of the reaction of pantoic acid nitrile **24** (generated *in situ* from hydroxypivaldehyde **15** (1.6 mM) and sodium cyanide (NaCN; 2.4 mM)) with β-alanine **3** (3.2 mM) in phosphate buffer (pH 9.0; 16 mM) at 20 ºC after: (A.) 30 mins; (B.) 5 hours; (C.) 1 day; (D.) 2 days; and (E.) 7 days.

Synthesis of pantothenic acid nitrile 9 by reaction of hydroxypivaldehyde 15 with β-alanine nitrile 8 and cyanide
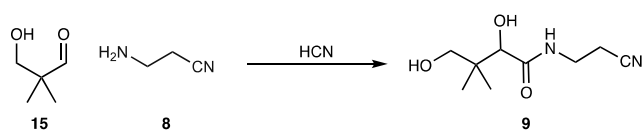
To a stirred solution of hydroxypivaldehyde **15** (1.6 or 3.1 mM), β-alanine nitrile **8** (3.1 - 6.3 mM), in phosphate buffer (pH 7.0; 16 - 31 mM) at 20 ºC was added sodium cyanide (1.1 equiv.). The solution was adjusted to the required pH by the addition of 1-8 M NaOH and the reaction was monitored periodically by NMR spectroscopy. Quantification of species are reported in Supplementary Table 19. Representative ¹H NMR spectra are shown in Supplementary Figure 73–Supplementary Figure 76.

Concentration (mM)		pH	Time (days)	Yield (%)				
15	8			9	5	5-NH₂	24	2
3.1	6.3	7	6	0	6	0	66	26
3.1	6.3	7.5	11	13	15	3	4	54
3.1	6.3	8	11	23	37	2	0	24
3.1	6.3	9	6	44	41	3	0	tr.
3.1	3.1	9	5	35	54	3	0	5
1.6	3.1	9	7	21	46	6	0	0

Supplementary Table 19. Quantification of species observed from the reaction of hydroxypivaldehyde **15** (1.6 or 3.1 mM) with sodium cyanide (NaCN; 1.1 equiv.) with β-alanine nitrile **8** (1 or 2 equiv.) in phosphate buffer (16 or 31 mM) at 20 ºC and specified pH.
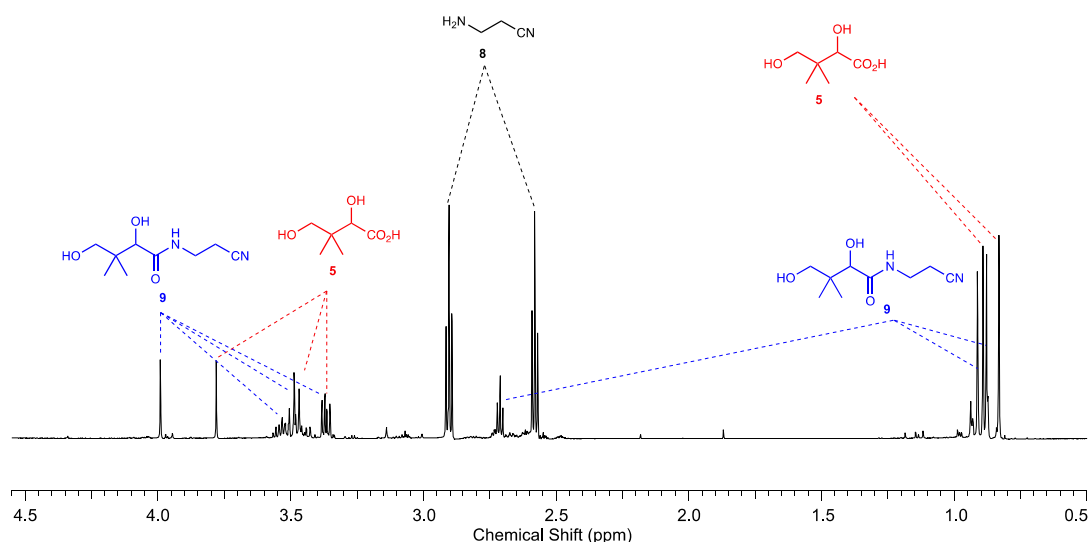
Supplementary Figure 73. ¹H NMR (700 MHz, H₂O, noesygppr1d, 0.5–3.5 ppm) spectrum of the reaction of pantoic acid nitrile **24** (generated *in situ* from hydroxypivaldehyde **15** (3.1 mM) and sodium cyanide (NaCN; 3.4 mM)) with β-alanine nitrile **8** (6.3 mM) in phosphate buffer (pH 9.0; 31 mM) at 20 ºC after 11 days.
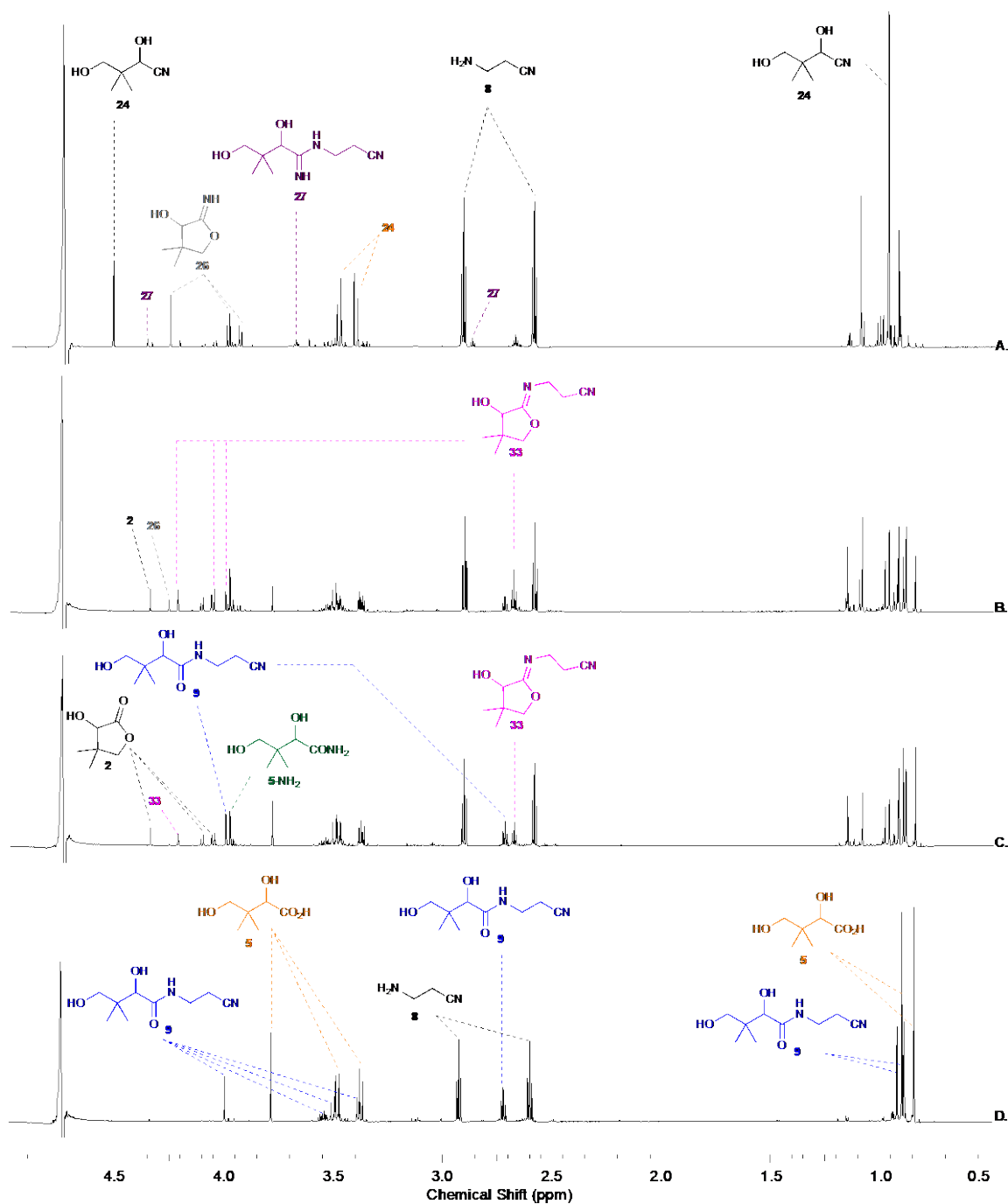
Supplementary Figure 74. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–5.0 ppm) spectra of the reaction of pantoic acid nitrile **24** (generated *in situ* from hydroxypivaldehyde **15** (3.1 mM) and sodium cyanide (NaCN; 3.4 mM)) with β-alanine nitrile **8** (3.1 mM) in phosphate buffer (pH 9.0; 31 mM) at 20 ºC after: (A.) 90 mins; (B.) 12 hours; (C.) 1 day; and (D.) 5 days.
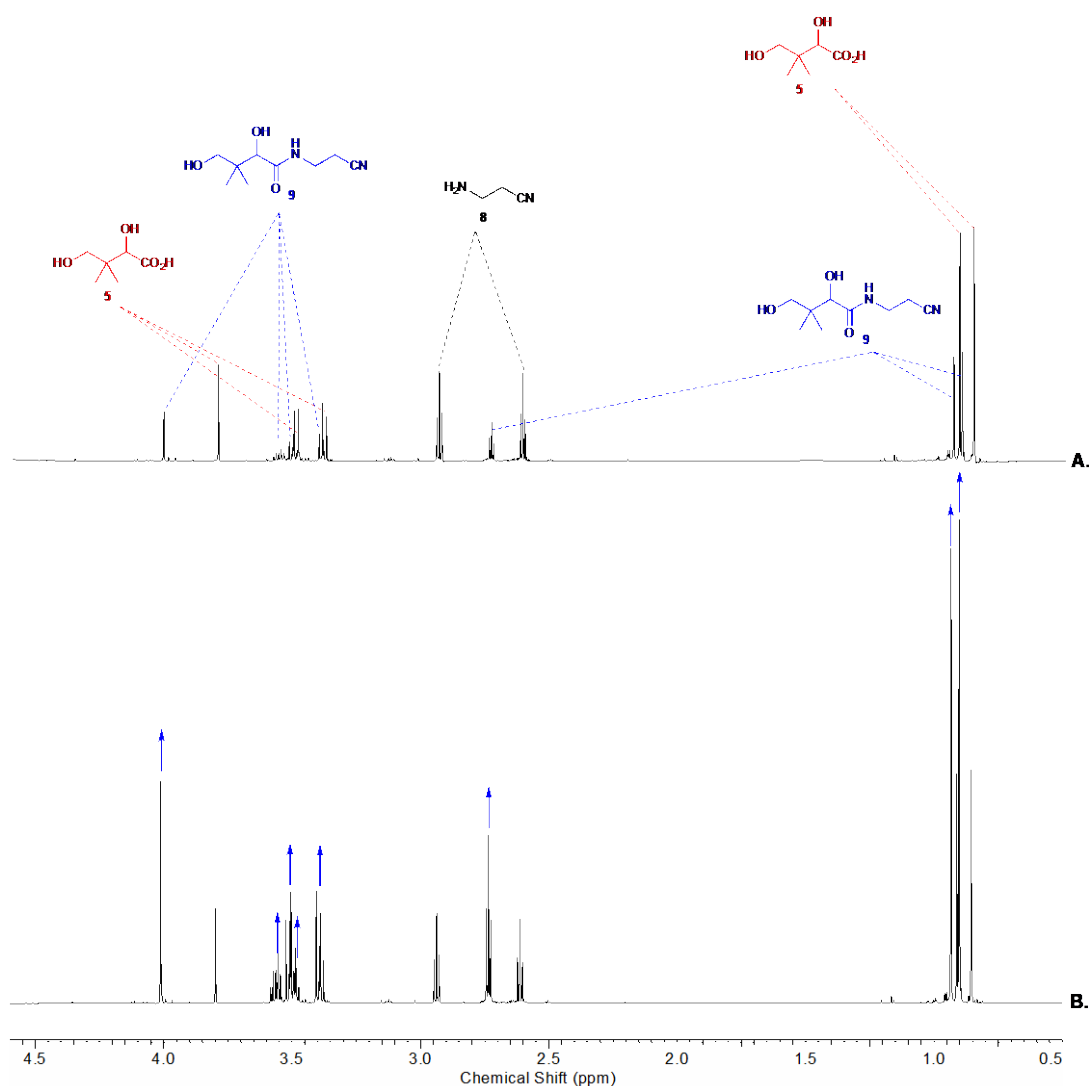
Supplementary Figure 75. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–5.0 ppm) spectra of the reaction of pantoic acid nitrile (generated *in situ* from hydroxypivaldehyde **15** (3.1 mM) and sodium cyanide (NaCN; 3.4 mM)) with β-alanine nitrile **8** (3.1 mM) in phosphate buffer (pH 9.0; 31 mM) at 20 ºC after: (A.) 5 days; and (B.) spiking with an authentic sample of pantothenic acid nitrile **9**.
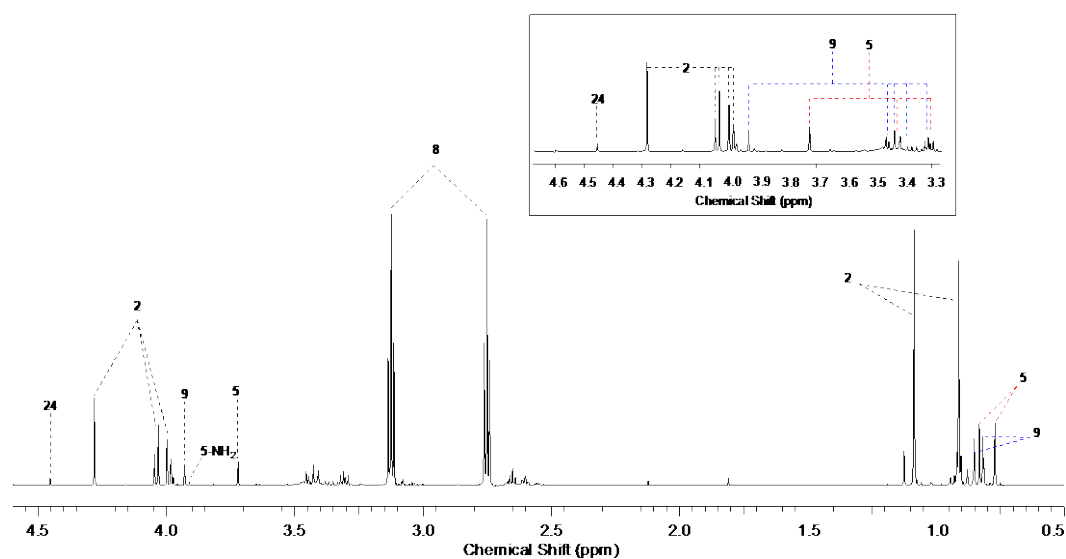
Supplementary Figure 76. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–5.0 ppm) spectra of the reaction of pantoic acid nitrile (generated *in situ* from hydroxypivaldehyde **15** (3.1 mM) and sodium cyanide (NaCN; 3.4 mM)) with β-alanine nitrile **8** (6.3 mM) in phosphate buffer (pH 7.5; 31 mM) at 20 ºC after 11 days, showing pantolactone **2** (54%) as the major pantoyl species.

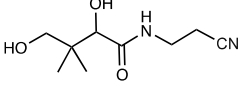*Competitive reaction of β-alanine nitrile **8** and glycine nitrile **7G** with hydroxypivaldehyde **15** and cyanide*
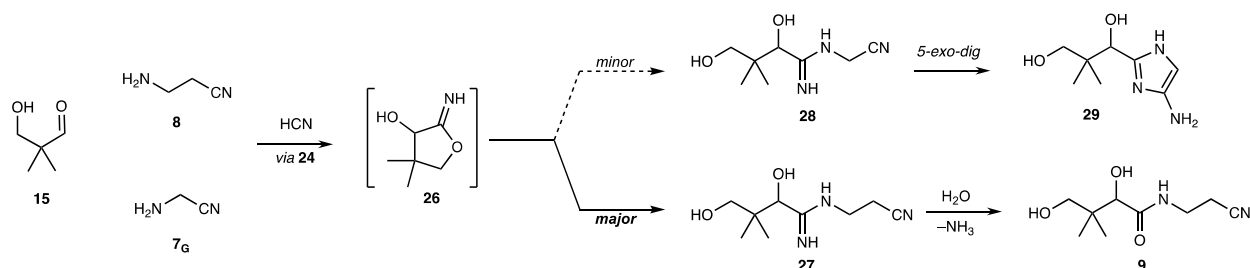
To a stirred solution of hydroxypivaldehyde **15** (3.1 mM), β-alanine nitrile **8** (6.3 mM), glycine nitrile **7G** (6.3 mM) in phosphate buffer (pH 7.0; 31 mM) at 20 °C was added sodium cyanide (1.5 equiv., 4.7 mM). The solution was adjusted from pH 7.0 to pH 9.0 by the addition of 1-8 M NaOH and the reaction was monitored periodically by NMR spectroscopy (Supplementary Figure 77 and Supplementary Figure 78) at 20 °C. Pantothenic acid nitrile **9** (44%) was observed as the major pantoylamide after 3 days, and pantoic acid **5** (44%). Pantoylglycine nitrile **30** was not observed because its formation was blocked by aminoimidazole **29** (<4%) formation, and no transamidation of **9** by **7G** to give **30** was observed*. The mechanism of **29** formation is proposed to go through the amidine **28** intermediate, which underwent rapid intramolecular cyclisation and aromatisation to **29**. The blocking of pantoylglycine nitrile **30** synthesis was further verified by the reaction of **15**, sodium cyanide, and glycine nitrile **7G** alone – see Supplementary Page 107 for further details.

---

* The transamidation of pantothenic acid nitrile **9** by glycine nitrile **7G** is sluggish and low-yielding, whereas the transamidation of pantoylglycine nitrile **30** by β-alanine nitrile **8** to give pantothenic acid nitrile **9** under comparable reaction conditions is high-yielding. See Supplementary Pages S57–S62.
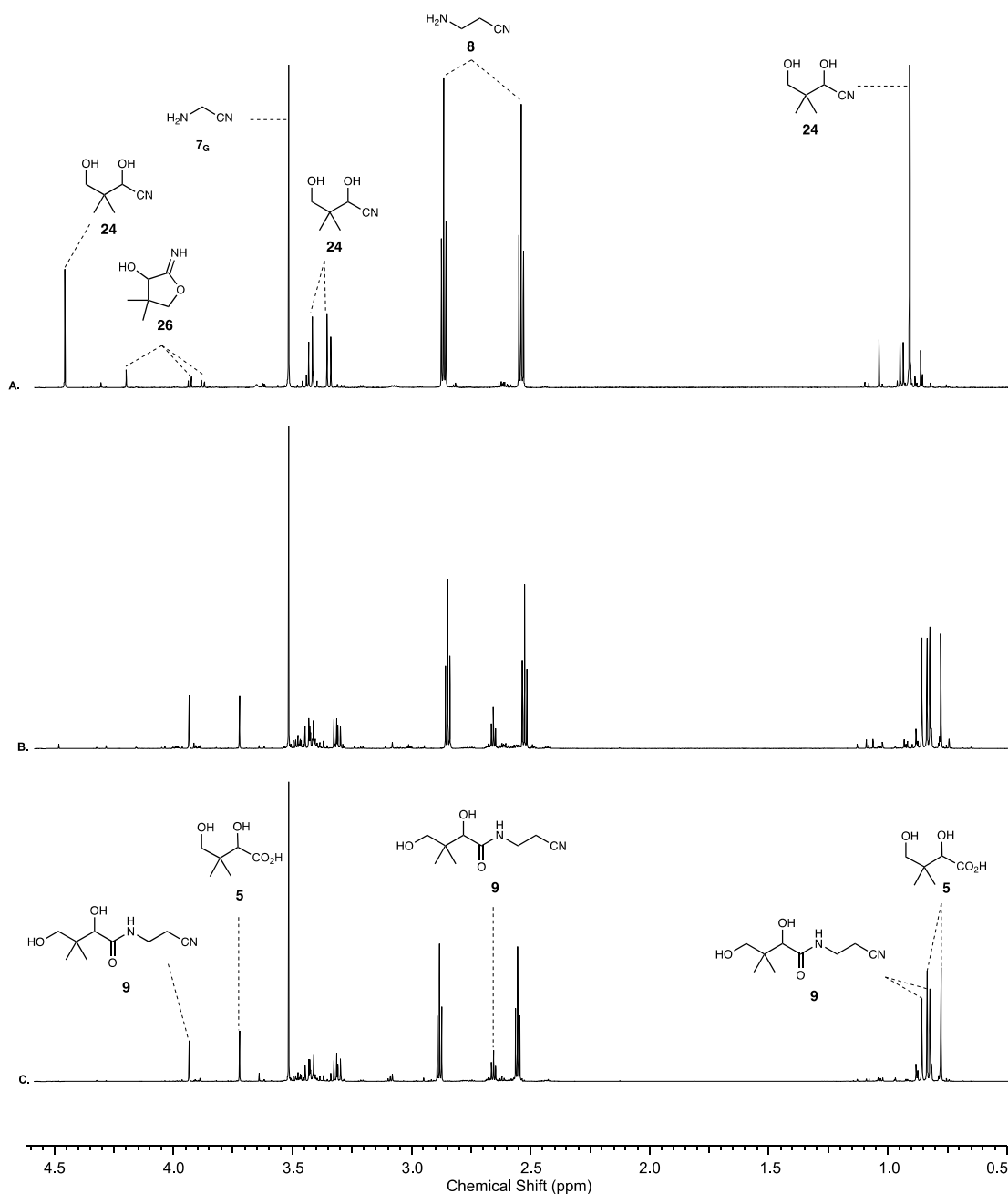
Supplementary Figure 77. ^{1}H NMR (700 MHz, H_2O, noesygppr1d, 0.5–4.5 ppm) spectra of the competitive reaction of pantoic acid nitrile (generated *in situ* from hydroxypivaldehyde **15** (3.1 mM) and sodium cyanide (NaCN; 4.7 mM)) with β-alanine nitrile **8** (6.3 mM) in the presence of glycine nitrile **7$_G$** (6.3 mM) in phosphate buffer (pH 9.0; 31 mM) at 20 °C after: (A.) 10 mins; (B.) 3 days; and (C.) 9 days, showing that pantothenic acid nitrile **9** does not undergo transamidation with **7$_G$** to give pantoylglycine nitrile **14**. See Supplementary Figure 78 for an expanded ^{1}H NMR spectrum.
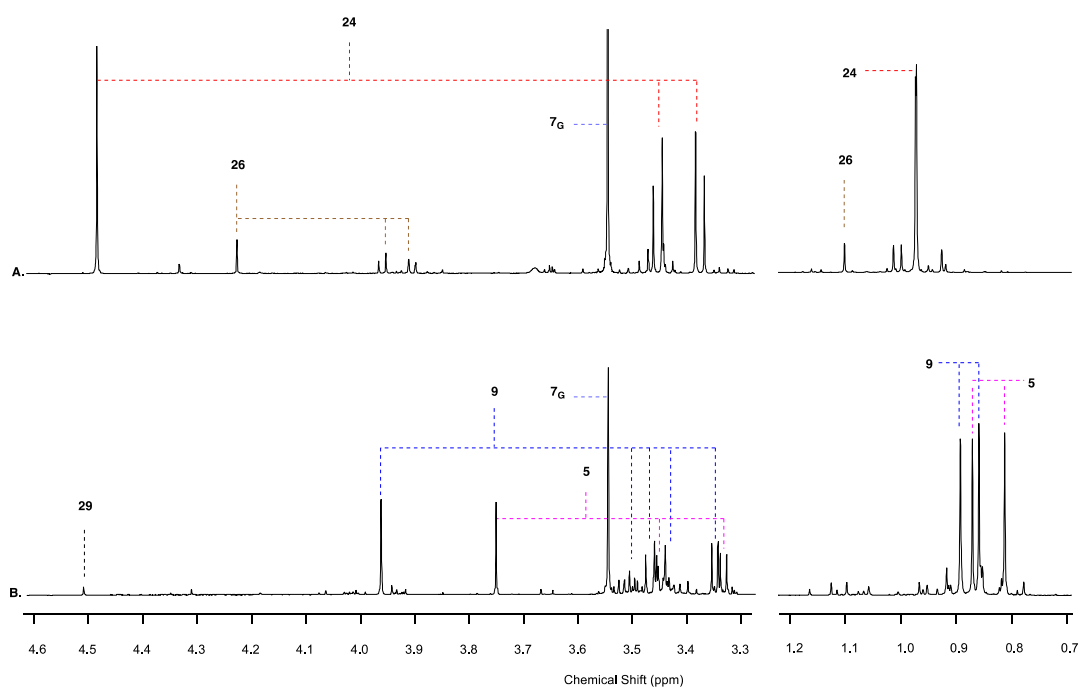
Supplementary Figure 78. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d) expanded spectra (0.7–1.2 ppm and 3.3–4.6 ppm) of the competitive reaction of pantoic acid nitrile (generated *in situ* from hydroxypivaldehyde **15** (3.1 mM) and sodium cyanide (NaCN; 4.7 mM)) with β-alanine nitrile **8** (6.3 mM) in the presence of glycine nitrile **7$_G$** (6.3 mM) in phosphate buffer (pH 9.0; 31 mM) at 20 ºC after: (A.) 10 mins; and (B.) 3 days.

Attempted synthesis of pantoylglycine nitrile 14 by reaction of hydroxypivaldehyde 15 with glycine nitrile 7$_G$ and cyanide leads to aminoimidazole formation
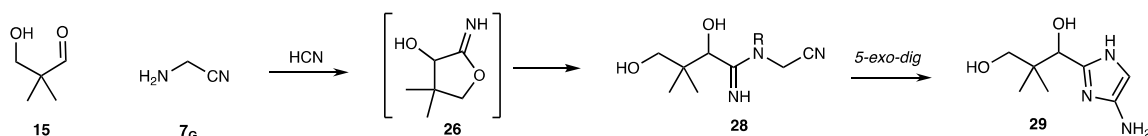
To a stirred solution of hydroxypivaldehyde **15** (50 mM), glycine nitrile **7$_G$** (2 equiv.), in phosphate buffer (pH 7.0; 500 mM) at 20 °C was added sodium cyanide (1.1 equiv., 55 mM). The solution was adjusted to pH 9.0 by the addition of 1-8 M NaOH and the reaction was monitored periodically by NMR spectroscopy. The conversion of pantoic acid nitrile **24** to new aromatic species assigned as the aminoimidazole **29** (42%) was observed after 16 hours[*]. Pantoylglycine nitrile **14** and amidine **28** were not observed by NMR spectroscopy (Supplementary Figure 79.).

Data for aminoimidazole 29
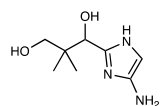
^{1}H NMR (600 MHz, noesygppr1d, H_2O) δ 6.32 (s, 1H, (C5$_{Ar}$)–H), 4.45 (s, 1H, (C2$_{Ar}$)–CH), 3.38 (AB, J = 11.3 Hz, 1H, CH$_{2a}$OH), 3.26 (AB, J = 11.3 Hz, 1H, CH$_{2b}$OH), 0.74 (s, 3H, CH$_3'$), 0.70 (s, 3H, CH$_3''$). ^{13}C NMR (151 MHz, noesygppr1d, H_2O) δ 145.2 (C2$_{Ar}$), 142.4 (C4$_{Ar}$), 100.8 (C5$_{Ar}$), 73.2 ((C2$_{Ar}$)–CH), 68.9 (CH$_2$OH), 40.1 (C(CH$_3$)$_2$), 21.3 (CH$_3$), 19.5 (CH$_3'$). HRMS-ESI [M+H]$^+$ calculated for $C_8H_{16}N_3O_2$ 186.1237; observed 186.1237.

---

[*] An aromatic resonance at 6.34 ppm in the ^{1}H NMR spectrum was observed, and assumed to be a second imidazole **29′** (15%) species. However, unambiguous characterisation of **29′** could not be made. See Supplementary Figure 79.
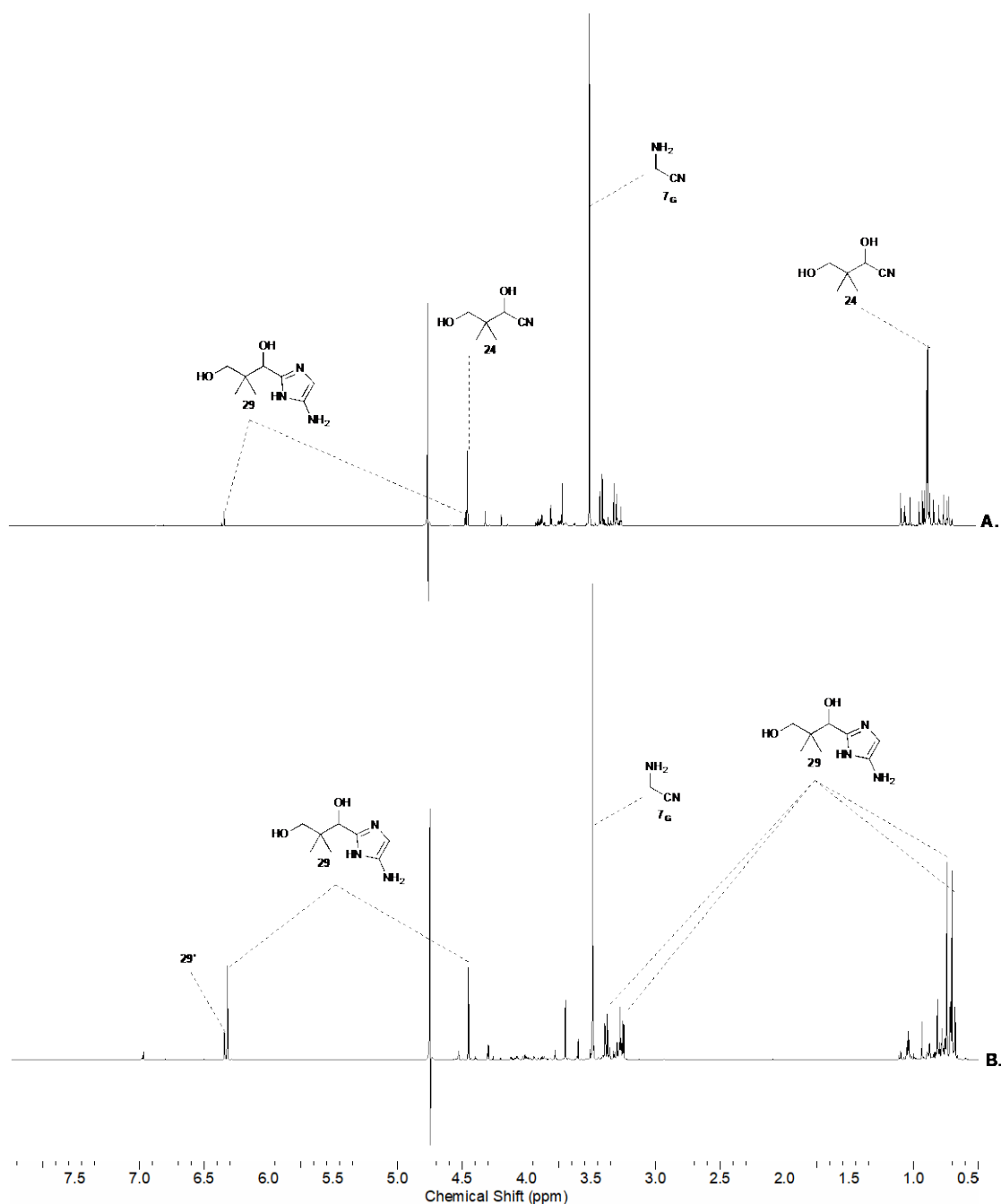
Supplementary Figure 79. ^{1}H NMR (600 MHz, H$_2$O, noesygppr1d, 0.5–8.0 ppm) spectra of the reaction of pantoic acid nitrile **24** (generated *in situ* from hydroxypivaldehyde **15** (50 mM) and sodium cyanide (NaCN; 55 mM)) with glycine nitrile **7$_G$** (2 equiv.) in phosphate buffer (pH 9.0; 500 mM) at 20 ºC after: (A.) 1 hour; and (B.) 16 hours.

Competitive reaction of β-alanine 3 and glycine Gly with hydroxypivaldehyde 15 and cyanide
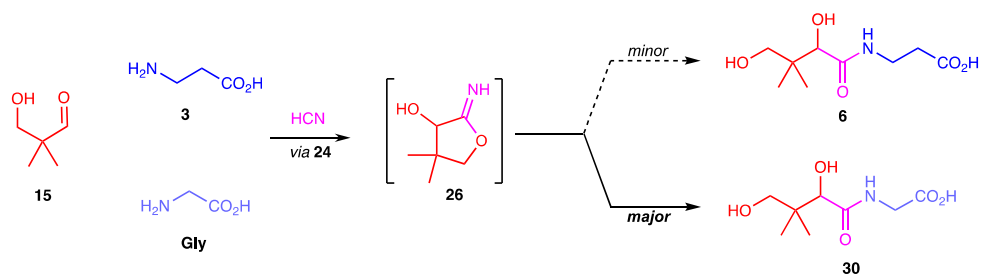
To a stirred solution of hydroxypivaldehyde **15** (3.1 mM), β-alanine **3** (2 equiv.), and glycine **Gly** (2 equiv.) in phosphate buffer (pH 7.0; 31 mM) at 20 ºC was added sodium cyanide (1.5 equiv., 4.7 mM). The solution was adjusted to pH 9.0 by the addition of 1-8 M NaOH and the reaction was monitored periodically NMR spectroscopy (Supplementary Figure 80). Pantoylglycine **30** (16%) was observed as the major pantoylamide species alongside pantothenic acid **6** (9%) as the minor species, and pantoic acid **5** (65%) after 8 days.

Supplementary Figure 80. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–4.5 ppm) spectrum of the competitive reaction of pantoic acid nitrile **24** (generated *in situ* from hydroxypivaldehyde **15** (3.1 mM) and sodium cyanide (NaCN; 4.7 mM)) with β-alanine **3** (2 equiv.) and glycine **Gly** (2 equiv.) in phosphate buffer (pH 9.0; 31 mM) at 20 ºC after: (A.) 10 mins, and (B.) 8 days. Inset: expanded ^{1}H NMR (3.6–4.6 ppm) spectral region.

*Competitive reaction of β-alanine **3** and β-alanine nitrile **8** with hydroxypivaldehyde **15** and cyanide*
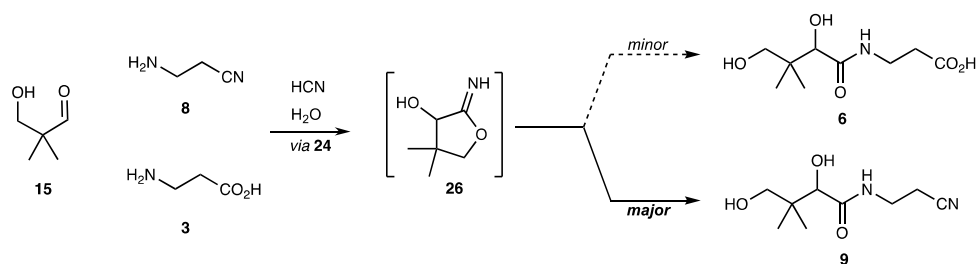
To a stirred solution of hydroxypivaldehyde **15** (3.1 mM), β-alanine **3** (1 equiv.), and β-alanine nitrile **8** (1 equiv.) in phosphate buffer (pH 7.0; 31 mM) at 20 ºC was added sodium cyanide (1.5 equiv., 4.7 mM). The solution was adjusted to pH 9.0 by the addition of 1-8 M NaOH. The reaction was monitored periodically by NMR spectroscopy (Supplementary Figure 81). The formation of pantothenic acid **6** (12%) and pantothenic acid nitrile **9** (25%) was observed after 10 days.
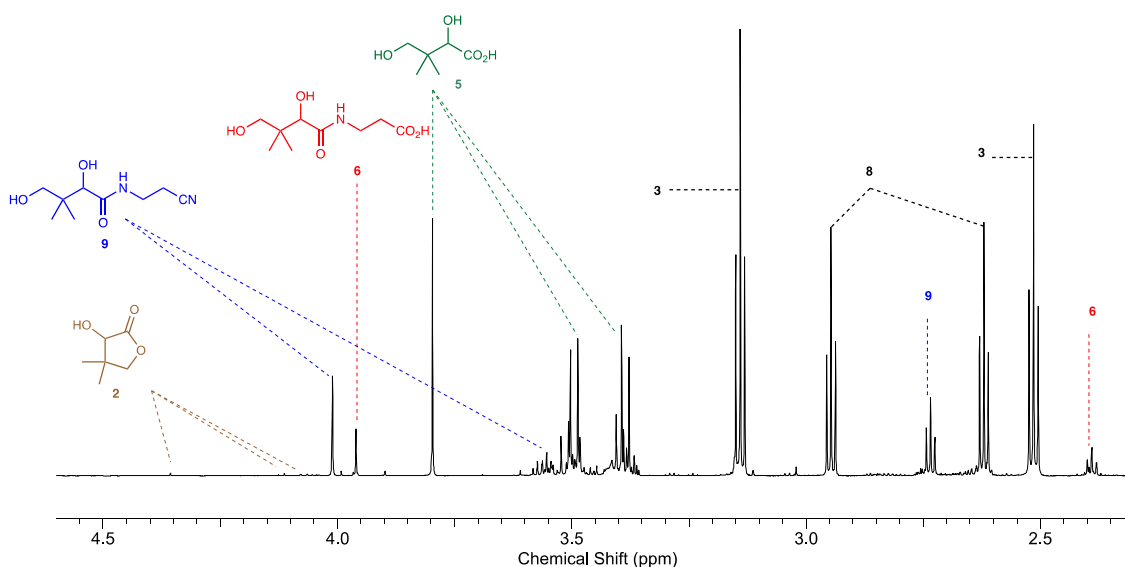
Supplementary Figure 81. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 2.3–4.6 ppm) spectrum of the competitive reaction of pantoic acid nitrile **24** (generated *in situ* from hydroxypivaldehyde **14** (3.1 mM) and sodium cyanide (NaCN; 4.7 mM)) with β-alanine **3** (1 equiv.) and β-alanine nitrile **8** (1 equiv.) in phosphate buffer (pH 9.0; 31 mM) at 20 ºC after 10 days.

Conversion of pantoamidine 25 (formed under Strecker conditions) to pantothenic acid nitrile 9 by β-alanine nitrile 8 in the presence of a large excess of ammonia
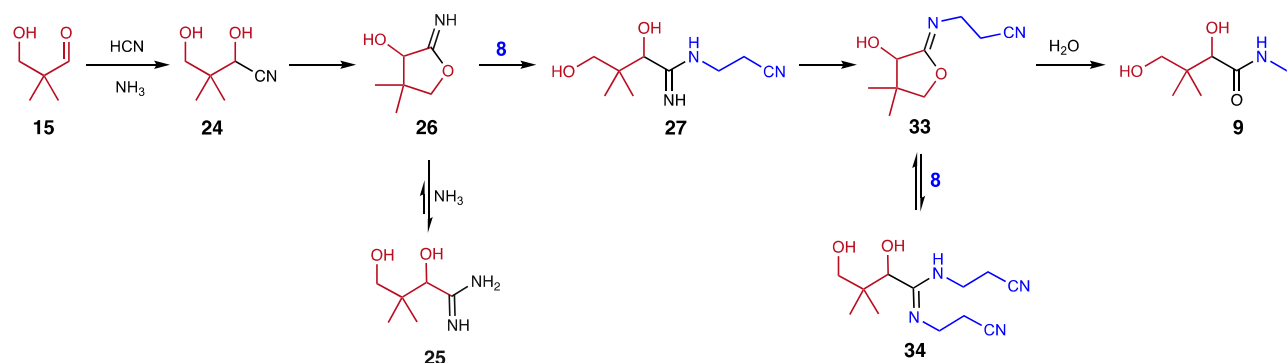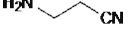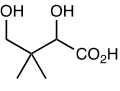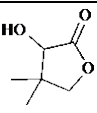
Hydroxypivaldehyde 15 at 20 mM

To a stirred solution of hydroxypivaldehyde **15** (20 mM), ammonium chloride (5 or 25 equiv.) in phosphate buffer (pH 7.0; 0.5 M) at 20 ºC was added sodium cyanide (1.5 equiv.). The solution was adjusted to pH 9.5 with 1-8 M NaOH and the reaction was monitored by NMR spectroscopy. After 4.5 hours, pantoamidine **25** (82%) was observed as the major product alongside iminopantolactone **26** (5%), pantoic acid nitrile **24** (3%) and pantolactone **2** (3%) (Supplementary Figure 82). β-Alanine nitrile **8** (1–5 equiv.) was then added, the solution was adjusted from pH 9.5 to pH 9.0 with 1-4 M HCl. The reaction was monitored periodically by NMR spectroscopy for 6 days at 20 ºC. Yields of products are reported in Supplementary Table 20. ¹H NMR spectra are shown in Supplementary Figure 82–Supplementary Figure 84. Spiking with authentic pantothenic acid nitrile **9** confirmed its identification (Supplementary Figure 85).

Hydroxypivaldehyde 15 at 100 mM

To a stirred solution of hydroxypivaldehyde **15** (100 mM), ammonium chloride (5 equiv.) in phosphate buffer (pH 7.0; 0.5 M) at 20 ºC was added sodium cyanide (1.5 equiv.). The solution was adjusted to pH 9.5 with 1-8 M NaOH and the reaction was monitored by NMR spectroscopy. After 4 hours, pantoamidine **25** (86%) was observed as the major product alongside iminopantolactone **26** (6%), pantoic acid nitrile **24** (1%) and pantolactone **2** (4%). β-Alanine nitrile **8** (1–5 equiv.) was then added, the solution was adjusted from pH 9.5 to pH 9.0 with 1-4 M HCl. The reaction was monitored periodically by NMR spectroscopy at 20 ºC. Yields of products are reported in Supplementary Table 20.

Entry	Concentration (mM)			Time (day)	Yield (%)		
	15	NH₃	8		9	5	2
1	20	500	20	6	32	59	3
2	20	500	40	6	46	45	2
3	20	500	100	6	64	29	2
4	100	500	100	12	25	61	1
5	100	500	200	12	38	44	2
6	100	500	500	11	58	24	1

Supplementary Table 20. Quantification of species observed from the reaction of hydroxypivaldehyde **15** (20 or 100 mM), sodium cyanide (NaCN; 1.5 equiv.) and ammonium chloride (5 or 25 equiv.) in phosphate buffer (pH 9.5; 0.5 M) followed by the addition of β-alanine nitrile **8** (1-5 equiv.; 4.5 hours after the reaction was initiated) and then incubated at pH 9.0 and 20 ºC for the specified time.
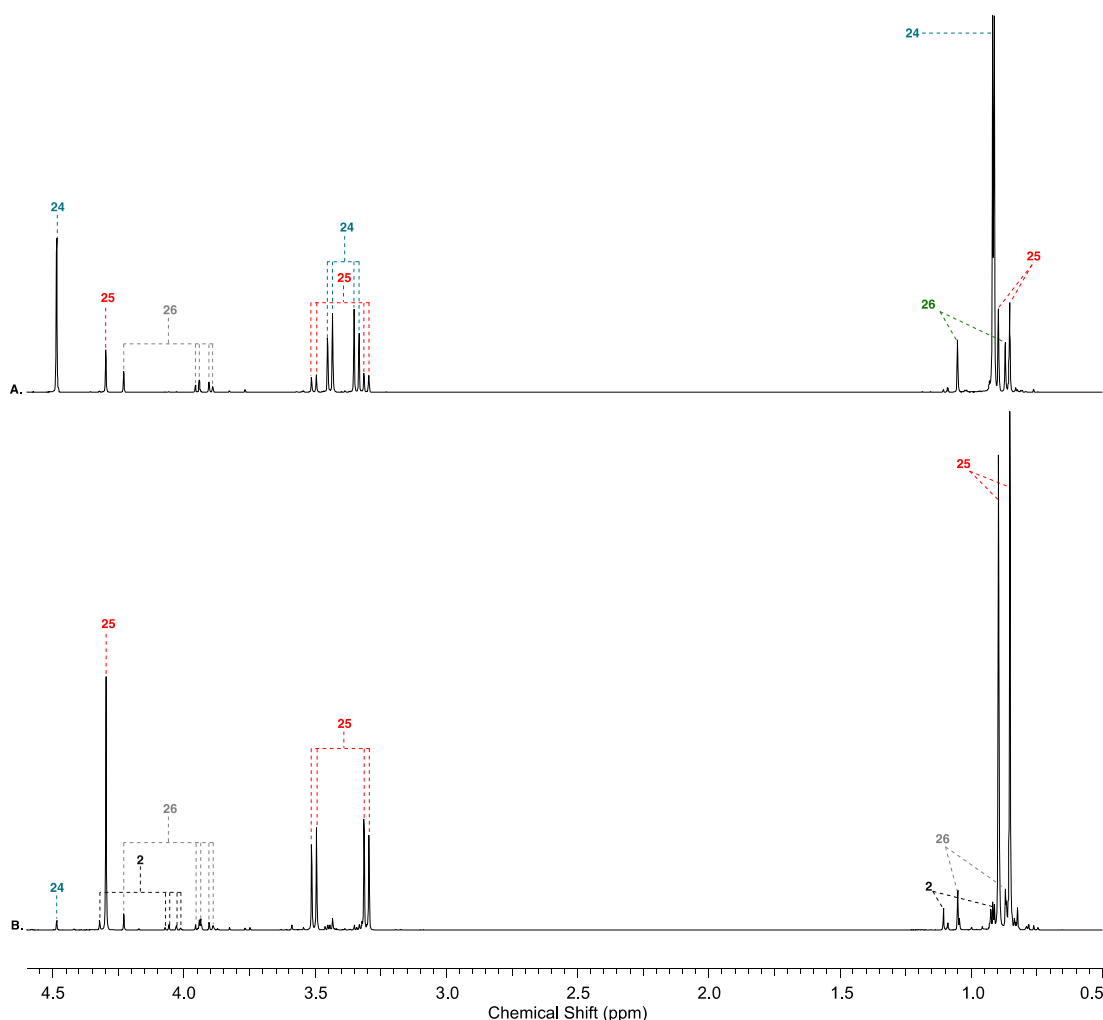
Supplementary Figure 82. ^{1}H NMR (600 MHz, H₂O, noesygppr1d, 0.5–4.5 ppm) spectra to show the reaction of hydroxypivaldehyde **15**, sodium cyanide (NaCN; 1.5 equiv.) and ammonium chloride (25 equiv.) in phosphate buffer (pH 9.5; 0.5 M) at 20 ºC after: (A.) 10 mins; and (B.) 4.5 hours.

*Data for pantoamidine **25** (Supplementary Figure 82)*
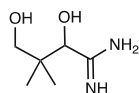
^{1}H NMR (600 MHz, H$_2$O, noesygppr1d) δ_H 4.30 (s, 1H, (C2)–H), 3.50 (AB, J = 11.4 Hz, 1H, (C4)–H),

3.31 (AB, J = 11.4 Hz, 1H, (C4)–H′), 0.90 (s, 3H, CH$_3$), 0.85 (s, 3H, CH$_3$′).

*Data for pantoic acid nitrile **24** (Supplementary Figure 82)*
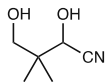
^{1}H NMR (600 MHz, H$_2$O, noesygppr1d) δ_H 4.49 (s, 1H, (C2)–H), 3.44 (AB, J = 11.5 Hz, 1H, (C4)–H), 3.34 (d, J = 11.5 Hz, 1H, (C4)–H′), 0.92 (s, 3H, CH$_3$), 0.91 (s, 3H, (CH$_3$′).

*Data for iminopantolactone **26** (Supplementary Figure 82)*
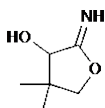
^{1}H NMR (600 MHz, H$_2$O, noesygppr1d) δ_H 4.23 (s, 1H, (C2)–H), 3.95 (AB, J = 8.8 Hz, 1H, (C4)–H), 3.90 (d, J = 8.8 Hz, 1H, (C4)–H′), 1.05 (s, 3H, CH$_3$), 0.87 (s, 3H, CH$_3$′).

*Data for pantolactone **2** (Supplementary Figure 82)*
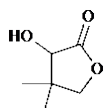
^{1}H NMR (600 MHz, H$_2$O, noesygppr1d) δ_H 4.32 (s, 1H, (C2)–H), 4.07 (AB, J = 9.0 Hz, 1H, (C4)–H),

4.02 (AB, J = 9.0 Hz, 1H, (C4)–H′), 1.11 (s, 3H, CH$_3$), 0.92 (s, 3H, CH$_3$′).
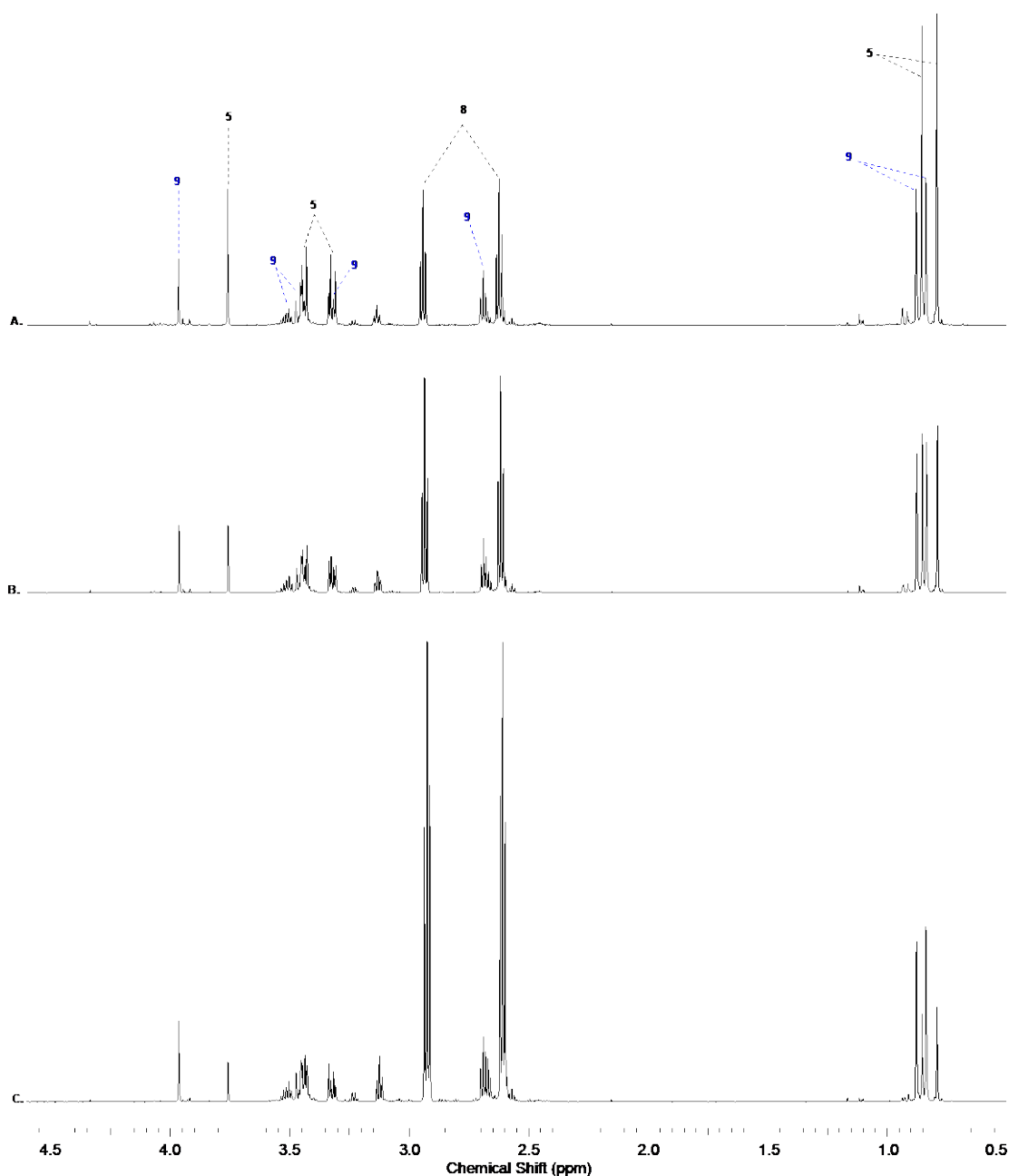
Supplementary Figure 83. The formation of pantothenic acid nitrile **9** at pH 9.0 and 20 ºC after the addition of β-alanine nitrile **8** (5 equiv.) to the mixture generated by the reaction of hydroxypivaldehyde **15** (20 mM), sodium cyanide (NaCN; 1.5 equiv.), and ammonium chloride (25 equiv.) in phosphate buffer (0.5 M) (see Supplementary Figure 82B for the NMR spectrum of the reaction mixture before the addition of **8**, where pantoamidine **25** (82%) was the major pantoyl species). The formation of pantothenic acid nitrile **9** was observed at 20 ºC after 6 days, even in the presence of excess ammonia (>24 equiv.) ^{1}H NMR (600 MHz, H$_2$O, noesygppr1d, 0.5–4.5 ppm) spectra show the synthesis of pantothenic acid nitrile **9** from amidine **25**, six days after the addition of β-alanine nitrile **8**: (A.) 1 equiv. of **8** → **9** (32%); (B.) 2 equiv. of **8** → **9** (46%); or (C.) 5 equiv. of **8** → **9** (64%). See text for further experimental details. See Supplementary Figure 84 for representative time course spectra for the reaction shown in spectrum B.
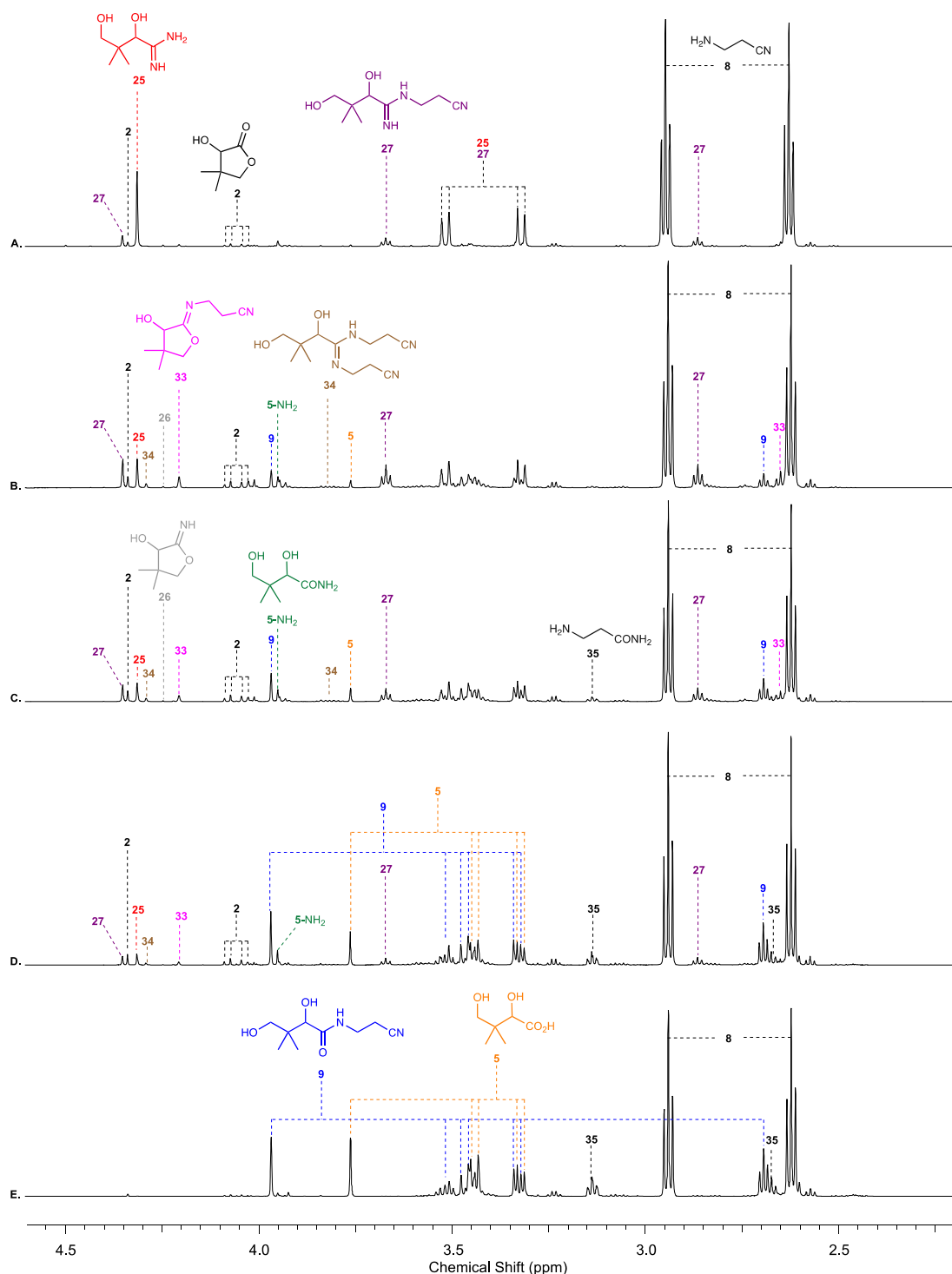
Supplementary Figure 84. ^{1}H NMR (600 MHz, H$_2$O, noesygppr1d, 2.2–4.5 ppm) spectra to show intermediates observed during the formation of pantothenic acid nitrile **9** at pH 9.0 and 20 ºC after the addition of β-alanine nitrile **8** (2 equiv.) to the reaction mixture generated by the reaction of hydroxypivaldehyde **15** (20 mM), sodium cyanide (NaCN; 1.5 equiv.), and ammonium chloride (25 equiv.) in phosphate buffer (0.5 M) (see Supplementary Figure 82B for reaction mixture before the addition of **8**, where pantoamidine **25** (82%) was the major pantoyl species before addition of **8**). The formation of pantothenic acid nitrile **9** was observed at 20 ºC after 6 days, even in the presence of excess ammonia (>24 equiv.). ^{1}H NMR spectra showing the reaction after: (A.) 10 mins; (B.) 12 hours; (C.)1 day; (D.) 2 days; and (E.) 6 days. See text for further experimental details.
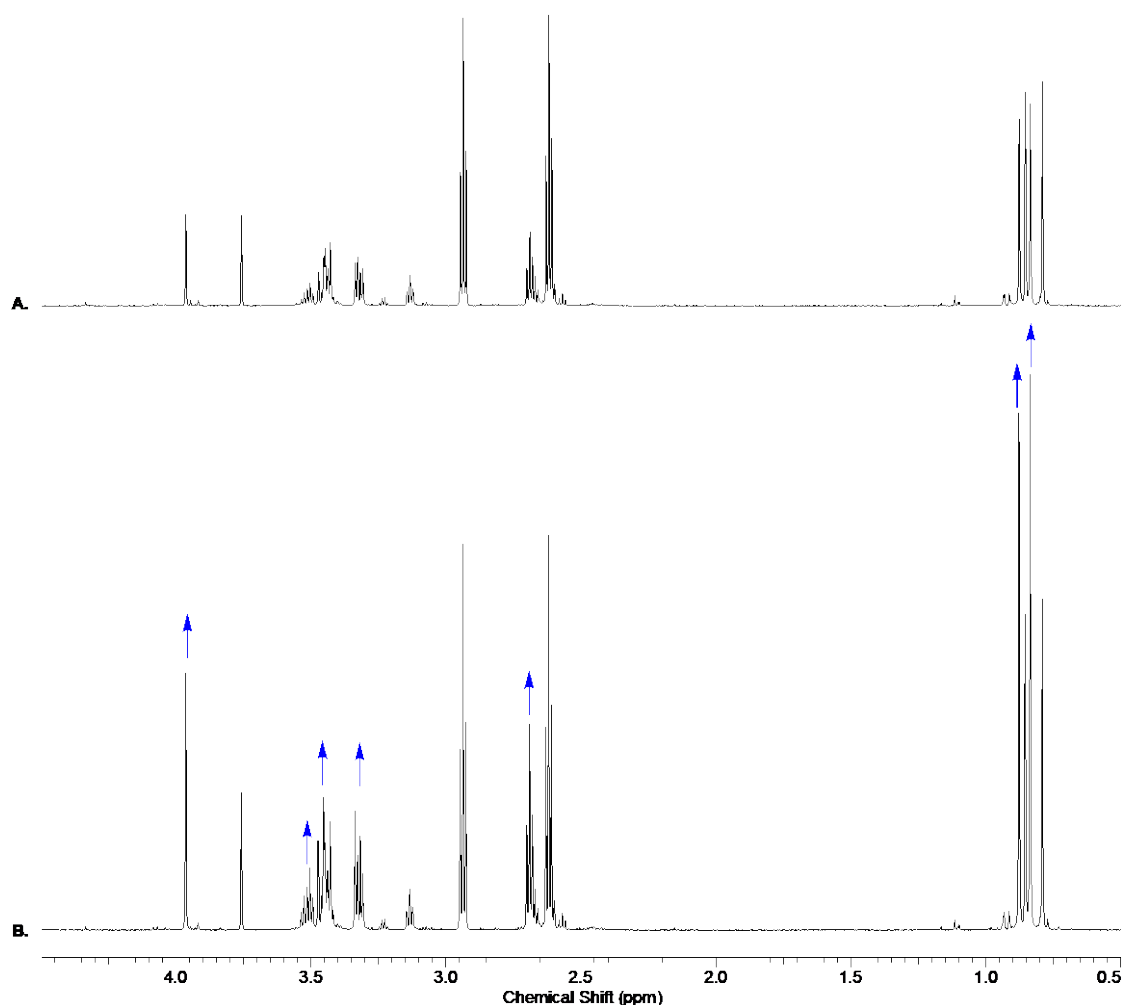
Supplementary Figure 85. ¹H NMR (600 MHz, H₂O, noesygppr1d, 0.5–4.5 ppm) spectra to show the confirmation of the identity of pantothenic acid nitrile **9** formation by spiking with an authentic sample of **9**. (A.) The crude reaction mixture generated at pH 9.0 and 20 ºC after the addition of β-alanine nitrile **8** (2 equiv.) to the reaction mixture generated by the reaction of hydroxypivaldehyde **15** (20 mM), sodium cyanide (NaCN; 1.5 equiv.), and ammonium chloride (25 equiv.) in phosphate buffer (0.5 M) (see Supplementary Figure 82B for reaction mixture before the addition of **8**, where pantoamidine **25** (82%) was the major pantoyl species before addition of **8**). The formation of pantothenic acid nitrile **9** was observed at 20 ºC after 6 days, even in the presence of excess ammonia (>24 equiv.). (B.) Spiking with an authentic sample of **9** showing an enhancement of ¹H NMR resonances corresponding to **9** (blue arrows).

Data for pantothenic acid nitrile **9** (Supplementary Figure 84)
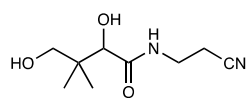
¹H NMR (600 MHz, H₂O, noesygppr1d) δ$_H$ 3.96 (s, 1H, (C2)–H), 3.55 - 3.48 (m, 1H, CH₂CH₂CN), 3.46 (AB, J = 11.3 Hz, 1H, (C4)–H), 3.47 - 3.42 (m, 1H, CH₂CH₂CN), 3.33 (AB, J = 11.3 Hz, 1H, (C4)–H′), 2.69 (t, J = 6.3 Hz, 2H, CH₂CH₂CN), 0.88 (s, 3H, CH₃), 0.84 (s, 3H, CH₃′).

Data for pantothenic acid **5** (Supplementary Figure 84)
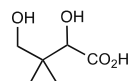
¹H NMR (600 MHz, H₂O, noesygppr1d) δ$_H$ 3.76 (s, 1H, (C2)–H), 3.44 (AB, J = 11.3 Hz, 1H, (C4)–H), 3.32 (AB, J = 11.3 Hz, 1H, (C4)–H′), 0.85 (s, 3H, CH₃), 0.79 (s, 3H, CH₃′).

Data for pantoamidine **25** (Supplementary Figure 84)
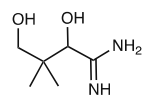
¹H NMR (600 MHz, H₂O, noesygppr1d) δ_H 4.31 (s, 1H, (C2)–H), 3.51 (AB, J = 11.4 Hz, 1H, (C4)–H), 3.32 (AB, J = 11.4 Hz, 1H, (C4)–H′), 0.90 (s, 3H, CH₃), 0.86 (s, 3H, CH₃′).

*Data for amidine **27** (Supplementary Figure 84)*
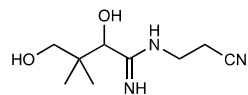
¹H NMR (600 MHz, H₂O, noesygppr1d) δ_H 4.35 (s, 1H, (C2)–H), 3.67 (t, J = 6.3 Hz, 2H, C$\underline{H_2}$CH₂CN), 3.51 (AB, J = 11.4 Hz, 1H, (C4)–H), 3.32 (AB, J = 11.4 Hz, 1H, (C4)–H′), 2.86 (t, J = 6.3 Hz, 1H, CH₂$\underline{CH_2}$CN), 0.91 (s, 3H, CH₃), 0.87 (s, 3H, CH₃′).

*Data for pantolactone **2** (Supplementary Figure 84)*
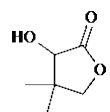
¹H NMR (600 MHz, H₂O, noesygppr1d) δ_H 4.33 (s, 1H, (C2)–H), 4.08 (AB, J = 8.9 Hz, 1H, (C4)–H), 4.03 (d, J = 8.9 Hz, 1H, (C4)–H′), 1.11 (s, 3H, CH₃), 0.93 (s, 3H, CH₃′).

*Data for pantoamide **5**-NH₂ (Supplementary Figure 84)*
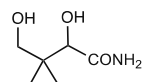
¹H NMR (600 MHz, H₂O, noesygppr1d, partial assignment) δ_H 3.95 (s, 1H, (C2)–H).

*Data for iminopantolactone **26** (Supplementary Figure 84)*
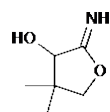
¹H NMR (600 MHz, H₂O, noesygppr1d, partial assignment) δ_H 4.24 (s, 1H, (C2)–H).

*Data for iminopantolactone **33** (Supplementary Figure 84)*
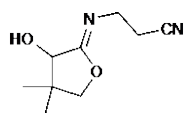
¹H NMR (600 MHz, H₂O, noesygppr1d, partial assignment) δ_H 4.20 (s, 1H, (C2)–H), 4.01 (AB, J = 8.9 Hz, 1H, (C4)–H), 3.93 (AB, J = 8.9 Hz, 1H, (C4)–H′), 2.64 (t, J = 6.3 Hz, 2H, CH₂$\underline{CH_2}$CN), 1.05 (s, 3H, CH₃).

*Data for amidine **34** (Supplementary Figure 84)*
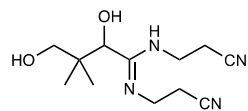
¹H NMR (600 MHz, H₂O, noesygppr1d, partial assignment) δ_H 4.28 (s, 1H, (C2)–H), 3.86 - 3.77 (m, 2H, NCH₂).

*Data for β-alaninamide **35** (Supplementary Figure 84)*
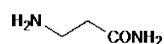
¹H NMR (600 MHz, H₂O, noesygppr1d) δ_H 3.13 (t, J = 6.9 Hz, 2H, (C3)–H₂), 2.67 (t, J = 6.9 Hz, 2H, (C2)–H₂).

Conversion of pantoamidine 25 (formed under Strecker conditions) to pantetheine 1 by β-alanine nitrile 8 and cysteamine 4 in the presence of a large excess ammonia
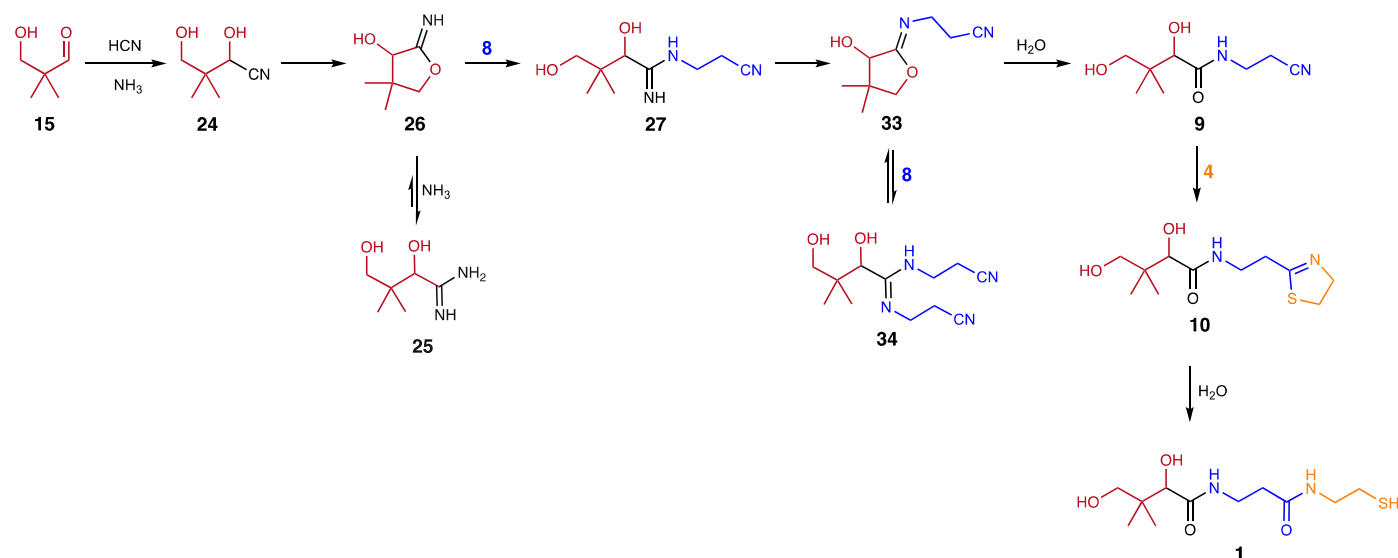
To a stirred solution of hydroxypivaldehyde **15** (100 mM), ammonium chloride (5 equiv.) in phosphate buffer (pH 7.0; 0.5 M) at 20 ºC was added sodium cyanide (1.5 equiv.). The solution was adjusted to pH 9.5 with 1-8 M NaOH and the reaction was monitored by NMR spectroscopy. After 4 hours, pantoamidine **25** (86%) was observed as the major product alongside iminopantolactone **26** (6%), pantoic acid nitrile **24** (1%) and pantolactone **2** (4%). β-Alanine nitrile **8** (5 equiv.) was then added, the solution was adjusted from pH 9.5 to pH 9.0 with 1–4 M HCl. The reaction was monitored periodically by NMR spectroscopy at 20 ºC. Pantothenic acid nitrile **9** (58%) and pantoic acid **5** (24%) were observed as the major pantoyl species (Supplementary Table 20, entry 6). Cysteamine **4** (10 equiv.) (Supplementary Figure 86A) was added and the reaction was then incubated at pH 7.0 and room temperature. After 21 days, **9** underwent conversion to thiazoline **10** and pantetheine **1** (**9** + **4** → **10** + **1**; 86%)* (Supplementary Figure 86C). The formation of **10** was confirmed by adjusting the pH from 7.0 to 5.0 with 4 M HCl and incubating the reaction at 20 ºC for 7 days, leading to **10** hydrolysis to pantetheine **1**. The solution pH was readjusted to 7.0 (Supplementary Figure 86D) and spiked with authentic pantetheine **1**, which led to enhancement [1]H (Supplementary Figure 87) and ^{13}C NMR resonances (Supplementary Figure 88) to confirm the formation of pantetheine **1** (51% based on **15**). Residual pantothenic acid nitrile **9** (5%), pantoic acid **5** (25%), pantolactone **2** (3%) and β-alanylcysteamine **36**[†] were also observed. Pantoylcysteamine **37** was not observed.[‡]

---

* Combined yield of **10** and **1** given. The quantification of pantetheine **1** and **10** could not be carried at this stage due to multiple ^{1}H NMR signal overlap. However, **10** was observed to predominate over **1** at 21 days. Pantetheine **1** is formed by concurrent hydrolysis of **10** – see Supplementary Page S43–55 for the synthesis of **1** from pantothenic acid nitrile **9** and cysteamine **4** at pH 7).

† β-Alanylcysteamine **36** was formed upon hydrolysis of 2-aminoethylthiazoline **32** (**32** was formed by the reaction of cysteamine **4** with residual β-alanine nitrile **8**) but could not be quantified due to multiple ^{1}H NMR signal overlap. The presence of **36** was confirmed by spiking with an authentic sample synthesised by hydrolysis of authentic **32**. See *Preparative synthesis of authentic standards* section (Supplementary Pages S133–S157) for the preparative synthesis of **32** and **36**.

‡ The absence of pantoylcysteamine **37** was confirmed by spiking with an authentic sample synthesised by preparative synthesis. See *Preparative synthesis of authentic standards* section (Supplementary Pages S133–S157) for the preparative synthesis of pantoylcysteamine **37**.

*Data for pantetheine **1** (Supplementary Figure 86D and Supplementary Figure 88)*
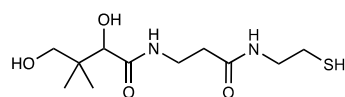

^{1}H NMR (700 MHz, H_2O, noesygppr1d, partial assignment) δ_H 3.91 (s, 1H, (C2)–H), 3.43 (AB, J = 11.3 Hz, 1H, (C4)–H), 2.69 (t, J = 6.6 Hz, 2H, NHCH$_2$C$\underline{H}_2$SH), 2.47 (t, J = 6.5 Hz, 2H, NHCH$_2$C$\underline{H}_2$CO), 0.83 (s, 3H, (C3)–CH$_3$), 0.80 (s, 3H, (C3)–CH$_3'$). ^{13}C NMR (176 MHz, H_2O) δ_C 175.8 (C1), 174.7 (NHCH$_2$CH$_2$$\underline{C}$O), 76.4 (C2), 69.1 (C4), 43.0 (NH$\underline{C}$H$_2$CH$_2$SH), 39.2 (C3), 36.2 (NH$\underline{C}$H$_2$CH$_2$CO), 36.0 (NHCH$_2$$\underline{C}H_2$CO), 23.9 (CH$_2$SH), 21.2 ((C3)–$\underline{C}H_3$), 19.8 ((C3)–$\underline{C}H_3'$).

*Data for pantoic acid **5** (Supplementary Figure 86D and Supplementary Figure 88)*
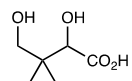
^{1}H NMR (700 MHz, H_2O, noesygppr1d, partial assignment) δ_H 3.75 (s, 1H, (C2)–H), 3.44 (AB, J = 11.0 Hz, 1H, (C4)–H'), 0.85 (s, 3H, CH$_3$), 0.80 (s, 3H, CH$_3'$). ^{13}C NMR (176 MHz, H_2O) δ_C 180.4 (C1), 78.1 (C2), 69.6 (C4), 38.8 (C3), 20.3 (CH$_3$), 19.8 (CH$_3'$).

*Data for cysteamine **4** (Supplementary Figure 86D and Supplementary Figure 88)*

^{1}H NMR (700 MHz, H_2O, noesygppr1d) δ_H 3.14 (t, J = 6.6 Hz, 2H, NH$_2$C$\underline{H}_2$CH$_2$SH), 2.77 (t, J = 6.6 Hz, 2H, NH$_2$CH$_2$C$\underline{H}_2$SH). ^{13}C NMR (176 MHz, H_2O) δ_C 42.9 (NH$_2$$\underline{C}H_2CH_2$SH), 22.2 (NH$_2CH_2$$\underline{C}H_2$SH).

*Data for cysteamine disulfide [**4**]$_2$ (Supplementary Figure 86D and Supplementary Figure 88)*
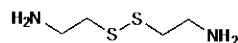
^{1}H NMR (700 MHz, H_2O, noesygppr1d) δ_H 3.20 (t, J = 6.5 Hz, 4H, NH$_2$C$\underline{H}_2$CH$_2$SH), 2.97 (t, J = 6.5 Hz, 4H, NH$_2$CH$_2$C$\underline{H}_2$SH). ^{13}C NMR (176 MHz, H_2O) δ_C 38.7 (NH$_2$$\underline{C}H_2CH_2$SH × 2), 34.2 (NH$_2CH_2$$\underline{C}H_2$SH × 2).

*Data for β-alanylcysteamine **36** (Supplementary Figure 86D and Supplementary Figure 88)*
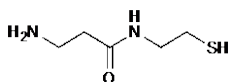
^{1}H NMR (700 MHz, H_2O, noesygppr1d, partial assignment) δ_H 3.20 (t, J = 6.7 Hz, 2H, C$\underline{H}_2$CH$_2$SH), 2.64 (t, J = 6.7 Hz, 2H, NH$_2$CH$_2$C$\underline{H}_2$), 2.61 (t, J = 6.7 Hz, 2H, CH$_2$C$\underline{H}_2$SH). ^{13}C NMR (176 MHz, H_2O) δ_C 173.0 (CO), 42.9 ($\underline{C}$H$_2$CH$_2$SH), 36.7 (NH$\underline{C}$H$_2$CH$_2$), 32.9 (NH$_2$CH$_2$$\underline{C}H_2$), 23.9 (CH$_2$$\underline{C}H_2$SH).

*Data for pantolactone **2** (Supplementary Figure 86D)*
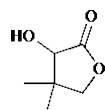
^{1}H NMR (700 MHz, H_2O, noesygppr1d) δ_H 4.33 (s, 1H, (C2)–H), 4.08 (AB, J = 8.9 Hz, 1H, (C4)–H), 4.03 (AB, J = 8.9 Hz, 1H, (C4)–H'), 1.12 (s, 3H, CH$_3$), 0.94 (s, 3H, CH$_3'$).

*Data for pantothenic acid nitrile **9** (Supplementary Figure 86D)*
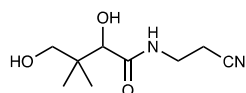
^{1}H NMR (700 MHz, H_2O, noesygppr1d, partial assignment) δ_H 3.96 (s, 1H, (C2)–H), 0.88 (s, 3H, CH$_3$), 0.84 (s, 3H, CH$_3'$).
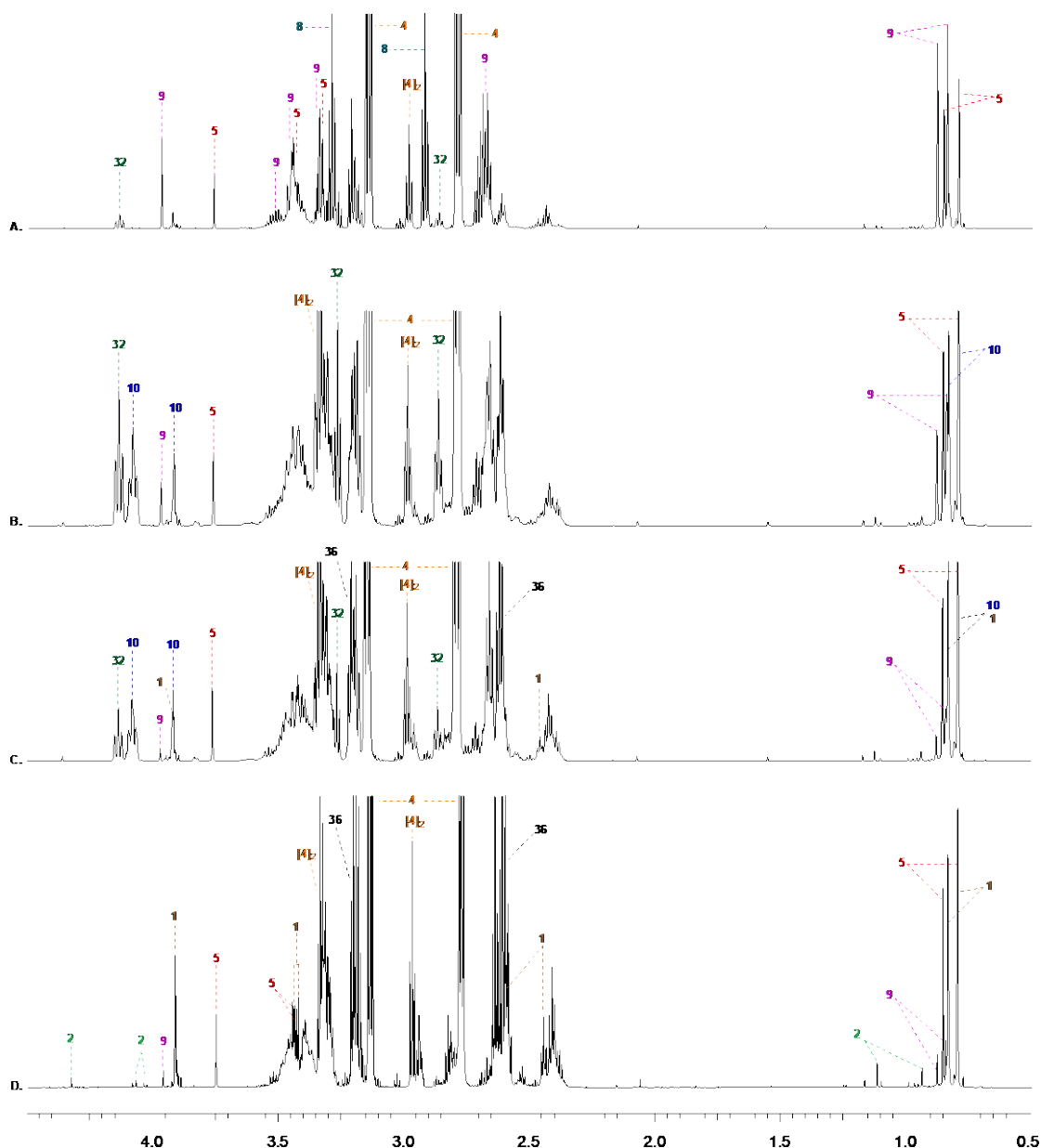
Supplementary Figure 86. ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, 0.5–4.5 ppm) spectra to show the one-pot multicomponent synthesis of pantetheine **1** in the presence of ammonia at 20 ºC in phosphate buffer (pH 7; 0.5 M): (**A.**) 30 mins after the addition of cysteamine **4** (10 equiv.) to a crude reaction mixture containing pantothenic acid nitrile **9** (58%); generated after incubation of hydroxypivaldehyde **15** (100 mM), NaCN (1.5 equiv), and β-alanine nitrile (5 equiv.) in the presence of ammonia (5 equiv.); (**B.**) 10 days after the addition of **4**; (**C.**) 21 days after the addition of **4**; (**D.**) the crude reaction mixture at pH 7 after incubation of the crude reaction mixture shown in spectrum (**C.**) at pH 5 and 20 ºC for 7 days. Pantetheine **1** (51% based on aldehyde **15**) is observed alongside pantothenic acid nitrile **9** (5%), pantoic acid **5** (25%), pantolactone **2** (3%) and β-alanylcysteamine **36***.

---

* β-Alanylcysteamine **36** was formed upon hydrolysis of 2-aminoethylthiazoline **32.** Thiazoline **32** was formed by the reaction of cysteamine **4** with residual β-alanine nitrile **8**. **36** was not quantified due to multiple ^{1}H NMR signal overlap, and no further attempt was made to quantity this byproduct which is not derived from aldehyde **15**. The presence of **36** was however confirmed by spiking with an authentic sample generated by hydrolysis of an authentic and isolated sample of **32**. See *Preparative synthesis of authentic standards* section (Supplementary Pages S133–S157) for the preparative synthesis of **32** and **36**.
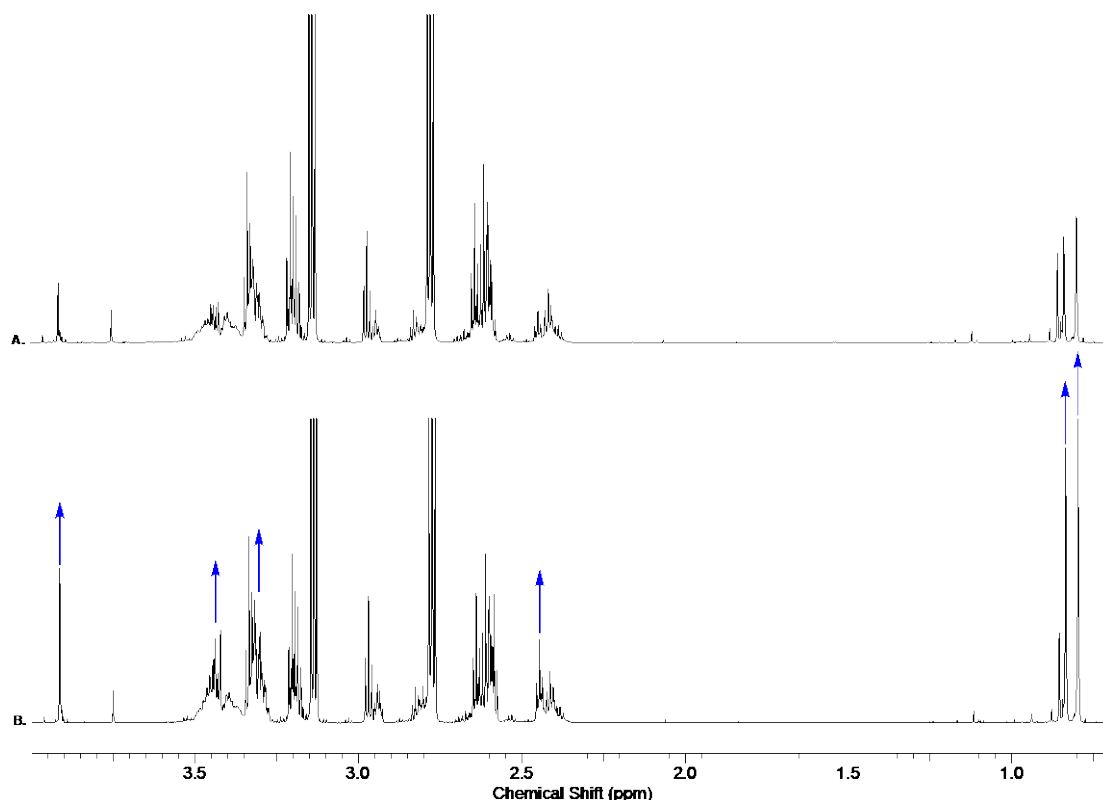
Supplementary Figure 87. ^{1}H NMR (700 MHz, H₂O, pH 7.0, noesygppr1d, 0.7–4.2 ppm) spectra to show the sample spiking of crude reaction mixture at pH 7.0 after the one-pot multicomponent synthesis of pantetheine **1** in the presence of ammonia at 20 ºC in phosphate buffer (0.5 M) after: (**A.**) the reaction of hydroxypivaldehyde **15** (100 mM), NaCN (1.5 equiv), β-alanine nitrile (5 equiv.) and ammonia (5 equiv.) at pH 9.0 and 20 ºC after 11 days, followed by cysteamine **4** (10 equiv.) addition and incubations at pH 7.0 for 21 day, and then incubation at pH 5.0 for 7 days. The crude reaction mixture contains pantetheine **1** (51% based on **15**), is observed alongside pantothenic acid nitrile **9** (5%), pantoic acid **5** (25%), pantolactone **2** (3%) and β-alanylcysteamine **36**[*]. (**B.**) Spiking of the crude reaction mixture shown in spectrum (**A**.) with authentic pantetheine **1**, leading to enhancement of characteristic ^{1}H NMR resonances that confirm the presence and identity of **1** as the major pantoyl species, which was further confirmed by high resolution ^{13}C NMR spectroscopy (see Supplementary Figure 88).

---

[*] β-Alanylcysteamine **36** was formed upon hydrolysis of 2-aminoethylthiazoline **32**. Thiazoline **32** was formed by the reaction of cysteamine **4** with residual β-alanine nitrile **8**. **36** was not quantified due to multiple ^{1}H NMR signal overlap, no further attempt was made to quantity this byproduct which is not derived from aldehyde **15**. The presence of **36** was however confirmed by spiking with an authentic sample generated by hydrolysis of an authentic and isolated sample of **32**. See *Preparative synthesis of authentic standards* section (Supplementary Pages S133–S157) for the preparative synthesis of **32** and **36**.

Supplementary Figure 88. ¹³C NMR (176 MHz, H₂O, pH 7.0) spectra to show all observed ¹³C NMR resonances (19–25, 32–44, 68–80, and 172–181 ppm) in the crude reaction mixture at pH 7.0 after the one-pot multicomponent synthesis of pantetheine **1** in the presence of ammonia at 20 °C in phosphate buffer (0.5 M) after: (**A.**) the reaction of hydroxypivaldehyde **15** (100 mM), NaCN (1.5 equiv), β-alanine nitrile (5 equiv.) and ammonia (5 equiv.) at pH 9.0 and 20 °C after 11 days, followed by cysteamine **4** (10 equiv.) addition and incubations at pH 7.0 for 21 day, and the incubation at pH 5.0 for 7 days. The crude reaction mixture contains pantetheine **1** (51% based on aldehyde **15**), is observed alongside pantothenic acid nitrile **9** (5%), pantoic acid **5** (25%), pantolactone **2** (3%) and β-alanylcysteamine **36**[*]; (**B.**) spiking of the crude reaction mixture shown in spectrum (**A.**) with authentic pantetheine **1**, leading to enhancement of characteristic ¹³C NMR resonances that confirm the presence and identity of **1** as the major pantoyl species, which was also confirmed by high resolution ¹H NMR spectroscopy (see Supplementary Figure 87)

---

[*] β-Alanylcysteamine **36** was formed upon hydrolysis of 2-aminoethylthiazoline **32.** Thiazoline **32** was formed by the reaction of cysteamine **4** with residual β-alanine nitrile **8**. **36** was not quantified due to multiple ¹H NMR signal overlap, no further attempt was made to quantity this byproduct which is not derived from aldehyde **15**. The presence of **36** was however confirmed by spiking with an authentic sample generated by hydrolysis of an authentic and isolated sample of **32**. See *Preparative synthesis of authentic standards* section (Supplementary Pages S133–S157) for the preparative synthesis of **32** and **36.**

5. Preparative syntheses of authentic standards

D-Pantetheine D-1
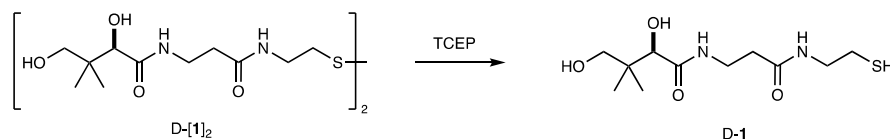
D-Pantethine (D-**[1]**$_2$; 116 mg, 0.21 mmol, 1 equiv.) and (tris(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl; 72 mg, 0.25 mmol, 1.2 equiv.) were dissolved in water (1 mL). The solution pH was adjusted to 7 with 4 M NaOH and the reaction mixture was incubated at 20 °C. After 1 hour, the aqueous reaction mixture was extracted with isopropanol:dichloromethane (70:30; (3 × 5 mL). The combined organic layer was dried with anhydrous sodium sulfate, filtered, and concentrated *in vacuo* to give D-pantetheine D-**1** (85 mg, 73%) as colourless oil. ^{1}H NMR (600 MHz, D$_2$O, noesygppr1d) δ_H 3.94 (s, 1H, (C2)–H), 3.52-3.44 (m, 2H, NH-C$\underline{H}_2$CH$_2$CO), 3.47 (AB, *J* = 11.3 Hz, 1H, (C4)–H), 3.34 (AB, *J* = 11.3 Hz, 1H, (C4)–H′), 3.36-3.30 (m, 2H, NH-C$\underline{H}_2$CH$_2$SH), 2.60 (t, *J* = 6.6 Hz, 2H, NHCH$_2$C$\underline{H}_2$SH), 2.47 (t, *J* = 6.6 Hz, 2H, NHCH$_2$C$\underline{H}_2$CO), 0.88 (s, 3H, (C3)-CH$_3$), 0.84 (s, 3H, (C3)-CH$_3$′). ^{13}C NMR (151 MHz, D$_2$O) δ_C 175.8 (C1), 174.7 (NHCH$_2$CH$_2$$\underline{C}$O), 76.4 (C2), 69.0 (C4), 42.9 (NH$\underline{C}$H$_2$CH$_2$SH), 39.2 (C3), 36.1 (NH$\underline{C}$H$_2$CH$_2$CO), 35.9 (NHCH$_2$$\underline{C}H_2$CO), 23.8 (CH$_2$SH), 21.1 ((C3)-$\underline{C}H_3$), 19.7 ((C3)-$\underline{C}H_3$′). IR (oil, cm^{-1}) 3292, 3086, 2960, 2931, 2876, 1639, 1518, 1418, 1241, 1075, 978 cm^{-1}. HRMS-ESI [M+H]$^+$ calculated for C$_{11}$H$_{23}$N$_2$O$_4$S 279.1373; observed 279.1363.

2-Aminoethylthiazoline dihydrochloride 32·2HCl
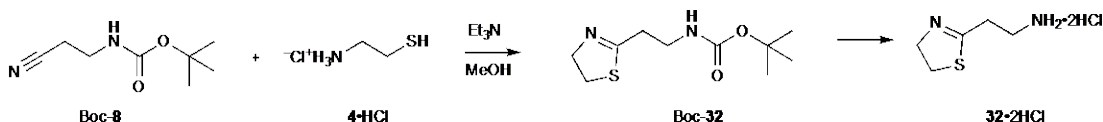
N-tert-Butoxycarbonyl-β-alanine nitrile Boc-**8** (7.70 g, 45.3 mmol), cysteamine hydrochloride **4**·HCl (7.19 g, 63.4 mmol) and triethylamine (8.84 mL, 63.4 mmol) were stirred in methanol (50 mL) at 70 °C for 18 hours. The reaction was cooled, concentrated *in vacuo*, and the residue was stirred vigorously in toluene (200 mL) and water (90 mL). The layers were separated, and the aqueous layer was extracted with toluene (200 mL × 3). The combined organic layers were dried over sodium sulfate, filtered, and concentrated *in vacuo* to give Boc-**32** as a pale white semi-solid that was used without further purification (10.4 g). Boc-**32** (8.2 g, 35.5 mmol) was suspended in 4 M HCl in 1,4-dioxane (140 mL) and stirred vigorously for 3 hours. The persisting suspension was cooled in an ice bath for 10 min, and the reaction was filtered. The precipitate was washed with diethyl ether and dried *in vacuo* to give 2-aminoethylthiazoline dihydrochloride **32**·2HCl (7.3 g, >95%) as a white powder. NMR spectral data was obtained by dissolving a solid sample (10 mg) in phosphate buffer (pH 7.0; 1 M). ^{1}H NMR (600 MHz, phosphate buffer (pH 7.0; 1 M), noesygppr1d) δ 4.12 (t, *J* = 8.5 Hz, 2H, (C4)−H$_2$), 3.32 (t, *J* = 8.5 Hz, 2H, (C5)−H$_2$), 3.25 (t, *J* = 6.9 Hz, 2H, CH$_2$C$\underline{H}_2$NH$_3$$^+$), 2.83 (t, *J* = 6.9 Hz, 2H, C$\underline{H}_2$CH$_2$NH$_3$$^+$). ^{13}C NMR (151 MHz, phosphate buffer (1 M, pH 7)) δ 171.4 (C2), 63.7 (C4), 37.7 (CH$_2$$\underline{C}H_2NH_3$$^+$), 33.9 (C5), 31.4 ($\underline{C}H_2CH_2NH_3$$^+$).

IR (solid, cm^{-1}) 3382, 3357, 3009, 2858, 2814, 2719, 2662, 2575, 1616, 1516, 1412, 1186, 1027, 930, 874, 782, 668. HRMS-ESI [M+H]$^+$ calculated for $C_5H_{11}N_2S$ 131.0648; observed 131.0645.

β-Alanylcysteamine 36

2-Aminoethylthiazoline dihydrochloride **32**·2HCl (813 mg, 4 mmol, 200 mM) was dissolved in water (20 mL) and incubated at 20 ºC for 6 hours. The solution was then adjusted from pH 2.0 to pH 7.0 with 8 M NaOH and NMR spectra were acquired. The conversion of **32** to β-alanylcysteamine **36** (90%) was observed. An aliquot of this crude solution of **32** (1 mL) was then adjusted to pH 10.0 with 2 M NaOH and further NMR spectra were periodically acquired. Partial oxidation of **36** to β-alanylcysteamine disulfide [**36**]$_2$ (5%) was observed.

Data for β-alanylcysteamine 36

^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, pH 7.0) δ$_H$ 3.41 (q, *J* = 6.5 Hz, 2H, $^+$NH$_3$C$\underline{H}_2$CH$_2$), 3.29 (t, *J* = 6.7 Hz, 2H, C$\underline{H}_2$CH$_2$SH), 2.72 (t, *J* = 6.5 Hz, 2H, $^+$NH$_3$CH$_2$C$\underline{H}_2$), 2.69 (t, *J* = 6.7 Hz, 2H, CH$_2$C$\underline{H}_2$SH). ^{13}C NMR (176 MHz, H$_2$O, pH 7.0) δ$_C$ 173.0 (CO), 42.9 (C̲H$_2$CH$_2$SH), 36.7 ($^+$NH$_3$C̲H$_2$CH$_2$), 32.8 ($^+$NH$_3$CH$_2$C̲H$_2$), 24.0 (CH$_2$C̲H$_2$SH). ^{1}H NMR (700 MHz, H$_2$O, noesygppr1d, pH 10.0) δ$_H$ 3.30 (t, *J* = 7.2 Hz, 2H, C$\underline{H}_2$CH$_2$SH), 3.03 (t, *J* = 6.7 Hz, 2H, NH$_2$C$\underline{H}_2$CH$_2$), 2.60 (t, *J* = 7.2 Hz, 2H, CH$_2$C$\underline{H}_2$SH), 2.50 (t, *J* = 6.7 Hz, 2H, NH$_2$CH$_2$C$\underline{H}_2$). ^{13}C NMR (176 MHz, H$_2$O, pH 10.0, partial assignment) δ$_C$ 174.4 (CO), 44.8 (C̲H$_2$CH$_2$SH), 37.8 (NH$_2$C̲H$_2$CH$_2$), 37.2 (NH$_2$CH$_2$C̲H$_2$), 24.3 (CH$_2$C̲H$_2$SH).

Pantothenic acid nitrile 9

Pantolactone **2** (2.80 g, 21.5 mmol) and β-alanine nitrile **8** (1.00 mL, 1.00 g, 14.27 mmol) were heated neat at 50 °C for 10 hours. The reaction was cooled to 20 ºC, and the residue was purified by flash column chromatography (SiO$_2$; eluting with a gradient of MeOH/CH$_2$Cl$_2$ 0:100→0:10) to give pantothenic acid nitrile **9** (2.99 g, 14.95 mmol, >95%) as a colourless oil that then solidified to a white crystalline solid. Nitrile **9** was recrystallised from EtOAc for single crystal X-ray crystallographic analysis (See Supplementary Page S129 for crystallographic data). ^{1}H NMR (700 MHz, D$_2$O, noesygpprd1d) δ$_H$ 4.04 (s, 1H, (C2)–H), 3.54 (AB, *J* = 11.3 Hz, 1H, (C4)–H), 3.43 (m, 2H, C$\underline{H}_2$CH$_2$CN), 3.43 (AB, *J* = 11.3 Hz, 1H, (C4)–H′), 2.76 (t, *J* = 6.4 Hz, 2H, CH$_2$C$\underline{H}_2$CN), 0.97 (s, 3H, (C3)–CH$_3$), 0.94 (s, 3H, (C3)–CH$_3$′). ^{13}C NMR (176 MHz, D$_2$O) δ$_C$ 170.4 (C1), 120.6 (CH$_2$CH$_2$C̲N), 76.6 (C2), 69.1 (C4), 39.4 (C3), 35.4 (C̲H$_2$CH$_2$CN), 21.3 (CH$_3$), 19.9 (CH$_3$′), 18.5 (CH$_2$C̲H$_2$CN). IR (solid, cm^{-1}) 3390, 3292, 3211,3203, 2991, 2973, 2930, 2915, 2889, 2251, 1646, 1519, 1317, 1246, 1082, 1052, 675, 628, 572. HRMS-ESI [M+H]$^+$ calculated for $C_9H_{17}N_2O_3$ 201.1234; observed 201.1234. m.p. 79-81 °C (EtOAc), lit. m.p 82-84 °C (EtOAc)[23].

Pantoylglycine nitrile **14**

Pantolactone **2** (2.11 g, 16.21 mmol) and glycine nitrile hydrochloride **7G**·HCl (1.02 g, 11.02 mmol) were dissolved in triethylamine (2.2 mL) and incubated at 50 °C for 5 hours. The solution was then cooled to room temperature and concentrated *in vacuo*. The residue was purified by flash column chromatography (SiO$_2$; eluting with a gradient of MeOH/CH$_2$Cl$_2$ 0:100→10:90) to give pantoylglycine nitrile **14** (1.19 g, 6.40 mmol, 59%) as a yellow oil. ^{1}H NMR (600 MHz, D$_2$O) δ_H 4.21 (AB, J = 17.6 Hz, 1H, NHCH$_2$), 4.17 (AB, J = 17.6 Hz, 1H, NHCH$_2$′) 4.02 (s, 1H, C2), 3.47 (AB, J = 11.3 Hz, 1H, (C4)–H), 3.37 (AB, J = 11.3 Hz, 1H, (C4)–H′), 0.90 (s, 3H, (C3)–CH$_3$), 0.87 (s, 3H, (C3)–CH$_3$′). ^{13}C NMR (151 MHz, D$_2$O) δ_C 176.7 (C1), 117.5 (CN), 76.1 (C2), 68.7 (C4), 39.6 (C3), 27.7 (CH$_2$CN), 21.1 (CH$_3$), 19.7 (CH$_3$′). IR (oil, cm^{-1}) 3203, 2966, 2910, 2716, 2243, 1657, 1583, 1513, 1417, 1328, 1128, 1077. HRMS-ESI [M+H]$^+$: calculated for C$_8$H$_{15}$N$_2$O$_3$ 187.1077 observed 187.1075.

γ-Amidonitrile **12**

Pantolactone **2** (352 mg, 2.70 mmol), γ-aminobutyronitrile hydrochloride **11**·HCl (216 mg, 1.79 mmol) and triethylamine (350 µL) were dissolved in methanol (2 mL) and incubated at 60 °C for 16 hours. The solution was then cooled to room temperature and concentrated *in vacuo*. The residue was purified by flash column chromatography (SiO$_2$; eluting with a gradient of MeOH/EtOAc 0:100→20:90) to give γ-amidonitrile **12** (232 mg, 1.08 mmol, 65%) as a colourless oil. ^{1}H NMR (600 MHz, D$_2$O) δ_H 4.02 (s, 1H, (C2)–H), 3.54 (AB, J = 11.2 Hz, 1H, (C4)–H), 3.42 (AB, J = 11.2 Hz, 1H, (C4)–H′), 3.38 (t, J = 6.7 Hz, 2H, NHCH$_2$), 2.54 (t, J = 7.2 Hz, 2H, CH$_2$CH$_2$CN), 1.92 (tt, J = 6.7, 7.2 Hz, 2H, CH$_2$CH$_2$CN), 0.96 (s, 3H, (C3)–CH$_3$), 0.92 (s, 3H, (C3)–CH$_3$′). ^{13}C NMR (151 MHz, D$_2$O) δ_C 176.2 (C1), 122.0 (CH$_2$CH$_2$CN), 76.6 (C2), 69.1 (C4), 39.4 (C3), 38.3 (NHCH$_2$CH$_2$), 25.2 (CH$_2$CH$_2$CN), 21.3 (CH$_3$), 19.9 (CH$_3$′), 14.9 (CH$_2$CH$_2$CN). IR (oil, cm^{-1}) 3308, 2971, 2939, 2877, 2601, 1645, 1528, 1422, 1073, 1038. HRMS-ESI [M+H]$^+$ calculated for C$_{10}$H$_{19}$N$_2$O$_3$ 215.1390; observed 215.1389.

Thiazoline 10

Pantothenic acid nitrile **9** (951 mg, 4.75 mmol), cysteamine hydrochloride **4**·HCl (776 mg, 6.83 mmol) and triethylamine (930 µL, 675 mg, 6.67 mmol) were dissolved in methanol (5 mL) and the resulting solution was refluxed for 16 hours. The solution was then cooled to room temperature and concentrated *in vacuo.* The residue was purified by flash column chromatography (SiO$_2$; eluting with a gradient of MeOH/EtOAc 0:100→100:0) to give thiazoline **10** (1.00 g, 3.86 mmol, 81%) as a colourless oil. ^{1}H NMR (600 MHz, D$_2$O, noesygppr1d) δ_H 4.18 (t, J = 8.5 Hz, 2H, (C4′)–H$_2$), 3.93 (s, 1H, C2), 3.55-3.45 (m, 2H, NHC$\underline{H}_2$CH$_2$), 3.46 (d, J = 11.3 Hz, 1H, (C4)–H), 3.33 (d, J = 11.3 Hz, 1H, (C4)–H′), 3.30 (t, J = 8.5 Hz, 2H, (C5′)–H$_2$), 2.73 (t, J = 6.6 Hz, 2H, NHCH$_2$C$\underline{H}_2$), 0.87 (s, 3H, (C3)–CH$_3$), 0.83 (s, 3H, (C3)–CH$_3$′). ^{13}C NMR (151 MHz, D$_2$O) δ_C 175.8 (C1), 173.7 (C2′), 76.4 (C2), 69.0 (C4), 63.6 (C4′), 39.2 (C3), 37.5 (NHC$\underline{H}_2$CH$_2$), 33.9 (1C)*, 33.8 (1C)*, 21.1 (CH$_3$). 19.7 (CH$_3$′). IR (oil, cm^{-1}) 3366, 2972, 2954, 2930, 2873, 1742, 1716, 1524, 1216, 1161, 1046, 745, 702. HRMS-ESI [M+H]$^+$ calculated for C$_{11}$H$_{21}$N$_2$O$_3$S 261.1267; observed 261.1263.

γ-Aminobutyronitrile 11

4-Azidobutyronitrile

Sodium azide (480 mg, 7.38 mmol) and 4-bromobutanenitrile (610 µL, 908 mg, 6.14 mmol) were suspended in DMF (10 mL) under an argon atmosphere and the mixture was stirred at room temperature for 18 hours. The reaction mixture was diluted with H$_2$O (25 mL) and extracted with Et$_2$O (3 x 20 mL). The organic extracts were combined and then washed with brine (20 mL). The organics were dried over MgSO$_4$, filtered and concentrated *in vacuo* to give 4-azidobutyronitrile (666 mg, 6.05 mmol, 99%) as a pale-yellow oil, which was used without further purification. ^{1}H NMR (600 MHz, CDCl$_3$) δ_H 3.50 (t, J = 6.3 Hz, 2H, N$_3$C$\underline{H}_2$), 2.48 (t, J = 7.0 Hz, 2H, C$\underline{H}_2$CN), 1.92 (tt, J = 6.3, 7.0 Hz, 2H, CH$_2$C$\underline{H}_2$CH$_2$CN). ^{13}C NMR (151 MHz, CDCl$_3$) δ_C 118.8 ($\underline{C}$N), 49.6 (N$_3$$\underline{C}H_2$), 25.1 ($\underline{C}H_2CH_2$CN), 14.7 ($\underline{C}H_2$CN). IR (oil, cm^{-1}) 2936, 2874, 2249, 2096, 1673, 1349, 1287, 1091. HRMS-ESI [M+H]$^+$ calculated for C$_4$H$_7$N$_4$ 111.0665; observed 111.0664.
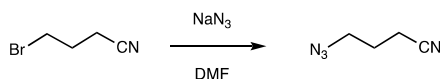
---

* NHCH$_2$$\underline{C}H_2$ and C5′ ^{13}C resonances for **10** were not unambiguously assigned due to signal overlap.

γ-Aminobutyronitrile hydrochloride 11·HCl[24]
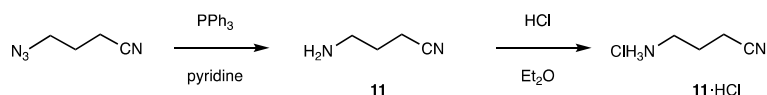
4-Azidobutyronitrile (598 mg, 5.43 mmol) was dissolved in pyridine (11 mL) under an argon atmosphere. Then triphenylphosphine (1.64 g, 6.26 mmol) was added and the mixture was stirred at room temperature for 2 hours. The reaction mixture was then diluted with H_2O (10 mL) and stirred for 1 hour. The mixture was extracted with EtOAc (3 x 20 mL) and the aqueous layer was concentrated *in vacuo*. The residue was then purified by flash column chromatography (SiO_2; eluting with a gradient of $MeOH/CH_2Cl_2/NEt_3$ 0:99:1→10:89:1) to give γ-aminobutyronitrile **11** (297 mg, 3.53 mmol, 65%). γ-Aminobutyronitrile **11** (297 mg, 3.53 mmol) was then dissolved in Et_2O (1 mL). The solution was stirred and HCl (2 M in Et_2O, 5 mL) was added dropwise. A white precipitate was observed to form during the addition of HCl. Once the addition was complete, the heterogeneous mixture was stirred at room temperature for 30 mins. The sample was then centrifuged, and the precipitate triturated with Et_2O. The solids were dried *in vacuo* to give γ-aminobutyronitrile hydrochloride **11**·HCl (308 mg, 2.55 mmol, 72%) as a white solid, which was **11**·HCl was recrystallised from EtOH.

Data for γ-aminobutyronitrile **11**

^{1}H NMR (600 MHz, D_2O, noesygppr1d) δ_H 2.96 (t, *J* = 7.8 Hz, 2H, NC$\underline{H}_2$), 2.52 (t, *J* = 7.2 Hz, 2H, C$\underline{H}_2$CN), 1.91 (tt, *J* = 7.2, 7.8 Hz, 2H, C$\underline{H}_2$CH_2CN).

Data for γ-aminobutyronitrile hydrochloride **11**·HCl

^{1}H NMR (700 MHz, D_2O, noesygppr1d) δ_H 3.16 (t, *J* = 7.8 Hz, 2H, NC$\underline{H}_2$), 2.67 (t, *J* = 7.1 Hz, 2H, C$\underline{H}_2$CN), 2.09 (tt, *J* = 7.2, 7.8 Hz, 2H, CH_2C$\underline{H}_2$CH_2CN). ^{13}C NMR (176 MHz, D_2O) δ_C 120.8 ($\underline{C}$N), 38.7 (N$\underline{C}$H_2), 23.3 ($\underline{C}$H_2CH_2CN), 14.6 ($\underline{C}$H_2CN). IR (solid, cm^{-1}) 3202, 2961, 2909, 2715, 2243, 1583, 1513, 1417, 1328, 1128, 1050, 976, 954, 854. HRMS-ESI [M+H]$^+$ calculated for $C_4H_9N_2$ 85.0760; observed 85.0764. m.p. 140-141 °C (EtOH), lit. m.p 142-144 °C (EtOH)[25].

Pantoamide 5-NH₂
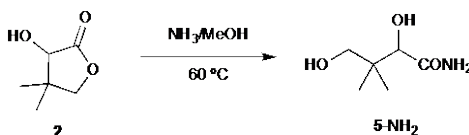
Pantolactone **2** (500 mg, 3.84 mmol) was heated in methanolic ammonia (7 N, 10 mL) for 24 hours. The reaction was concentrated *in vacuo* to give pantoamide **5**-NH_2 (94%) and residual pantolactone **2** (6%) as an off-white solid that was used without further purification. R_f 0.12 (100 % EtOAc). ^{1}H NMR (600 MHz, D_2O) δ_H 3.92 (s, 1H, (C2)−H), 3.45 (AB, *J* = 11.2 Hz, 1H, (C4)−H, AB), 3.32 (AB, *J* = 11.2 Hz, 1H, (C4)−H′), 0.87 (s, 3H, (C3)−CH_3), 0.84 (s, 3H, (C3)−CH_3′). ^{13}C NMR (151 MHz, D_2O) δ_C 179.1 (C1), 76.2 (C2), 69.0 (C4), 39.0 (C3), 21.1 ((C3)−CH_3), 19.7 ((C3)−CH_3′). HRMS-ESI [M+H]$^+$ calculated for $[C_6H_{14}NO_3]^+$ 148.0968; measured 148.0970.

Sodium pantoate 5-Na⁺
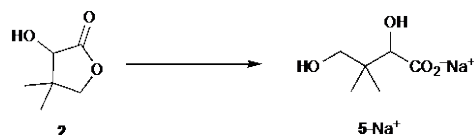
To a solution of pantolactone **2** (65 mg, 0.5 mmol) in water (5 mL) was added sodium hydroxide (4 M; 0.25 mL, 1 mmol). The solution was adjusted from pH 12.6 to pH 9.8 with 4 M HCl, and analysed by NMR spectroscopy. Quantitative conversion of pantolactone **2** to sodium pantoate (**5**-Na⁺) was observed. ^{1}H NMR (600 MHz, H_2O, pH 9.8) δ_H 3.73 (s, 1H, (C2)−H), 3.43 (AB, J = 11.1 Hz, 1H, (C4)−H), 3.31 (AB, J = 11.1 Hz, 1H, (C4)−H′), 0.84 (s, 3H, (C3)−CH$_3$), 0.78 (s, 3H, (C3)−CH$_3$′). ^{13}C NMR (151 MHz, H_2O, pH 9.8) δ_C 179.4 (C1), 77.1 (C2), 68.6 (C4), 37.7 (C3), 20.1 ((C3)−CH$_3$), 19.2 ((C3)−CH$_3$′). HRMS-ESI [M+H]⁺ calculated for [$C_6H_{13}O_4$]⁺ 148.0808; measured 148.0810.

Pantoylcysteamine 37
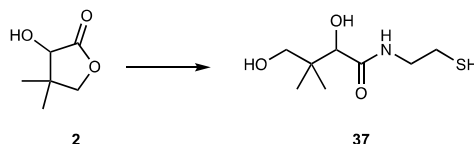
Pantolactone **2** (860 mg, 6.61 mmol) and cysteamine hydrochloride **4** (506 mg, 4.45 mmol) were dissolved in triethylamine (1 mL) and heated at 60 °C in for 16 h. The reaction was cooled to room temperature and concentrated *in vacuo*. The crude residue was purified by flash column chromatography (SiO$_2$; eluting with a gradient of MeOH/CH$_2$Cl$_2$ 0:100→10:90) to give pantoylcysteamine **37** (496 mg, 2.39 mmol, 54%) as a colourless oil. ^{1}H NMR (600 MHz, D_2O) δ_H 3.95 (s, 1H, C2), 3.46 (AB, J = 11.3 Hz, 1H, (C4)–H, 3.40-3.32 (m, 2H, NHC$\underline{H}_2$), 3.35 (AB, J = 11.3 Hz, 1H, (C4)–H′), 2.62 (t, J = 6.6 Hz, 1H, C$\underline{H}_2$SH), 0.89 (s, 3H, (C3)–CH$_3$), 0.87 (s, 3H, (C3)–CH$_3$′). ^{13}C NMR (151 MHz, D_2O) δ_C 176.0 (C1), 76.4 (C2), 69.0 (C4), 42.4 (NH$\underline{C}$H$_2$), 39.3 (C3), 23.8 ($\underline{C}$H$_2$SH), 21.2 ((C3)–CH$_3$), 19.8 ((C3)–CH$_3$′). IR (cm⁻¹, oil) 3339, 2960, 2933, 2874, 1642, 1527, 1472, 1441, 1296, 1249, 1073, 1035, 607. HRMS-ESI [M+H]⁺ calculated for $C_8H_{18}O_3NS$ 208.1002; obs. 208.1000.

Single crystal X-ray diffraction data

The diffraction data for *rac*-pantothenic acid nitrile **9** was collected on a four-circle *Agilent SuperNova* (Dual Source) single crystal X-ray diffractometer using a micro-focus CuK$_\alpha$ X-ray beam (λ = 1.54184 Å) and a *HyPix-Arc 100* curved hybrid photon counting X-ray detector'. The sample temperature was controlled with an *Oxford Instruments* cryojet. All data were processed using the *CrysAlis^{Pro}* programme package from *Rigaku Oxford Diffraction*[26]. The crystal structures were solved with the *SHELXT* programme[27], used within the *Olex2* software suite[28], and refined by least squares on the basis of F^2 with the *SHELXL*[29] programme using the *ShelXle* graphical user interface[30]. All non-hydrogen atoms were refined anisotropically by the full-matrix least-squares method. Hydrogen atoms associated with carbon atoms were refined isotropically [U_{iso}(H) = 1.2U_{eq}(C)] in geometrically constrained positions. The substitutional disorder involving two enantiomers of pantothenic acid nitrile **9** occupying the same position was modelled using SADI restraint and the EADP and EXYZ constraints in *SHELX*L[29].

empirical formula	$C_9H_{16}N_2O_3$
M_r / g mol^{-1}	200.24
crystal system	monoclinic
space group	$P2_1/c$
a / Å	10.97392(6)
b / Å	10.43970(6)
c / Å	17.43660(10)
α / °	90
β / °	92.8045(5)
γ / °	90
V / Å^3	1995.221(19)
Z	8
ρ_{calc} / g cm^{-3}	1.333
T / K	150.0(1)
μ / mm^{-1}	0.832
$F(000)$	864
crystal size / mm^3	0.34 × 0.26 × 0.13
radiation	CuK$_\alpha$ (λ = 1.54184 Å)
2θ range for data collection / °	8.066–133.172
index ranges	$-13 \leq h \leq 12$ $-12 \leq k \leq 12$ $-20 \leq l \leq 20$
number of collected reflections	43515
unique reflections	3512
number of unique reflections	3253 [$I > 2\sigma(I)$]
R_{int}	0.0355
$R(F)$, $F > 2\sigma$(F)	0.0335
$wR(F^2)$, $F > 2\sigma$(F)	0.0355
$R(F)$, all data	0.0949
$wR(F^2)$, all data	0.0967
Δ_r (max., min.) e Å^{-3}	0.284/−0.278
CCDC deposition number	2216395

Supplementary Table 21. Crystallographic and refinement parameters of *rac*-pantothenic acid nitrile **9**.
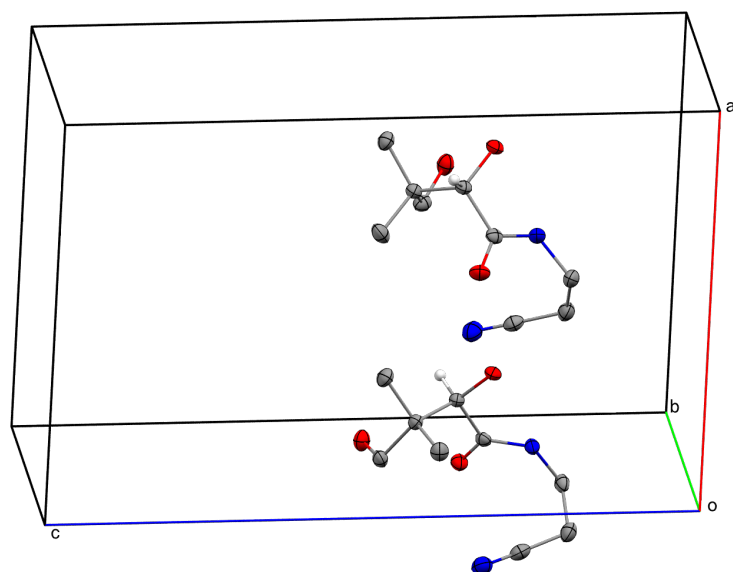
Supplementary Figure 93. The asymmetric unit of *rac*-pantothenic acid nitrile **9**. The thermal ellipsoids are drawn at the 50% probability level. All atoms belonging to the minor occupation site of the substitutional disorder, and hydrogen atoms affiliated with non-chiral carbon atoms, are omitted for clarity. Colour scheme: carbon – dark grey, nitrogen – blue, oxygen – red, hydrogen – white.

Supplementary References

1.   Miller, S. L. & Schlesinger, G. Prebiotic syntheses of vitamin coenzymes: II. Pantoic acid, pantothenic acid, and the composition of coenzyme A. *J. Mol. Evol.* **36**, 308–314 (1993).

2.   Keefe, A. D., Newton, G. L. & Miller, S. L. A possible prebiotic synthesis of pantetheine, a precursor to coenzyme A. *Nature* **373**, 683–685 (1995).

3.   Williams, R. J., Mitchell, H. K., Weinstock, H. H., Jr. & Snell, E. E. Pantothenic acid. VII. Partial and total synthesis studies. *J. Am. Chem. Soc.* **62**, (7), 1784–1785 62 (1940).

4.   Ferris, J. P. Life at the margins. *Nature* **373**, 659–659, (1995).

5.   Fridman, Y. D., Sarbaev, D. & Kebets, N. Salts of pantoic acid. *Russian J. Inorg. Chem.* **30**, 16–19 (1985).

6.   Canavelli, P., Islam, S. & Powner, M. W. Peptide ligation by chemoselective aminonitrile coupling in water. *Nature* **571**, 546–549, (2019).

7.   Thoma, B. & Powner, M. W. Selective synthesis of lysine peptides and the prebiotically plausible synthesis of catalytically active diaminopropionic acid peptide nitriles in Water. *J. Am. Chem. Soc.* **145**, 3121–3130 (2023).

8.   Foden, C. S. *et al.* Prebiotic synthesis of cysteine peptides that catalyze peptide ligation in neutral water. *Science* **370**, 865–869 (2020).

9.   Singh, J. et al. Prebiotic catalytic peptide ligation yields proteinogenic peptides by intramolecular amide catalyzed hydrolysis facilitating regioselective lysine ligation in neutral water. *J. Am. Chem. Soc.* **144**, 10151–10155 (2022).

10.   Liu, R. & Orgel, L. E. Polymerization of β-amino acids in aqueous solution. *Orig. Life Evol. Biosph.* 28, 47–60 (1998).

11.   Pascal, R., Boiteau, L. & Commeyras, A. in *Prebiotic chemistry* (Ed. Walde, P.) 69–122 (Springer Berlin Heidelberg, 2005).

12.   Danger, G., Plasson, R. & Pascal, R. Pathways for the formation and evolution of peptides in prebiotic environments. *Chem. Soc. Rev.* **41**, 5416–5429, (2012).

13.   Danger, G. *et al.* 5(4H)-Oxazolones as intermediates in the carbodiimide- and cyanamide-promoted peptide activations in aqueous solution. *Angew. Chem. Int. Ed.* **52**, 611–614 (2013).

14.   Ajram, G., Rossi, J.-C., Boiteau, L. & Pascal, R. A reassessment of prebiotically relevant chemical agents for the activation of α-amino acids and peptides. *J. Syst. Chem.* **7**, 19–28 (2019).

15.   Danger, G., d'Hendecourt, L. L. S. & Pascal, R. On the conditions for mimicking natural selection in chemical systems. *Nat. Rev. Chem.* **4**, 102–109 (2020).

16.   Whicher, A., Camprubi, E., Pinna, S., Herschy, B. & Lane, N. Acetyl phosphate as a primordial energy currency at the origin of life. *Orig. Life Evol. Biosph.* **48**, 159–179 (2018).

17.   Griesser, H., Bechthold, M., Tremmel, P., Kervio, E. & Richert, C. Amino acid-specific, ribonucleotide-promoted peptide formation in the absence of enzymes. *Angew. Chem. Int. Ed.* **56**, 1224–1228 (2017).

18.   Schimpl, A., Lemmon, R. M. & Calvin, M. Cyanamide formation under primitive earth conditions. *Science* 147, 149–150 (1965).

19.   Islam, S., Bučar, D.-K. & Powner, M. W. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nat. Chem.* **9**, 584–589 (2017).

20.   Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).

21.   Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301–307 (2015).

22.   Cieślak, J., Ausín, C., Grajkowski, A. & Beaucage, S. L. The 2-cyano-2,2-dimethylethanimine-N-oxymethyl group for the 2′-hydroxyl protection of ribonucleosides in the solid-phase synthesis of RNA sequences. *Chem. Eur. J.* **19**, 4623–4632 (2013).

23.   Shimizu, M., Ohta, G., Nagase, O., Okada, S. & Hosokawa, Y. Investigations on pantothenic acid and its related compounds. I. Chemical studies. (1). A novel synthesis of pantethine. *Chem. Pharm. Bull.* **13**, 180–188 (1965).

24.   Capon, P. K., Avery, T. D., Purdey, M. S. & Abell, A. D. An improved synthesis of 4-aminobutanenitrile from 4-azidobutanenitrile and comments on room temperature stability. *Synth. Commun.* **51**, 428–436 (2021).

25.   McKay, A. F. *et al.* Bacteriostats. II.1 The chemical and bacteriostatic properties of isothiocyanates and their derivatives. *J. Am. Chem. Soc.* **81**, 4328–4335 (1959).

26.   *CrysAllisPro*, *Agilent Technologies Inc.*, (2022).

27.   Sheldrick, G. SHELXT - Integrated space-group and crystal-structure determination. *Acta Crystallogr. A* **71**, 3–8 (2015).

28.   Dolomanov, O. V., Bourhis, L. J., Gildea, R. J., Howard, J. A. K. & Puschmann, H. OLEX2: a complete structure solution, refinement and analysis program. *J. Appl. Crystallogr.* **42**, 339–341 (2009).

29.   Sheldrick, G. Crystal structure refinement with SHELXL. *Acta Crystallogr. C* **71**, 3–8 (2015).

30.   Hubschle, C. B., Sheldrick, G. M. & Dittrich, B. ShelXle: a Qt graphical user interface for SHELXL. *J. Appl. Crystallogr.* **44**, 1281–1284 (2011).

1   Miller, S. L. & Schlesinger, G. Prebiotic syntheses of vitamin coenzymes: II. Pantoic acid, pantothenic acid, and the composition of coenzyme A. *Journal of Molecular Evolution* **36**, 308-314, doi:10.1007/BF00182178 (1993).

2   Keefe, A. D., Newton, G. L. & Miller, S. L. A possible prebiotic synthesis of pantetheine, a precursor to coenzyme A. *Nature* **373**, 683, doi:10.1038/373683a0 (1995).

3   Williams, R. J., Mitchell, H. K., Weinstock, H. H., Jr. & Snell, E. E. Pantothenic Acid. VII. Partial and Total Synthesis Studies. *Journal of the American Chemical Society* **62**, 1784-1785, doi:10.1021/ja01864a038 (1940).

4   Ferris, J. P. Life at the margins. *Nature* **373**, 659-659, doi:10.1038/373659a0 (1995).

5   Fridman, Y. D., Sarbaev, D. & Kebets, N. Salts of pantoic acid. *Rusian J Inorg Chem* **30**, 16-19 (1985).

6   Canavelli, P., Islam, S. & Powner, M. W. Peptide ligation by chemoselective aminonitrile coupling in water. *Nature* **571**, 546-549, doi:10.1038/s41586-019-1371-4 (2019).

7   Thoma, B. & Powner, M. W. Selective Synthesis of Lysine Peptides and the Prebiotically Plausible Synthesis of Catalytically Active Diaminopropionic Acid Peptide Nitriles in Water. *Journal of the American Chemical Society* **145**, 3121-3130, doi:10.1021/jacs.2c12497 (2023).

8   Foden, C. S. *et al.* Prebiotic synthesis of cysteine peptides that catalyze peptide ligation in neutral water. *Science* **370**, 865-+, doi:10.1126/science.abd5680 (2020).

9   Singh, J. *et al.* Prebiotic Catalytic Peptide Ligation Yields Proteinogenic Peptides by Intramolecular Amide Catalyzed Hydrolysis Facilitating Regioselective Lysine Ligation in Neutral Water. *Journal of the American Chemical Society* **144**, 10151-10155, doi:10.1021/jacs.2c03486 (2022).

10   Liu, R. & Orgel, L. E. Polymerization of β-amino Acids in Aqueous Solution. *Orig. Life Evol. Biosph.* **28**, 47-60, doi:10.1023/A:1006580918298 (1998).

11   Pascal, R., Boiteau, L. & Commeyras, A. in *Prebiotic Chemistry*   (ed Peter Walde) 69-122 (Springer Berlin Heidelberg, 2005).

12   Danger, G., Plasson, R. & Pascal, R. Pathways for the formation and evolution of peptides in prebiotic environments. *Chem Soc Rev* **41**, 5416-5429, doi:10.1039/C2CS35064E (2012).

13   Danger, G. *et al.* 5(4H)-Oxazolones as Intermediates in the Carbodiimide- and Cyanamide-Promoted Peptide Activations in Aqueous Solution. *Angewandte Chemie International Edition* **52**, 611-614, doi:10.1002/anie.201207730 (2013).

14   Ajram, G., Rossi, J.-C., Boiteau, L. & Pascal, R. A reassessment of prebiotically relevant chemical agents for the activation of α-amino acids and peptides. *Journal of Systems Chemistry* **7**, 19-28 (2019).

15   Danger, G., d'Hendecourt, L. L. S. & Pascal, R. On the conditions for mimicking natural selection in chemical systems. *Nature Reviews Chemistry* **4**, 102-109, doi:10.1038/s41570-019-0155-6 (2020).

16   Whicher, A., Camprubi, E., Pinna, S., Herschy, B. & Lane, N. Acetyl Phosphate as a Primordial Energy Currency at the Origin of Life. *Origins of Life and Evolution of Biospheres* **48**, 159-179, doi:10.1007/s11084-018-9555-8 (2018).

17   Griesser, H., Bechthold, M., Tremmel, P., Kervio, E. & Richert, C. Amino Acid-Specific, Ribonucleotide-Promoted Peptide Formation in the Absence of Enzymes. *Angewandte Chemie International Edition* **56**, 1224-1228, doi:10.1002/anie.201610651 (2017).

18   Schimpl, A., Lemmon, R. M. & Calvin, M. Cyanamide Formation under Primitive Earth Conditions. *Science* **147**, 149-150, doi:doi:10.1126/science.147.3654.149 (1965).

19   Islam, S., Bučar, D.-K. & Powner, M. W. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nature Chemistry* **9**, 584, doi:10.1038/nchem.2703

https://www.nature.com/articles/nchem.2703#supplementary-information (2017).

20   Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239-242, doi:10.1038/nature08013 (2009).

21   Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nature Chemistry* **7**, 301, doi:10.1038/nchem.2202

https://www.nature.com/articles/nchem.2202#supplementary-information (2015).

22   Cieślak, J., Ausín, C., Grajkowski, A. & Beaucage, S. L. The 2-Cyano-2,2-dimethylethanimine-N-oxymethyl Group for the 2′-Hydroxyl Protection of Ribonucleosides in the Solid-Phase Synthesis of RNA Sequences. *Chemistry – A European Journal* **19**, 4623-4632, doi:https://doi.org/10.1002/chem.201204235 (2013).

23   Shimizu, M., Ohta, G., Nagase, O., Okada, S. & Hosokawa, Y. Investigations on Pantothenic Acid and its Related Compounds. I. Chemical Studies. (1). A Novel Synthesis of Pantethine. *CHEMICAL & PHARMACEUTICAL BULLETIN* **13**, 180-188, doi:10.1248/cpb.13.180 (1965).

24   Capon, P. K., Avery, T. D., Purdey, M. S. & Abell, A. D. An improved synthesis of 4-aminobutanenitrile from 4-azidobutanenitrile and comments on room temperature stability. *Synthetic Commun* **51**, 428-436, doi:10.1080/00397911.2020.1832527 (2021).

25   McKay, A. F. *et al.* Bacteriostats. II.1 The Chemical and Bacteriostatic Properties of Isothiocyanates and their Derivatives. *Journal of the American Chemical Society* **81**, 4328-4335, doi:10.1021/ja01525a057 (1959).

26   .

27   Sheldrick, G. SHELXT - Integrated space-group and crystal-structure determination. *Acta Crystallographica Section A* **71**, 3-8, doi:doi:10.1107/S2053273314026370 (2015).

28   Dolomanov, O. V., Bourhis, L. J., Gildea, R. J., Howard, J. A. K. & Puschmann, H. OLEX2: a complete structure solution, refinement and analysis program. *Journal of Applied Crystallography* **42**, 339-341, doi:doi:10.1107/S0021889808042726 (2009).

29   Sheldrick, G. Crystal structure refinement with SHELXL. *Acta Crystallographica Section C* **71**, 3-8, doi:doi:10.1107/S2053229614024218 (2015).

30   Hubschle, C. B., Sheldrick, G. M. & Dittrich, B. ShelXle: a Qt graphical user interface for SHELXL. *Journal of Applied Crystallography* **44**, 1281-1284, doi:doi:10.1107/S0021889811043202 (2011).

Appendix 2

Experimental Data for the Photochemistry of Thioacids

1.   UV Irradiation of Thioacids

Irradiation of α-Amino Nucleophiles in the Presence of Thioacid Substrates
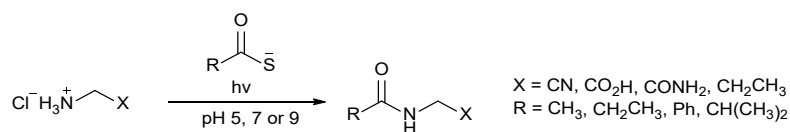
General procedure 1: A stock solution of the specified amino substrate (**AA**-X) (50 mM) was made up in either degassed H_2O or the required buffer solution prepared in degassed H_2O and the pH of was adjusted to pH 5, 7 or 9, as required. 3 mL of the stock solution was transferred into a cuvette with thioacid substrate (3 equiv.), then sealed with argon and parafilm. The reactions were then irradiated at 254 nm with RPR-2547A lamps for 24 hours, unless stated otherwise. The progress of the reaction was monitored by ^{1}H NMR with aliquots taken at specified time points and the reaction was monitored by ^{1}H NMR. The resulting reaction mixtures were analysed by 1D and 2D NMR spectroscopy (^{1}H–^{13}C HMBC, ^{1}H–^{1}H COSY, and ^{1}H–^{13}C HSQC). The products were quantified by ^{1}H NMR spectroscopy against the internal MSM standard of the aliquot.

*Irradiation of **Gly**-CN with AcSK at pH 5*
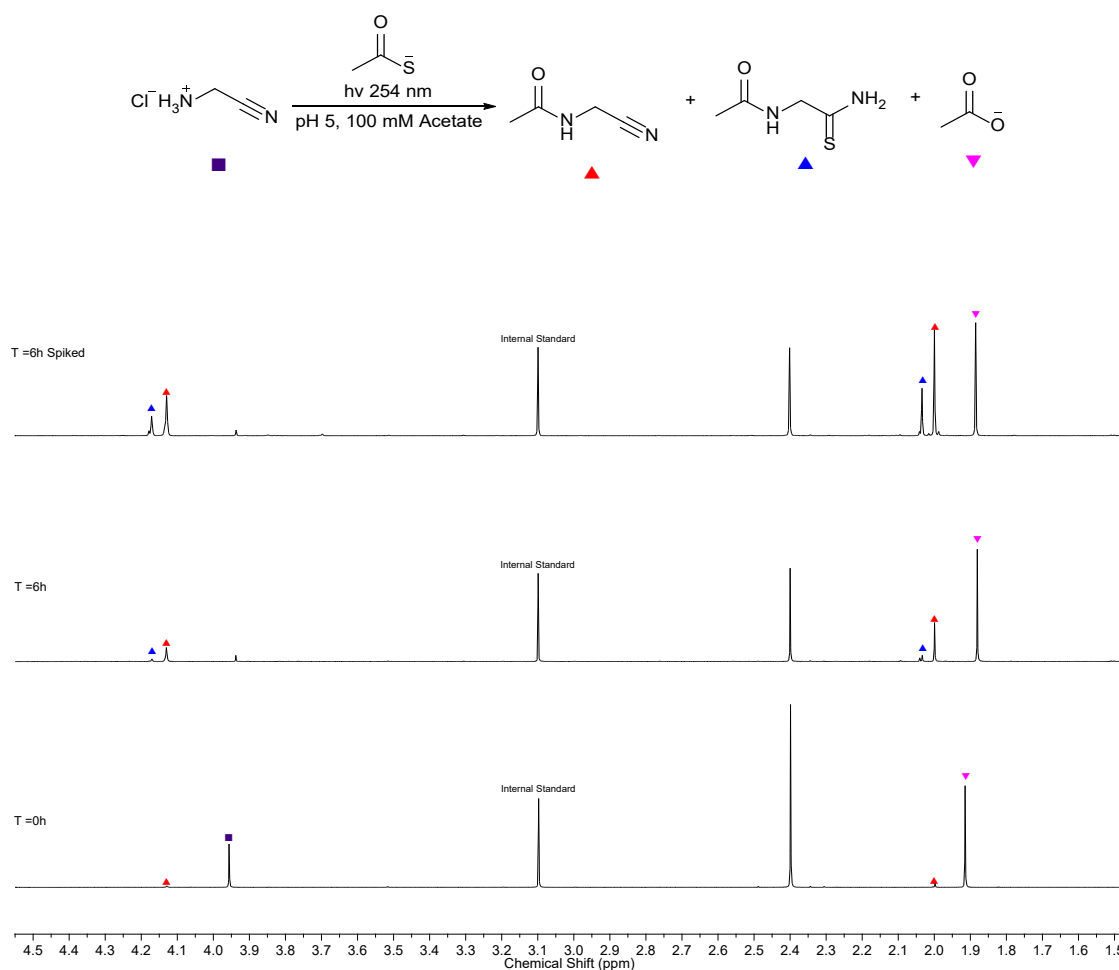
Figure 1. ^{1}H NMR (700 MHz, D$_2$O, 1.5-4.5 ppm, noseygppr1d) spectrum to show the acetylation of *glycine nitrile* (**Gly**-CN, 50 mM) with *potassium thioacetate* (**AcSK**, 150 mM), with UV light (254 nm) after 6 h at pH 5. See **general procedure 1** for method.

^{1}H NMR (700 MHz, D$_2$O) *N-Acetyl glycine nitrile*, Ac-**Gly**-CN (▲): δ 4.13 (s, 2H, (C2)-H$_2$), 2.00 (s, 3H, COCH$_3$); *N-Acetyl glycine thioamide*, Ac-**Gly**-SNH$_2$ (▲): δ 4.17 (s, 2H, (C2)-H$_2$), 2.04 (s, 3H, COCH$_3$); *Acetate*, **AcO**$^-$ (▼): δ 1.88 (s, COCH$_3$).

Time /h	Limiting Reagent	Average % Yield by NMR Internal Standard		Total Acetylation %	SD
	■	▲	▲		
0	92	6	0	6	0.34
0.25	59	21	0	21	1.59
0.5	56	27	0	27	0.15
1	33	45	2	45	5.96
2	18	65	4	65	2.12
4	4	77	9	77	4.04
6	0	82	14	82	1.69

Table 1. Average yields of **Gly**-CN (50 mM) with **ACSK** (150 mM), with UV light (254 nm) over 6 hours at pH 5 with MSM (25 mM) added as the internal standard. See **general procedure 1** for method.

*Irradiation of **Gly**-CN with AcSK at pH 7*
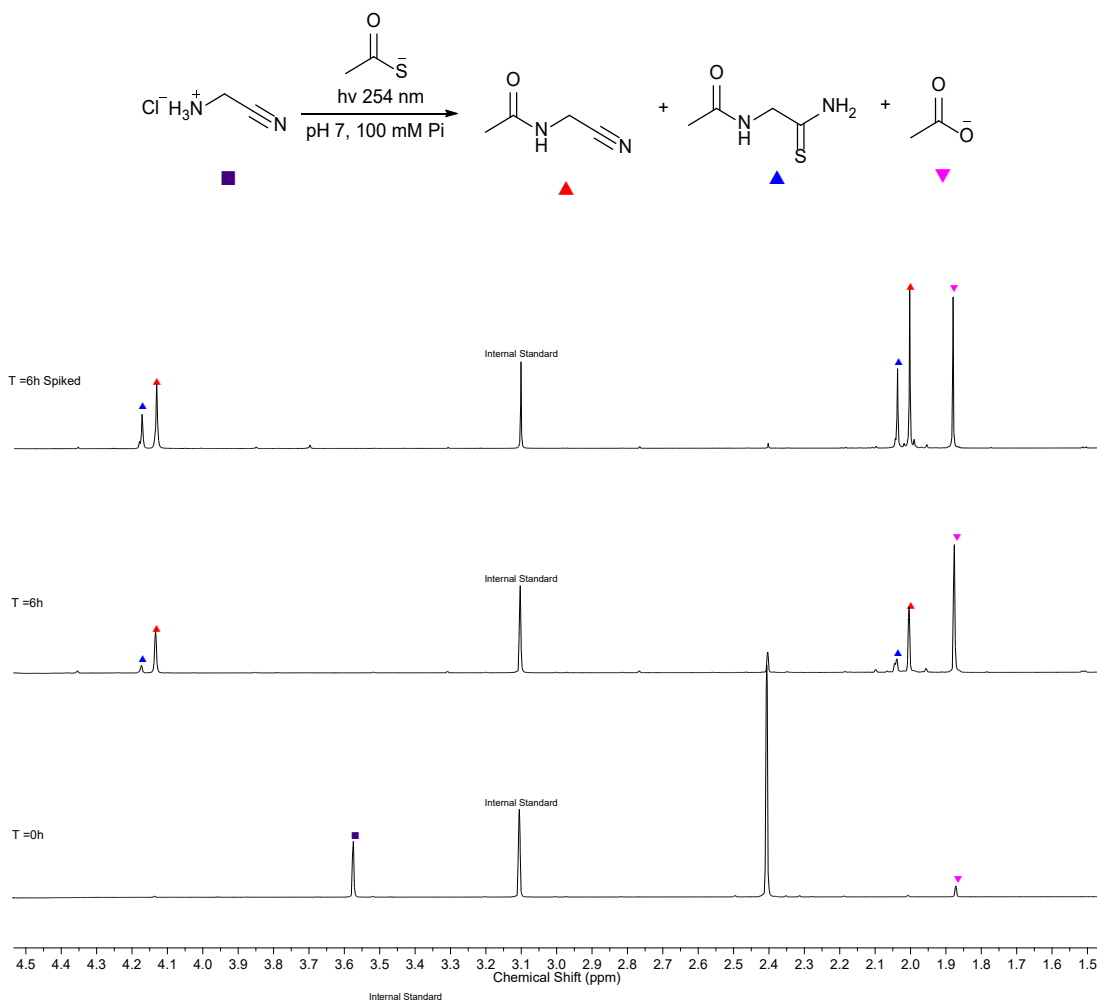
Figure 3. ^{1}H NMR (700 MHz, D$_2$O, 1.5-4.5 ppm, noseygppr1d) spectrum to show the acetylation of *glycine nitrile* (**Gly**-CN, 50 mM) with *potassium thioacetate* (**AcSK**, 150 mM), with UV light (254 nm) after 24 h at pH 7. See **general procedure 1** for method.

^{1}H NMR (700 MHz, D$_2$O) *N-Acetyl glycine nitrile*, Ac-**Gly**-CN (▲): δ 4.19 (s, 2H, (C2)-H$_2$), 2.07 (s, 3H, COCH$_3$); *N-Acetyl glycine thioamide*, Ac-**Gly**-SNH$_2$ (▲): δ 4.23 (s, 2H, (C2)-H$_2$), 2.11 (s, 3H, COCH$_3$); *Acetate*, **AcO**$^-$ (▼): δ 1.93 (s, 3H, COCH$_3$).

Time /h	Limiting Reagent	Average % Yield by NMR Internal Standard		Total Acetylation %	SD
	■	▲	▲		
0	96	0	0	0	0.00
0.25	80	15	0	15	0.38
0.5	71	25	1	26	1.18
1	48	44	3	47	1.85
2	20	64	11	75	1.83
4	0	79	18	97	5.58
6	0	85	18	99	5.45
24	0	19	68	87	-

Table 2. Average yields of **Gly**-CN (50 mM) with **ACSK** (150 mM), with UV light (254 nm) over 24 hours at pH 7 with MSM (25 mM) added as the internal standard. See **general procedure 1** for method.

*Irradiation of **Gly**-CN with AcSK at pH 9*
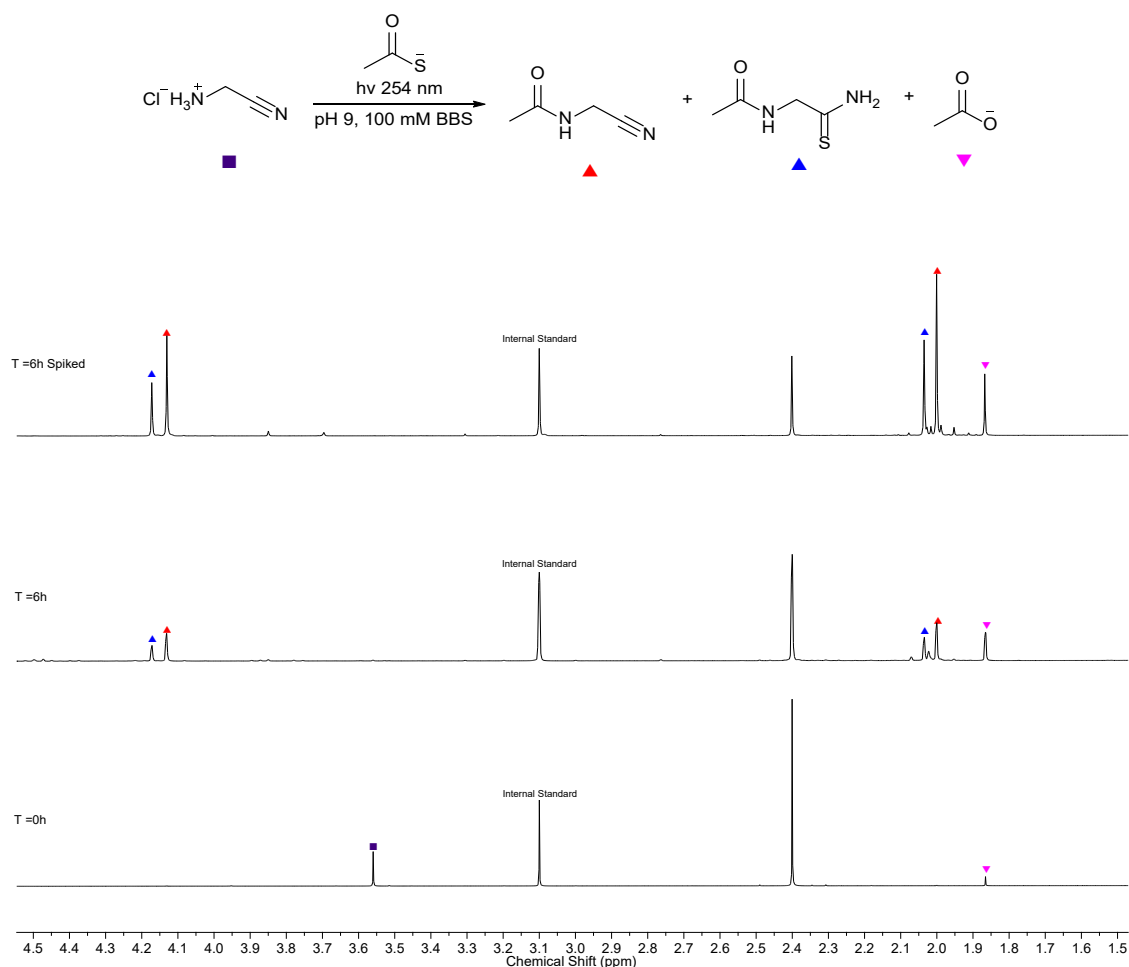
Figure 5. ^{1}H NMR (700 MHz, D$_2$O, 1.0-5.0 ppm, noseygppr1d) spectrum to show the acetylation of *glycine nitrile* (**Gly**-CN, 50 mM) with *potassium thioacetate* (**AcSK**, 150 mM), with UV light (254 nm) after 8 h at pH 9. See **general procedure 1** for method.

^{1}H NMR (700 MHz, D$_2$O) *N-Acetyl glycine nitrile*, Ac-**Gly**-CN (▲): δ 4.12 (s, 2H, (C2)-H$_2$), 1.99 (s, 3H, COCH$_3$); *N-Acetyl glycine thioamide*, Ac-**Gly**-SNH$_2$ (▼): δ 4.16 (s, 2H, (C2)-H$_2$), 2.03 (s, 3H, COCH$_3$); *Acetate*, **AcO$^-$** (◆): δ 1.86 (s, 3H, COCH$_3$).

Time /h	Limiting Reagent ■	Average % Yield by NMR Internal Standard ▲	▲	Total Acetylation %	SD
0	101	0	0	0	0.00
0.25	80	19	0	19	2.40
0.5	65	33	0	33	3.60
1	48	53	3	56	7.14
2	23	62	14	76	7.35
4	8	65	23	88	9.01
6	0	53	33	86	2.85
8	0	32	44	76	-

Table 3. Average yields of **Gly**-CN (50 mM) with **ACSK** (150 mM), with UV light (254 nm) over 8 hours at pH 9 with MSM (25 mM) added as the internal standard. See **general procedure 1** for method.

Acylation of α-Amino Nucleophiles with UV Irradiation

N-(Cyanomethyl)propionamide
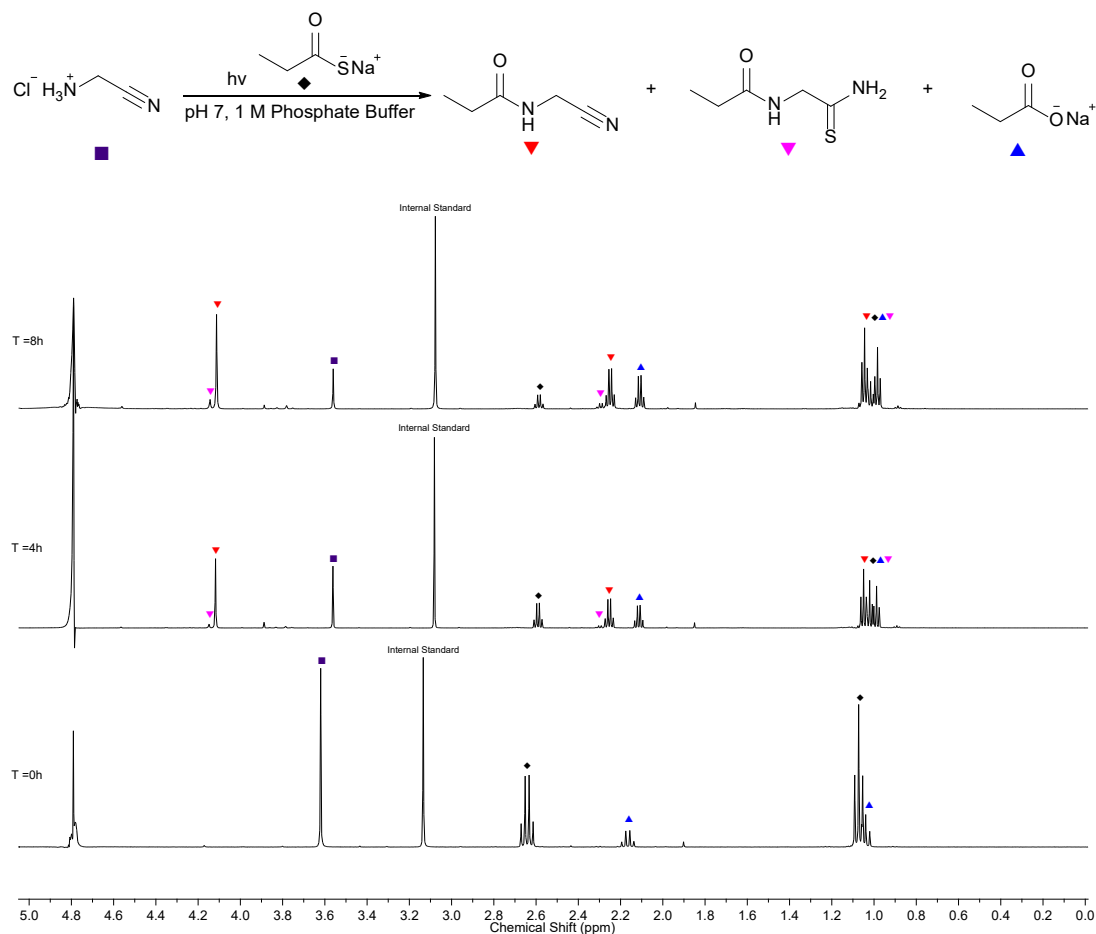
Figure 7. ^{1}H NMR (600 MHz, D$_2$O, 0-5.0 ppm, noseygppr1d) spectrum to show the acylation of *2-aminoacetonitrile hydrochloride salt* (**Gly**-CN, 50 mM) and *sodium propanethioate* (EtCOS[-]Na[+], 150 mM) with UV light (254 nm) after 8 h at pH 7. See **general procedure 1** for method.

^{1}H NMR (600 MHz, D$_2$O, noseygppr1d) *N-(Cyanomethyl)propionamide*, (▼): δ 4.12 (s, 2H, (C2)-H$_2$, 2.25 (q, *J* = 7.7 Hz, 2H, COCH$_2$), 1.05 (t, *J* = 7.7 Hz, 3H, CH$_3$); *N-(2-Amino-2-thioxoethyl)propionamide* (▼): δ 4.15 (s, 2H, (C2)-H$_2$, 2.31 (obs. q, J = 7.6 Hz, 2H, COCH$_2$), obs. (s, 3H, CH$_3$); *Propionate*, **EtCOO**[-] (▲): δ 2.59 (q, J = 7.6 Hz, 2H, COCH$_2$), 0.99 (t, J = 7.6 Hz, 3H, CH$_3$).

Time /h	Limiting Reagent	% Yield by NMR Internal Standard		% Thioacid Hydrolysis
	■	▼	▼	
0	104	-	-	25
0.5	87	14	-	27
1	56	28	1	27
2	47	41	2	27
4	36	52	5	34
6	30	55	7	39
8	23	65	9	47

Table 4. Reaction yields of **Gly**-CN (50 mM) with **sodium propanethioate** (EtCOS[-]Na[+], 150 mM), with UV light (254 nm) over 8 hours at pH 7 with MSM (16.6 mM) added as the internal standard. See **general procedure 1** for method.

N-(Cyanomethyl)isobutyramide
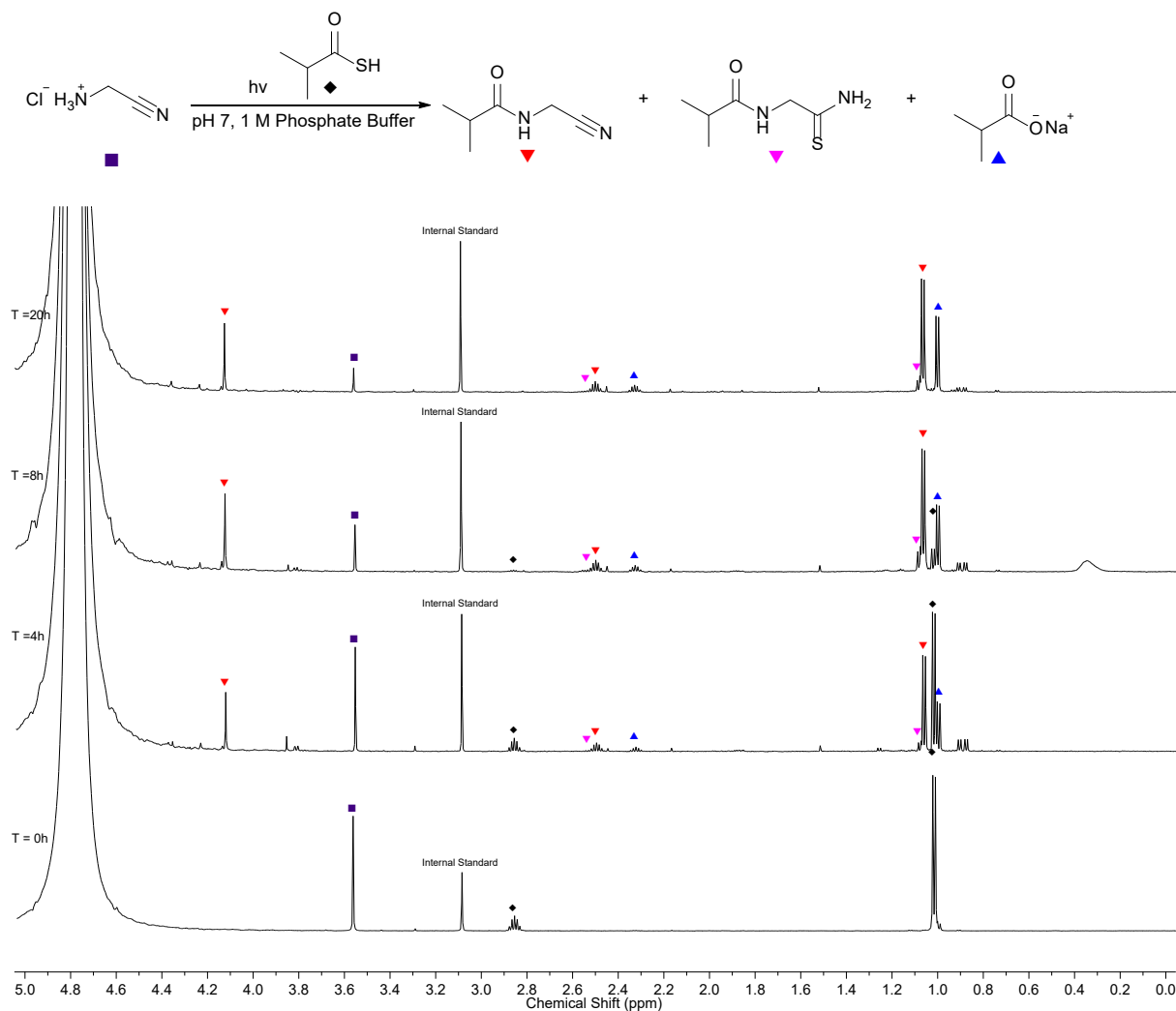
Figure 8. ^{1}H NMR (600 MHz, D$_2$O, 0-5.0 ppm, noseygppr1d) spectrum to show the acylation of *2-aminoacetonitrile hydrochloride salt* (**Gly**-CN, 50 mM) and *2-methyl propanethioic acid* (*^{i}*BuCOSH, 50 mM) with UV light (254 nm) after 20 h at pH 7. See **general procedure 1** for method.

^{1}H NMR (600 MHz, D$_2$O) *N-(Cyanomethyl)propionamide*, (▼): δ 4.13 (s, 2H, (C2)-H$_2$, 2.50 (hept., *J* = 6.9 Hz, 1H, C**H**(CH$_3$)$_2$, 1.07 (d, *J* = 6.9 Hz, 6H, (CH$_3$)$_2$; *N-(2-Amino-2-thioxoethyl)propionamide* (▼): δ 4.14 (s, 2H, (C2)-H$_2$, 2.56 (obs. hept., *J* = 6.9, 1H, CH(CH$_3$)$_2$, 1.09 (obs. d, *J* = 6.9 Hz, 6H, (CH$_3$)$_2$. *Sodium propionate* (▲): δ 2.34 (hept., *J* = 7.0 Hz, 1H, C**H**(CH$_2$)$_3$, 1.00 (d, *J* = 7.0 Hz, 6H, (CH$_3$)$_2$.

Time /h	Limiting Reagent	% Yield by NMR Internal Standard		% Thioacid Hydrolysis
	■	▼	▼	▲
0	102	-	-	3
0.5	87	16	-	3
2	67	26	3	5
6	34	53	13	9
20	20	66	10	18

Table 5. Reaction yields of **Gly**-CN (50 mM) with **2-methyl propanethioic acid** (*^{i}*BuCOSH, 50 mM), with UV light (254 nm) over 20 hours at pH 7 with MSM (16.6 mM) added as the internal standard. See **general procedure 1** for method.

One equivalent of the 2-methyl propanethioic acid was used in this experiment.

N-(Cyanomethyl)benzamide
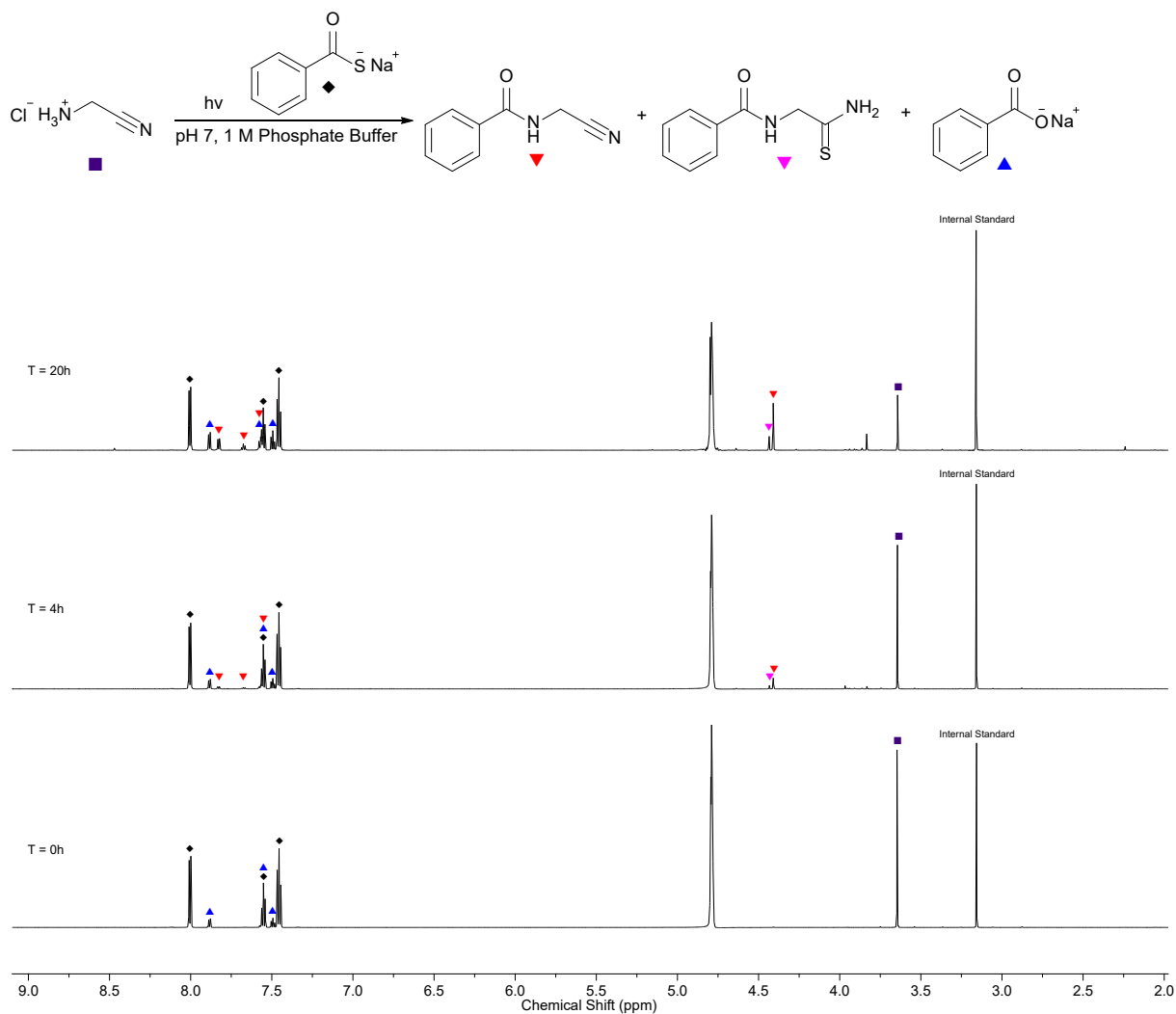
Figure 9. ^{1}H NMR (700 MHz, D$_2$O, 2.0-9.0 ppm, noseygppr1d) spectrum to show the acylation of *2-aminoacetonitrile hydrochloride salt* (**Gly**-CN, 50 mM) and *sodium benzothioate* (PhCOSH, 150 mM) with UV light (254 nm) after 20 h at pH 7. See **general procedure 1** for method.

^{1}H NMR (700 MHz, D$_2$O, noseygppr1d) *N-(Cyanomethyl)benzamide*, (▼): δ 7.82 (dd, *J* = 1.2, 8.4 Hz, 2H, Ar), 7.67 (t, *J* = 7.5 Hz, 1H, Ar), 7.57 − 7.53 (obs., 2H, Ar), 4.40 (s, 2H, (C2)-H$_2$; *N-(2-Amino-2-thioxoethyl)benzamide*, (▼): δ 8.05 − 7.57 (obs., 5H, Ar), .43 (s, (C2)-H$_2$; *Benzoate*, (▲): δ 7.88 (dd, *J* = 1.3, 8.3 Hz, 2H, Ar), 7.57 − 7.55 (obs., 1H, Ar), 7.49 (t, *J* = 7.6 Hz, 2H, Ar).

Time /h	Limiting Reagent	% Yield by NMR Internal Standard		% Thioacid Hydrolysis
	■	▲	▼	▲
0	104	-	-	25
2	88	5	1	25
6	75	12	3	30
8	66	16	4	32
20	25	32	7	43

Table 6. Reaction yields of **Gly**-CN (50 mM) with **sodium benzothioate** (PhCOS$^-$Na$^+$, 150 mM), with UV light (254 nm) over 20 hours at pH 7 with MSM (16.6 mM) added as the internal standard. See **general procedure 1** for method.

N-(Cyanomethyl)benzamide is sparing soluble in water, so mass was lost when comparing integrals in the NMR spectra. Will repeat experiment and analyse the precipitate.

Irradiation of Ac-Gly-SNH$_2$
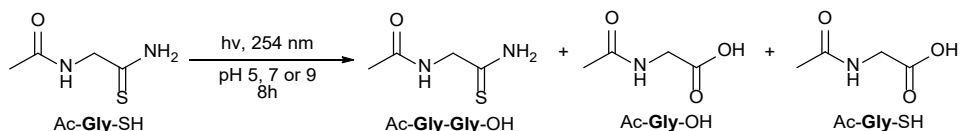
A stock solution of *N*-acetyl-glycine thioamide Ac-**Gly**-SNH$_2$ (50 mM) was made up in degassed H$_2$O/D$_2$O 9:1. Then 3 mL of the stock solution was transferred into a vial with appropriate buffering agent (Acetate; 25 mg, 0.30 mmol, Phosphate; 43 mg, 0.30 mmol, Borate; 19 mg, 0.30 mmol). The samples were adjusted to the required pH with HCl/NaOH and transferred into quartz cuvettes and sealed. The reactions were then irradiated at 254 nm and the progress of the reaction was monitored by ^{1}H NMR. At each timepoint, aliquots were taken (200 µL), MSM (50 µL, 0.1 M) was added as an internal standard and the sample was diluted with D$_2$O (500 µL). The resulting reaction mixtures were analysed by 1D and 2D NMR spectroscopy (^{1}H–^{13}C HMBC, ^{1}H–^{1}H COSY, and ^{1}H–^{13}C HSQC) and the products were quantified by ^{1}H NMR spectroscopy against the internal MSM standard.

Entry	pH	Starting Material	Products	
		Ac-Gly-SNH$_2$ (%)	Ac--Gly-OH (%)	Ac-AA^1-SH (%)
1	5	>99	<1	n.d
2	7	94	6	n.d
3	9	73	4	6

Table 14. Table of the results of the irradiation at 254 nm of Ac-Gly-SNH$_2$ at pH 5, 7 and 9 over 8h.

Irradiation of Ac-Gly-SH and Gly-OH
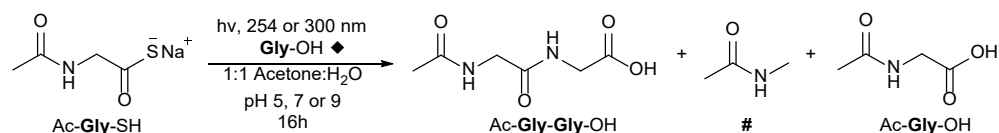
A stock solution of glycine nitrile **Gly**-OH (50 mM) was made up in degassed H_2O/D_2O 9:1. Then 3 mL of the stock solution was mixed with sodium 2-acetamidoethanethioate, Ac-**Gly**-SH (11.63 mg, 0.075 mmol). The sample was transferred to a quartz cuvette and the pH was adjusted to the required value with HCl/NaOH. The sample was sealed with nitrogen and parafilm, then irradiated at either 254 nm or 300 nm. The progress of the reaction was monitored by ^{1}H NMR with aliquots (200µL) taken at specified time points. To each aliquot was added MSM (16.6 µL, 0.1 M) as an internal standard, then diluted with D_2O (500 µL)   The resulting reaction mixtures were analysed by 1D and 2D NMR spectroscopy (^{1}H–^{13}C HMBC, ^{1}H–^{1}H COSY, and ^{1}H–^{13}C HSQC) and the products were quantified by ^{1}H NMR spectroscopy against the internal MSM standard. The identity of the products was confirmed by spiking with authentic samples of Ac-**Gly-Gly**-OH.

Entry	Wavelength (nm)	pH	Yield (%)		
			Ac-Gly-Gly-OH	#	Ac-AA-OH
1	254	5	26	13	23
2	254	7	28	12	21
3	254	9	37	13	33
4	300	5	44	7	5
5	300	7	44	4	19
6	300	9	43	7	24

Table 15. The dipeptide yields of the UV irradiation (254 nm or 300 nm) of **Gly**-CN (50mM) with Ac-**Gly**-SH (50 mM) at pH 5, 7 or 9.

2. Dethiocarboxylation of Aminothioacids with UV Light

Irradiation of Ac-AA-SH
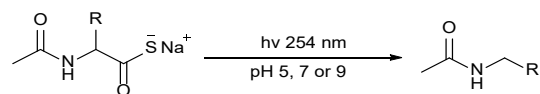
General procedure 3: A stock solution of the specified *N*-acetylated amino thioacid (Ac-**AA**-SH) (50 mM) was made up in either degassed H_2O or the required buffer solution prepared at 0.1 M in degassed H_2O and the pH of was adjusted to pH 5, 7 or 9, as stated. 3 mL of the stock solution was transferred into a cuvette, sealed with argon and parafilm. The reactions were then irradiated at 254 nm with RPR-2547A lamps until the starting material was consumed, unless stated otherwise. The progress of the reaction was monitored by ^{1}H NMR with aliquots taken at specified time points and the reaction was stopped when the starting material was consumed by ^{1}H NMR. The resulting reaction mixtures were analysed by 1D and 2D NMR spectroscopy (^{1}H-^{13}C HMBC, ^{1}H-^{1}H COSY, and ^{1}H-^{13}C HSQC). The products were quantified by ^{1}H NMR spectroscopy against the internal MSM standard of the aliquot.

Entry	Ac-**AA**-SH	pH	R	Yield (%)	Time (h)
1	Ac-**Gly**-SH	5	H	39	8
2	Ac-**Gly**-SH	7	H	42	8
3	Ac-**Gly**-SH	9	H	33	8
4	Ac-**β-Ala**-SH	7	Me	n.d	8
5	Ac-**Ala**-SH	7	Me	64	4
6	Ac-**Met**-SH	7	CH_2CH_2SMe	73	6
7	Ac-**Phe**-SH	7	CH_2Ph	89	2
8	Ac-**Val**-SH	7	$CH(CH_3)_2$	95	6

Table 16. Yields of the dethiocarboxylation of Ac-**AA**-SH (50 mM) with UV irradiation at 254 nm.

Irradiation of Ac-Gly-SH at pH 5

The UV irradiation experiment was performed following general procedure 3 from sodium 2-acetamidoethanethioate, Ac-**Gly**-S⁻Na⁺ (50 mM). An aliquot (0.2 mL) was taken periodically to monitor the progress of the experiment. To each aliquot (0.2 mL) was added 0.5 mL D_2O (0.5 mL) and then methylsulfonylmethane solution (MSM; 50 μL, 0.1 M). The sample w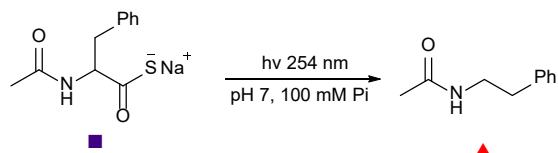as then submitted for analysis by 1D and 2D NMR spectroscopy and the product were confirmed by spiking with synthetically prepared methylacetamide.
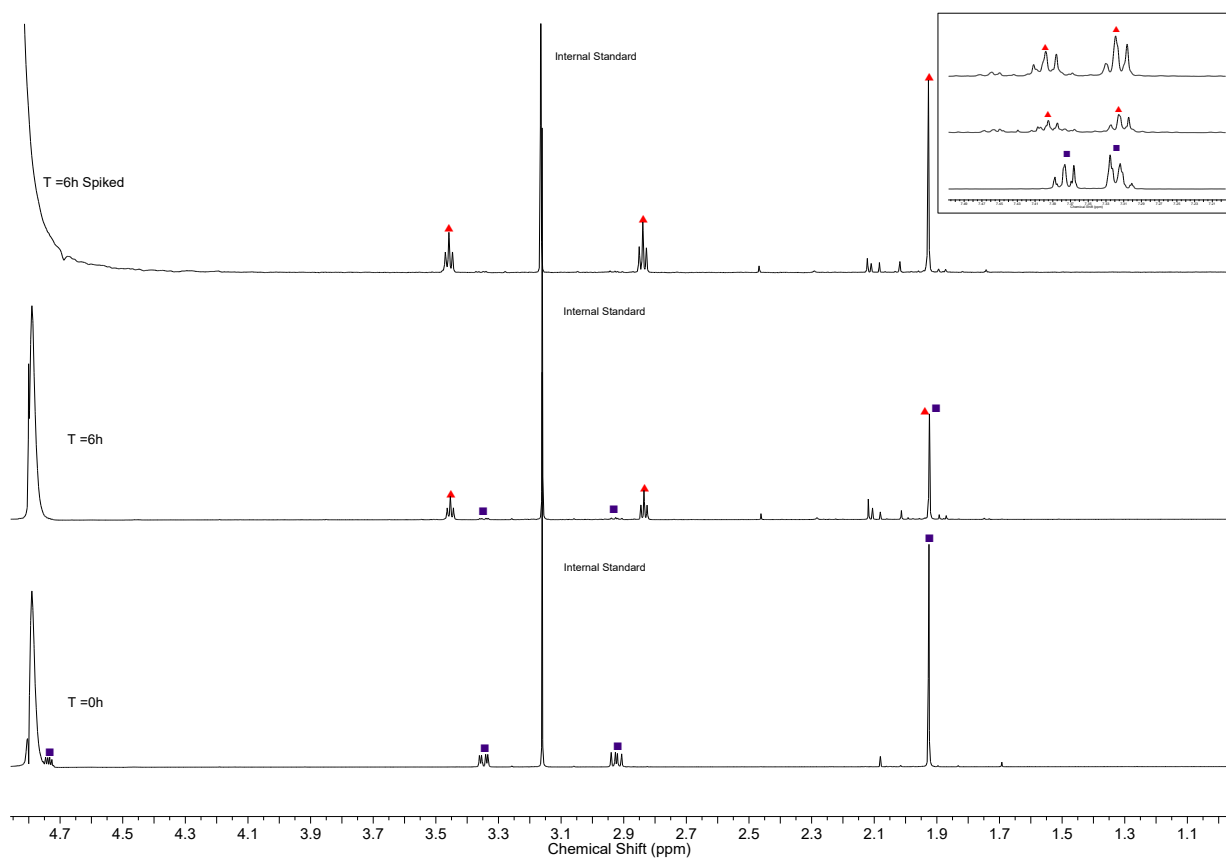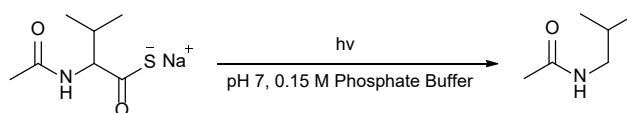

Figure 18. ¹H NMR (700 MHz, D_2O, 1.5-4.5 ppm) spectrum to show the dethiocarboxylation of *Sodium 2-acetamidoethanethioate* (Ac-**Gly**-S⁻Na⁺, 50 mM), with UV light (254 nm) after 8 h at pH 5. See **general procedure 3** for method.

¹H NMR (700 MHz, D_2O) *N-Methylacetamide*, **AcNHMe** (▲): δ 2.72 (s, 3H, NHCH₃), 1.99 (s, 3H, COCH₃); *N-Acetyl-Glycine*, Ac-**Gly**-OH (▼): δ 3.76 (s, 2H, (C2)-H₂), obs. (s, 3H, COCH₃); *Sodium 2-acetamidoethanethioate*, Ac-**Gly**-S⁻Na⁺ (■)

Irradiation of Ac-Gly-SH at pH 7
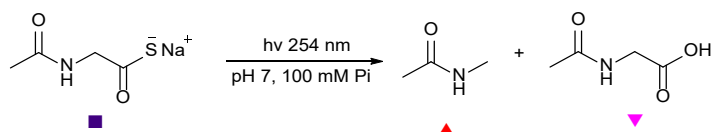
The UV irradiation experiment was performed following general procedure 3 from sodium 2-acetamidoethanethioate, Ac-**Gly**-S⁻Na⁺ (50 mM). An aliquot (0.2 mL) was taken periodically to monitor the progress of the experiment. To each aliquot (0.2 mL) was added 0.5 mL D$_2$O (0.5 mL) and then methylsulfonylmethane solution (MSM; 50 µL, 0.1 M). The sample was then submitted for analysis by 1D and 2D NMR spectroscopy and the product were confirmed by spiking with synthetically prepared methylacetamide.
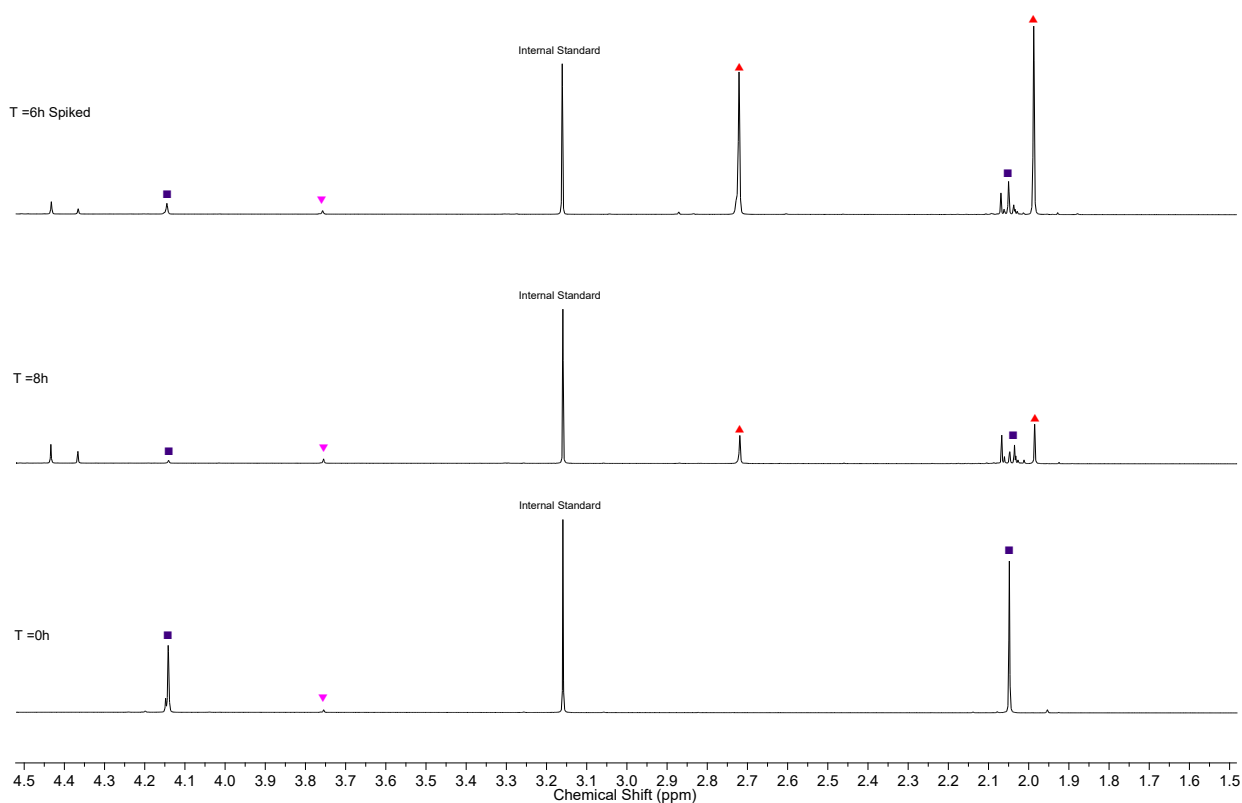
Figure 19. ¹H NMR (700 MHz, D$_2$O, 1.5-4.5 ppm) spectrum to show the dethiocarboxylation of *Sodium 2-acetamidoethanethioate* (Ac-**Gly**-S⁻Na⁺, 50 mM), with UV light (254 nm) after 8 h at pH 7. See **general procedure 3** for method.

¹H NMR (700 MHz, D$_2$O) *N-Methylacetamide*, **AcNHMe** (▲): δ 2.72 (s, 3H, NHCH$_3$), 1.99 (s, 3H, COCH$_3$); *N-Acetyl-Glycine*, Ac-**Gly**-OH (▼): δ 3.76 (s, 2H, (C2)-H$_2$), obs. (s, 3H, COCH$_3$); *Sodium 2-acetamidoethanethioate*, Ac-**Gly**-S⁻Na⁺ (■)

Irradiation of Ac-Gly-SH at pH 9
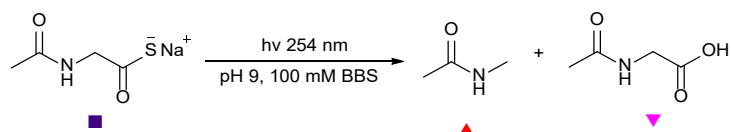
The UV irradiation experiment was performed following general procedure 3 from sodium 2-acetamidoethanethioate, Ac-**Gly**-S⁻Na⁺ (50 mM). An aliquot (0.2 mL) was taken periodically to monitor the progress of the experiment. To each aliquot (0.2 mL) was added 0.5 mL D$_2$O (0.5 mL) and then methylsulfonylmethane solution (MSM; 50 µL, 0.1 M). The sample was then submitted for analysis by 1D and 2D NMR spectroscopy and the product were confirmed by spiking with synthetically prepared methylacetamide.
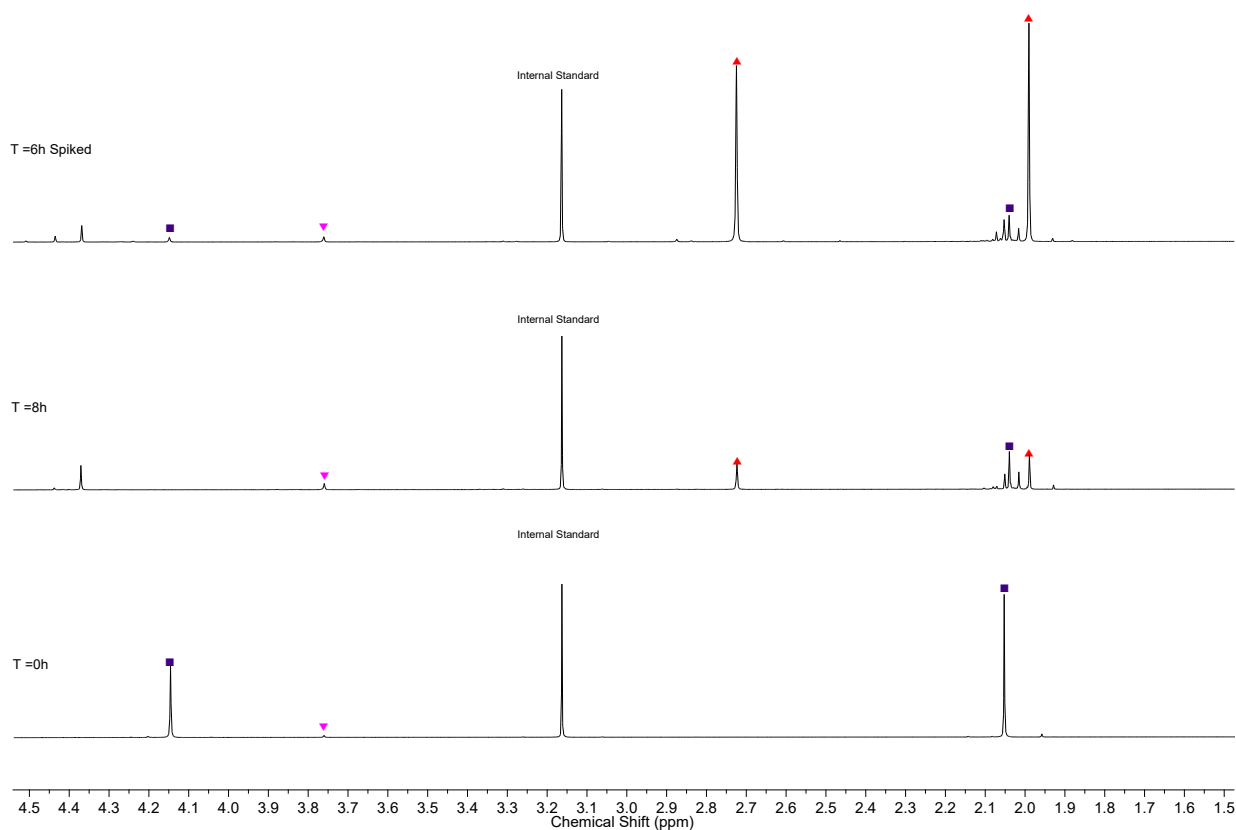

Figure 20. ^{1}H NMR (700 MHz, D$_2$O, 1.5-4.5 ppm) spectrum to show the dethiocarboxylation of *Sodium 2-acetamidoethanethioate* (Ac-**Gly**-S⁻Na⁺, 50 mM), with UV light (254 nm) after 8 h at pH 9. See **general procedure 3** for method.

^{1}H NMR (700 MHz, D$_2$O) *N-Methylacetamide*, **AcNHMe** (▲): δ 2.72 (s, 3H, NHCH$_3$), 1.99 (s, 3H, COCH$_3$); *N-Acetyl-Glycine*, Ac-**Gly**-OH (▼): δ 3.76 (s, 2H, (C2)-H$_2$), obs. (s, 3H, COCH$_3$); *Sodium 2-acetamidoethanethioate*, Ac-**Gly**-S⁻Na⁺ (■)

Irradiation of Ac-Ala-SH at pH 7
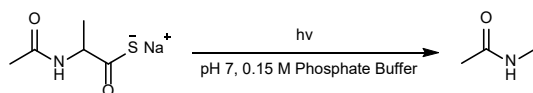
The UV irradiation experiment was performed following general procedure 3 from *N*-acetyl-alanine thioacid, Ac-**Ala**-S⁻Na⁺ (50 mM). An aliquot (0.2 mL) was taken periodically to monitor the progress of the experiment. To each aliquot (0.2 mL) was added 0.5 mL D$_2$O (0.5 mL) and then methylsulfonylmethane solution (MSM; 16.6 μL, 0.1 M). The sample was subsequently submitted for analysis by 1D and 2D NMR spectroscopy.
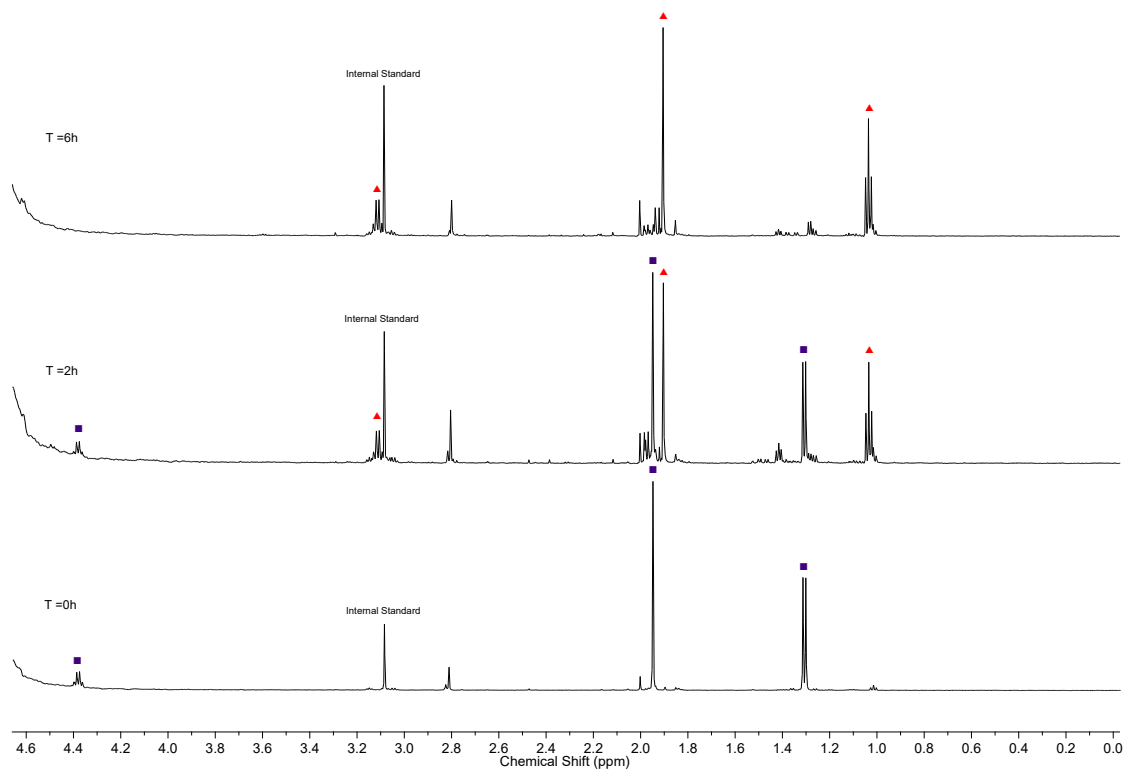
Figure 21. ^{1}H NMR (600 MHz, D$_2$O, 0-4.6 ppm) spectrum to show the dethiocarboxylation of *Sodium 2-acetamidopropanethioate* (Ac-**Ala**-S⁻Na⁺, 50 mM), with UV light (254 nm) after 6 h at pH 7. See ***general procedure 3*** for method.

^{1}H NMR (600 MHz, D$_2$O) *N-Ethylacetamide*, **AcNHEt** (▲): δ 3.11 (q, J = 7.3 Hz, 2H, NHCH$_2$), 1.91 (s, 3H, COCH$_3$), 1.04 (t, *J* = 7.3 Hz, 3H, CH$_3$); *Sodium 2-acetamidopropanethioate*, Ac-**Ala**-S⁻Na⁺ (■).

Irradiation of Ac-Met-SH at pH 7
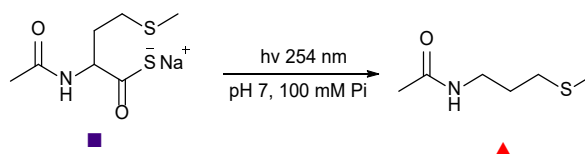
The UV irradiation experiment was performed following general procedure 3 from *N*-acetyl-methionine thioacid, Ac-**Met**-S⁻Na⁺ (50 mM). An aliquot (0.2 mL) was taken periodically to monitor the progress of the experiment. To each aliquot (0.2 mL) was added 0.5 mL D_2O (0.5 mL) and then methylsulfonylmethane solution (MSM; 50 µL, 0.1 M). The sample was then submitted for analysis by 1D and 2D NMR spectroscopy.
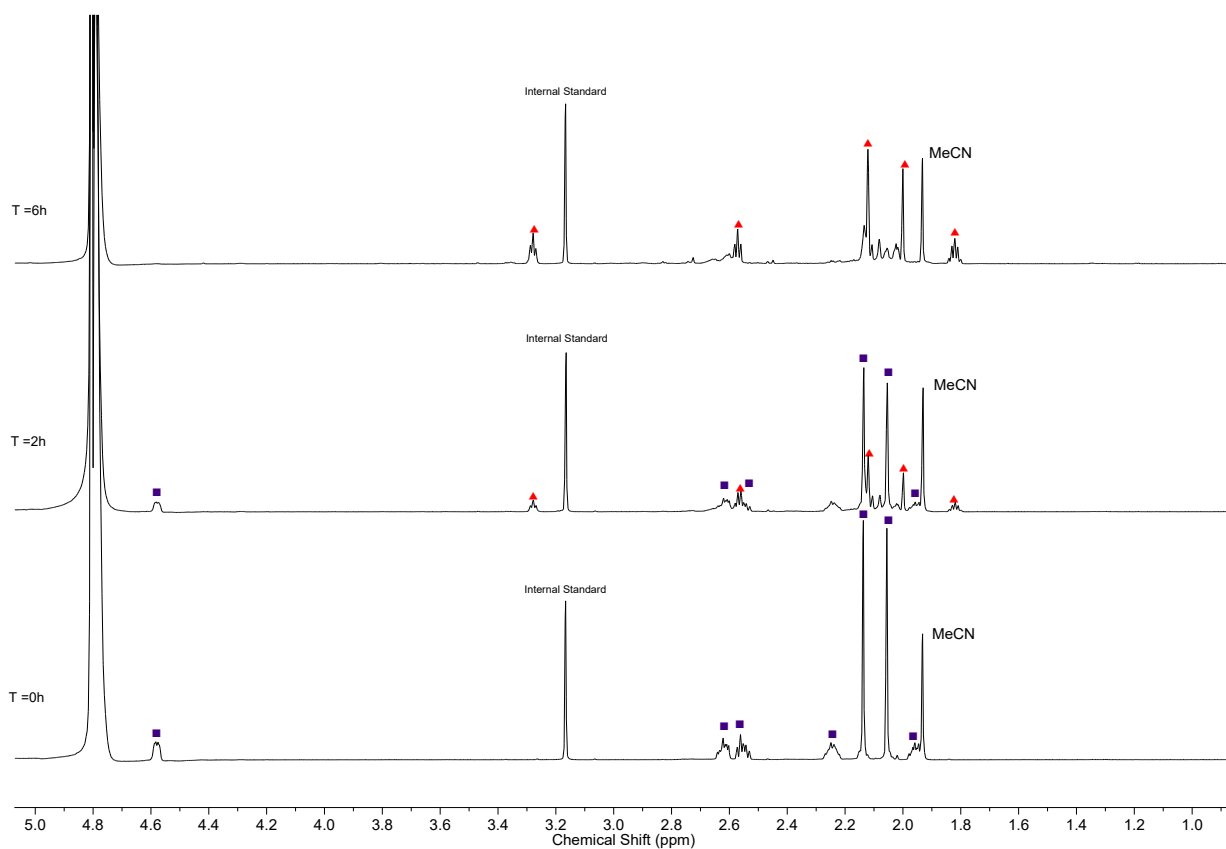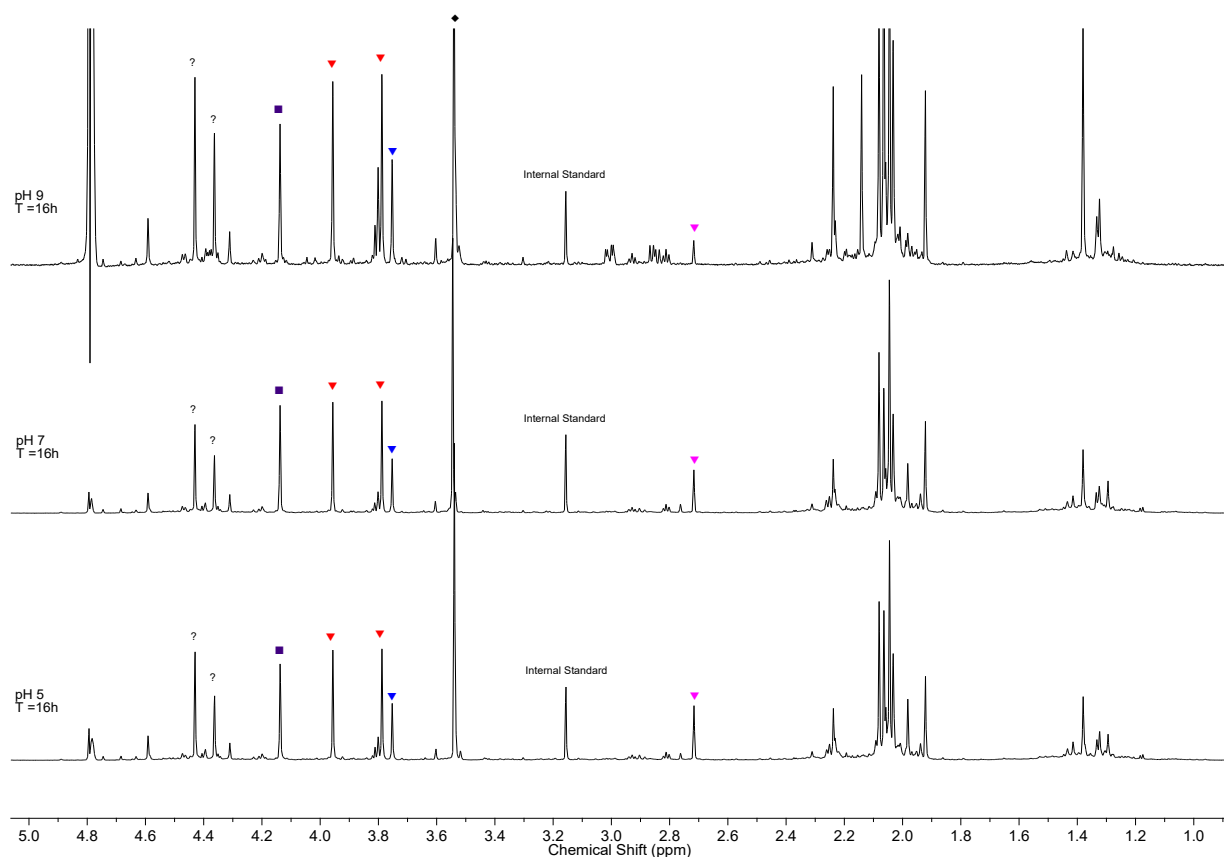

Figure 22. ^{1}H NMR (700 MHz, D_2O, 0-5.0 ppm) spectrum to show the dethiocarboxylation of *Sodium 2-acetamidoethanethioate* (Ac-**Met**-S⁻Na⁺, 50 mM), with UV light (254 nm) after 6 h at pH 7. See **general procedure 3** for method.

^{1}H NMR (700 MHz, D_2O) *N-(3-(Methylthio)propyl)acetamide*, **AcNHCH₂CH₂CH₂SCH3** (▲): δ 3.10 (t, *J* = 6.6 Hz, 2H, NHCH₂), 2.39 (t, *J* = 7.2 Hz, 2H, CH₂SMe); 1.67 – 1.60 (m, 2H, CH₂), 1.94 (s, 3H, COCH₃), 1.82 (s, 3H, SCH₃);; *Sodium 2-acetamido-4-(methylthio)butane thioate*, Ac-**Met**-S⁻Na⁺ (■)

Irradiation of Ac-Phe-SH at pH 7

The UV irradiation experiment was performed following general procedure 3 from *N*-acetyl-phenylalanine thioacid, Ac-**Phe**-S⁻Na⁺ (50 mM). An aliquot (0.2 mL) was taken periodically to monitor the progress of the experiment. To each aliquot (0.2 mL) was added 0.5 mL D$_2$O (0.5 mL) and then methylsulfonylmethane solution (MSM; 50 µL, 0.1 M). The sample was subsequently submitted for analysis by 1D and 2D NMR spectroscopy and the product were confirmed by spiking with synthetically prepared phenylethylacetamide.

Figure 23. ¹H NMR (700 MHz, D$_2$O, 1.0-4.8 ppm) spectrum to show the dethiocarboxylation of *Sodium 2-acetamido-3-phenylpropanethioate* (Ac-**Phe**-S⁺, 50 mM), with UV light (254 nm) after 6 h at pH 7. See ***general procedure 3*** for method.

¹H NMR (700 MHz, D$_2$O) *N-Phenethylacetamide*, **AcNHCH$_2$CH$_2$Ph** (▲): δ 7.42-7.27 (m, 5H, Ar), 3.45 (t, *J* = 6.8 Hz, 2H, NHCH$_2$), 2.83 (t, *J* = 6.9 Hz, 2H, CH$_2$Ar), 1.92 (s, 3H, COCH$_3$); *Sodium 2-acetamido-3-phenylpropanethioate*, Ac-**Phe**-S⁻Na⁺ (■).

The UV irradiation experiment was performed following general procedure 3 from *N*-acetyl-valine thioacid, Ac-**Val**-S⁻Na⁺ (50 mM). An aliquot (0.2 mL) was taken periodically to monitor the progress of the experiment. To each aliquot (0.2 mL) was added 0.5 mL D_2O (0.5 mL) and then methylsulfonylmethane solution (MSM; 8.3 µL, 0.1 M). The sample was subsequently submitted for analysis by 1D and 2D NMR spectroscopy.
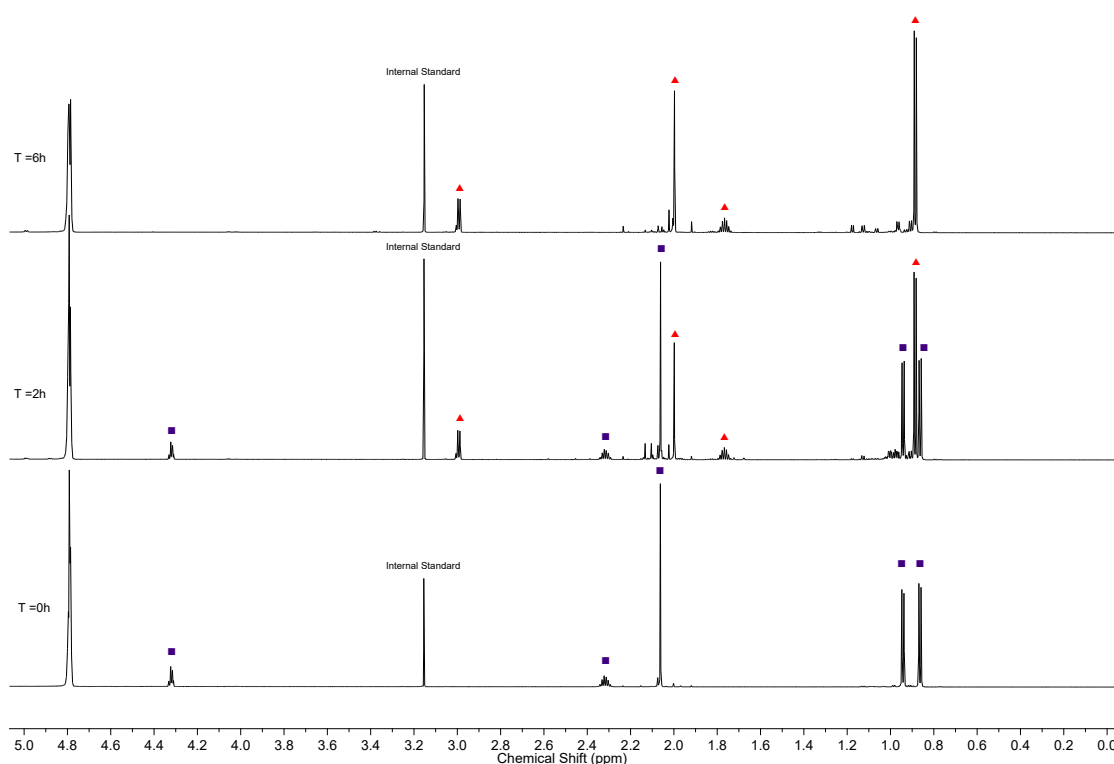

Figure 24. ¹H NMR (700 MHz, D_2O, 0-5.0 ppm) spectrum to show the dethiocarboxylation of *Sodium 2-acetamido-3-methylbutanethioate* (Ac-**Val**-S⁻Na⁺, 50 mM), with UV light (254 nm) after 6 h at pH 7. See ***general procedure 3*** for method.

¹H NMR (700 MHz, D_2O) *N-Isobutylacetamide*, **AcNHCH₂CH(CH₃)₂** (▲): δ 2.83 (d, *J* = 6.82 Hz, 2H, NHCH₂CH), 1.83 (s, 3H, COCH₃), 1.60 (dp, *J* = 13.51, 6.76 Hz, 1H, CH2CH(CH₃)₂), 0.72 (d, J = 6.72 Hz, 6H, CH(CH₃)₂); *Sodium 2-acetamido-3-methylbutanethioate*, Ac-**Val**-S⁻Na⁺ (■).

Irradiation of Ac-Gly-SH and Gly-OH with Acetone as a Triplet Sensitiser
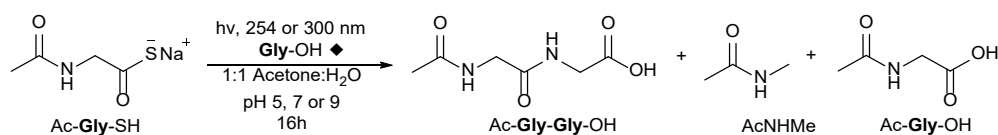
A stock solution of glycine nitrile **Gly**-OH (100 mM) was made up in degassed H_2O/D_2O 9:1. Then 1.5 mL of the stock solution was mixed with *N*-acetyl glycine thioacid Ac-**Gly**-SH (11.63 mg, 0.075 mmol). Degassed acetone (1.5 mL) was added to the mixture. The sample was transferred to a quartz cuvette and the pH was adjusted to the required value with HCl/NaOH. The sample was sealed with nitrogen and parafilm, then irradiated at either 254 nm or 300 nm. The progress of the reaction was monitored by ${}^1H$ NMR with aliquots (200μL) taken at specified time points. The aliquots were lyophilized and to each sample was added MSM (16.6 μL, 0.1 M) as an internal standard, then diluted with D_2O (500 μL) The resulting reaction mixtures were analysed by 1D and 2D NMR spectroscopy (${}^1H$–${}^{13}C$ HMBC, ${}^1H$–${}^1H$ COSY, and ${}^1H$–${}^{13}C$ HSQC) and the products were quantified by ${}^1H$ NMR spectroscopy against the internal MSM standard.

Entry	Wavelength (nm)	pH	Yield (%)		
			Ac-Gly-Gly-OH	AcNHMe	Ac-AA-OH
1	254	5	26	13	23
2	254	7	28	12	21
3	254	9	37	13	33
4	300	5	44	7	5
5	300	7	44	4	19
6	300	9	43	7	24

Table 17. Table of the yields of Ac-**Gly-Gly**-OH formed from **Gly**-OH (50 mM) and Ac-**Gly**-SH (50 mM) at either 254 nm or 300 nm at pH 5, 7 or 9 over 16 h with acetone acting as a triplet sensitizer.
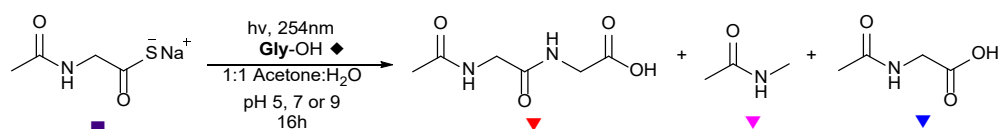
Glycine, **Gly**-OH (50 mM) and sodium 2-acetamidoethanethioate, Ac-**Gly**-SH (50 mM) was irradiated at 254 nm. An aliquot (0.2 mL) was taken periodically to monitor the progress of the experiment. To each aliquot (0.2 mL) was added 0.5 mL D$_2$O (0.5 mL) and then methylsulfonylmethane solution (MSM; 16.6 µL, 0.1 M). The sample was then submitted for analysis by 1D and 2D NMR spectroscopy.
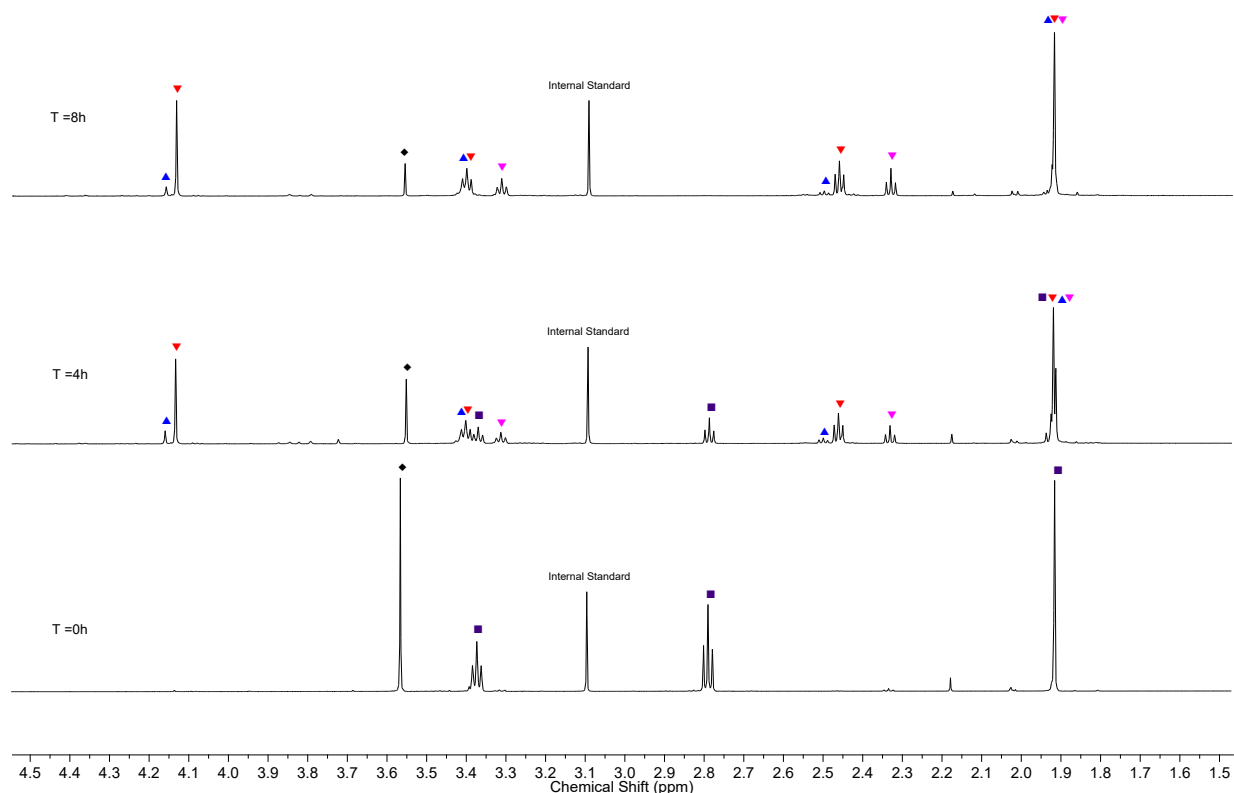

Figure 25. ^{1}H NMR (700 MHz, D$_2$O, 1.0-5.0 ppm, noseygppr1d) spectrum to show the ligation of *glycine* (**Gly**-OH, 50 mM) with *sodium 2-acetamidoethanethioate* (Ac-**Gly**-SH, 50 mM), with UV light (254 nm) and acetone after 16 h at pH 5, 7 and 9.

Irradiation at 300 nm of Ac-Gly-SH and Gly-OH with Acetone as a Triplet Sensitiser
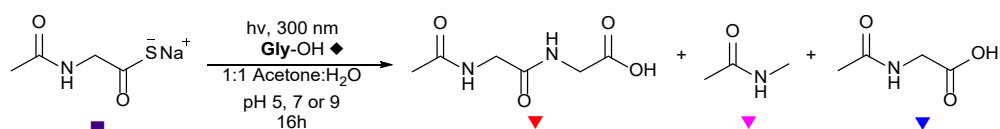
Glycine, **Gly**-OH (50 mM) and sodium 2-acetamidoethanethioate, Ac-**Gly**-SH (50 mM) was irradiated at 300 nm. An aliquot (0.2 mL) was taken periodically to monitor the progress of the experiment. To each aliquot (0.2 mL) was added 0.5 mL D_2O (0.5 mL) and then methylsulfonylmethane solution (MSM; 16.6 µL, 0.1 M). The sample was then submitted for analysis by 1D and 2D NMR spectroscopy.
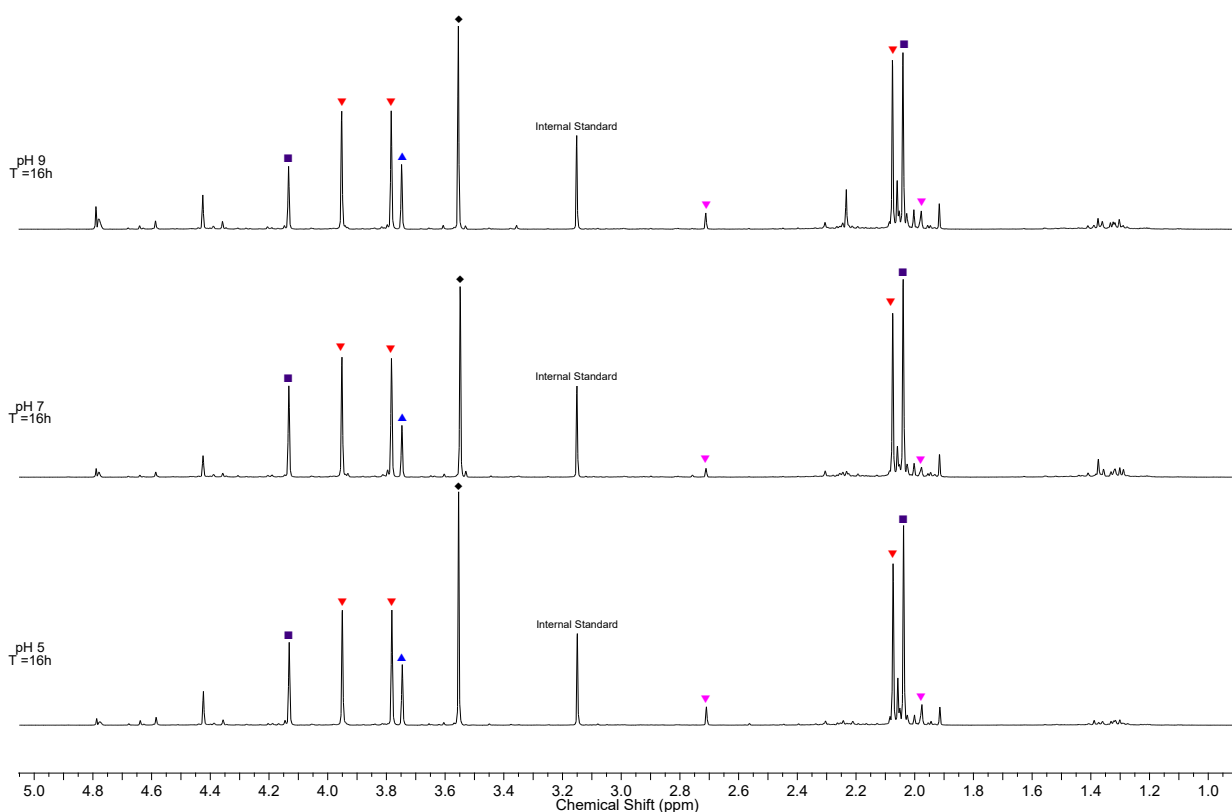
Figure 26. ^{1}H NMR (700 MHz, D_2O, 1.0-5.0 ppm, noseygppr1d) spectrum to show the ligation of *glycine* (**Gly**-OH, 50 mM) with *sodium 2-acetamidoethanethioate* (Ac-**Gly**-SH, 50 mM), with UV light (300 nm) and acetone after 16 h at pH 5, 7 and 9.

Irradiation at 254 nm of Ac-β-Ala-SH and a Amine to Form Dipeptde
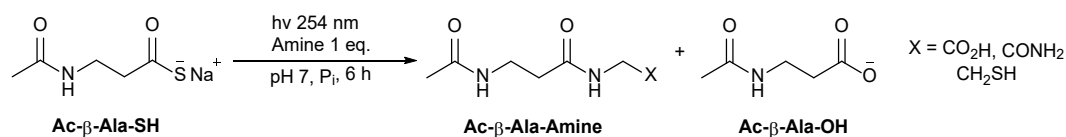
A stock solution of sodium 2-acetamidopropanethioate, Ac-**β-Ala**-OH (50 mM) and sodium phosphate tribasic (100 mM) was made up in degassed H_2O/D_2O 9:1. Then 3 mL of the stock solution was mixed with the amine (50 mM). The sample was transferred to a quartz cuvette and adjusted to pH 7 with HCl/NaOH. The sample was sealed with nitrogen and parafilm, then irradiated at 254 nm. The progress of the reaction was monitored by ^1H NMR with aliquots (200µL) taken at specified time points. To each aliquot was added MSM (50 µL, 0.1 M) as an internal standard, then diluted with D_2O (500 µL) The resulting reaction mixtures were analysed by 1D and 2D NMR spectroscopy (^1H–^{13}C HMBC, ^1H–^1H COSY, and ^1H–^{13}C HSQC) and the products were quantified by ^1H NMR spectroscopy against the internal MSM standard.
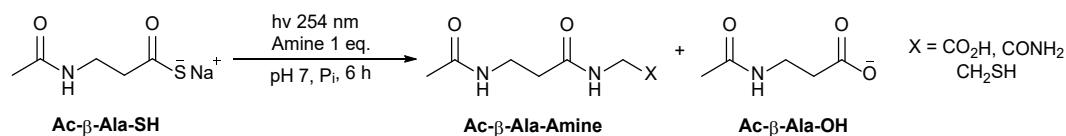
Entry	Amine	Ligation Product	Yield (%)		
			Ac-Gly-Amine	Ac-Gly-AA-SNH₂	Ac-β-Ala-OH
1	**Gly**-OH	Ac-**β-Ala-Gly**-OH	9	-	37
2	**Gly**-CN	Ac-**β-Ala-Gly**-CN	56	9	31
3	Cysteamine	Ac-**β-Ala**-Cysteamine	48	-	38

Table 18. Table of the yields of dipeptide formed from Ac-β-Ala-SH and an amine at pH 7 under 254 nm UV irradiation for 6h.

Irradiation at 254 nm of Ac-β-Ala-SH and Gly-CN to Dipeptide Ac-β-Ala-Gly-CN
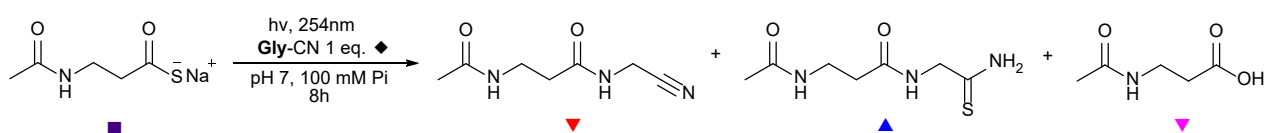
Sodium 2-acetamidopropanethioate, Ac-**β-Ala**-SH (50 mM) and glycine nitrile, **Gly**-CN (50 mM) was irradiated at 254 nm. An aliquot (0.2 mL) was taken periodically to monitor the progress of the experiment. To each aliquot (0.2 mL) was added 0.5 mL D$_2$O (0.5 mL) and then methylsulfonylmethane solution (MSM; 50 µL, 0.1 M). The sample was then submitted for analysis by 1D and 2D NMR spectroscopy.

Figure 27. ^{1}H NMR (700 MHz, D$_2$O, 1.5-4.5 ppm) spectrum to show the ligation of *glycine nitrile* (**Gly**-CN, 50 mM) with *sodium 2-acetamidopropanethioate* (Ac-**β-Ala**-SH, 50 mM), with UV light (254 nm) after 8 h at pH 7.

Irradiation of Ac-β-Ala-SH and Cysteamine to Dipeptide Ac-β-Ala-Cysteamine
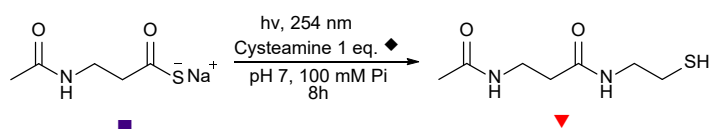
Sodium 2-acetamidopropanethioate, Ac-**β-Ala**-SH (50 mM) and cysteamine, cysteamine (50 mM) was irradiated at 254 nm. An aliquot (0.2 mL) was taken periodically to monitor the progress of the experiment. To each aliquot (0.2 mL) was added 0.5 mL D$_2$O (0.5 mL) and then methylsulfonylmethane solution (MSM; 50 μL, 0.1 M).   The sample was then submitted for analysis by1D and 2D NMR spectroscopy.
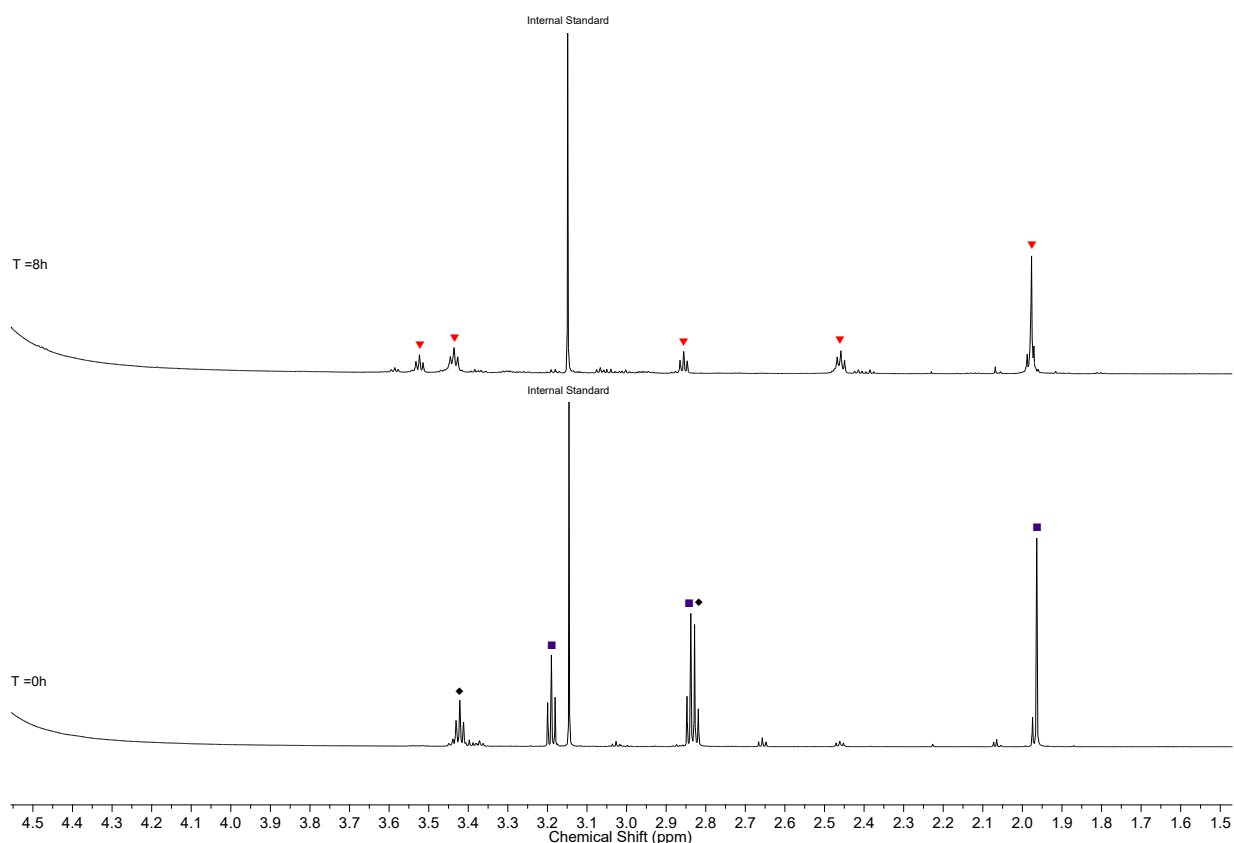
Figure 28. ^{1}H NMR (700 MHz, D$_2$O, 1.5-4.5 ppm) spectrum to show the ligation of *cysteamine* (**Cysteamine**, 50 mM) with *sodium 2-acetamidopropanethioate* (Ac-**β-Ala**-SH, 50 mM), with UV light (254 nm) after 8 h at pH 7.

2. Acid Catalysed Peptide Ligation

Reaction of Gly-CN with AcSK Without Irradiation
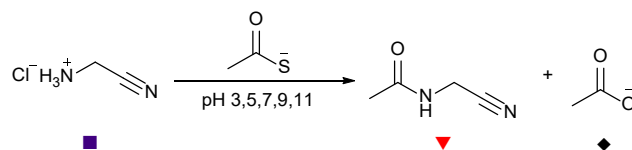
General procedure 2: A stock solution of glycine nitrile **Gly**-CN (1 M) with methylsulfonylmethane (MSM; 0.166 M) was made up in degassed H_2O/D_2O 9:1 and the pH were adjusted to pH 7. Then 50 µL of the stock solution was transferred into a vial with potassium thioacetate (17 mg, 0.15 mmol) and the mixture was made up to 1 mL with degassed H_2O/D_2O 9:1. The samples were adjusted to the required pH with 0.1 M HCl or NaOH, transferred into NMR tubes and sealed with parafilm and nitrogen. The progress of the reaction was monitored by ^1H NMR in the absence UV light over 24 hours. The resulting reaction mixtures were analysed by 1D and 2D NMR spectroscopy (^1H–^{13}C HMBC, ^1H–^1H COSY, and ^1H–^{13}C HSQC) and the products were quantified by ^1H NMR spectroscopy against the internal MSM standard.
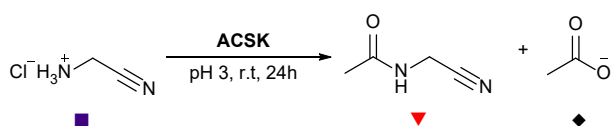
The control experiment was performed following general procedure 2 from glycine nitrile, **Gly**-CN (50 mM), potassium thioacetate, **AcSK** (17.6 mg, 150 mM) and methylsulfonylmethane (MSM; 8.3 mM). The sample was adjusted to pH 3 and then submitted periodically for analysis by 1D and 2D NMR spectroscopy to monitor the progress of the experiment.
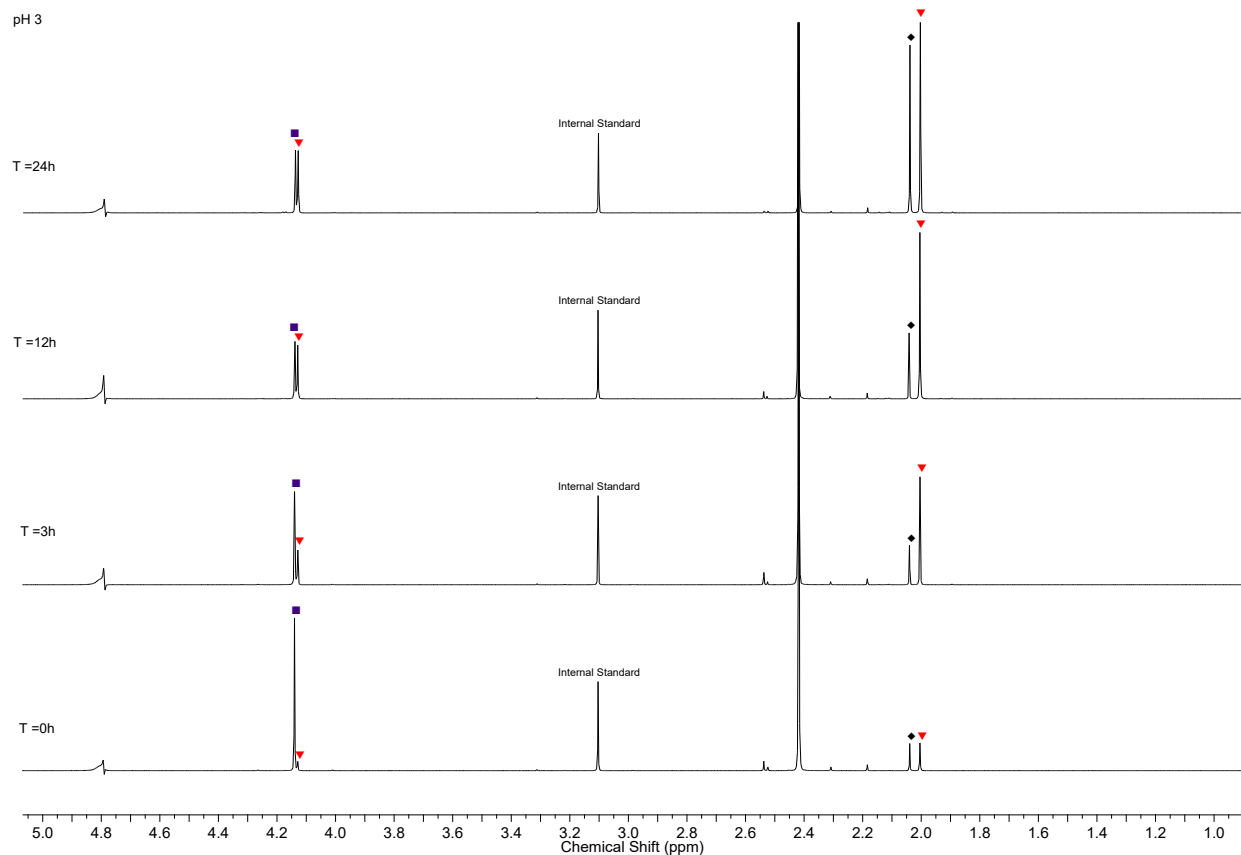
Figure 10. ^{1}H NMR (600 MHz, H$_2$O/D$_2$O 9:1, 1.0-5.0 ppm, noseygppr1d) spectrum to show the acetylation of *glycine nitrile* (**Gly**-CN, 50 mM) with *potassium thioacetate* (**AcSK**, 150 mM) after 24 h at pH 3. See **general procedure 2** for method.

^{1}H NMR (600 MHz, H$_2$O/D$_2$O 9:1, noseygppr1d) *N-Acetyl glycine nitrile*, Ac-**Gly**-CN (▼): δ 4.13 (s, 2H, (C2)-H$_2$), 2.00 (s, 3H, COCH$_3$); *Acetate*, **AcO**$^-$ (◆): δ 2.04 (s, 3H, COCH$_3$).
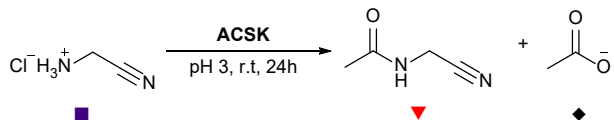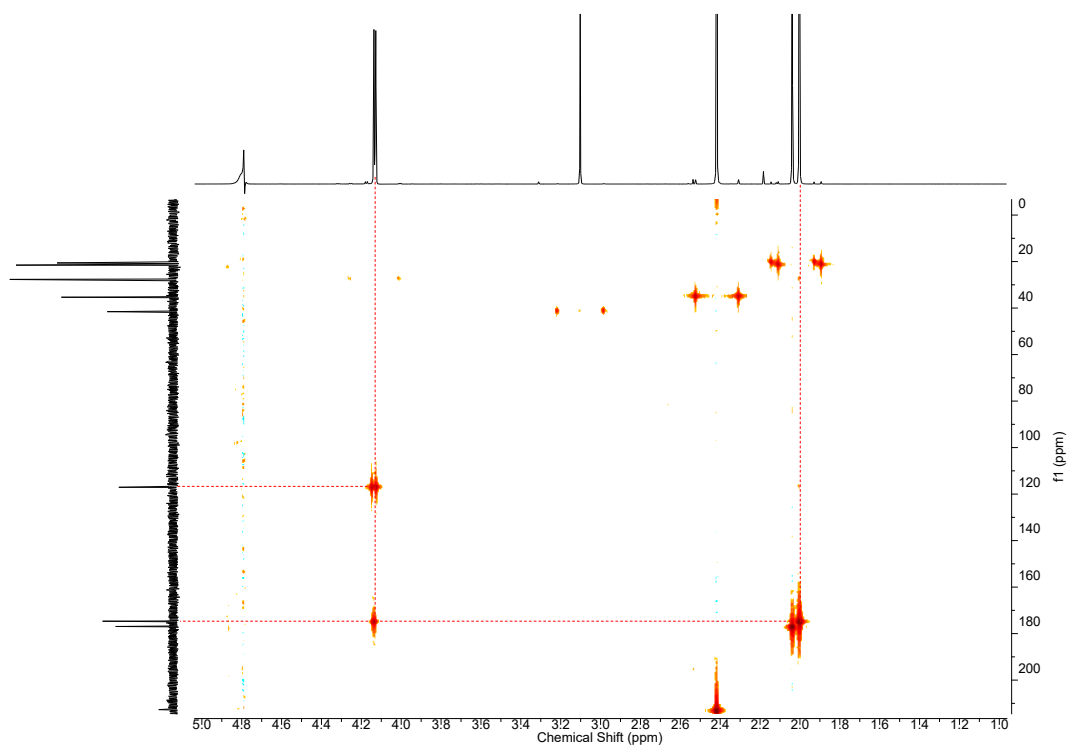
Figure 11. ^1H–^{13}C HMBC (^1H: 600 MHz [1.0–5.0 ppm], ^{13}C: 150 MHz [0–220 ppm], H_2O, noseygppr1d) spectrum showing the diagnostic $^2J_{CH}$ and $^3J_{CH}$ coupling of the coupling of Ac-**Gly**-CN.

Time /h	Limiting Reagent	% Yield by NMR Internal Standard	% Thioacetate Hydrolysis
	■	▼	◆
0	92	7	3
3	72	28	5
6	63	38	7
12	56	44	8
24	48	52	23

Table 7. Reaction yields of **Gly**-CN (50 mM) with **ACSK** (150 mM) at pH 3 over 24 hours without irradiation with MSM (8.3 mM) as the internal standard. See **general procedure 2** for method.

Reaction of Gly-CN with AcSK at pH 5
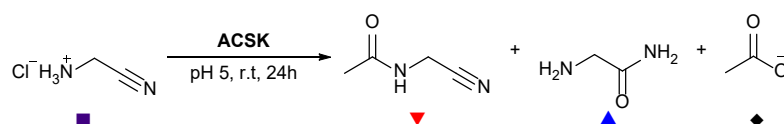
The control experiment was performed following general procedure 2 from glycine nitrile, **Gly**-CN (50 mM), potassium thioacetate, **AcSK** (17.2 mg, 150 mM) and methylsulfonylmethane (MSM; 8.3 mM). The sample was adjusted to pH 5 and then submitted periodically for analysis by 1D and 2D NMR spectroscopy to monitor the progress of the experiment.
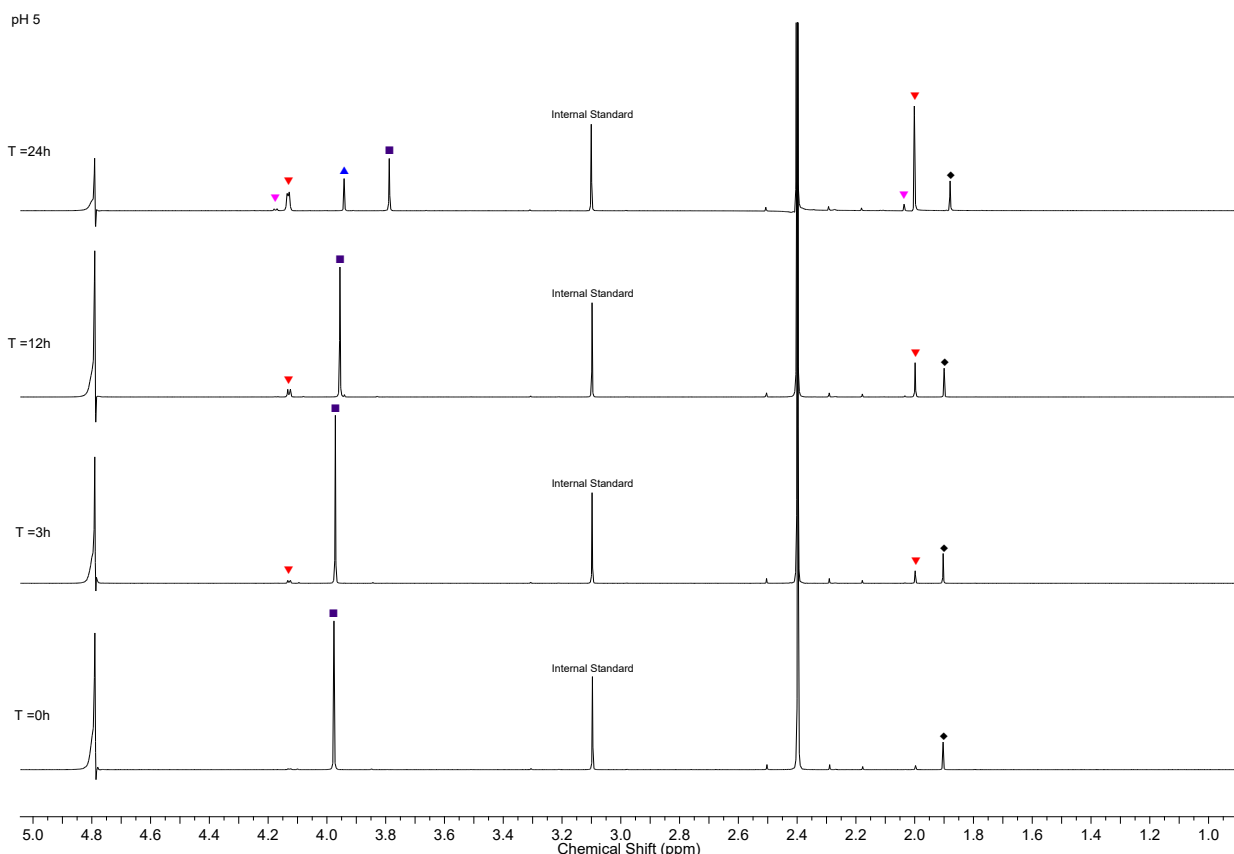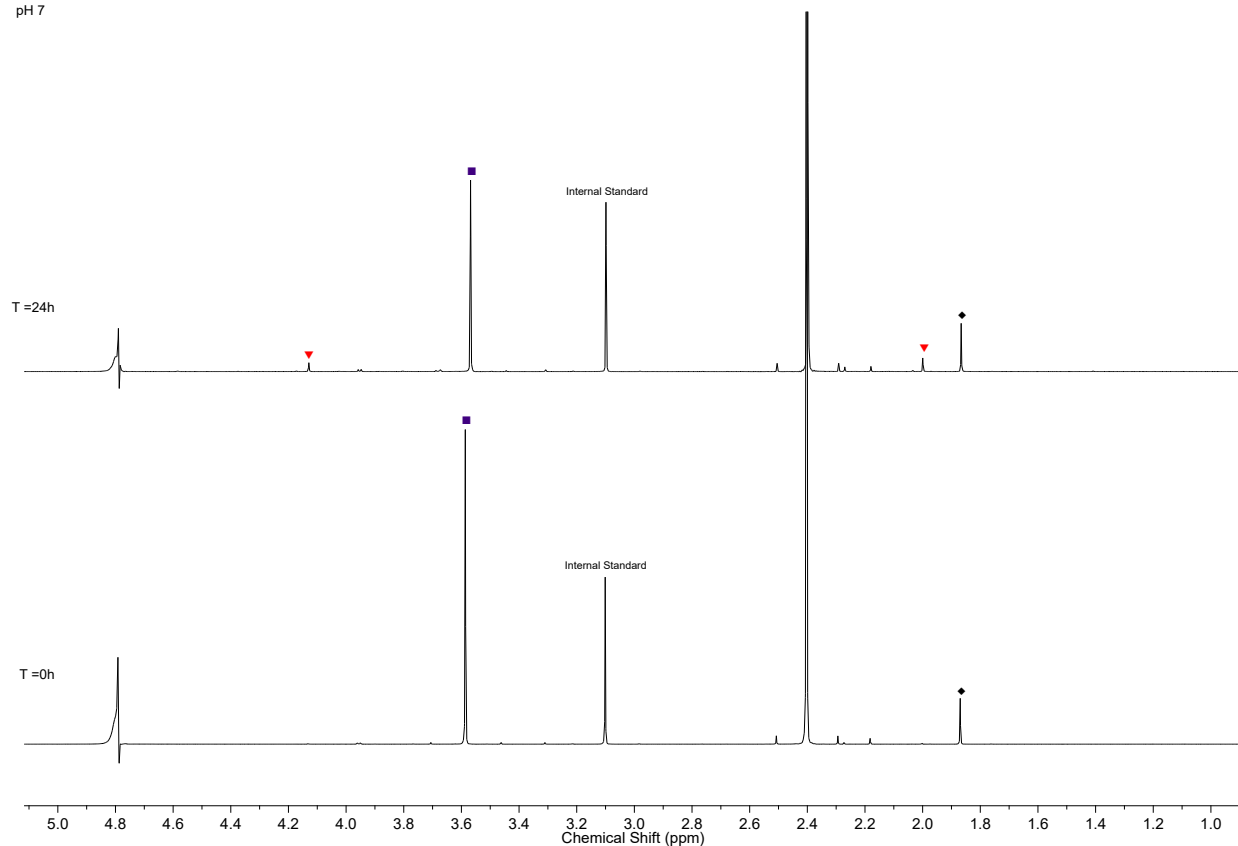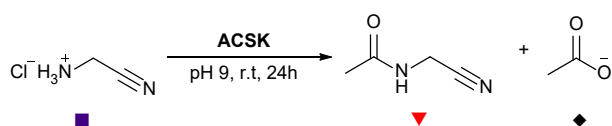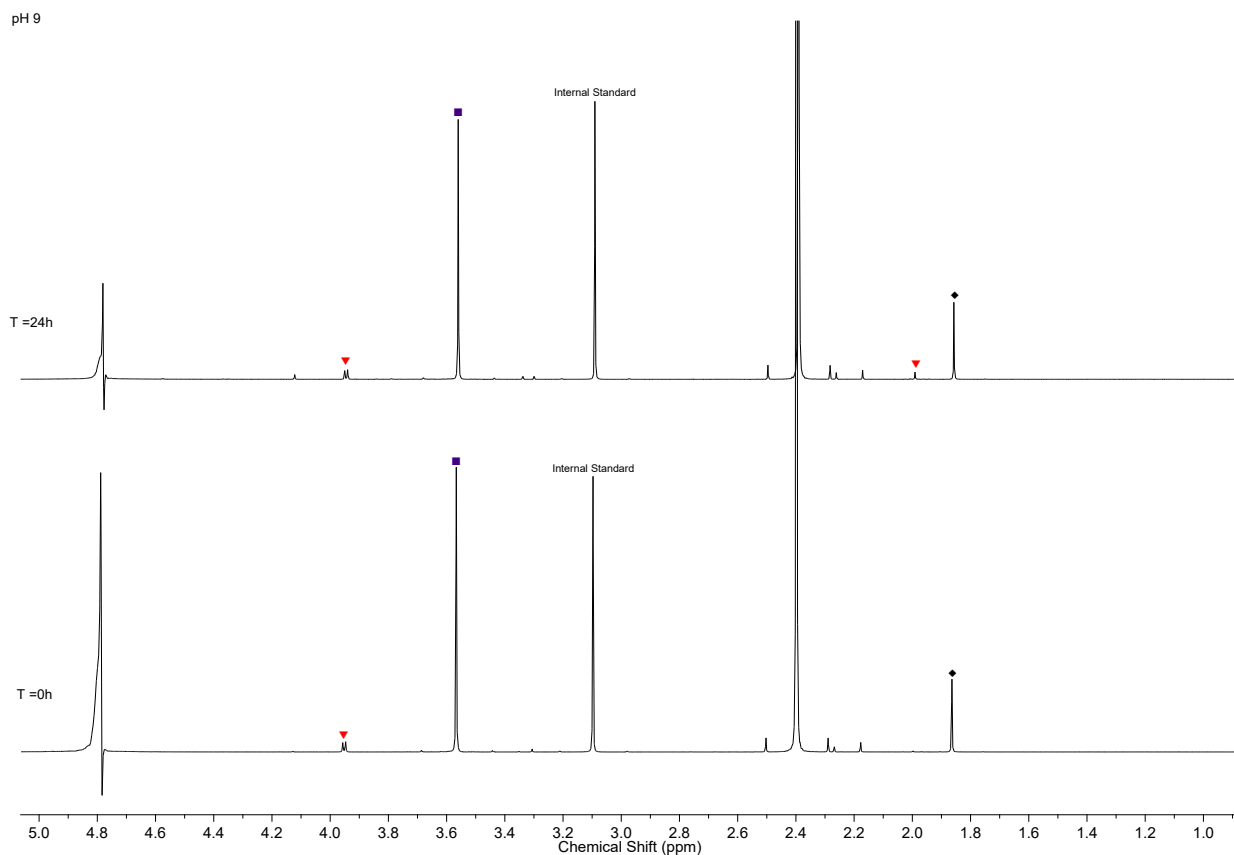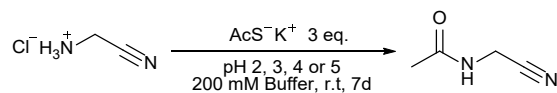

Figure 12. ^{1}H NMR (600 MHz, H$_2$O/D$_2$O 9:1, 1.0-5.0 ppm, noseygppr1d) spectrum to show the acetylation of *glycine nitrile* (**Gly**-CN, 50 mM) with *potassium thioacetate* (**AcSK**, 150 mM) after 24 h at pH 5. See **general procedure 2** for method.

^{1}H NMR (600 MHz, H$_2$O/D$_2$O 9:1, noseygppr1d) *N-Acetyl glycine nitrile*, Ac-**Gly**-CN (▼): δ 4.13 (d, *J* = 3.6 Hz, 2H, (C2)-H$_2$), 2.00 (s, 3H, COCH$_3$); *N-Acetyl glycine thioamide*, Ac-**Gly**-SNH$_2$ (▼): δ 4.17 (d, *J* = 6.0 Hz, 2H, (C2)-H$_2$), 2.04 (s, 3H, COCH$_3$); *Glycinamide*, **Gly**-NH$_2$ (▲): δ 3.94 (s, 2H, (C2)-H$_2$); *Acetate*, **AcO**$^-$ (♦): δ 1.88 (s, 3H, COCH$_3$).

Chemical shift of glycine nitrile caused by a change in pH (possible pH increase)
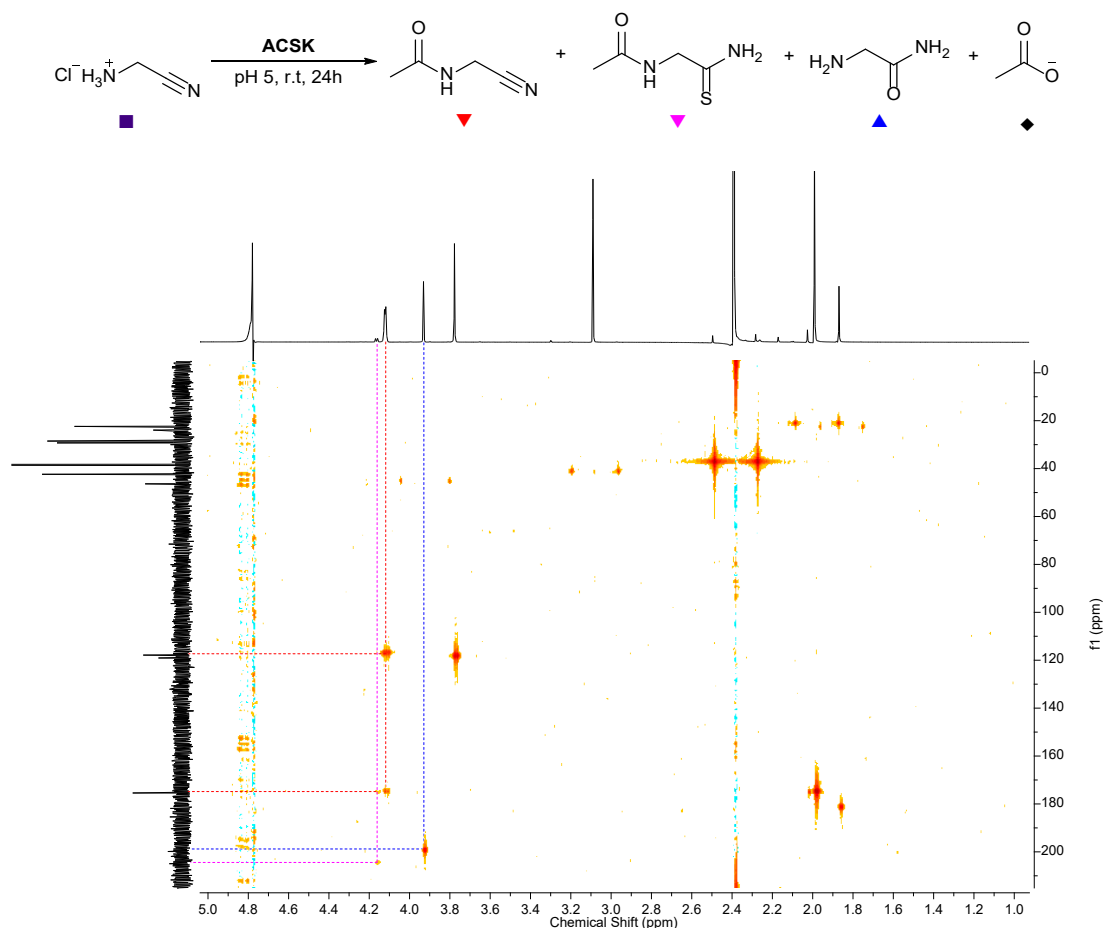
Figure 13. ^1H–^{13}C HMBC (^1H: 600 MHz [1.0–5.0 ppm], ^{13}C: 150 MHz [0–220 ppm], H_2O, noseygppr1d) spectrum showing the diagnostic $^2J_{CH}$ and $^3J_{CH}$ coupling of the coupling of Ac-**Gly**-CN, Ac-**Gly**-SNH$_2$ and **Gly**-NH$_2$.

Time /h	Limiting Reagent	% Yield by NMR Internal Standard			% Thioacetate Hydrolysis
	■	▼	▼	▲	◆
0	98	2	-	-	5
3	94	6	-	-	5
6	90	10	-	-	5
12	87	16	-	-	5
24	34	47	3	19	6

Table 8. Reaction yields of **Gly**-CN (50 mM) with **ACSK** (150 mM) at pH 5 over 24 hours without irradiation with MSM (8.3 mM) as the internal standard. See **general procedure 2** for method.

The control experiment was performed following general procedure 2 from glycine nitrile, **Gly**-CN (50 mM), potassium thioacetate, **AcSK** (17.4 mg, 150 mM) and methylsulfonylmethane (MSM; 8.3 mM). The sample was adjusted to pH 7 and submitted periodically for analysis by 1D and 2D NMR spectroscopy to monitor the progress of the experiment.

Figure 14. ^{1}H NMR (600 MHz, H$_2$O/D$_2$O 9:1, 1.0-5.0 ppm, noseygppr1d) spectrum to show the acetylation of *glycine nitrile* (**Gly**-CN, 50 mM) with *potassium thioacetate* (**AcSK**, 150 mM) after 24 h at pH 7. See **general procedure 2** for method.

^{1}H NMR (600 MHz, H$_2$O/D$_2$O 9:1) *N-Acetyl glycine nitrile*, Ac-**Gly**-CN (▼): δ 4.13 (s, 2H, (C2)-H$_2$), 2.00 (s, 3H, COCH$_3$), 5%; *Acetate*, **AcO**⁻ (◆): δ 1.87 (s, 3H, COCH$_3$), 7%.

Reaction of Gly-CN with AcSK at pH 9

The control experiment was performed following general procedure 2 from glycine nitrile, **Gly**-CN (50 mM), potassium thioacetate, **AcSK** (16.9 mg, 150 mM) and methylsulfonylmethane (MSM; 8.3 mM). The sample was adjusted to pH 9 and submitted periodically for analysis by 1D and 2D NMR spectroscopy to monitor the progress of the experiment.

Figure 15. ^{1}H NMR (600 MHz, H_2O/D_2O 9:1, 1.0-5.0 ppm, noseygppr1d) spectrum to show the acetylation of *glycine nitrile* (**Gly**-CN, 50 mM) with *potassium thioacetate* (**AcSK**, 150 mM) after 24 h at pH 9. See **general procedure 2** for method..

^{1}H NMR (600 MHz, H_2O/D_2O 9:1, noseygppr1d) *N-Acetyl glycine nitrile*, Ac-**Gly**-CN (▼): δ 3.96 (d, *J* = 6.1 Hz, 2H, (C2)-H$_2$), 2.00 (s, 3H, COCH$_3$), 8%; *Acetate*, **AcO**$^-$ (♦): δ 1.87 (s, 3H, COCH$_3$), 8%).

Reaction of Gly-CN with AcSK at pH 2, 3, 4 or 5

A stock solution of glycine nitrile **Gly**-CN (50 mM) with methylsulfonylmethane (MSM; 25 mM) was made up in degassed H_2O/D_2O 9:1. Then 1 mL of the stock solution was mixed with potassium thioacetate (17 mg, 0.15 mmol) and the appropriate buffering agent (0.2 mmol). The samples were adjusted to the required pH with HCl/NaOH and transferred into NMR tubes and sealed with nitrogen and parafilm. The progress of the reaction was monitored by ^{1}H NMR. The resulting reaction mixtures were analysed by 1D and 2D NMR spectroscopy (^{1}H–^{13}C HMBC, ^{1}H–^{1}H COSY, and ^{1}H–^{13}C HSQC) and the products were quantified by ^{1}H NMR spectroscopy against the internal MSM standard.

Entry	pH	Buffer	Ac-Gly-CN (%)	Ac-Gly-SNH₂ (%)	Total Acetylation (%)	Time (d)
1	2	KCl	45	4	49	7
2	3	Acetate	99	3	>99	5
3	3	Citrate	94	9	>99	2
4	4	Acetate	95	3	98	3
5	5	Acetate	83	9	92	5
6	7	Phosphate	<5	n.d	<5	7

Table 9. Reaction yields of **Gly**-CN (50 mM) with thioacetate (150 mM), at pH 2, 3, 4, 5, 7 with MSM (25 mM) added as the internal standard.

Entry	Time (h)	Total Acetylation of Gly-CN (Ac-Gly-CN + Ac-Gly-SNH₂) (%)				
		pH 2 (HCl/KCl)	pH 3 (Acetate)	pH 3 (Citrate)	pH 4 (Acetate)	pH 5 (Acetate)
1	0	0	12	8	5	0
2	1	5	24	20	12	5
3	3	9	44	26	24	7
4	6	15	55	38	40	22
5	12	22	71	49	60	27
6	24	30	84	73	81	45
7	48	36	90	98	93	68
8	72	37	93	>99	95	82
9	120	41	>99	>99	98	92
10	168	45	>99	>99	>99	92

Table 10. Yields of the acetylation of Gly-CN 7d in buffered solutions at pH 2, 3, 4 and 5. (Data plotted as a graph in the PhD upgrade report Figure 7, page 28)

*Optimisation of the Reaction of Gly-CN with AcSK*
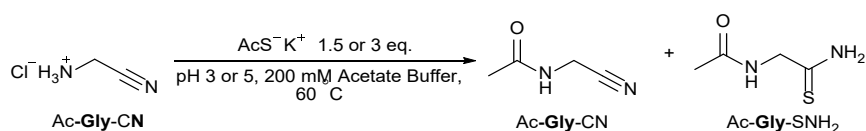
A stock solution of glycine nitrile **Gly**-CN (50 mM) with methylsulfonylmethane (MSM; 25 mM) was made up in degassed H_2O/D_2O 9:1. Then 1 mL of the stock solution was mixed with potassium thioacetate (AcSK, 17 mg, 0.15 mmol) and the appropriate buffering agent (0.2 mmol). The samples were adjusted to the required pH with HCl/NaOH and transferred into NMR tubes and sealed with nitrogen and parafilm. The NMR tubes were heated at 60 °C and the progress of the reaction was monitored by ^{1}H NMR over 24 h. The resulting reaction mixtures were analysed by 1D and 2D NMR spectroscopy (^{1}H–^{13}C HMBC, ^{1}H–^{1}H COSY, and ^{1}H–^{13}C HSQC) and the products were quantified by ^{1}H NMR spectroscopy against the internal MSM standard.

Entry	AcSK (eq.)	pH	Buffer	Ac-Gly-CN (%)	Ac-Gly-SNH$_2$ (%)	Total Acetylation (%)	Time (h)
1	1.5	3.0	Acetate	78	5	82	24
2	3	3.0	Acetate	93	9	>99	20
3	1.5	3.0	Citrate	94	7	>99	24
4	3	3.0	Citrate	91	10	>99	24
5	1.5	5.0	Acetate	51	22	73	24
6	3	5.0	Acetate	62	41	>99	24
7	1.5	7.0	Phosphate	12	5	17	24
8	3	7.0	Phosphate	14	11	25	24

Table 11. The acetylation yields of **Gly**-CN at pH 3 or 5 with 1.5 or 3 eq. of thioacetate at rt. or at 60 °C.

Reaction of Aminonitriles AA-CN with AcSK at pH 3
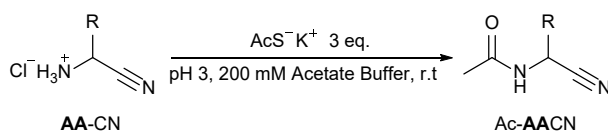
The aminonitrile substrate **AA**-CN (0.1 mmol) was weighed into a vial with methylsulfonylmethane (MSM; 4.71 mg, 0.05 mmol) and dissolved in 2 mL degassed H_2O/D_2O 9:1. Then 1 mL of the stock solution was transferred into a vial with potassium thioacetate (AcSK, 17 mg, 0.15 mmol) and sodium acetate (16.41 mg, 0.2 mmol). The samples were adjusted to the required pH with HCl/NaOH, transferred into NMR tubes and sealed with nitrogen and parafilm. The progress of the reaction was monitored by ^1H NMR. The resulting reaction mixtures were analysed by 1D and 2D NMR spectroscopy (^1H–^{13}C HMBC, ^1H–^1H COSY, and ^1H–^{13}C HSQC) and the products were quantified by ^1H NMR spectroscopy against the internal MSM standard.

Entry	AA-CN	Ac-AA-CN (%)	Time (d)
1	**Gly**-CN	>99	5
2	**Ala**-CN	68	8
3	**Met**-CN	>99	8
4	**Phe**-CN	84	8
5	**Ser**-CN	91	8
6	**Val**-CN	69	7
7	**β-Ala**-CN	44	7

Table 12. The acetylation yields of different AA-CN substrates with thioacetate (3 eq.) at rt. and pH 3..

Ligation of Ac-Gly-SH and Gly-CN to Dipeptide Ac-Gly-Gly-CN at pH 3
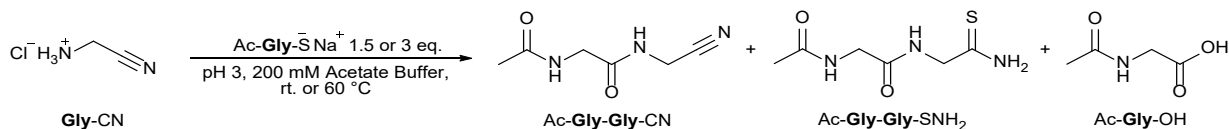
A stock solution of glycine nitrile **Gly**-CN (50 mM) with methylsulfonylmethane (MSM; 25 mM) was made up in degassed H_2O/D_2O 9:1. Then 1 mL of the stock solution was mixed with sodium 2-acetamidoethanethioate, Ac-**Gly**-SH (11.63 mg, 0.075 mmol or 23.27 mg, 0.15 mmol) and the sodium acetate (16.41 mg, 0.2 mmol). The samples were adjusted to pH 3 with HCl/NaOH and transferred into NMR tubes and sealed with nitrogen and parafilm. The NMR tubes were either left at room temperature or heated at 60 °C. The progress of the reaction was monitored by ^{1}H NMR. The resulting reaction mixtures were analysed by 1D and 2D NMR spectroscopy (^{1}H–^{13}C HMBC, ^{1}H–^{1}H COSY, and ^{1}H–^{13}C HSQC) and the products were quantified by ^{1}H NMR spectroscopy against the internal MSM standard. The identity of the products was confirmed by spiking with authentic samples of Ac-**Gly-Gly**-CN.

Entry	AcSK (eq.)	Ac-Gly-Gly-CN (%)	Ac-Gly-Gly-SNH₂ (%)	Total Dipeptide (%)	Ac-Gly-OH (%)	Temperature (°C)	Time (h)
1	1.5	35	2	37	27	rt.	5d
2	3	56	3	59	38	rt.	5d
3	1.5	62	7	69	58	60	24h
4	3	83	8	91	90	60	24h

Table 13. The dipeptide yields of the reaction of Gly-CN with either 1.5 or 3 eq. Ac-Gly-SH at rt. or heating at 60 °C, in acetate buffer at pH 3.

Ligation of Gly-CN with 1.5 equiv. Ac-Gly-SH at pH 3
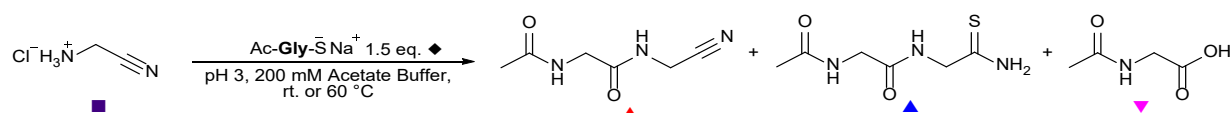
Two samples of glycine nitrile, **Gly**-CN (50 mM), sodium 2-acetamidoethanethioate, Ac-**Gly**-SH (75 mM), methylsulfonylmethane (MSM 25 mM) and sodium acetate (200 mM) were prepared in a solution of 9:1 H_2O/D_2O at pH 3. Each sample was transferred to an NMR tube and either left at room temperature or heated at 60 °C. The progress each of the reactions was monitored periodically by ^1H NMR. The sample was then submitted for analysis by 1D and 2D NMR spectroscopy. The experiments were spiked with authentic samples of Ac-**Gly**-**Gly**-CN.
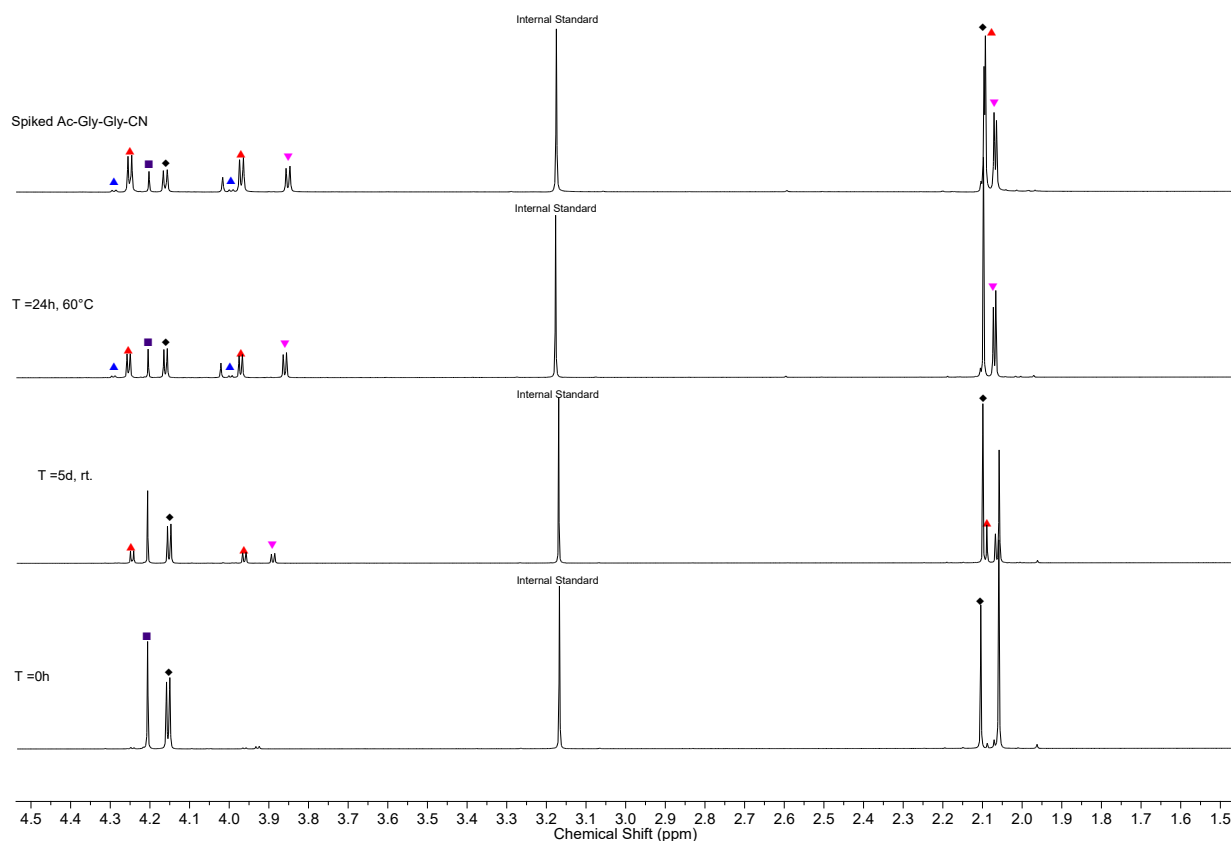
Figure 17. ^1H NMR (700 MHz, D_2O, 1.5-4.5 ppm, noseygppr1d) spectrum to show the ligation of *glycine nitrile* (**Gly**-CN, 50 mM) with *sodium 2-acetamidoethanethioate* (Ac-**Gly**-SH, 75 mM) at pH 3.

Ligation of Gly-CN with 3 equiv. Ac-Gly-SH at pH 3
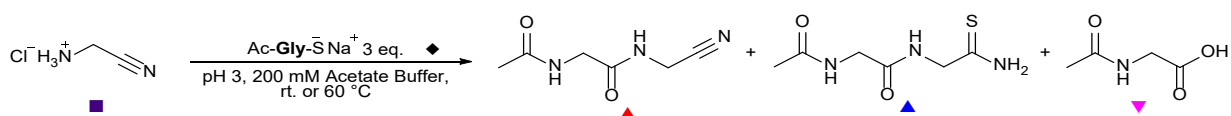
Two samples of glycine nitrile, **Gly**-CN (50 mM), sodium 2-acetamidoethanethioate, Ac-**Gly**-SH (150 mM), methylsulfonylmethane (MSM 25 mM) and sodium acetate (200 mM) were prepared in a solution of 9:1 H_2O/D_2O at pH 3. Each sample was transferred to a NMR tube and either left at room temperature or heated at 60 °C. The progress each of the reactions was monitored periodically by ^1H NMR. The sample was then submitted for analysis by 1D and 2D NMR spectroscopy. The experiments were spiked with authentic samples of Ac-**Gly**-**Gly**-CN.
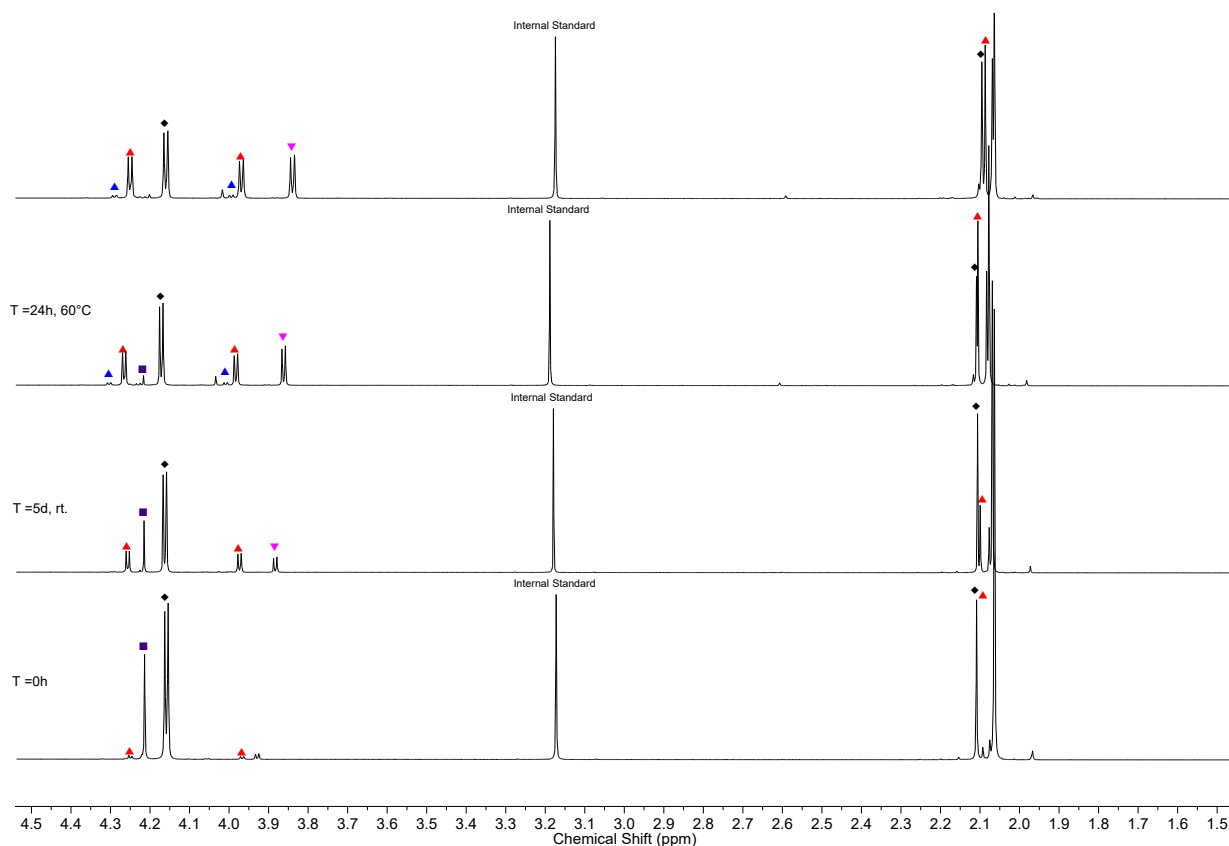
Figure 18. ^1H NMR (700 MHz, D_2O, 1.5-4.5 ppm, noseygppr1d) spectrum to show the ligation of *glycine nitrile* (**Gly**-CN, 50 mM) with *sodium 2-acetamidoethanethioate* (Ac-**Gly**-SH, 150 mM) at pH 3.

3. Preprative Syntheses

Perfluorophenyl propionate
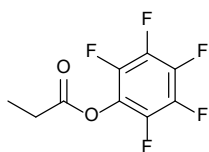
N-(3-Dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC·HCl; 7.78 g, 40.50 mmol) was added to a stirring solution of propionic acid (1.02 mL, 13.67 mmol) and pentafluorophenol (2.75 g, 14.58 mmol) in CH_2Cl_2 (10 mL) at 0 °C. The reaction mixture was left to warm to room temperature and stirred for 19 h. The reaction mixture was washed with H_2O (2 × 20 mL) and the aqueous layer was back extracted with Et_2O (3 × 10 mL). The organic extracts were combined and washed with brine (10 mL). The organic extract was dried over $MgSO_4$ and the solvent was removed *in vacuo*. The crude residue was purified with a silica plug eluting with 30% EtOAc/Petroleum Ether (40-60) to give perfluorophenyl propionate, **S1** (1.78 g, mmol, 55%) as a colourless oil. R_f = 0.61 (100% EtOAc). ^{1}H NMR (700 MHz, $CDCl_3$) δ 2.70 (q, *J* = 7.6 Hz, 2H, CH_2), 1.30 (t, *J* = 7.5 Hz, 3H, CH_3). ^{13}C NMR (176 MHz, $CDCl_3$) δ 170.4 (**C**OCH_3), 142.1 (ArF), 140.7 (ArF), 140.3 (ArF), 138.9 (ArF), 138.8 (ArF), 137.4 (ArF), 125.4 (ArF), 26.9 (CH_2). 8.8 (CH_3). ^{19}F NMR (282 MHz, $CDCl_3$) δ -153.0 (2F, m), -158.2 (1F, t, *J* = 21.7 Hz), 162.5 (2F, m). IR (cm^{-1}) 1789, 1515, 1106, 1071, 998.

Perfluorophenyl 2-cyclopropylacetate
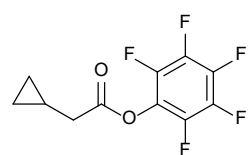
N-(3-Dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC·HCl; 1.15 g, 5.99 mmol) was added to a stirring solution of 2-cyclopropylacetic acid (186 μL, 2.00 mmol) and pentafluorophenol (1.10 g, 5.99 mmol) in CH_2Cl_2 (10 mL) at 0 °C. The reaction mixture was left to warm to room temperature and stirred for 5 h before concentrating *in vacuo*. The residue was purified by flash column chromatography (SiO2; eluting with 100% Petroleum Ether (40-60)) to give the titled compound, **S2** (440 mg, 1.65 mmol, 83%) as a clear oil. R_f = 0.52 (100% Petroleum Ether (40-60)). ^{1}H NMR (700 MHz, $CDCl_3$) δ 2.56 (d, *J* = 7.2 Hz, 2H, CH_2CO), 1.18 (m, 1H, CH), 0.66 (m, 2H, (C4)–CH_2), 0.33 (m, 2H, (C4')–CH_2). ^{13}C NMR (176 MHz, $CDCl_3$) δ 169.1 (C1), 142.0 (ArF), 140.6 (ArF), 140.3 (ArF), 138.8 (ArF), 137.3 (ArF), 125.3 (ArF), 38.6 (C2), 6.9 (C3). 4.7 (C4). ^{19}F NMR (659 MHz, $CDCl_3$) δ -152.8 (2F, m), -158.2 (1F, t, *J* = 21.7 Hz), -162.5 (2F, m). HRMS-ESI [M+H]$^+$ calcd. for $C_{11}H_7F_5O_2$ 267.0439; obs. 267.0437. IR (cm^{-1}) 1788, 1515, 1081, 994.

Perflurophenyl 2-acetamidoacetate, Ac-Gly-PFP
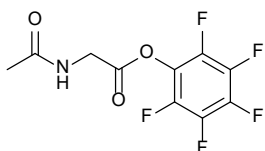
N-(3-Dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC·HCl; 2.46 g, 12.83 mmol) was added to a stirring mixture of *N*-acetyl glycine (1.01 g, 7.70 mmol) and pentafluorophenol (2.36 g, 12.82 mmol) in CH_2Cl_2 (20 mL) at 0 °C. The reaction mixture was left to warm to room temperature and stirred for 18 h. The reaction mixture was diluted with CH_2Cl_2 (20 mL) and washed with H_2O (2 x 20 mL), $NaHCO_3$ (sat., 2 x 20 mL) and brine (20 mL). The organic layer was dried over $MgSO_4$, filtered, and concentrated *in vacuo* to give the titled compound perfluorophenyl 2-acetamidoacetate, Ac-**Gly**-PFP (1.53 g, 5.40 mmol, 63%) as a white solid. R_f

= 0.36 (50% EtOAc/Petroleum Ether (40-60)). ^{1}H NMR (700 MHz, CDCl$_3$) δ 6.06 (bs, 1H, NH), 4.43 (d, J = 5.6 Hz, 2H, (C2)–CH$_2$), 2.10 (s, 3H, COCH$_3$). ^{13}C NMR (176 MHz, CDCl$_3$) δ 170.4 (**C**OCH$_3$), 166.6 (C1), 141.8 (ArF), 140.7 (ArF), 140.4 (ArF), 139.2 (ArF), 138.8 (ArF), 137.3 (ArF), 124.7 (ArF), 40.9 (C2). 23.0 (CO**C**H$_3$). ^{19}F NMR (659 MHz, CDCl$_3$) δ -152.3 (2F, m), -157.0 (1F, t, J = 21.7 Hz), -161.7 (2F, m). HRMS-ESI [M+H]$^+$ calcd. for C$_{10}$H$_6$F$_5$NO$_3$ 284.0341; obs. 284.0341. IR (cm^{-1}) 3265, 3082, 1796, 1788, 1649, 1515, 985.

Perflurophenyl 3-acetamidopropanoate, Ac-β-Ala-PFP
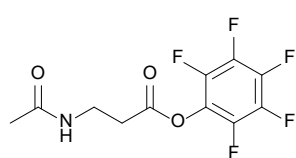
N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC·HCl; 2.11 g, 11.01 mmol) was added to a stirring mixture of 3-acetamidopropanoic acid (545 mg, 4.16 mmol) and pentafluorophenol (772 mg, 4.19 mmol) in CH$_2$Cl$_2$ (20 mL). The reaction mixture was stirred for 16 h before concentrating *in vacuo*. The residue was purified by flash column chromatography (SiO2; eluting with a gradient of EtOAc/Petroleum Ether (40-60) 0:100 → 70:30) to give the titled compound perfluorophenyl 3-acetamidopropanoate, Ac-**β-Ala**-PFP (858 mg, 2.89 mmol, 76%) as a white solid. R$_f$ = 0.27 (70% EtOAc/Petroleum Ether (40-60)). ^{1}H NMR (700 MHz, CDCl$_3$) δ 6.03 (bs, 1H, NH), 3.63 (dd, J = 6.1, 12.1 Hz, 2H, (C3)–H$_2$), 2.95 (t, J = 6.1 Hz, 2H, (C2)–H$_2$), 2.00 (s, 3H, COCH$_3$). ^{13}C NMR (176 MHz, CDCl$_3$) δ 170.7 (**C**OCH$_3$), 168.8 (C1), 141.9 (ArF), 140.5 (ArF), 140.4 (ArF), 139.2 (ArF), 138.8 (ArF), 137.3 (ArF), 124.9 (ArF), 35.0 (C3), 33.5 (C2). 23.3 (CO**C**H$_3$). ^{19}F NMR (659 MHz, CDCl$_3$) δ -152.7 (2F, m), -157.4 (1F, t, J = 21.7 Hz), -162.0 (2F, m). HRMS-ESI [M+H]$^+$ calcd. for C$_{11}$H$_8$F$_5$NO$_3$ 298.0497; obs. 298.0497. IR (cm^{-1}) 3285, 3082, 2959, 1777, 1644, 1548, 1005.

Perflurophenyl (*tert*-butoxycarbonyl)glycinate, Boc-Gly-PFP
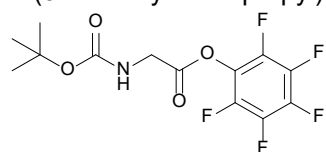
N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC·HCl; 3.31 g, 17.27 mmol) was added to a stirring mixture of *tert*-butoxycarbonyl glycine (1.03 g, 5.91 mmol) and pentafluorophenol (3.18 g, 17.12 mmol) in CH$_2$Cl$_2$ (20 mL). The reaction mixture was stirred for 1 h. The reaction mixture was diluted with CH$_2$Cl$_2$ (20 mL) and washed with H$_2$O (3 x 40 mL), NaHCO$_3$ (sat., 2 x 20 mL) and brine (20 mL). The organic extract was dried over MgSO$_4$, filtered, and concentrated *in vacuo*. The residue was purified by flash column chromatography (SiO2; eluting with a gradient of EtOAc/Petroleum Ether (40-60) 0:100 → 0:20) to give the titled compound perfluorophenyl (*tert*-butoxycarbonyl)glycinate, Boc-**Gly**-PFP (1.59 g, 4.66 mmol, 82%) as a white crystaline solid. ^{1}H NMR (700 MHz, CDCl$_3$) δ 5.08 (bs, 1H, NH), 4.29 (d, J = 5.8 Hz, 2H, (C2)–H$_2$), 1.47 (s, 9H, C(CH$_3$)$_3$). ^{19}F NMR (659 MHz, CDCl$_3$) δ -157.3 (2F, m), -157.3 (1F, t, J = 21.7 Hz), -161.9 (2F, m). IR (cm^{-1}) 3393, 2986, 1778, 1716, 1518.
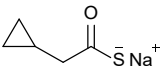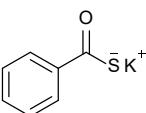
Sodium propanethioate
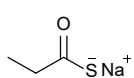
Finely ground sodium sulfide hydrate (Na$_2$S·xH$_2$O; 74 mg, 0.95 mmol) was added to a solution of perfluorophenyl propionate (206 mg, 0.86 mmol) in anhydrous MeCN (4 mL) at room temperature. The suspension was stirred under an argon

atmosphere for 1 h. The resultant precipitate was isolated by centrifugation and dried *in vacuo* to afford sodium propanethioate, **S3** (85 mg) as a white solid which used without further purification. ^{1}H NMR (700 MHz, D$_2$O, noseygpprd1) δ 2.64 (q, *J* = 7.5 Hz, 2H, CH$_2$), 1.30 (t, *J* = 7.5 Hz, 3H, CH$_3$). ^{13}C NMR (176 MHz, CDCl$_3$) δ 227.3 (COS), 44.6 (CH$_2$), 12.0 (CH$_3$). IR (cm[-1]) 3413, 3285, 3229, 234, 1619, 1564, 1502.
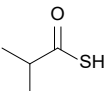
Sodium 2-cyclopropylethanethioate

Finely ground sodium hydrosulfide hydrate (NaSH·xH$_2$O; 296 mg, 5.28 mmol) was added to a solution of perfluorophenyl 2-cyclopropylacetate (151 mg, 0.57 mmol) in anhydrous MeCN (5 mL) at room temperature. The suspension was stirred under an argon atmosphere for 3 h. The solution was filtered, and the solvent was removed *in vacuo* to afford sodium 2-cyclopropylethanethioate, **S4** (177 mg, Purity 82%) as a white solid which used without further purification. ^{1}H NMR (700 MHz, D$_2$O, noseygppr1d) δ 2.57 (d, *J* = 7.1 Hz, 2H, CH$_2$CO), 1.03 (m, 1H, CH), 0.49 (m, 2H, CHCH$_2$), 0.15 (m, 2H, CHCH$_2$'). ^{13}C NMR (176 MHz, D$_2$O) δ 225.5 (COS), 55.9 (**C**H$_2$CO), 9.2 (CH). 4.1 (**C**H$_2$CH).
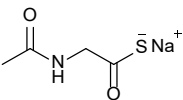
Potassium benzothioate

Thiobenzoic acid (1.00 g, 0.85 mL, 7.24 mmol) was dissolved in a solution of KOH (370 mg, 6.52 mmol), MeOH (7.5 mL) and EtOH (7.5 mL). The mixture was heated at 50 ºC and left to stir for 2.5 h. The solvent was removed *in vacuo* and the solid was washed with Et$_2$O (30 mL) and dried to give potassium benzothioate, **S5** (1.03 g, 5.84 mmol, 81%, lit.[4] 99%) as a yellow solid. ^{1}H NMR (700 MHz, D$_2$O) δ 8.02 - 7.98 (m, 2H, Ar), 7,57 – 7.53 (m, 1H, Ar), 7.47 – 7.43 (m, 2H, Ar). ^{13}C NMR (176 MHz, D$_2$O) δ 216.4 (COS), 144.2 (Ar), 132.0 (Ar), 128.6 (Ar), 128.2 (Ar).

2-Methylpropane thioacid
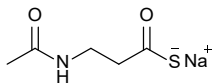
To a solution of sodium hydrosulfide (483 mg, 8.62 mmol) in anhydrous MeOH (5 mL) was added *iso*-butyryl chloride (396 mg, 0.39 mL, 3.72 mmol) and stirred at 0 °C for 1 h. The solution was left to stir at room temperature for a further 2 h. The reaction mixture was centrifuged, and the supernatant decanted off. The solvent was removed *in vacuo* to give 2-methylpropane thioacid, **S6** (443 mg) which was used without further purification. ^{1}H NMR (700 MHz, CDCl$_2$) δ 2.97 (spt, *J* = 6.9 Hz, 1H, CH), 1.29 (d, *J* = 6.9 Hz, 6H, CH(CH$_3$)$_2$). ^{13}C NMR (176 MHz, CDCl$_2$) δ 197.8 (COS), 42.8 (CH), 19.4 (CH(CH$_3$)$_2$). LRMS-ESI [M-H][-] calcd. for C$_4$H$_7$OS 103.02: obs. 103.02. HRMS-ESI [M-H][-] calcd. for C$_4$H$_7$OS 103.0212; obs. 103.0179. IR (cm[-1]) 3073, 2974, 2933, 1720, 1520, 1455, 1094, 1069

Sodium 2-acetamidoethanethioate, Ac-Gly-S[-]Na[+]

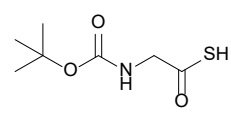Finely ground sodium hydrosulfide hydrate (NaSH·xH$_2$O; 404 mg, 7.21 mmol) was added to a solution of *N*-acetylglycine pentafluorophenyl ester (1.00 g, 3.53 mmol) in anhydrous MeCN (35 mL). The solution was stirred at room temperature under Argon for 1 h. The resulting precipitate was isolated by centrifugation, washed

with MeCN (10 mL), then triturated with Et₂O (10 mL). The solid was dried *in vacuo* to afford sodium 2-acetamidoethanethioate, Ac-**Gly**-S⁻Na⁺ (554 mg) as a white solid which was used without further purification. ¹H NMR (700 MHz, D₂O, noseygppr1d) δ 4.14 (s, 2H, (C2)–H₂), 2.05 (s, 3H, COCH₃). ¹³C NMR (176 MHz, D₂O) δ 216.8 (C1), 174.5 (**C**OCH₃), 54.1 (C2), 22.4 (CO**C**H₃). HRMS-ESI [M+H]⁺ calcd. for $C_4H_7NO_2S$ 134.0270; obs. 134.0272. IR (cm⁻¹) 3256, 2987, 1629, 1598, 1520, 1301, 1090.
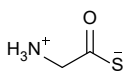
Sodium 3-acetamidopropanethioate, Ac-β-Ala-S⁻Na⁺ (JF-04-230/JF-05-329)

Finely ground sodium hydrosulfide hydrate (NaSH·xH₂O; 97 mg, 1.73 mmol) was added to a solution of pentafluorophenyl 3-acetamidopropanoate (258 mg, 0.87 mmol) in anhydrous MeCN (8 mL). The solution was stirred at room temperature under Argon for 1 h. The resulting precipitate was isolated by centrifugation, washed with MeCN (20 mL), then triturated with Et₂O (20 mL). The solid was dried *in vacuo* to afford sodium 3-acetamidopropanethioate, Ac-**β-Ala**-S⁻Na⁺ (148 mg) as a white solid which was used without further purification. ¹H NMR (700 MHz, D₂O) δ 3.43 (t, *J* = 6.7 Hz, 2H, (C3)–H₂), 2.85 (t, *J* = 6.7 Hz, 2H, (C2)–H₂), 1.98 (s, 3H, COCH₃). ¹³C NMR (176 MHz, D₂O) δ 221.9 (C1), 174.5 (**C**OCH₃), 49.7 (C2), 38.0 (C3), 22.4 (CO**C**H₃). HRMS-ESI [M+H]⁺ calcd. for $C_5H_9NO_2S$ 148.0427; obs. 148.0427. IR (cm⁻¹) 3341, 3238, 3080, 1631 1514, 1018.

2-((*tert*-Butoxycarbonyl)amino)ethanethioic *S*-acid, Boc-Gly-SH (JF-05-351)

Finely ground sodium hydrosulfide hydrate (NaSH·xH₂O; 146 mg, 2.60 mmol) was added to a solution of perfluorophenyl (*tert*-butoxycarbonyl)glycinate (452 mg, 1.32 mmol) in anhydrous MeCN (15 mL). The solution was stirred at room temperature under Argon for 3 h. The solvent was removed *in vacuo* and the residue was triturated with Et₂O. The residue was dissolved in EtOAc (20 mL) and cooled to 0 °C. To the solution was added dropwise, chilled citric acid solution (982 mg in 50 mL) until the solution reached pH 3. The organic layer was partitioned, and the aqueous layer was extracted with EtOAc (2 x 25 mL). The organic extracts were combined and washed with brine (20 mL), dried over MgSO₄, filtered, and concentrated *in vacuo* to give the titled compound 2-((*tert*-butoxycarbonyl)amino)ethanethioic *S*-acid, Boc-**Gly**-SH (171 mg, 0.89 mmol, 68%) as a pale yellow oil which was taken to the next step without purification. ¹H NMR (600 MHz, CDCl₃) δ 5.14 (bs, 1H, NH), 4.04 (d, *J* = 5.7 Hz, 2H, (C2)–H₂), 1.47 (s, 9H, C(CH₃)₃). IR (cm⁻¹) 3332, 2978, 2933, 1686, 1510,1159.

Glycine thioacid, Gly-SH

A solution of butoxycarbonyl)amino)ethanethioic *S*-acid (160 mg, 0.84 mmol) was cooled to 0 °C, then chilled TFA (5 mL) was added. The solution was stirred at 0 °C for 30 min and then stirred for a further 1.5 h at room temperature. The solvent was removed *in vacuo* and the residue resuspended with Et₂O (20 mL). The resulting precipitate was dried *in vacuo* to afford glycine thioacid, **Gly**-SH (72 mg, 0.79 mmol, 95%) as a off white solid which was

used without further purification. ^{1}H NMR (700 MHz, D$_2$O, noseygppr1d) δ 3.92 (s, 2H, (C2)–H$_2$). ^{13}C NMR (176 MHz, D$_2$O) δ 210.9 (C1), 51.1 (C2). IR (cm^{-1}) 3362, 3069, 2946, 1685, 1518.

N-Acetyl glycine nitrile, Ac-Gly-CN
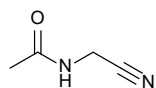
Aminoacetonitrile hydrochloride (927 mg, 10.02 mmol) was dissolved in degassed H$_2$O (50 mL) and the solution was adjusted to pH 9.0 with 6 M NaOH. Acetic anhydride (3.08 g, 2.85 mL, 30.15 mmol) was added and the solution was readjusted to pH 9.0 with 6 M NaOH. The solution was left to stir for 1 h, after which a ^{1}H NMR showed complete conversion to the desired product. The reaction mixture was concentrated *in vacuo* and the residue was suspended in EtOAc. The undissolved white solid was filtered off and washed with EtOAc. The organic extracts were combined, and the solvent was removed under reduced pressure to afford *N*-acetylglycine nitrile, Ac-**Gly**-CN (879 mg, 9.05 mmol, 90%, (lit.[1] 99%) as a white crystalline solid. ^{1}H NMR (700 MHz, D$_2$O, noseygppr1d) δ 4.20 (s, 2H, (C2)–H$_2$), 2.07 (s, 3H, COCH$_3$). ^{13}C NMR (176 MHz, D$_2$O) δ 175.2 (**C**OCH$_3$), 117.6 (C1), 28.1 (C2), 22.0 (CO**C**H$_3$). HRMS-ESI [M+H]$^+$ calcd. for C$_4$H$_6$N$_2$O 99.0553; obs. 99.0556. IR (cm^{-1}) 3237, 3200, 3060, 2250, 1643, 1553, 1298, 1279.

N-(2-Cyanoethyl) acetamide, Ac-β-Ala-CN
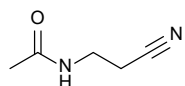
3-Aminopropanenitrile (1.01 g, 1.05 mL, 14.27 mmol) was dissolved in degassed H$_2$O (5 mL, 2.9 M) and the solution was adjusted to pH 9.0 with 4 M NaOH. Acetic anhydride (4.40 g, 4.20 mL, 43.10 mmol) was added dropwise, whilst maintaining the solution at pH 9.0 with 4 M NaOH. The solution was left to stir for 1 h under an argon atmosphere, after which a ^{1}H NMR showed quantitative conversion to the desired product. After this time the reaction mixture was adjusted to pH 4 with 4 M HCl and the solution was concentrated *in vacuo*. The resulting solid was recrystallised from EtOAc to afford *N*-(2-cyanoethyl) acetamide Ac-**β-Ala**-CN (1.23 g, 10.97 mmol, 77%, lit.[1] 83%) as clear crystals. ^{1}H NMR (700 MHz, CDCl$_3$) δ 6.07 (bs, 1H, NH), 3.50 (t, *J* = 6.2 Hz, 2H, (C3)–H$_2$), 2.63 (t, *J* = 6.2 Hz, 2H, (C2)–H$_2$), 2.02 (s, 3H, COCH$_3$). ^{13}C NMR (176 MHz, CDCl$_3$) δ 170.8 (**C**OCH$_3$), 118.4 (C1), 35.9 (C3), 23.2 (C2), 18.6 (CO**C**H$_3$). LRMS-ESI [M+H]$^+$ calcd. for C$_5$H$_9$N$_2$O 113.07: obs. 113.07. HRMS-ESI [M+H]$^+$ calcd. for C$_5$H$_9$N$_2$O 113.0719; obs. 113.0715. IR (cm^{-1}) 3255, 3073, 2243, 1653, 1557, 1361, 1294.

2-Acetamido-*N*-(cyanomethyl) acetamide, Ac-Gly$_1$-Gly$_2$-CN (JF-03-262)
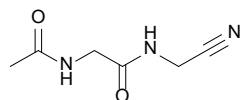
To a solution of aminoacetonitrile hydrochloride (279 mg, 3.02 mmol) in CH$_2$Cl$_2$ (25 mL) was added triethylamine (609 mg, 840 µL, 6.03 mmol). The solution was stirred, then perfluorophenyl 2-acetamidoacetate (711 mg, 2.51 mmol) was added. The reaction mixture was left to stir for 16 h, after which the solvent was removed *in vacuo*. The crude product was triturated with Et$_2$O (2 x 20 mL), then purified by flash column chromatography (SiO2; with a gradient of MeOH/CH$_2$Cl$_2$ 0:100 → 10:90) to give the titled compound, 2-acetamido-*N*-(cyanomethyl) acetamide (95 mg, 0.61 mmol, 25%) as a white solid. R$_f$ = 0.30 (100% MeOH/CH$_2$Cl$_2$). ^{1}H NMR (700 MHz, D$_2$O, noseygppr1d) δ 4.25 (s, 2H, Gly$_1$-

(C2)–H$_2$), 3.96 (s, 2H, Gly$_2$-(C2)–H$_2$), 2.09 (s, 3H, COCH$_3$). ^{13}C NMR (176 MHz, D$_2$O) δ 175.6 (**C**OCH$_3$), 172.7 Gly$_1$-(C1), 117.4 Gly$_2$-(C1), 43.0 Gly$_1$-(C2), 28.0 Gly$_2$-(C2), 22.2 (CO**C**H$_3$). HRMS-ESI [M+H]$^+$ calcd. for C$_6$H$_9$N$_3$O$_2$ 156.0768; obs. 156.0769.

3-Acetamidopropionic acid, Ac-β-Ala-OH
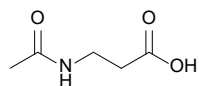
To a solution of 3-Aminopropanoic acid (510 mg, 5.72 mmol) in H$_2$O (7 mL) was added acetic anhydride (3.46 g, 3.2 mL, 33.85 mmol). The solution was sonicated for 15 min, after which ^{1}H NMR showed quantitative conversion to the desired product. The solution was concentrated *in vacuo* to afford 3-acetamidopropanoic acid, Ac-**β-Ala**-CN (785 mg, 5.99 mmol, > 99%) as a clear oil. ^{1}H NMR (700 MHz, D$_2$O) δ 3.44 (t, *J* = 6.5, 2H, (C3)–H$_2$), 4.20 (t, *J* = 6.5, 2H, (C2)–H$_2$), 1.98 (s, 3H, COCH$_3$). ^{13}C NMR (176 MHz, D$_2$O) δ 176.8 (C1), 174.7 (**C**OCH$_3$), 35.7 (C3), 34.0 (C2), 22.3 (CO**C**H$_3$). HRMS-ESI [M+H]$^+$ calcd. for C$_5$H$_{10}$NO$_3$ 132.0655; obs. 132.0657. IR (cm^{-1}) 3333, 2918, 2592, 1706, 1615, 1542, 1194.

N-Acetyl-*S*-methyl cysteine, Ac-Cys^{Me}-OH
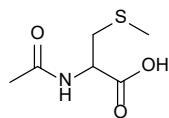
To a solution of *S*-methyl cysteine (252 mg, 1.86 mmol) in H$_2$O (3 mL) was added acetic anhydride (1.51 g, 1.51 mL, 14.81 mmol). The solution was sonicated for 15 min, after which ^{1}H NMR showed quantitative conversion to the desired product. The reaction mixture was concentrated *in vacuo* to give the desired product *N*-acetyl-*S*-methyl cysteine, Ac-**Cys^{Me}**-OH (367 mg, 2.07 mmol, >99%) as a white solid. ^{1}H NMR (700 MHz, D$_2$O, noseygppr1d) δ 4.60 (dd, *J* = 4.8, 8.2 Hz, 1H, (C2)–H), 3.07 (ABX, *J* = 4.7, 14.1 Hz, 1H, (C3)–H), 2.93 (ABX, *J* = 8.3, 14.1 Hz, 1H, (C3')–H), 2.16 (s, 3H, SCH$_3$) 2.08 (s, 3H, COCH$_3$). ^{13}C NMR (176 MHz, D$_2$O) δ 175.1 (**C**OCH$_3$), 174.9 (C1), 52.9 (C2), 35.3 (C3), 22.2 (CO**C**H$_3$), 15.3 (SCH$_3$). HRMS-ESI [M+H]$^+$ calcd. for C$_6$H$_{11}$NO$_3$S 178.0532; obs. 178.0531. IR (cm^{-1}) 3369, 3318, 2916, 2490, 1698, 1609, 1536, 1421.

N-Methylacetamide, AcNHMe

Methylamine hydrochloride (255 mg, 3.77 mmol) was dissolved in H$_2$O (6 mL) and the acetic anhydride (2.38 g, 2.2 mL, 23.27 mmol) was added. The solution was adjusted to pH 9.0 with 6 M NaOH and left to stir for 1 h, after which ^{1}H NMR showed complete conversion to the desired product. The reaction mixture was concentrated *in vacuo* and the residue was extracted with EtOAc. The undissolved solid was filtered and washed with EtOAc. The organic extracts were combined, and the solvent was removed under reduced pressure to afford *N*-methylacetamide, AcNHMe (195 mg, 2.67 mmol, 72%) as a clear oil. ^{1}H NMR (700 MHz, CDCl$_3$) δ 5.61 (bs, 1H, NH), 2.79 (d, *J* = 4.9 Hz, 3H, CH$_3$), 2.97 (s, 3H, COCH$_3$). ^{13}C NMR (176 MHz, CDCl$_3$) δ 171.0 (**C**OCH$_3$), 26.6 (CH$_3$), 23.3 (CO**C**H$_3$). IR (cm^{-1}) 3360, 3079, 2919, 2851, 1629, 1556, 1529.

N-Phenylethylacetamide, AcNHCH$_2$CH$_2$Ph

2-Phenylethylamine (250 mg, 260 µL, 2.06 mmol) was dissolved in H_2O (6 mL) and the acetic anhydride (1.30 g, 1.2 mL, 12.69 mmol) was added. The solution was adjusted to pH 9.0 with 6 M NaOH and left to stir for 1 h, after which ^1H NMR showed complete conversion to the desired product. The reaction mixture was concentrated *in vacuo* and the residue was extracted with EtOAc. The undissolved solid was filtered and washed with EtOAc. The organic extracts were combined, and the solvent was removed under reduced pressure to afford 2-phenylethylacetamide, $AcNHCH_2CH_2Ph$ (320 mg, 1.96 mmol, 95%) as a white crystalline solid. ^1H NMR (700 MHz, $CDCl_3$) δ 7.33 – 7.30 (m, 2H, Ar), 7.25 – 7.24 (m, 1H, Ar), 7.21 – 7.17 (m, 2H, Ar), 5.46 (bs, 1H, NH), 3.52 (dd, J = 6.9, 12.9 Hz, 2H, ($NHCH_2$), 2.82 (t, J = 6.9 Hz, 2H, (CH_2), 1.94 (s, 3H, $COCH_3$). ^{13}C NMR (176 MHz, $CDCl_3$) δ 170.1 (**C**OCH_3), 139.0 (Ar), 128.9 (Ar × 2), 128.8 (Ar × 2), 126.7 (Ar), 40.8 ($NHCH_2$), 35.8 (CH_2), 23.5 (CO**C**H_3). IR (cm^{-1}) 3282, 3066, 3028, 2930, 2871, 1638, 1538.

Acetyl-glyclglycine, Ac-Gly-Gly-OH

To a solution of glycylglycine (502 mg, 3.80 mmol) in H_2O (8 mL) was added acetic anhydride (2.7 g, 2.5 mL, 26.45 mmol). The solution was sonicated for 15 min, after which ^1H NMR showed quantitative conversion to the desired product. The reaction mixture was concentrated *in vacuo* to give the desired product acetyl-glycylglycine, Ac-**Gly-Gly**-OH (646 mg, 3.71 mmol, 98%) as a white solid. ^1H NMR (700 MHz, D_2O, noseygppr1d) δ 4.02 (s, 2H, Gly_1-(C2)–H_2), 3.97 (s, 2H, Gly_2-(C2)–H_2), 2.07 (s, 3H, $COCH_3$). ^{13}C NMR (176 MHz, D_2O) δ 175.5 (**C**OCH_3), 173.9 Gly_2-(C1), 172.7 Gly_1-(C1), 42.9 Gly_2-(C2), 41.6 Gly_1-(C2), 22.3 (CO**C**H_3). HRMS-ESI [M+H]$^+$ calcd. for $C_6H_{10}N_2O_4$ 175.0713; obs. 175.0714. IR (cm^{-1}) 3336, 3303, 3088, 2525, 1704, 1659, 1611, 1548, 1414, 1377, 1231.

## *N*-Acetyl-glycine thioamide, Ac-Gly-SNH$_2$ (JF-05-343)

N-acetylglycine nitrile (254 mg, 2.90 mmol) was dissolved in BBS (25 mL, 1 M). The finely ground sodium hydrosulfide hydrate (NaSH·xH_2O; 1.43 g, 25.51 mmol) was added to the stirring solution and the pH was adjusted to pH 9 with 4 M NaOH. The solution was left to stir for 24 h under an argon atmosphere, after which ^1H NMR showed >96% conversion to the desired product. The pH was adjusted to pH 7 and the solution was purged of H_2S by bubbling nitrogen through the solution into bleach traps. The solvent was removed *in vacuo* and the crude product was extracted from the solids with EtOAc. The organic extracts were filtered, and the solvent removed *in vacuo*. The residue was purified by flash column chromatography (SiO2; eluting with a gradient of MeOH/CH_2Cl 0:100 → 10:90) to give the titled compound *N*-Acetyl-glycine thioamide, Ac-**Gly**-SNH$_2$ (181 mg, 1.37 mmol, 54%) as a yellow crystalline solid. ^1H NMR (600 MHz, D_2O, noseygppr1d) δ 4.16 (s, 2H, (C2)–H_2), 2.02 (s, 3H, $COCH_3$). ^{13}C NMR (150 MHz, D_2O) δ 204.9 (C1), 175.6 (**C**OCH_3), 50.0 (C2), 22.5 (CO**C**H_3). IR (cm^{-1}) 3342, 3287, 3081, 1640, 1525.

Serine Nitrile, Ser-CN
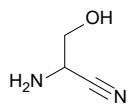
Glycolaldehyde (1.20 g, 20.00 mmol), sodium cyanide (1.08 g, 22.06 mmol) were dissolved in H_2O/D_2O (9:1, 15 mL). Ammonium chloride (5.40 g, 100.95 mmol) was added and the solution adjusted to pH 10 with 6M NaOH. The reaction was monitored by ^1H NMR over 24 h and >90% conversion to Ser-CN was observed. The solvent was removed *in vacuo* and Ser-CN was extracted with the minimum amount of MeOH. The organic extract was concentrated *in vacuo* and purified by flash column chromatography (SiO2; eluting with a gradient of MeOH/EtOAc 0:100 → 20:80) to give the titled compound serine nitrile, **Ser**-CN (732 mg, 8.50 mmol, 43%) as a yellow solid. ^1H NMR (700 MHz, D_2O) δ 3.82 (dd, *J* = 4.7, 5.8 Hz, 1H, (C2)–H), 3.82 (ABX, *J* = 4.7, 11.5 Hz, 1H, (C3)–H), 3.74 (ABX, *J* = 5.8, 11.5 Hz, 1H, (C3')–H). ^{13}C NMR (176 MHz, D_2O) δ 121.8 (C1), 63.2 (C3), 45.2 (C2). HRMS-ESI [M+H]$^+$ calcd. for $C_3H_6N_2O$ 87.0553; obs. 87.0556. IR (cm^{-1}) 3332, 3282, 2839, 2233, 1607, 1037.

General Procedure for the Syntheses of α-Acetoamidothioacids
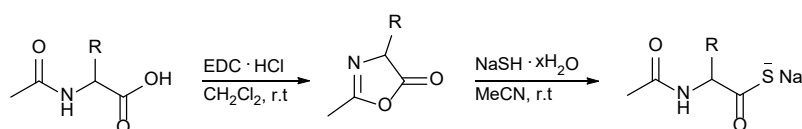
General procedure A for syntheses of 2-methyloxazol-5(*4H*)-ones

N-(3-Dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC·HCl; 550 mg, 2.87 mmol) was added to a suspension of *N*-acetyl amino acid (Ac-**AA**-OH; 2.11 mmol) in CH_2Cl_2 (20 mL) at 0 °C and then stirred for 1 h. The resultant homogenous solution was diluted with CH_2Cl_2 (20 mL) and washed with water (2 x 20 mL), $NaHCO_3$ (sat., 2 x 20 mL) and brine (20 mL). The organic layer was dried over $MgSO_4$, filtered, and concentrated *in vacuo* to give 2-methyloxazol-5(*4H*)-one, **S7-AA** as an oil, which was used immediately without further purification.[1]

General procedure B for syntheses of a-acetamidothioacids from 2-methyl-5(*4H*)-oxazolones

Finely ground sodium hydrosulfide hydrate ($NaSH×·H_2O$, 2 equiv.) was added to a solution of 2-methyl-5(*4H*)-oxazolone, **S7-AA** (1 equiv.) in anhydrous MeCN (0.1 M) at room temperature. The suspension was stirred under an argon atmosphere and the resultant precipitate was isolated by centrifugation and dried in vacuo to afford Ac-**AA**-S⁻Na⁺ as a white solid which was used without further purification.[1]
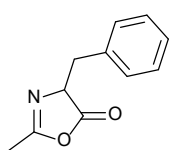
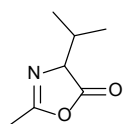## 2,4-Dimethyloxazol-5(4H)-one
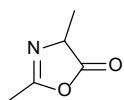
Prepared by following general procedure A from *N*-acetyl-*L*-alanine, Ac-**Ala**-OH (279 mg, 2.12 mmol). The alanine 2,4-dimethyloxazol-5(*4H*)-one, **S7-Ala** was isolated by careful concentration under reduced pressure (300 mbar, r.t) to give a colourless oil (607 mg, 89% by MSM Internal Standard). Due to its known volatility and instability, **S7-Ala** was taken onto the next step without further purification. ^{1}H NMR (600 MHz, CDCl$_3$) δ δ 4.19 (qq, *J* = 2.0, 7.6 Hz, 2H, (C4)–H), 2.21 (d, *J* = 2.0 Hz, 3H, (C2)–CH$_3$), 1.47 (d, *J* = 7.6 Hz, 3H, (C4)–CH$_3$). ^{13}C NMR (151 MHz, CDCl$_3$) δ 179.5 (C5), 162.9 (C2), 60.7 (C4), 16.6 (C1'), 15.5 (CH$_3$). LRMS-ESI calcd. for C$_5$H$_8$NO$_2$ 114.06; obs. 114.05. HRMS-ESI [M+H]$^+$ calcd. for C$_5$H$_8$NO$_2$ 114.0555; obs. 114.0561.
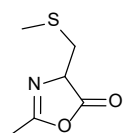
4-Benzyl-2-methyloxazol-5(4H)-one

Prepared by following general procedure A from *N*-acetyl-*L*-phenylalanine, Ac-**Phe**-OH (254 mg, 1.21 mmol). The phenylalanine 2-methyloxazol-5(*4H*)-one, **S7-Phe** was isolated as a colourless oil (196 mg, 1.04 mmol, 86%, lit.[1] 99%). ^{1}H NMR (700 MHz, CDCl$_3$) δ 7.31 – 7.27 (m, 2H, Ar), 7.26 – 7.23 (m, 1H, Ar), 7.22 – 7.20 (m, 2H, Ar), 4.44 (ddd, *J* = 2.1, 4.6, 6.8 Hz, 1H, (C4)–H), 3.25 (ABX, *J* = 4.8, 14.0 Hz, 1H, (C1')–H), 3.06 (ABX, J = 6.8, 14.0 Hz, 1H, (C1'')–H), 2.09 (d, *J* = 1.9 Hz, 3H, (C2)–CH$_3$). ^{13}C NMR (176 MHz, CDCl$_3$) δ 178.1 (C5), 163.1 (C2), 135.4 (Ar), 129.6 (Ar × 2), 128.6 (Ar × 2), 127.4 (Ar), 66.2 (C4), 37.1 (C1'), 15.2 (CH$_3$). HRMS-ESI [M+H]$^+$ calcd. for C$_{11}$H$_{12}$NO$_2$ 190.0863; obs. 190.0863.

## 4-Isopropyl-2-methyloxazol-5(4H)-one

Prepared by following general procedure A from *N*-acetyl-*L*-valine, Ac-**Val**-OH (336 mg, 2.11 mmol). The valine 2-methyloxazol-5(*4H*)-one, **S7-Val** was isolated as a colourless oil (177 mg, 1.25 mmol, 59%, lit.[1] 85%). ^{1}H NMR (600 MHz, CDCl$_3$) δ 4.03 (dq, *J* = 2.1, 4.2 Hz, 1H, (C4)–H), 2.29 - 2.23 (m, 1H, (C1')–H), 2.22 (d, *J* = 2.1 Hz, 3H, (C2)–CH$_3$), 1.08 (d, *J* = 6.9 Hz, 3H, (C2')–H), 0.95 (d, *J* = 6.9 Hz, 3H, (C2'')–H$_3$). ^{13}C NMR (151 MHz, CDCl$_3$) δ 178.2 (C5), 163.1 (C2), 70.3 (C4), 30.7 (C1'), 18.9 (C2'), 17.4 (C2''), 15.3 (CH$_3$).
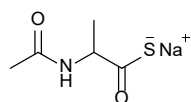
2-Methyl-4-((methylthio)methyl)oxazol-5(4H)-one

Prepared by following general procedure A from *N*-acetyl-*S*-methyl cysteine, Ac-**Cys^{Me}**-OH (374 mg, 2.11 mmol). The *S*-methyl cysteine 2-methyloxazol-5(*4H*)-one, **S7-Cys^{Me}** was isolated as a colourless oil (329 mg, 0.85 mmol, 98%). ^{1}H NMR (600 MHz, CDCl$_3$) δ 4.45 (ddd, *J* = 2.1, 4.2, 5.7 Hz, 1H, (C4)–H), 3.08 (ABX, *J* = 4.2, 14.2 Hz, 1H, (C1')–H), 2.90 (ABX, *J* = 5.6, 14.2 Hz, 1H, (C1'')–H), 2.25 (d, *J* = 2.1 Hz, 3H, (C2)–CH$_3$), 2.18 (s, 3H, SCH$_3$). ^{13}C NMR (151 MHz, CDCl$_3$) δ 177.4 (C5), 164.2 (C2), 66.2 (C4), 35.0 (C1'), 17.5 (SCH$_3$) 15.4 (CH$_3$).

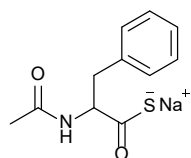### 2-Methyl-4-((2-(methylthio)ethyl)oxazol-5(4H)-one
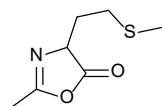
Prepared by following general procedure A from *N*-acetyl-methionine, Ac-**Met**-OH (409 mg, 2.14 mmol). The methionine-methyloxazol-5(*4H*)-one, **S11-Met** was isolated as a colourless oil (335 mg, 1.93 mmol, 92%). ^{1}H NMR (700 MHz, CDCl$_3$) δ 4.34 (ddq, *J* = 2.1, 6.1, 7.8 Hz, 1H (C4)–H), 2.67 (dd, *J* = 1.1, 7.1 Hz, 2H, (C2')–H$_2$), 2.21 (d, *J* = 2.1 Hz, 3H, (C2)–CH$_3$), 2.21 – 2.16 (m, 1H, (C1')–H), 2.10 (s, 3H, SCH$_3$), 2.03 – 1.97 (m, 1H, (C1'')–H). ^{13}C NMR (176 MHz, CDCl$_3$) δ 178.8 (C5), 163.5 (C2), 63.3 (C4), 30.2 (C2'), 30.1 (C1'), 15.4 (CH$_3$), 15.2 (SCH$_3$).

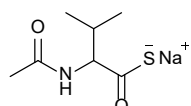### Sodium 2-acetamidopropanethioate, Ac-Ala-S⁻Na⁺

Prepared following a modification of general procedure B. Finely ground sodium hydrosulfide hydrate (NaSH·xH$_2$O; 153 mg, 2.72 mmol) was added to a solution of 2,4-dimethyloxazol-5-*4H*-one, **S7-Ala** (153 mg, 1.35 mmol) in anhydrous MeCN (5 mL) at room temperature. The suspension was stirred under an argon atmosphere for 16 h. The resulting solution was filtered and then concentrated *in vacuo* to afford sodium 2-acetamidopropanethioate Ac-**Ala**-S⁻Na⁺ as a white hydroscopic foam (221 mg) that was used without further purification. ^{1}H NMR (700 MHz, D$_2$O) δ 4.46 (q, *J* = 7.2 Hz, 1H, (C2)–H), 2.02 (s, 3H, COCH$_3$), 1.38 (d, *J* = 7.2 Hz, 3H, (C3)–H$_3$). ^{13}C NMR (176 MHz, D$_2$O) δ 222.5 (C1), 173.7 (**C**OCH$_3$), 60.5 (C2), 22.4 (CO**C**H$_3$), 19.5 (C3). LRMS-ESI+ calcd. for C$_5$H$_9$NO$_2$SNa 170.03; obs. 170.02. HRMS-ESI [M+H]⁺ calcd. for C$_5$H$_9$NO$_2$SNa 170.0252; obs. 170.0267.

Sodium 2-acetamido-3-phenylpropanethioate, Ac-Phe-S⁻Na⁺

Prepared following general procedure B from 4-benzyl-2-methyloxazol-5(4H)-one, **S8-Phe** (153 mg, 0.81 mmol). Sodium 2-acetamido-3-phenylpropanethioate Ac-**Phe**-S⁻Na⁺ was isolated after 1 h as a white solid (204 mg). ^{1}H NMR (700 MHz, D$_2$O) δ 7.40 – 7.36 (m, 2H, Ar), 7.34 – 7.30 (m, 3H, Ar), 4.74 (dd, *J* = 4.8, 9.3 Hz, 1H, (C2)–H), 3.35 (ABX, *J* = 4.8, 14.0 Hz, 1H, (C3)–H), 2.93 (ABX, *J* = 9.3, 14.0 Hz, 1H, (C3')–H), 1.90 (s, 3H, COCH$_3$). ^{13}C NMR (176 MHz, D$_2$O) δ 220.2 (C1), 173.7 (**C**OCH$_3$), 138.3 (Ar), 129.8 (Ar × 2), 129.1 (Ar × 2), 127.4 (Ar), 65.8 (C2), 39.7 (C3), 22.4 (CO**C**H$_3$). LRMS-ESI [M+H]⁺ calcd. for C$_{11}$H$_{12}$NO$_2$SNa 246.06; obs. 246.06. HRMS-ESI [M+H]⁺ calcd. for C$_{11}$H$_{12}$NO$_2$SNa 246.0559; obs. 246.0560. IR (cm^{-1}) 3298, 3029, 2965, 1652, 1535, 1023, 990.

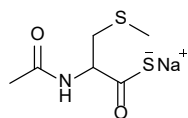### Sodium 2-acetamido-3-methylbutanethioate, Ac-Val-S⁻Na⁺

Prepared following general procedure B from 4-isopropyl-2-methyloxazol-5(4H)-one, **S9-Val** (90 mg, 0.64 mmol). Sodium 2-acetamido-3-methylbutanethioate Ac-**Val**-S⁻Na⁺ was isolated after 16 h as a white solid (85 mg). ^{1}H NMR (700 MHz, D$_2$O) δ 4.33 (d, J = 5.9 Hz, 1H, (C2)–H), 2.30 – 2.24 (m, 1H, (C3)–H), 2.07 (s, 3H, COCH$_3$), 0.95 (d, *J* = 6.9 Hz, 3H, (C4)–H$_3$), 0.87 (d, *J* = 6.9 Hz, 3H, (C4')–H$_3$). ^{13}C NMR (176 MHz, D$_2$O) δ 221.1 (C1), 174.2 (**C**OCH$_3$), 70.0 (C2), 31.9 (C3), 22.5 (CO**C**H$_3$), 19.6 (C4), 17.0 (C4'). LRMS-ESI [M]⁻ calcd. for C$_7$H$_{12}$NO$_2$S 174.05; obs. 175.06. HRMS-ESI [M+H]⁺

calcd. for $C_7H_{13}NO_2S$ 176.0740; obs. 176.0745. IR (cm^{-1}) 3396, 3228, 3153, 2962, 1623, 1546, 1514, 1134, 1001.

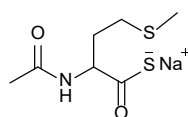### Sodium 2-acetamido-3-(methylthio)propanethioate, Ac-Cys^{Me}-S$^-$Na$^+$ (JF-02-174/03-205)

Prepared following general procedure B from 2-methyl-4-((methylthio)methyl)oxazol-5(4H)-one, **S10-Cys^{Me}** (301 mg, 1.88 mmol). Sodium 2-acetamido-3-(methylthio)propanethioate Ac-**Cys^{Me}**-S$^-$Na$^+$ was isolated after 21 h as a white solid (357 mg). ^{1}H NMR (700 MHz, D$_2$O, noseygppr1d) δ 4.65 (dd, J = 4.3, 8.8 Hz, 1H (C2)–H), 3.13 (ABX, J = 4.3, 13.9 Hz, 1H, (C3)–H), 2.87 (ABX, J = 8.9, 13.9 Hz, 1H, (C3')–H), 2.16 (s, 3H, SCH$_3$), 2.07 (s, 3H, COCH$_3$). ^{13}C NMR (176 MHz, D$_2$O) δ 218.9 (C1), 174.2 (**C**OCH$_3$), 63.4 (C2), 37.7 (C3), 22.5 (CO**C**H$_3$), 15.3 (SCH$_3$). HRMS-ESI [M+H]$^+$ calcd. for $C_6H_{11}NO_2S_2$ 194.0304; obs. 194.0305. IR (cm^{-1}) 3345, 3088, 2918, 1631, 1515.

### Sodium 2-acetamido-4-(methylthio)butanethioate, Ac-Met-S$^-$Na$^+$

Prepared following general procedure B from 2-methyl-4-((methylthio)methyl)oxazol-5(4H)-one, **S11-Met** (300 mg, 1.88 mmol). Sodium 2-acetamido-4-(methylthio)butanethioate Ac-**Met**-S$^-$Na$^+$ was isolated after 1 h as a off white solid (160 mg). ^{1}H NMR (700 MHz, D$_2$O) δ 4.57 (dd, J = 4.3, 9.3 Hz, 1H (C2)–H), 2.64 – 2.59 (m, 1H, (C3)–H), 2.57 – 2.51 (m, 1H, (C3')–H), 2.27 – 2.21 (m, 1H, (C4)–H), 2.13 (s, 3H, SCH$_3$), 2.05 (s, 3H, COCH$_3$), 1.97 – 1.91 (m, 1H, (C4')–H). ^{13}C NMR (176 MHz, D$_2$O) δ 220.7 (C1), 174.1 (**C**OCH$_3$), 63.7 (C2), 33.1 (C4), 30.3 (C3), 22.5 (CO**C**H$_3$), 14.7 (SCH$_3$). IR (cm^{-1}) 3370, 3239, 1628, 155, 1519, 1106.
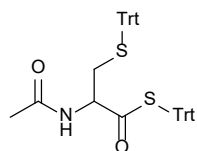
N-Acetyl-*S*-trityl-*L*-cysteine, Ac-Cys^{Trt}-OH
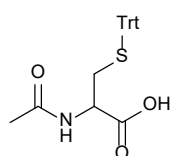
To a solution of *N*-acetyl-*L*-cysteine (1.07 g, 6.56 mmol) in anhydrous *N*-*N*-dimethylformamide (DMF; 8 mL) was added triphenylmethyl chloride (2.56 g, 9.18 mmol). The mixture was left to stir for 20 h at room temperature under an argon atmosphere. The solvent was removed under reduced pressure and the residue was purified by flash column chromatography (SiO2; eluting with a gradient of MeOH with 1% AcOH and CH$_2$Cl$_2$ 0:100 → 35:65). The solvent was removed under reduced pressure to afford *N*-acetyl-*S*-trityl-*L*-cysteine, Ac-**Cys^{Trt}**-OH (2.18 g, 5.6 mmol, 87%, (lit.[3] 82%)) as a white solid. ^{1}H NMR (700 MHz, CDCl$_3$) δ 7.43 – 7.39 (m, 6H, Ar), 7.31 – 7.27 (m, 6H, Ar), 7.25 – 7.20 (m, 3H, Ar), 5.83 (d, J = 7.0 Hz, 1H, NH), 4.37 (td, J = 4.6, 6.7 Hz, 1H, (C2)–H), 2.77 (ABX, J = 6.6, 13.0 Hz, 1H, (C3)–H), 2.69 (ABX, J = 4.6, 13.0 Hz, 1H, (C3')–H), 1.94 (s, 3H, COCH$_3$). ^{13}C NMR (176 MHz, CDCl$_3$) δ 172.4 (C1), 171.4 (**C**OCH$_3$), 144.3 (Ar), 129.6 (Ar), 128.3 (Ar), 127.1 (Ar), 67.3 (S**C**Ph$_3$), 51.7 (C2), 32.9 (C3), 23.1 (CO**C**H$_3$). LRMS-ESI [M-H]$^-$ calcd. for $C_{24}H_{22}NO_3S$ 404.12; obs. 404.13. HRMS-ESI [M-Na]$^+$ calcd. for $C_{24}H_{23}NO_3SNa$ 428.1291; obs. 428.1279. IR (cm^{-1}) 3333, 3056, 2929, 1723, 1629, 1595.

S-Trityl-2-acetamido-3-(tritylthio)propanethioate, Ac-Cys^Trt-STrt

N, *N'*-carbonyldiimidazole (CDI; 2.98 g, 18.39 mmol) was added to a solution of *N*-acetyl-*S*-trityl-*L*-cysteine (4.97 g, 12.26 mmol) in anhydrous MeCN (50 mL) and stirred for 30 min. Then triphenylmethanethiol (5.08 g, 18.39 mmol) and 4-dimethylaminopyridine (DMAP; 199 mg, 1.23 mmol) were added to the reaction mixture. The mixture was left to stir for 18 h under an argon atmosphere. The reaction was filtered, and solvent was removed *in vacuo*. The residue was purified by flash column chromatography (SiO2; eluting with a gradient of EtOAc/Petroleum Ether (40-60) 0:100 → 20:80) to give the titled compound, Ac-**Cys**^Trt-STrt (5.05 g, 7.61 mmol, 62%) as a white solid. R_f = 0.26 (30% EtOAc/Petroleum Ether). ^{1}H NMR (700 MHz, CDCl3) δ 7.38 – 7.35 (m, 6H, Ar), 7.28 – 7.25 (m, 6H, Ar), 7.24 – 7.18 (m, 18H, Ar), 5.69 (d, *J* = 8.2 Hz, 1H, (NH), 4.58 (ddd, *J* = 4.5, 6.5, 8.2 Hz, 1H, (C2)–H), 2.65 (ABX, *J* = 6.5, 12.6 Hz, 1H, (C3)–H), 2.39 (ABX, *J* = 4.5, 12.6 Hz, 1H, (C3')–H), 1.90 (s, 3H, COCH3). ^{13}C NMR (176 MHz, CDCl3) δ 196.1 (C1), 169.7 (**C**OCH3), 144.4 (Ar), 143.4 (Ar), 130.0 (Ar), 129.6 (Ar), 128.2 (Ar), 127.9 (Ar), 127.3 (Ar), 127.1 (Ar), 71.0 (COS**C**Ph3), 67.4 (CH2S**C**Ph3), 57.8 (C2), 34.5 (C3), 23.3 (CO**C**H3). HRMS-ESI [M+H]+ calcd. for $C_{43}H_{37}NO_2S_2$ 664.2338; obs. 664.2335. IR (cm^{-1}) 3263, 3019, 1623, 1650, 1537, 1489, 1443.

2-Acetamido-3-mercaptopropane thioacid, Ac-Cys-SH
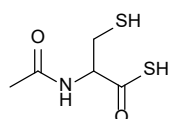
To a solution of *S*-trityl-2-acetamido-3-(tritylthio)propanethioate, Nac-Cys-(STrt)-STrt (49 mg, 73.66 µmol) in anhydrous CH2Cl2 (1.5 mL) was added triisopropylsilane (TIPS; 150 mg, 186 µL, 0.60 mmol) and then trifluoroacetic acid (TFA; 171 mg, 115 µL 1.50 mmol) dropwise over 5 min. The mixture was stirred for 1 h under an argon atmosphere. The solvent was removed *in vacuo* and the residue was triturated with Et2O (2 × 2 mL). The solid was dried *in vacuo* to give the titled compound, 2-acetamido-3-mercaptopropanethio acid, Ac-**Cys**-SH (8 mg, 41.84 µmol, 56%) as an off white solid. ^{1}H NMR (600 MHz, D2O) δ 4.60 (dd, *J* = 4.6, 7.2 Hz, 1H, (C2)–H), 2.97 (ABX, *J* = 4.6, 14.1 Hz, 1H, (C3)–H), 2.92 (ABX, *J* = 7.2, 14.1 Hz, 1H, (C3')–H), 2.03 (s, 3H, COCH3). ^{13}C NMR (151 MHz, D2O) δ 210.6 (C1), 174.8 (**C**OCH3), (C1), 64.6 (C2), 27.1 (C3), 22.4 (CO**H**3).

References

1   P. Canavelli, S. Islam and M. W. Powner, *Nature*, 2019, **571**, 546–549.

2   F. Freeman, B. G. Huang and R. I. S. Lin, *J. Org. Chem.*, 1994, **59**, 3227–3229.

3   C. Loukou, P. Changenet-Barret, M. N. Rager, P. Plaza, M. M. Martin and J. M. Mallet, *Org. Biomol. Chem.*, 2011, 9, 2209–2218.

4   J. Peng, H. Wang, Z. Shang Z, 2018, *Method for synthesizing novel meropenem chiral side chain*, CN107840815 (A) - 2018-03-27