

**Prebiotic peptide elongation by the chemoselective  
ligation of aminonitriles**

**Pierre Canavelli**

**Thesis submitted in partial fulfilment for the grade of**

**Doctor of Philosophy**

**Faculty of Mathematical and Physical Sciences**

**Department of Chemistry**

**Primary Supervisor: Dr M. W. Powner**

**Secondary Supervisor: Dr V. Chudasama**

**15<sup>th</sup> July 2019**

## **Declaration of authorship**

I, Pierre Canavelli, confirm that the work presented in this report is my own. Where information has been derived from other sources, I confirm that this has been indicated in the report.

Pierre Canavelli

15<sup>th</sup> July 2019

## Abstract

Elucidating the sequence of events that led to the emergence of life on Earth remains one of the most fundamental open problems in all of science. At its core, a living system requires a substrate able to store, replicate and modify information. Due to its ability to serve both as a genetic information carrier and biocatalyst, RNA was first proposed to be sufficient. However, decades of efforts have failed to overcome the difficulties met when attempting to experimentally realise pure RNA models.

A recent development was the proposal of a hybrid peptide-RNA world, which would tap into the enzymatic and structural capacity of peptides to complement the properties of RNA. While promising, this hypothesis found itself limited by similar hurdles. Chiefly, the formation of peptide bonds under prebiotic conditions remains a daunting challenge. Although amides are arguably the most important chemical group in all of biology and synthetic chemistry, their formation remains infamously difficult. These difficulties are compounded by the specificities of the prebiotic environment, which mandates that amidations happen in water, and precludes the use of sophisticated synthetic reagents.

Due to its biomimetic nature and the assumed prebiotic abundance of amino acid monomers, the field has traditionally focused on the free polymerisation of amino acids. However, all documented amino acid-based strategies have failed to produce sufficient amounts of peptides, and are limited to a small subset of the proteinogenic amino acids. This work marks a clear departure from such paradigms, and sets out to explore the reactivity of aminonitrile precursors to amino acids. Herein, we demonstrate that (i) aminonitriles can be converted to *N*-capped peptide thioacids and aminothioacids; (ii) those thioacids chemoselectively ligate with aminonitriles to afford proteinogenic structures; (iii) this methodology can operate as a continuous, one-pot elongation cycle, which affords *N*-capped peptides in unprecedentedly high yields. Those findings bring to light a direct link between prebiotic protometabolism and the canonical peptide structures of extant biology, and pave the way towards the experimental realisation of hybrid protein-RNA protobiological systems.

## **Acknowledgements**

I wish to thank M. W. Powner for welcoming me in his research group, and supporting me through challenging years.

Also, I would like to thank all the members of the Powner group for their friendship, assistance and encouragements.

Best of luck to all of you. It gets wild out there.

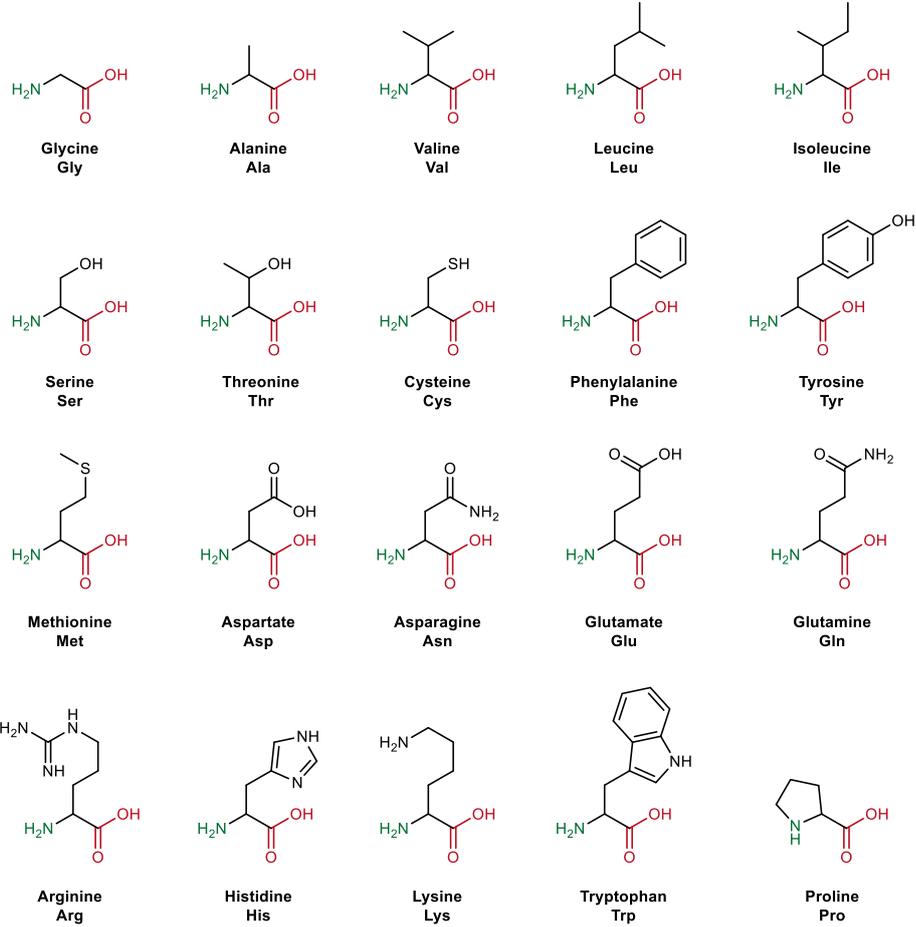
## Abbreviations

<b>ADP</b>	Adenosine diphosphate
<b>AMP</b>	Adenosine monophosphate
<b>ATP</b>	Adenosine triphosphate
<b>BBS</b>	Borate buffer saline
<b>CAA</b>	<i>N</i> -carbonylamino acids
<b>Calc.</b>	Calculated
<b>CDI</b>	1,1'-Carbonyldiimidazole
<b>DBU</b>	1, 8-Diazabicycloundec-7-ene
<b>DCC</b>	<i>N, N'</i> -Dicyclohexylcarbodiimide
<b>DCM</b>	Dichloromethane
<b>DIC</b>	<i>N, N'</i> -Diisopropylcarbodiimide
<b>DMAP</b>	4-Dimethylaminopyridine
<b>DMF</b>	Dimethylformamide
<b>DMSO</b>	Dimethylsulfoxide
<b>DNA</b>	Deoxyribonucleic acid
<b>DPQ</b>	3, 3', 5, 5'-Tetra- <i>tert</i> -butyldiphenone
<b>EDC</b>	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
<b>EG-F</b>	Elongation factor G
<b>EG-Tu</b>	Elongation factor thermo unstable
<b>EtOAc</b>	Ethyl acetate
<b>Ga</b>	Billion years ( <i>Giga annum</i> )
<b>GDP</b>	Guanosine diphosphate
<b>GOE</b>	Great Oxygenation Event
<b>GTP</b>	Guanosine triphosphate
<b>HATU</b>	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium3-oxid hexafluorophosphate
<b>HFIP</b>	Hexafluoroisopropanol
<b>HOAt</b>	1-Hydroxy-7-azabenzotriazole
<b>HOBt</b>	Hydroxybenzotriazole
<b>IgG</b>	Immunoglobulin G
<b>LHB</b>	Late Heavy Bombardment

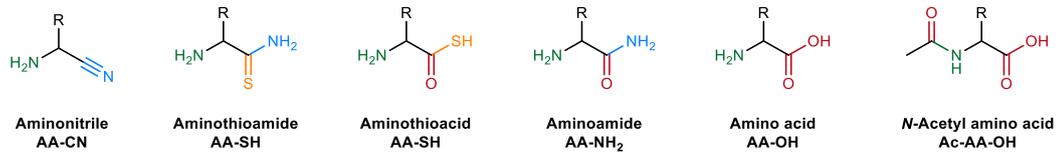
<b>NCA</b>	<i>N</i> -carboxyanhydrides
<b>NCL</b>	Native Chemical Ligation
<b>NHC</b>	<i>N</i> -Heterocyclic carbenes
<b>PFP</b>	Pentafluorophenol
<b>PNA</b>	Peptide nucleic acid
<b>PTSA</b>	<i>para</i> -Toluenesulfonic acid
<b>RaNi</b>	Raney Nickel
<b>RNA</b>	Ribonucleic acid
<b>rRNA</b>	Ribosomal ribonucleic acid
<b>S</b>	Svedberg
<b><i>t</i>BuOOH</b>	<i>tert</i> -Butyl hydroperoxide
<b>tRNA</b>	Transfer Ribonucleic acid

# Numbering and nomenclature

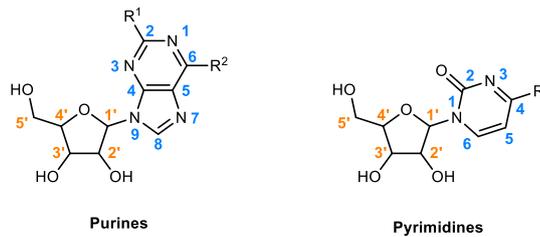
## Proteinogenic amino acid sidechains



## Amino acid precursors and derivatives



## Nucleotides



# Contents

<b>1. Introduction</b> .....	<b>12</b>
1. 1. Prebiotic chemistry .....	12
1. 1. 1. Geochemical context of the origins of life .....	12
1. 1. 2. The prebiotic environment .....	14
1. 1. 3. Principles of prebiotic chemistry .....	15
1. 1. 4. The Miller-Urey experiment.....	15
1. 1. 5. The RNA world hypothesis .....	17
1. 1. 6. The peptide-RNA world.....	18
1. 2. Making peptides.....	20
1. 2. 1. Amides from carboxylic acids and amines .....	21
1. 2. 2. Amides from amines and aldehydes.....	22
1. 2. 3. Amides from thioacids .....	23
1. 2. 4. Native chemical ligation.....	23
1. 2. 5. Biological and bio-inspired peptide synthesis.....	24
1. 2. 6. Prebiotic peptide synthesis.....	27
1. 2. 6. 1. Prebiotic peptide synthesis from amino acids .....	27
1. 2. 6. 2. Beyond biomimetism: making peptides from non-amino acid monomers .....	30
1. 2. 6. 2. 1. Peptides from aminothioacids .....	30
1. 2. 6. 2. 2. Peptides from aminonitriles.....	32
1. 3. Project aims .....	35
<b>2. Attempted oligomerisation of aminonitriles</b> .....	<b>36</b>
<b>3. Prebiotic synthesis of <i>N</i>-capped peptide thioacids</b> .....	<b>38</b>
3. 1. From Strecker mixtures to <i>N</i> -acetyl aminonitriles .....	39
3. 1. 1. Strecker synthesis of aminonitriles .....	40
3. 1. 2. Oxidative acetylation of aminonitriles .....	41
3. 1. 3. Preparative synthesis of <i>N</i> -acetyl aminonitriles .....	42
3. 2. Thiolysis of <i>N</i> -acetyl aminonitriles .....	43
3. 3. Attempted amidation of <i>N</i> -acetyl thioamides .....	44
3. 4. From <i>N</i> -acetyl thioamides to <i>N</i> -acetyl thioacids .....	47
3. 4. 1. Hydrolysis of <i>N</i> -acetyl thioamides.....	47
3. 4. 1. 1. Attempted suppression of the cyclisation of <i>N</i> -acetyl aminothioacids .....	51
3. 4. 1. 2. Attempted hydrolysis of methyl thiooxazoles .....	52
3. 4. 1. 3. Attempted oxidative ligation of methyl thiooxazoles.....	52

3. 4. 1. 4. Preparative synthesis of <i>N</i> -acetyl aminothioacids .....	53
3. 5. Conclusions .....	55
<b>4. Ligations of <i>N</i>-capped peptide thioacids.....</b>	<b>57</b>
4. 1. Optimisation studies .....	57
4. 2. Scope of the ligation .....	60
4. 3. Conclusions .....	63
<b>5. Chemoselectivity of the elongation process .....</b>	<b>64</b>
5. 1. Chemoselectivity of the thiolysis .....	64
5. 2. Chemoselectivity of the ligation .....	65
5. 3. Conclusions .....	70
<b>6. Enantioretention of the elongation process .....</b>	<b>71</b>
<b>7. One-pot experiments.....</b>	<b>74</b>
7. 1. Prebiotic, one-pot synthesis of <i>N</i> -acetyl dipeptide nitriles .....	74
7. 2. Iterative elongation of glycine nitrile.....	75
<b>8. Conclusions and future work.....</b>	<b>77</b>
8. 1. Photoredox regeneration of inorganic reagents .....	78
8. 2. Sulfites as amidonitrile activating agents .....	79
8. 3. Dehydroalanine and post-ligation sidechain synthesis .....	80
<b>9. Experimental.....</b>	<b>83</b>
9. 1. General experimental .....	83
9. 2. Prebiotic experiments.....	84
9. 2. 1. Attempted polymerisation of glycine nitrile .....	84
9. 2. 2. Oxidative acetylation of aminonitriles.....	86
9. 2. 3. Thiolysis of <i>N</i> -acetylamino nitriles.....	87
9. 2. 4. Prebiotic <i>N</i> -acetylaminoacyl thioacid synthesis .....	97
9. 2. 5. Ligation of <i>N</i> -acetyl thioacids .....	108
9. 2. 6. Selectivity studies .....	138
9. 2. 7. Enantioretention studies .....	153
9. 2. 8. One-pot experiments.....	160
9. 3. Preparation of intermediates and synthetic standards .....	169
9. 3. 1. Synthesis and characterisation of aminonitriles .....	169
9. 3. 2. Synthesis and characterisation of <i>N</i> -acetyl aminonitriles .....	179
9. 3. 3. Synthesis and characterisation of <i>N</i> -acetyl perfluorophenyl esters .....	190

9. 3. 4. Synthesis and characterisation of <i>N</i> -acetyl aminothioacids .....	192
9. 3. 5. Synthesis and characterisation of <i>N</i> -acetyl dipeptide nitriles .....	193
9. 3. 6. Synthesis and characterisation of competitor ligation byproducts .....	202
9. 3. 7. Synthesis and characterisation of glycine oligomers .....	208
9. 3. 8. Preparation of cyanoacetylene .....	210
<b>10. Bibliography. ....</b>	<b>211</b>

*“Everybody has a plan until they get punched in the mouth.”*

– Mike Tyson

## 1. Introduction

The formulation of a physico-chemical theory accounting for the emergence of living systems from the abiotic world is, to date, one of the major unsolved problems of modern science<sup>[1]</sup>. Although spectacular progress has been achieved in understanding the mechanisms through which primordial organisms evolved into the diversity of species present today<sup>[2]</sup>, the question of *how* these primordial organisms came into existence remains elusive<sup>[3]</sup>. With most of the key processes through which living systems store, replicate and transform information taking place at the molecular level and operating through chemical reactions<sup>[4]</sup>, the elucidation of the events that led to abiogenesis is an inherently chemical problem. This endeavour to understand the chemical origins of living systems has given birth to the field of prebiotic chemistry.

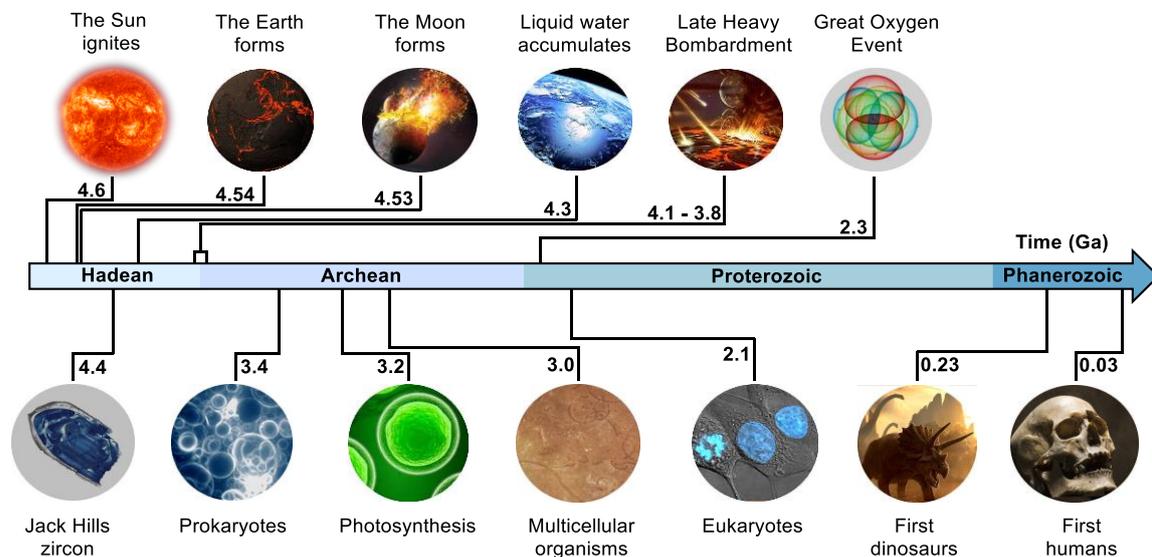
### 1. 1. Prebiotic chemistry

The field of prebiotic chemistry is equally defined by its aims and by its methods<sup>[5]</sup>. The fundamental objective of prebiotic chemistry is to elaborate models accounting for the emergence of life from abiotic environments. The chief obstacle to the realisation of this objective stems from the scarcity of reliable information regarding the environments in which life could have appeared. With the earliest fossilised traces of life dating back to 3.4 billion years ago<sup>[6]</sup>, determining the exact composition of the prebiotic milieu is faced with two major uncertainties: first, the impossibility of pinpointing the exact place and time when life arose; and second, the lack of solid, undisputed evidence concerning the geochemical state of the Earth at such remote epochs. Although most direct evidence has been lost to time, geological, astronomical and archeobiological data allow to conjecture a plausible timeframe during which life could have appeared on Earth, and to approximate the geochemical environment of the time.

#### 1. 1. 1. Geochemical context of the origins of life

The trove of geological and astronomical data accumulated over the last few decades, coupled with rapid progresses made in modelling planetary dynamics and cosmological processes, allow for assembling a timeline of the major events that shaped the Earth since its genesis (**Figure 1**). Formed from the collapse of a giant solar nebula<sup>[7]</sup>, mostly composed of helium and hydrogen, the Sun ignited at a date that computational studies and radiometric dating place at *ca.* 4.6 Ga ago<sup>[8-9]</sup>. Over the next millions of years, accretion of the remaining nebular mass around local density perturbations resulted in the formation of protoplanetary bodies<sup>[10-11]</sup>. Further accretion of one of these bodies, located *ca.* 150 Gm from the Sun, led to the formation of the proto-Earth *ca.* 4.54 Ga ago<sup>[12]</sup>. Separation of the heavy constituents of the proto-Earth, *e.g.* iron and nickel, led to the

differentiation of the mostly homogenous protoplanet into a dense metallic core and a lighter, silicates-rich primitive mantle<sup>[13]</sup>. It is believed that an oblique collision between a Mars-sized astronomical body and the newly-formed Earth, that would have occurred *ca.* 4.53 Ga ago, led to the ejection of a jet of matter that would later condense to form the Moon<sup>[14-15]</sup>. The cooling period that ensued allowed for the progressive accumulation of liquid water, whose earliest occurrence in notable quantities on the surface of the Earth is debated at *ca.* 4.3 Ga ago<sup>[16]</sup>. Due to the crucial role of water in living systems and its involvement in most proposed prebiotic reaction networks, this event is regarded as a turning point for the feasibility of life on Earth<sup>[17]</sup>. Another decisive event in the planetary timeline is the Late Heavy Bombardment (LHB), which is thought to have started 3.8 Ga ago and have lasted for about 300 million years<sup>[18]</sup>. During that period, an exceptionally intense barrage of asteroids from the Kuiper belt battered the Earth, scattering the surface with over 22,000 craters with a diameter above 20 km, and triggering environmental changes of cataclysmic magnitude every 100 years<sup>[19]</sup>. Any form of life that could have developed beforehand would most likely have been obliterated under such conditions; the end of the LHB is therefore often considered as the earliest plausible time when life as we know it could have appeared, and coincides with the end of the Hadean eon. This point, however, remains debated, as computational simulations suggest that thermophile lifeforms could have survived the LHB<sup>[20]</sup>.



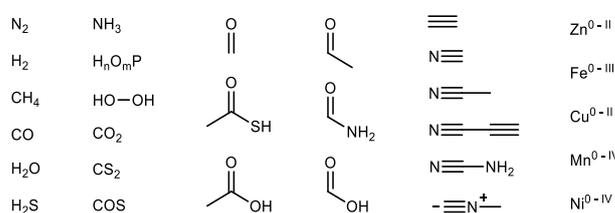
**Figure 1:** Timeline of the major planetary and biological events.

Due to paucity of exploitable fossils from the early Archean eon and the technical difficulties attached to the taphonomy of extremely old organisms<sup>[21]</sup>, the key dates of the evolution of life are much hazier and more mobile. With the earliest known terrestrial material uncovered in Jack Hills, Australia and dated at *ca.* 4.4 Ga<sup>[22]</sup> ago, any geochemical information predating this point is, as of today, out of reach. The oldest known fossils of prokaryote organisms have been dated

back to 3.4 Ga<sup>[23]</sup> ago, making them the oldest traces of life on Earth. The second half of the Archean eon saw the appearance of the first photosynthetic organisms and of primitive multicellular life, whose emergence has been dated at *ca.* 3.2<sup>[24]</sup> and 3.0 Ga<sup>[25]</sup> ago, respectively. Their proliferation led to the release of massive amounts of dioxygen into the atmosphere, which in turn triggered the Great Oxygenation Event (GOE) that is believed to have caused a mass extinction of most anaerobic organisms<sup>[26]</sup>. The Proterozoic eon saw the rise of eukaryotic organisms<sup>[27]</sup>, which would later evolve into the increasingly complex life forms: dinosaurs<sup>[28]</sup>, mammals<sup>[29]</sup>, and humans<sup>[30]</sup>.

### 1. 1. 2. The prebiotic environment

Although a variety of alternative contexts and scenarios for the emergence of life have been proposed, it is generally admitted that the timeframe relevant to prebiotic chemistry spans from the late stages of the LHB to the oldest known traces of life. The geochemical data currently at hand put these borders at around 3.8 to 3.4 Ga. In spite of the apparent intractability of determining the composition of the Earth at such remote times, geological models and the observation of other celestial bodies (*e.g.* comets, exoplanets, intergalactic clouds and asteroids) allow to outline the geochemical environment of the late Hadean and early Archean (**Figure 2**).



**Figure 2:** Representative feedstock compounds thought to be available in prebiotic environments.

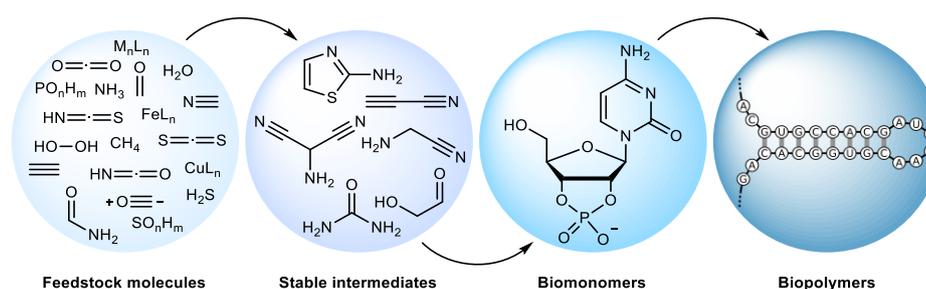
The prebiotic atmosphere is believed to have been mostly composed of hydrogen, nitrogen, water and carbon oxides<sup>[31]</sup>. The most notable difference with today's atmosphere lies in the quasi absence of free oxygen<sup>[32]</sup>. Basins of liquid water were present on and beneath the surface<sup>[33]</sup>, and are thought to have been rich in cyanide<sup>[34]</sup>, ammonia<sup>[35-36]</sup>, formaldehyde<sup>[37]</sup> and formamide<sup>[38]</sup>. Phosphates<sup>[39]</sup>, peroxides<sup>[40]</sup> and simple carbonyls<sup>[41]</sup> are also believed to have been present in notable quantities. Importantly, hydrogen sulfide and simple organosulfur compounds are plausible components of the prebiotic environment<sup>[42]</sup>, as well as byproducts of volcanic activity such as carbonyl sulfide and carbon disulfide<sup>[43]</sup>. Highly unsaturated alkynes, such as cyanoacetylene, are also commonly admitted prebiotic feedstock molecules<sup>[44]</sup>. The environment is also supposed to have been rich in soluble transition metal species, such as iron, copper, nickel, zinc and manganese, in various oxidation states<sup>[45-46]</sup>.

### 1. 1. 3. Principles of prebiotic chemistry

The leitmotiv of prebiotic chemistry is to uncover chemical networks based on plausible feedstock reagents leading to the spontaneous formation of the major constituents of living systems<sup>[5]</sup>. To be considered alive, an organism must present a number of hallmarks, including<sup>[47]</sup>:

- A capacity to store, edit and transmit information. In most known species, these functions are effected by ribo- and nucleic acid polymers.
- The ability to extract, store and exploit energy from their environment through metabolic networks. Example of such processes include the photosynthetic machinery and cellular respiration, and involve stable energy carriers such as nucleotide triphosphates.
- Compartmentalisation from the outer milieu and/or between internal subsystems. In modern biology, this property is commonly achieved by the formation of protein-supported lipid bilayers.
- The capability to realise energetically unfavourable processes involved in aforementioned functions, usually through the means of a specific set of biocatalysts.

Owing to their central role in all known species and their functional versatility, biopolymers such as DNA, RNA, proteins and their analogues have traditionally been a major focus of interest for prebiotic chemists. Both proteins and nucleic acid polymers are indeed well known for their capacity to serve as robust information carriers and highly effective catalysts<sup>[4]</sup>. For this reason, the synthesis of their monomers from plausible feedstock compounds and their polymerisation remain the principal objective of prebiotic chemistry (**Figure 3**).

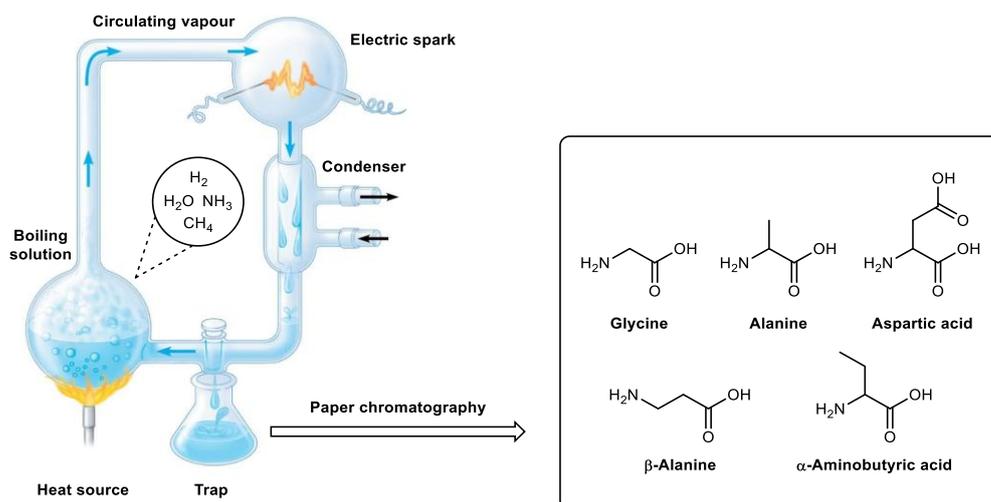


**Figure 3:** General objective of prebiotic chemistry.

### 1. 1. 4. The Miller-Urey experiment

In 1952, Stanley Miller and Harold Urey carried out a series of experiments that are widely regarded as a cornerstone of prebiotic chemistry<sup>[48]</sup>. It had been hypothesised at the time that under certain conditions, aqueous mixtures of simple gases, such as methane, hydrogen and ammonia, could react to form biogenic organic compounds. However, none of the many attempts

to validate this hypothesis had managed to form such compounds in significant concentrations. This series of failures led Miller and Urey to question the composition of the proposed prebiotic atmosphere, which was then believed to be strongly oxidising. Building upon Oparin's hypothesis of a reducing prebiotic atmosphere, Miller designed an apparatus circulating vaporised mixtures of simple feedstock molecules through electric arcs under an oxygen-free atmosphere (**Figure 4**).



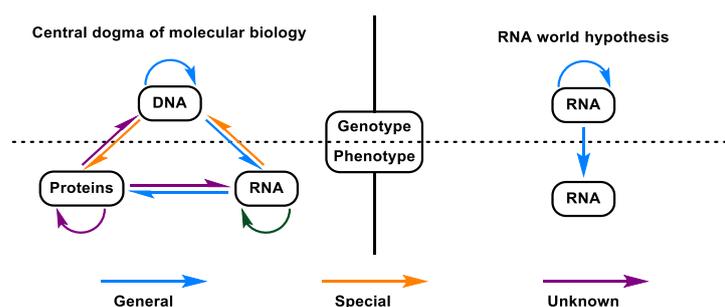
**Figure 4:** The Miller-Urey experiment.

Miller's apparatus was designed to mimic a plausible prebiotic environment. The aqueous mixture was boiled to replicate the hypothesised high temperatures of the Hadean Earth and form a gas phase mimicking the prebiotic atmosphere. The vapour thus produced was then passed through a secondary flask where continuous electric sparks were fired to simulate lightning and serve as an electrochemical energy source. The resulting mixture was finally cooled down and collected into a trap for analysis. When Miller used this apparatus to incubate water, hydrogen, methane and ammonia, he observed that an intensely pink solution formed after a day, suggesting the formation of conjugated organic compounds. After one week, a turbid, deep red solution was collected. Paper chromatography allowed Miller to identify five amino acids, including the proteinogenic glycine, alanine and aspartic acid. By demonstrating the possibility of forming biogenic compounds from simple feedstock gases, the Miller-Urey experiment ushered in a new era for the study of abiogenesis. Incidentally, this experiment underlined the critical limitation posed by the analytical equipment and techniques in the study of complex chemical systems: although Miller could only identify five amino acids in his time, later experiments and frozen samples from the original apparatus revealed the presence of over 20 amino acids when analysed using modern methods<sup>[49]</sup>. This observation stands true to this day, as even state-of-the-art

analytical techniques, such as mass spectrometry and multidimensional NMR, remain ill-suited to characterise dynamic chemical systems and complex mixtures of products.

### 1. 1. 5. The RNA world hypothesis

Modern forms of life operate through complex biochemical networks, built around strikingly different classes of molecules: proteins, nucleobases, sugars, lipids, and their combinations. In an effort to elucidate the main information transfer pathways, the hypothesis currently known as the *Central dogma of molecular biology* was formulated by Francis Crick<sup>[50]</sup> (**Figure 5**). According to this hypothesis, the genetic information of an organism is stored in its DNA. The self-replication and transcription of the genotypic information into RNA, followed by the translation of RNA into proteins, constitute the general ways of perpetuation of an organism's genotype and of its expression into a phenotype. This model has been amply validated with the unravelling of the mechanisms of transcription and translation<sup>[4]</sup>, and complemented by the observation of reverse transcription<sup>[51]</sup>, RNA replication<sup>[52]</sup> and direct DNA translation<sup>[53]</sup>.



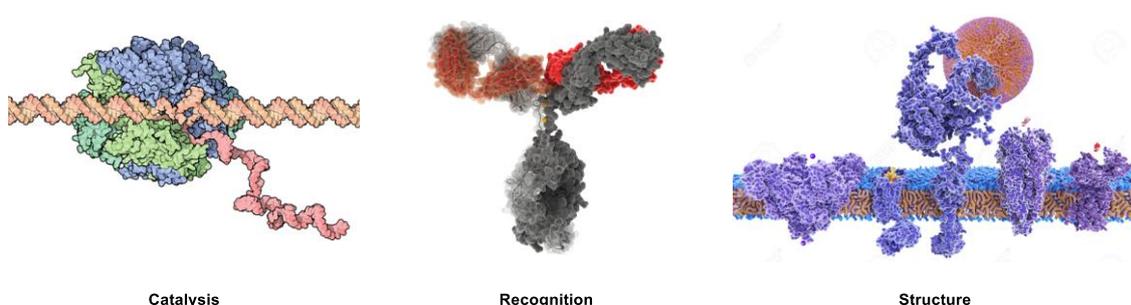
**Figure 5:** The central dogma of molecular biology and its RNA world counterpart.

Because of the vanishingly low probability of several classes of complex biomolecules arising coincidentally, and spontaneously assembling into a self-sustained reaction network, a number of simpler, hypothetical genetic networks have been proposed. The most widely accepted of these models is the RNA world hypothesis. Inspired by the folding capacities and catalytic role of rRNA in ribosomal transcription and the coding capabilities of mRNA, Crick and Orgel postulated that primitive life forms could have depended solely on RNA to both carry and replicate information<sup>[54–55]</sup>. The RNA world hypothesis dramatically simplifies the speculated interactome of primordial cells, thus making it far more likely to emerge. In addition, the RNA world hypothesis is supported by the functional versatility of RNA, which has been shown to catalyse nucleotide cleavage and ligation<sup>[56]</sup>; peptide coupling by the ribosome<sup>[57]</sup>; and to bind small metabolites<sup>[58]</sup> and catalytically active metal ions<sup>[59]</sup>. For these reasons, the RNA world hypothesis has been at the centre of most prebiotic models since the 1980's. Prebiotic syntheses of the four RNA

nucleobases<sup>[60 – 61]</sup> and, more recently, of the full RNA ribonucleotides<sup>[62]</sup> have been proposed, along with phosphorylation<sup>[63 – 64]</sup> and oligomerisation strategies<sup>[65 – 66]</sup>. However, these strategies remain plagued with hurdles, including poor nucleobases yields<sup>[60]</sup> and difficult nonenzymatic nucleotides elongation, which produce mixtures of 3'-5' and 2'-5' linkages<sup>[67]</sup>. These cumulated impediments drastically lower the plausibility of spontaneously arising, RNA-only organisms. By factoring in the sluggishness and inaccuracy of nonenzymatic RNA elongation, it has been estimated that exploring enough sequences to form the set of functional ribozymes needed to create a RNA-based metabolism would require a time far greater than the age of the universe<sup>[68]</sup>. Although important steps have been made by J. Szostak towards an efficient RNA non-templated polymerisation, notably through the use of lipid surfaces and nucleotide derivatisation<sup>[69]</sup>, these difficulties call for the investigation of alternative theories of the origins of life.

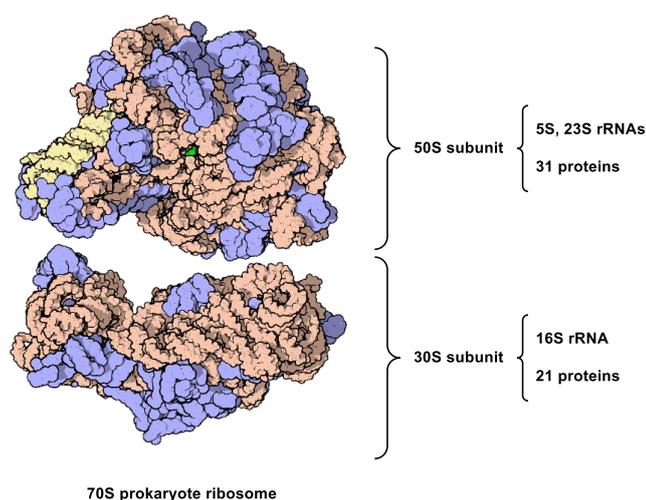
### 1. 1. 6. The peptide-RNA world

The many roadblocks encountered by the original RNA world hypothesis have led to the proposition of modified theories which, while remaining centred around RNA coding and catalysis, are augmented with other classes of functional biomolecules. Among these theories, a particularly strong case has been made for the possibility of a mixed peptide-RNA world<sup>[70 – 71]</sup>. This idea is supported by the central role proteins play in nearly all operations encountered in extant biology (**Figure 6**). The overwhelming majority of biological reactions are catalysed by enzymes, which include some of the highest catalytic activities known<sup>[4]</sup>. In addition, most complex functions found in eukaryotes, such as immune response and signal transduction, are performed by proteins and protein complexes. Some necessary properties of living cells, such as compartmentation and mechanical resistance, similarly rely on the physical properties of protein structures. This functional prolificacy suggests that peptides could have helped alleviate some of the difficulties faced by the RNA world, by accessing catalytic and autoreplicative activity.



**Figure 6:** 3D representations of key protein actors in extant biology; **left:** RNA polymerase on an RNA strand; **centre:** structure of a human IgG; **right:** the cytoskeleton scaffolding underlying cell membranes and intracellular organelles distribution.

Aside from the range of functionality peptides and proteins would bring to the RNA world, extant biology is laden with instances of protein-nucleic acid complementarity, which further support the notion of a joined evolution tracing back to the prebiotic era. First, proteins and nucleic acids play deeply intertwined roles in the storage and replication of genetic information. All stages of the replication, and transcription of DNA into mRNA are orchestrated by protein complexes displaying remarkable binding affinities for nucleic acid polymers<sup>[4]</sup>. This interplay culminates in the translation of mRNA into protein by the ribosome, which is itself composed of both RNA strands and a collection of ribosomal proteins. The ribosome is a highly preserved, macromolecular catalytic complex present in both prokaryotes and eukaryotes<sup>1</sup>. The prokaryote ribosome is composed of a small 30S subunit and a large 50S subunit<sup>[72]</sup>. The 50S subunit is comprised of two 5S and 23S strands or ribosomal RNA linked to a collection of 31 proteins, while the small subunit includes a single 16S rRNA strand and 21 proteins (**Figure 7**).

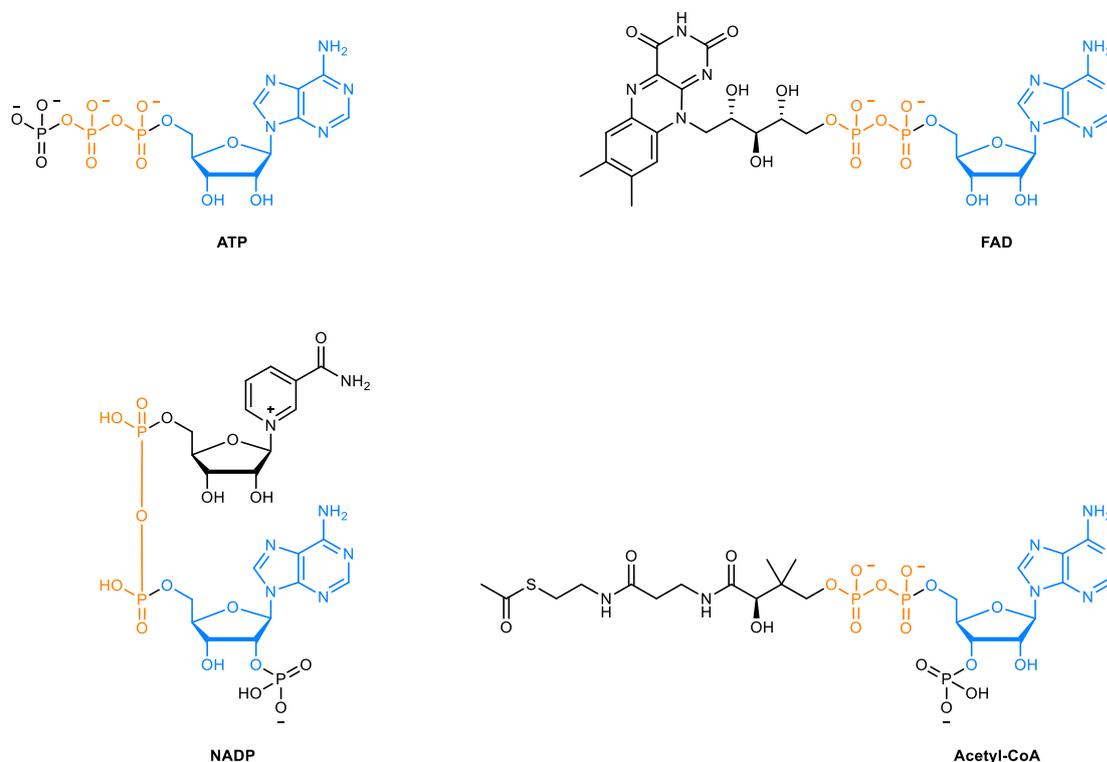


**Figure 7:** Structure of the prokaryote ribosome.

This complementarity is mirrored down to the macromolecular scale, as most enzymes require metabolites and cofactors to function, with the majority of the most important cofactors sharing nucleic acid subunits in their structure. The most striking of these are the adenoylated cofactors, including adenosine triphosphate **ATP**, the energetic currency of most know organisms<sup>[4]</sup>, and whose hydrolysis of the energy-rich phosphodiester bond provides the driving force of most endergonic biological processes. Other notable examples can be found in the structure of nicotinamide adenine dinucleotide phosphate (**NADP**), flavin adenine dinucleotide (**FAD**), and

<sup>1</sup> Despite variations in their structure and the cofactors involved, both the prokaryote and eukaryote ribosome operate according to a similar mechanism. For this reason, and for the sake of brevity, the following description will focus on the better-characterised prokaryote ribosome.

acetyl-CoA, which are involved in nearly all metabolic pathways<sup>[4]</sup>, and share an adenosine diphosphate moiety (**Figure 8**).



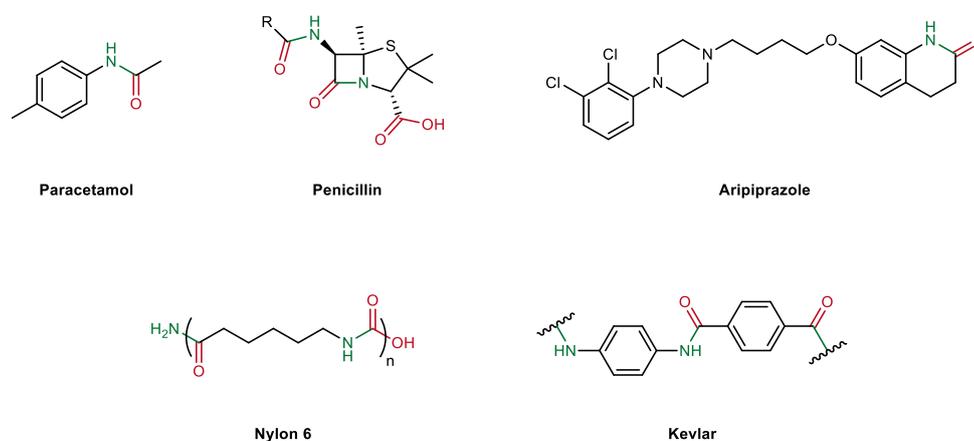
**Figure 8:** Structure of key nucleic acid metabolites and cofactors. The common adenosine moiety is highlighted in blue; the 5'-diphosphate phosphate groups, in orange.

This omnipresent relationship between nucleic acids and proteins has been interpreted as a possible vestige of common evolution. This idea is supported by the recent demonstration by C. W. Carter *et al.* that peptides mimicking the active core of class I and II aminoacyl-tRNA synthetases, while only 120 to 130 residues in length, retain *ca.* 60% of the catalytic activity of the full enzyme towards tRNA aminoacylation, the key reaction of codon-dependant translation<sup>[73]</sup>. These peptides, named *Urzymes* (from *Ur: first, primitive*), were designed in an effort to mirror the primitive enzymes from which modern translation aminoacyl-tRNA synthetases are likely to have evolved. The spectacular catalytic capabilities of these short peptides – which are believed to be still more complex than the minimal structures required for catalysis to emerge – further supports the hypothesis of a peptide-RNA coevolution.

## 1. 2. Making peptides

Due to their high stability, polarity and conformational flexibility, amide bonds are a ubiquitous motif in both naturally occurring and designed compounds. Many best-selling drugs including at least one amide linkage in their structure, with over 700 peptide-based therapeutics currently on

the market or in development. Moreover, amide bonds are one of the most commonly used linkers in the polymers industry, with best-selling polymers such as Kevlar® and the nylon family being based on amide linkages<sup>[74-76]</sup> (**Figure 9**).

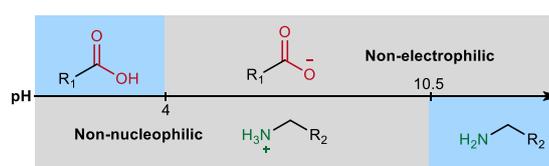


**Figure 9:** Amide-containing pharmaceuticals and polymers of high significance.

The omnipresence of amide bonds in modern chemistry might suggest that a certain degree of ease in synthesising them has been achieved. But, surprisingly, the formation of amides remains in many cases an open challenge<sup>[77]</sup>. Although much effort has been invested in developing methods to promote amidation reactions – leading to the development of an ever-growing collection of sophisticated and highly expensive reagents –, a general protocol for the efficient formation of amide bonds is yet to be achieved. In spite of their relative generality and robustness, the methods used both on the laboratory scale and in industrial peptide synthesis, relying heavily on derivatisation, protecting groups and activating agents, are widely regarded as ineffective and unsustainable.

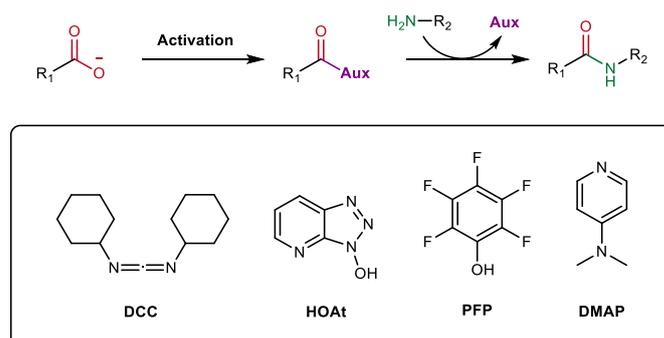
### 1. 2. 1. Amides from carboxylic acids and amines

The traditional and most obvious way of making amides is to condense an amine and a carboxylic acid<sup>[78]</sup>. While apparently simple, this reaction is rendered difficult by the  $pK_a$  of each reacting moiety. With an average value of *ca.* 4 for acid groups and *ca.* 10.5 for aliphatic ammonium salts, the two moieties are mutually unreactive under any pH in their native states (**Figure 10**).



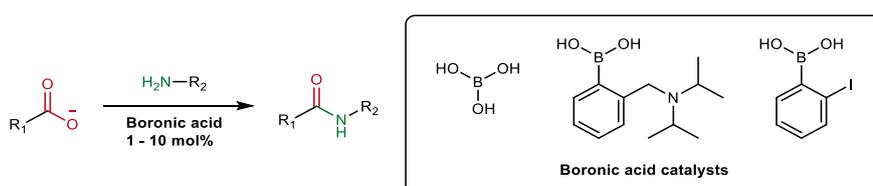
**Figure 10:** Diagram of the predominant protonation states of carboxylic acids and amines.

To circumvent this limitation, a wide and diverse range of strategies aiming to activate carboxylic acids have been developed. Most of these methods aim to convert the acid into an activated ester, thus facilitating the nucleophilic attack by the amine moiety and allowing amide formation *via* transesterification<sup>[79]</sup> (**Figure 11**). Activation of carboxylic acids is typically achieved by reaction with electrophilic auxiliaries such as *N,N'*-dicyclohexylcarbodiimide (DCC), with the assistance of nucleophilic catalysts such as 1-hydroxy-7-azabenzotriazole (HOAt), 4-dimethylaminopyridine (DMAP) or pentafluorophenol (PFP).



**Figure 11:** General carboxylic acid activation strategies and commonly employed reagents.

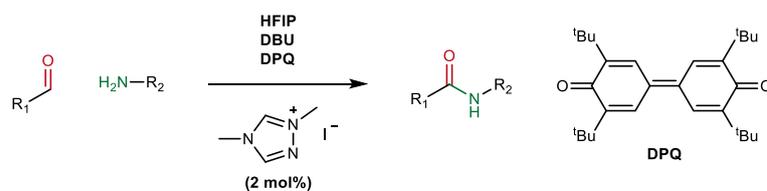
While prevalent in peptide synthesis, these activations strategies remain highly atom-inefficient, and tend to yield variable results depending on the substrate. For this reason, the development of alternative carboxylic acid – amine coupling strategies remains an active field of research. A remarkable product of the recent progress in this domain involves the use of boronic acid catalysts which, although significantly less expensive, are faced with limitations in substrate scope<sup>[80 – 81]</sup> (**Figure 12**).



**Figure 12:** Boronic acid-based amide coupling strategies.

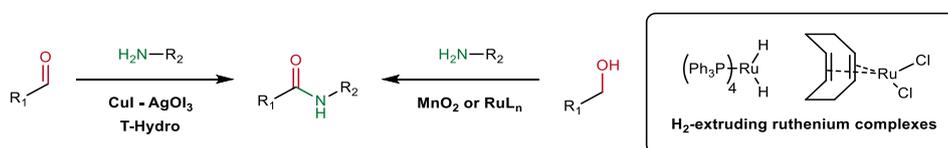
### 1. 2. 2. Amides from amines and aldehydes

Alternative amidation methods based on the redox coupling of amines and aldehydes have gained considerable traction over the last decade. These methods circumvent the problem of carboxylic acid activation by relying on the much more reactive aldehyde moiety. Oxidation is usually achieved through the use of strong oxidizing agents such as quinone derivatives, using *N*-heterocyclic carbenes (NHC) as catalysts<sup>[82]</sup> (**Figure 13**).



**Figure 13:** Oxidative amidation of aldehydes by the NHC – DPQ system.

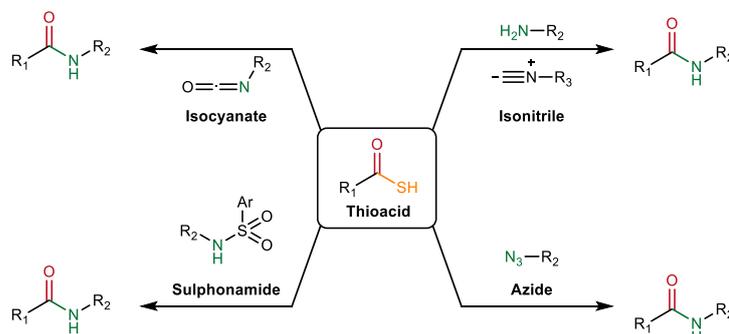
A considerable advantage of redox amidation strategies lies in their compatibility with abundant, inexpensive metal catalysts, such as Cu(I) complexes, and environmentally benign oxidants, such as peroxides and oxygen. Notable success has recently been achieved by employing a copper/silver iodate catalytic system<sup>[83]</sup>. Interestingly, primary alcohols can be used as an amine acceptor *via in situ* oxidation to aldehydes using manganese oxides<sup>[84]</sup> or H<sub>2</sub>-extruding ruthenium complexes<sup>[85]</sup> (**Figure 14**).



**Figure 14:** Metal-catalysed aldehyde activation strategies.

### 1. 2. 3. Amides from thioacids

An increasingly popular number of amidation strategies rely on the unique reactivity of thioacids. These strongly nucleophilic carbonyls can indeed be coupled with a variety of amine surrogates, such as alkyl azides<sup>[86]</sup>, isonitriles<sup>[87]</sup>, sulphonamides<sup>[88]</sup> and isocyanates<sup>[89]</sup> to yield amides using mild conditions (**Figure 15**).

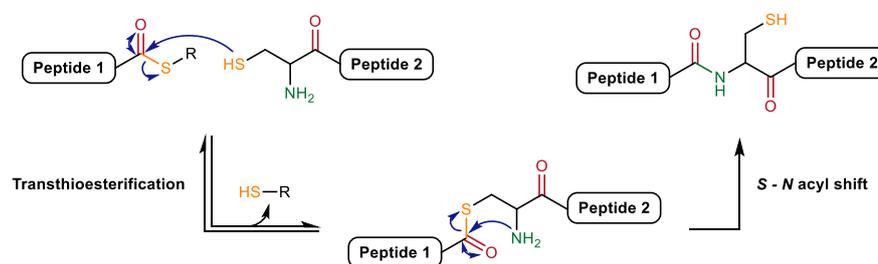


**Figure 15:** Thioacid-based amidation strategies.

### 1. 2. 4. Native chemical ligation

Motivated by the need for efficient amidation strategies compatible with the rapidly expanding field of automated and solid phase peptide synthesis, a series of peptide ligation methods were

developed during the 1990s. Among these, the native chemical ligation (NCL) method, developed by Kent and coworkers, has become a staple of peptide synthesis due to its high yields, its robustness and its compatibility with entirely unprotected peptides up to several hundred residues in length<sup>[90]</sup> (**Figure 16**).

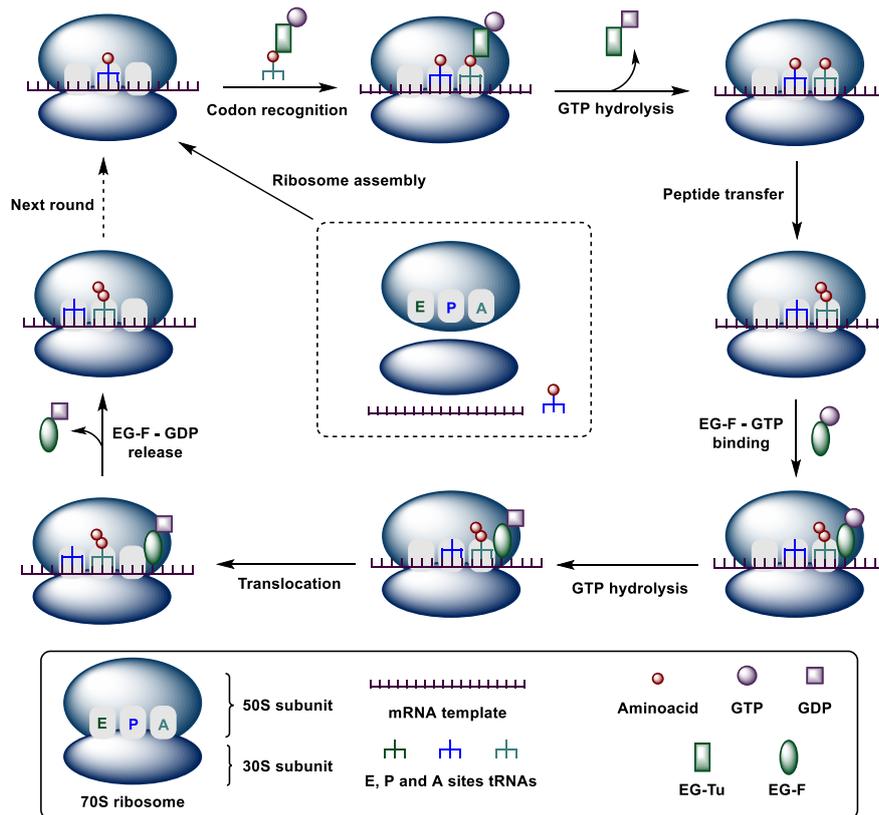


**Figure 16:** Native chemical ligation strategy.

This strategy relies on the rapid, reversible formation of cysteine thioesters, which are quickly transformed into the desired peptide bond *via* intramolecular  $S \rightarrow N$  acyl shift. NCL strategies have found numerous applications in the synthesis of long peptides. As traditional peptide synthesis suffers from diminishing yields as the length of the peptide increases, synthesising the final peptide in separate blocs that can be assemble through NCL has become routine practice. Despite its advantages, this method remains limited to specific cases, as only a peptide whose *N*-terminus bears a cysteine residue can be ligated to another fragment, which can lead to fragments stretching out to unpractical sizes in order to reach their next cysteine position. Furthermore, the formation of *C*-terminal thioesters remains a non-trivial exercise, especially in the case of large protein fragments.

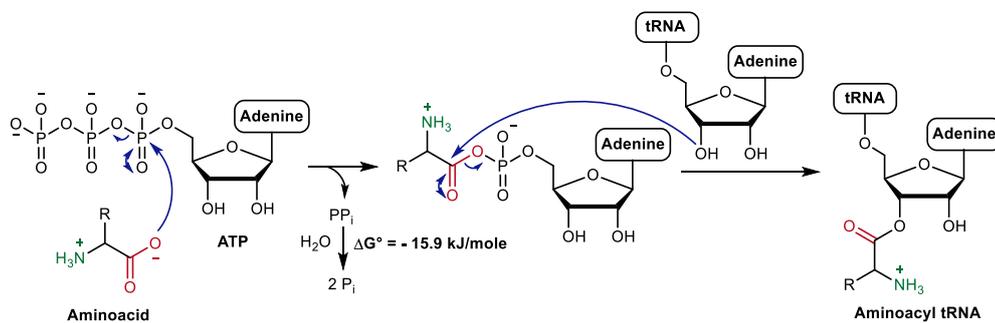
### 1. 2. 5. Biological and bio-inspired peptide synthesis

In biological systems, the synthesis of most peptides and proteins is achieved through the template-mediated translation of mRNA by the ribosome<sup>[4]</sup>. On a macromolecular level, the prokaryotic translation starts with the initiation factors-assisted sequential assembly of the ribosome-mRNA-tRNA complex. Upon recognition of the mRNA codon located in the A site of the ribosome, an aminoacyl tRNA – EG-Tu – GTP complex positions itself in the A site through mRNA – tRNA triplet hybridisation. GTP hydrolysis and release of EG-Tu liberate the A site aminoacyl tRNA, making it available for coupling. Binding of the EG-F – GTP complex to the A site induces the translocation of the tRNAs present in the P and A sites to the E and P sites, respectively. After release of the EG-F – GDP complex, the elongating peptide, which is left in the P site and can either undergo further elongation, or be hydrolysed to release the final peptide (**Figure 17**).



**Figure 17:** General description of the prokaryotic ribosomal translation.

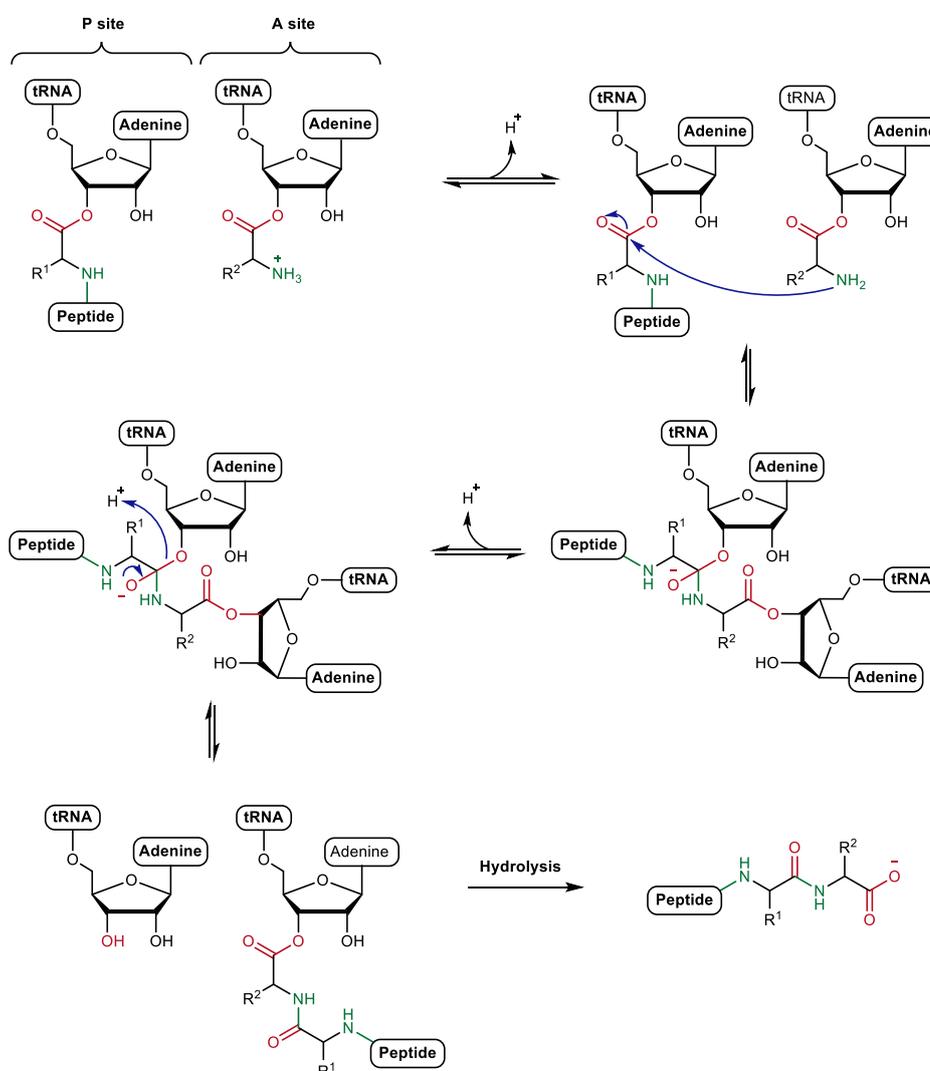
On an atomic scale, the ribosomal peptide coupling proceeds in a way remarkably similar to common carboxylic acid activation strategies. Both the elongating peptide and the amino acid monomer are derivatised as aminoacyl tRNAs through sequential reaction with ATP and selective transesterification by the 2'-hydroxyl (class I) or 3'-hydroxyl (class II) of the terminal adenosine of a tRNA's 3'-terminus. The reaction is thermodynamically favoured by the highly exergonic hydrolysis of the pyrophosphate byproduct<sup>[91]</sup> (**Figure 18**).



**Figure 18:** ATP-based aminoacyl formation.

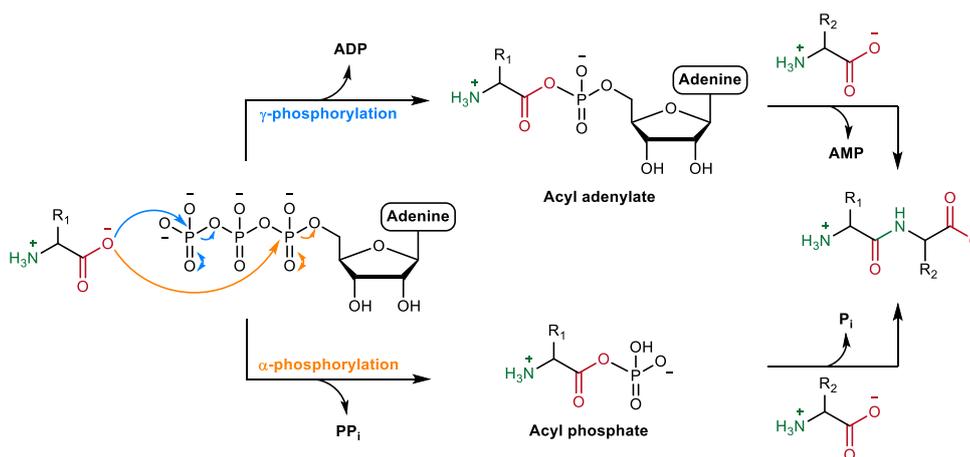
During the elongation process, aminoacyl tRNAs bearing the elongating peptide and activated monomers are kept in contact distance in the P and A sites of the ribosome. Deprotonation of the ammonium moiety of the amino acid monomer allows the formation of a tetrahedral aminal

intermediate through nucleophilic attack of the C-terminal ester of the elongating peptide. An additional proton transfer step is followed by the collapsing of the tetrahedral transition state into the coupled peptide aminoacyl tRNA. The elongation is terminated and the final peptide released *via* hydrolysis upon recognition of a stop codon by the ribosome<sup>[92]</sup> (**Figure 19**).



**Figure 19:** Mechanism of ribosomal peptide elongation.

Ever since the elucidation of the mechanisms at work in the ribosome, bio-inspired peptide coupling strategies have been an active field of research. These strategies, whose ultimate goal is to provide environmentally benign, water-compatible amidation methodologies, have notably investigated the possibility of activating carboxylic acids through the use of phosphorylated nucleotides and polyphosphates. Because of its biological role as amino acid activator, the use of ATP as potential coupling agent has gathered strong interest. The reaction of amino acids with ATP in the presence of divalent cations has indeed been shown to yield biomimetic acyl adenylate and acyl phosphates, depending on which phosphate is attacked<sup>[93]</sup> (**Figure 20**).



**Figure 20:** ATP-mediated activation of amino acids.

Although significant activation could be observed by trapping the activated intermediates through reaction with hydroxylamine, these strategies result in poor peptide yields. The suggested reason to these low yields stems from the strong influence of the pH on the reaction: while the activation step benefits from acidic conditions and has been observed down to pH 5, nucleophile attack by the amine requires alkaline pHs to happen at a significant rate. At such high pHs (typically > 10), the concentration of hydroxyl ions becomes so high as for the hydrolysis of the activated intermediates to outcompete the coupling reaction.

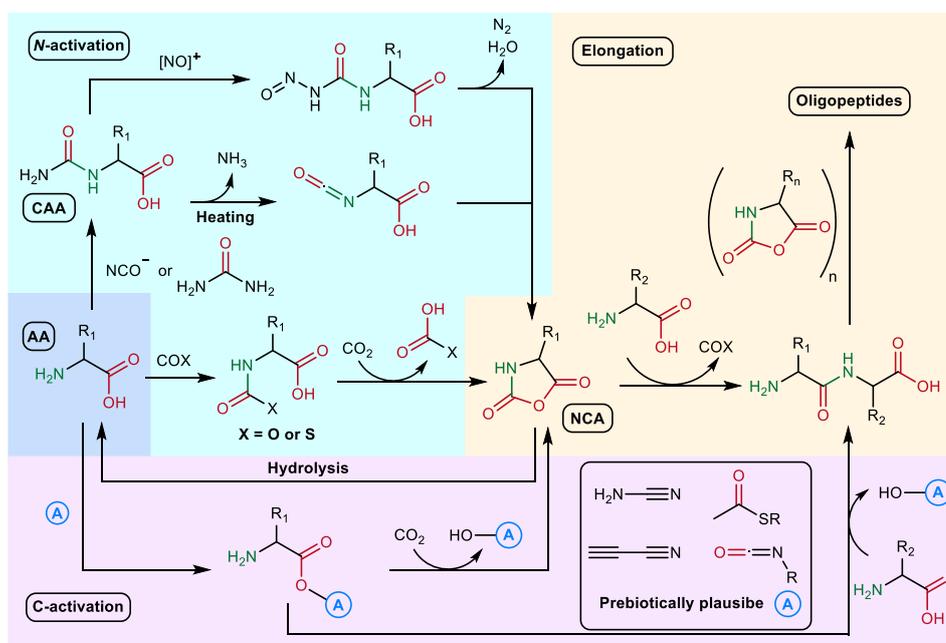
## 1. 2. 6. Prebiotic peptide synthesis

### 1. 2. 6. 1. Prebiotic peptide synthesis from amino acids

Amino acids and peptides are regarded as highly plausible components of the prebiotic environment since the Miller-Urey experiment. Several other pathways leading to the synthesis of amino acids have also been brought to light<sup>[94 - 96]</sup>. Interestingly, a cyanosulfidic network recently described by J. Sutherland and coworkers has been shown to yield precursors to both nucleotides and a variety of amino acids<sup>[97]</sup> (**Figure 30**), thus reinforcing the hypothesis of a common, cooperative origin of RNA and peptides. Furthermore, traces of amino acids have been observed in comets<sup>[98]</sup>, meteorites<sup>[99]</sup>, and the interstellar medium<sup>[100]</sup>, thus complementing the possibility of amino acids arising from the terrestrial environment with potential delivery from extraterrestrial sources. In addition, a plethora of amino acid polymerisation pathways have been proposed and shown to yield oligopeptides under prebiotic conditions (see section 1. 2. 6.).

Although a number of plausible pathways leading to the formation of amino acids and their precursors have been proposed, the prebiotic synthesis of peptides remains a mostly unanswered question. Indeed, the necessity of relying on reagents and conditions that match prebiotic

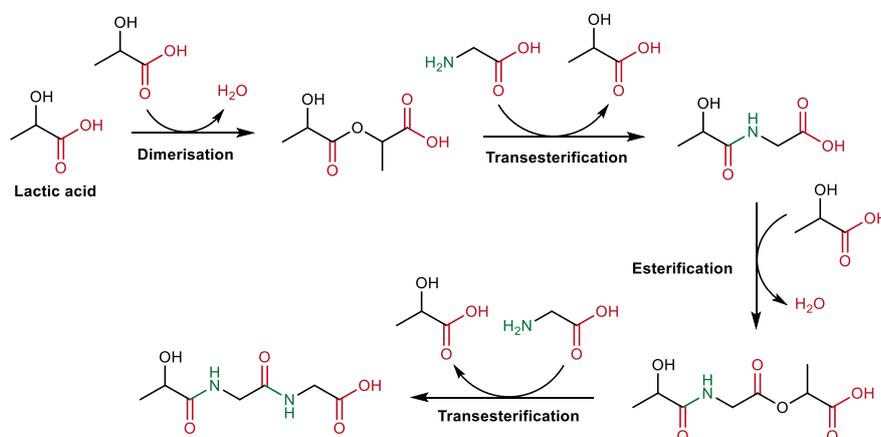
scenarios forbids the use of anhydrous conditions and of most of the usual coupling agents traditionally employed by amidation strategies. Building upon the reasonable assumption that the shortest and most direct reaction pathways are the most plausible, the vast majority of efforts have focused on the polymerisation of amino acids. The direct condensation of amino acids being extremely inefficient, especially in aqueous media, an array of strategies aiming at activating either the amine or carboxylic acid moieties has been investigated. Predominantly, these strategies explore the possibility of turning the free amine into an activating group that is reactive enough to trigger the intramolecular activation of the geminal carboxylic acid moiety (**Figure 21**). Notable examples of such strategies include formation of *N*-carboxylamino acids (CAA) *via* formamidation, followed by dehydration<sup>[101]</sup> or nitrosation<sup>[102]</sup>, respectively forming isocyanate and *N*-nitroso-CAA derivatives. These highly reactive intermediates readily cyclise to form *N*-carboxyanhydrides (NCA), which can in turn react with the *N*-terminal amine of free amino acid or elongating peptides to form oligopeptides. In addition, the amine group can be reacted with carbon oxides and sulphides, such as COS<sup>[103]</sup> and CS<sub>2</sub><sup>[104]</sup>, to form *N*-carboxyamino acids. Upon reaction with another equivalent of dicarbonyl, these intermediates have been shown to also form NCAs. Alternatively, some strategies seek to activate the carboxylic acid moiety through reaction with prebiotic activating groups mimicking the properties of traditional activating agents<sup>[105]</sup>. These C-activated amino acids can subsequently either be directly reacted with free amines, or react with *N*-activators such as CO<sub>2</sub> to form NCAs.



**Figure 21:** Overview of the amino acid-based prebiotic peptide coupling strategies.

Although these routes have the advantage of requiring very few steps, and converge to the central, high-energy intermediate NCA, they remain faced with severe limitations. The first of these is their generally poor yields, which are especially problematic in the case of polymeric species such as peptides, where the overall yield decreases exponentially with the peptide length. This problem is brought to light when one considers that the smallest known peptides displaying enzymatic activity have a length of *ca.* 100 amino acids<sup>[106]</sup>. Consequently, even coupling yields as high as 90% would only lead to the formation of active peptides in 0.003% total yield. Furthermore, most aforescribed reactions proceed with rather slow kinetics, generally taking days to weeks to reach 50% conversion of the starting materials. These slow kinetics put the polymerisation process on the same timescale as the hydrolysis of amide bonds, thus further limiting the probability of forming catalytic peptides. An additional hurdle comes from the incompatibility of NCA formation with free *N*-terminal amine moieties, which rapidly react to form elongation-stopping hydantoin species<sup>[103]</sup>. This side-reaction imposes that the activation and polymerisation be kept separated at all times, which further degrades the plausibility of such scenarios. Finally, these strategies offer no selectivity against other prebiotic amines (*e.g.* ammonia, lysine sidechains), which could irreversibly form a collection of non-proteinogenic byproducts.

Recently, a team led by Hud and Krishnamurthy has developed a novel approach to amino acids polymerisation based on the use of lactic acid as transient activator<sup>[107]</sup>. While exposed to a wet – dry cycle intended to mimic terrestrial night – day patterns, it has indeed been observed that glycine monomers heated between 65 °C (night) and 85 °C (day) yielded depsipeptides of up to ten glycine units (**Figure 22**). Although simple in principle, such strategies rely on the use of high-temperature, wet-dry cycles, which drastically reduce the number of suitable environments, and thus lower their plausibility.



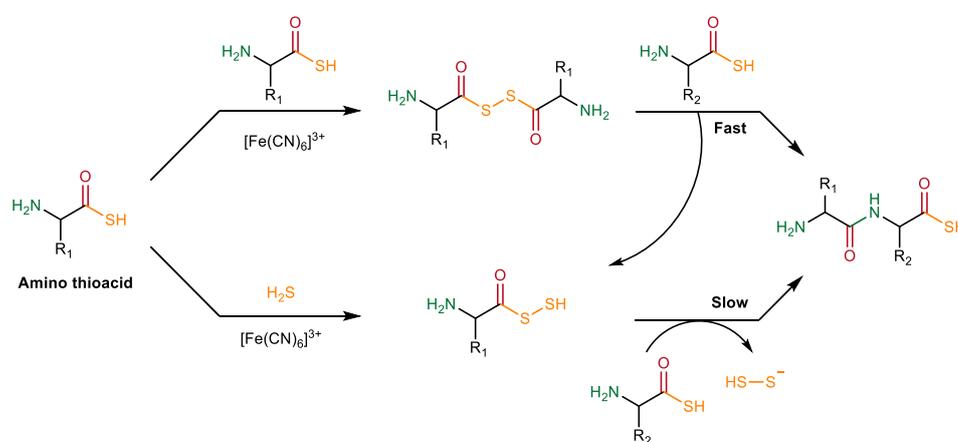
**Figure 22:** Lactate-mediated amino acid polymerisation.

### 1. 2. 6. 2. Beyond biomimetism: making peptides from non-amino acid monomers

In spite of their apparent plausibility as prebiotic monomers and of decades of investigation, the polymerisation of amino acids has, to date, produced unsatisfactory results at best. In an attempt to avoid the limitations of amino acid-based pathways, reaction networks focused on alternative monomers have been investigated.

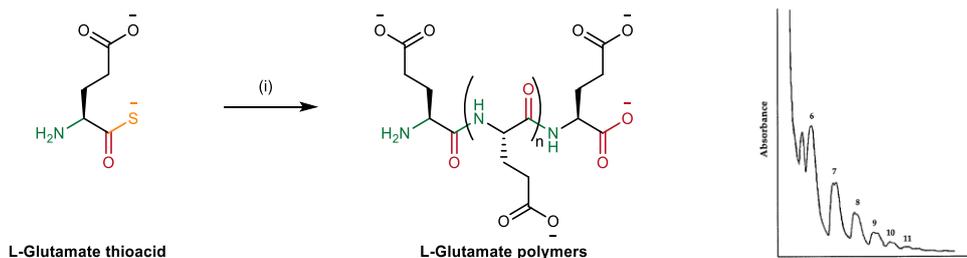
#### 1. 2. 6. 2. 1. Peptides from aminothioacids

Due to their high reactivity and prebiotic plausibility, the possibility of forming prebiotic peptides from aminothioacids has been investigated<sup>[108]</sup> (**Figure 23**). Owing to their unique ability to form disulfides when oxidised in the presence of thiols of other thiocarbonyls, thioacids are readily activated under prebiotic conditions. Upon aqueous oxidation with ferricyanide, aminothioacids have been shown to form highly electrophilic dimers, which quickly react with amine nucleophiles to afford amides in high yields. Additionally, the carbonyl disulfane species generated either as leaving groups from the amidation of aminothioacid dimers, or by cross-oxidation of thioacid and hydrogen sulfide, can also form amides through amine addition, albeit with significantly slower kinetics.



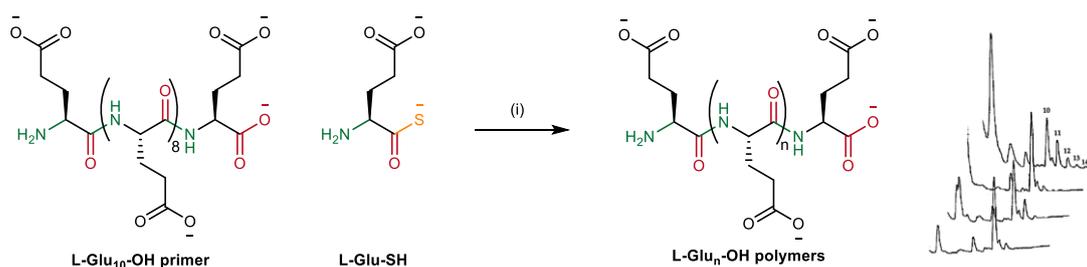
**Figure 23:** Thioacid-based prebiotic peptide coupling strategies.

In spite of such promising reactivity, all efforts to polymerise aminothioacids have led to unsatisfactory results. Orgel demonstrated that the polymerisation of L-glutamate thioacid **L-Glu-SH** was achieved through extended incubation with potassium ferricyanide, sodium carbonate, and magnesium chloride<sup>[108]</sup>. However, those experiments only afforded trace amounts of oligomers longer than 6 residues. In addition, the complex mixtures of products obtained were analysed using HPLC traces alone, which renders accurate yield estimations perilous at best (**Figure 24**).



**Figure 24:** Ferricyanide-mediated polymerisation of **L-Glu-SH** and the associated HPLC trace; (i) **L-Glu-SH** (200 mM),  $K_3[Fe(CN)_6]$  (100 mM),  $NaHCO_3$  (250 mM),  $MgCl_2$  (50 mM), pH 8.2, 50 °C, 14 d.

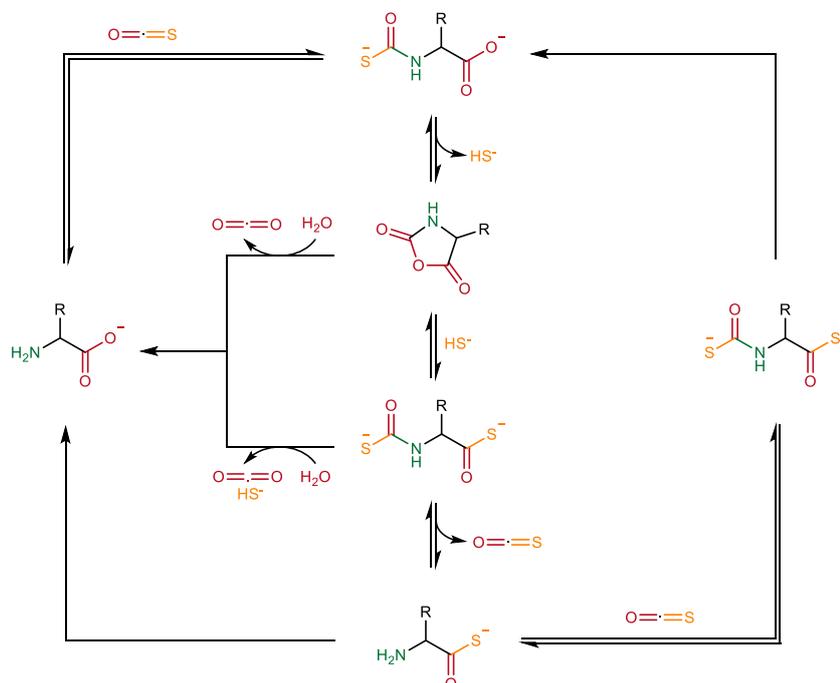
Further attempts to improve polymerisation yields were made by supplementing the mixture with **L-Glu<sub>10</sub>-OH** decamer primers. Unfortunately, no significant improvement resulted from those revised conditions, even in the presence of a 500-fold excess of thioacid monomer. While traces of oligomers up to 15 residues were observed, the total amount of ligations achieved between starting material and longest product remains unchanged at around 5 residues. Furthermore, the assumption made by the authors that no cross-polymerisation involving the glutamate sidechain occurred, and that all traces correspond to linear peptides remains unverified (**Figure 25**).



**Figure 25:** Ferricyanide-mediated polymerisation of **L-Glu-SH** in the presence of **L-Glu<sub>10</sub>-OH** and the associated HPLC trace; (i) **L-Glu<sub>10</sub>-OH** (0.4 mM), **L-Glu-SH** (200 mM),  $K_3[Fe(CN)_6]$  (100 mM),  $NaHCO_3$  (250 mM),  $MgCl_2$  (50 mM), pH 8.2, 50 °C, 24 h.

In addition to poor polymerisation yields, the use of aminothioacids in a prebiotic peptide synthesis context faces another severe limitation: as of the start of our project, a satisfactory prebiotic synthetic pathway to aminothioacids had yet to be proposed. Although the synthesis of simple thioacids from hydrogen sulfide and carboxylates has been reported<sup>[109]</sup>, only minute amounts of thioacid were obtained through such routes. Recently, more sophisticated approaches involving carbonyl sulfide were proposed by Lemán and Ghadiri<sup>[110]</sup>. However, even highly constrained conditions afforded aminothioacid yields no higher than 1.4%. These low yields are attributed to the fact that aminothioacids are highly unstable in the presence of the carbonyl sulfide used for their synthesis (**Figure 26**). Furthermore, the instability of carbonyl sulfide renders

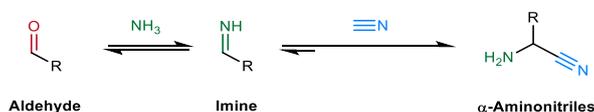
the accumulation of prebiotically relevant amounts of aminothioacids through such pathways implausible at best.



**Figure 26:** Carbonyl sulfide-mediated synthesis of aminothioacids.

### 1. 2. 6. 2. 2. Peptides from aminonitriles

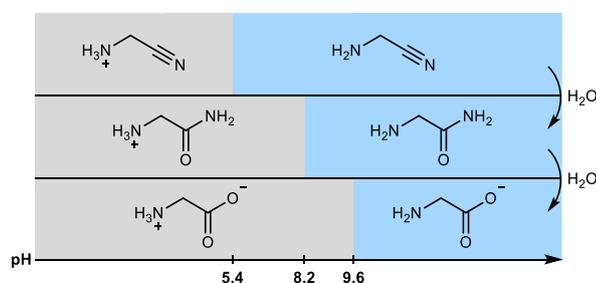
The formation and reactivity of  $\alpha$ -aminonitriles has been of particular interest to prebiotic chemists since the early 1970's<sup>[111]</sup>. A central appeal of aminonitriles comes from their spontaneous emergence from simple, prebiotic building blocks. Indeed, aminonitrile precursors to most prebiotically relevant amino acids have been shown to form in excellent to quantitative yields from ammonia, cyanide and aldehydes through the well-documented Strecker reaction<sup>[112]</sup> (**Figure 27**).



**Figure 27:** Synthesis of  $\alpha$ -aminonitriles *via* the Strecker reaction.

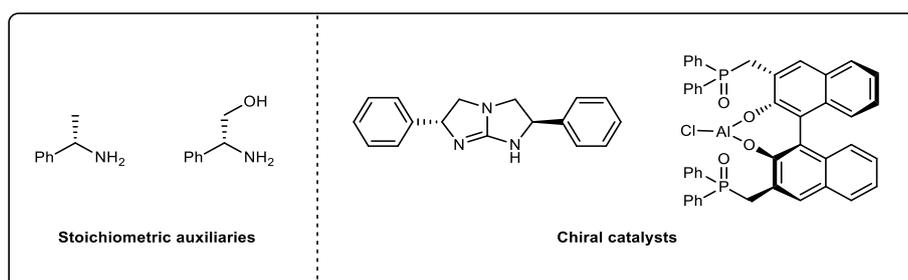
Aminonitriles have been regarded as particularly relevant to abiogenesis due to their ability to yield canonical amino acids through successive hydrolysis (**Figure 28**). Although the hydrolysis of aminonitriles is relatively slow under physiological conditions, it is significantly accelerated by prebiotically plausible catalysts, including simple aldehydes and carbohydrates<sup>[113]</sup>. However, such methods typically stop at the aminoamide stage, which requires prolonged heating and incubation

under strongly acidic or basic conditions to further hydrolyse to amino acids. Although not implausible in the prebiotic environment, such conditions are known to rapidly degrade peptides and proteins into amino acids, which drastically limits the compatibility of such pathways with the formation of catalytic peptides in the absence of protective mechanisms, such as compartmentalisation.



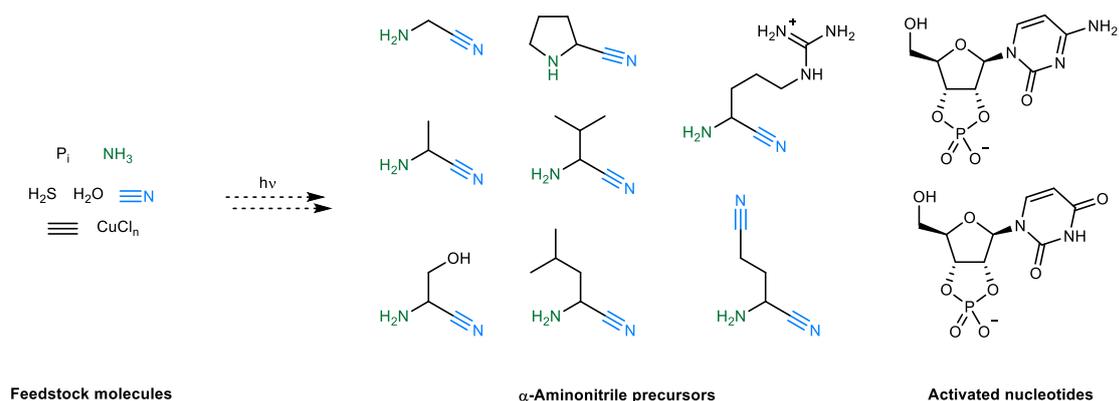
**Figure 28:** Hydrolysis of aminonitriles into aminoamides and amino acids.

Since the early 1960s, the potential of the Strecker reaction for the production of amino acids has ignited an intense search for chiral variants of the reaction<sup>[114]</sup>. This research has produced a range of reagents and catalysts, able to produce aminonitriles with enantiomeric excesses often up to 99% ee. The first generation of reagents was mostly focused on chiral amine auxiliaries that can be used instead of ammonia to direct the attack by cyanide onto a specific face of the imine intermediate. These auxiliaries, while simple and inexpensive, need to be used in stoichiometric quantities and subsequently reduced to afford a free amine moiety. For this reason, the most recent research on the subject has shifted towards the development of low-loading asymmetric catalysts (**Figure 29**).



**Figure 29:** Chiral auxiliaries and catalysts for chiral Strecker synthesis.

Due to the plausible availability of cyanide, ammonia and simple aldehydes on the early Earth,  $\alpha$ -aminonitriles are regarded as plausible components of the prebiotic environment, and are known to spontaneously arise from aqueous cyanide solutions<sup>[115]</sup>. Furthermore,  $\alpha$ -aminonitrile analogues of nine essential amino acids have been shown to emerge from a model cyanosulfidic protometabolism<sup>[97]</sup> (**Figure 30**).

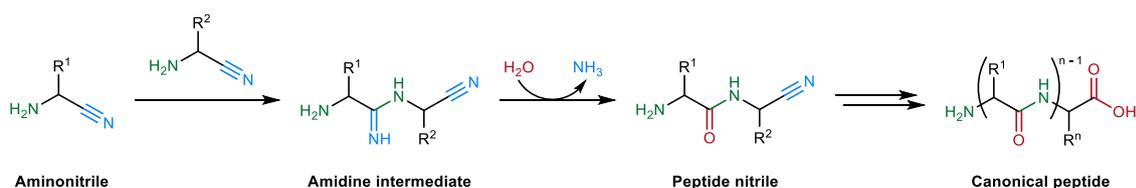


**Figure 30:** Formation of aminonitriles and activated nucleotides from a cyanosulfidic protometabolism.

Besides their prebiotic availability,  $\alpha$ -aminonitriles exhibit a number of unique chemical properties. The geminal electron-withdrawing nitrile group significantly lowers the basicity of the amine moiety, resulting in a strongly suppressed  $pK_{aH}$  value of *ca.* 5.3<sup>[116]</sup>. This shift in  $pK_{aH}$  has profound implications of the reactivity of  $\alpha$ -aminonitriles:

- By expanding the range of pH in which  $\alpha$ -aminonitriles exist as nucleophilic amines down to *ca.* 4.3, it broadens the range of conditions under which peptide couplings can occur, and closes the gap of mutual reactivity between carboxylic acids and amines.
- By permitting the use of neutral and acidic pH, where hydroxyl anions are present in lower concentrations,  $\alpha$ -aminonitriles allow for the use of activation strategies relying on intermediate species that would be rapidly hydrolysed under alkaline conditions, and hence have found little application in prebiotic peptide coupling studies.

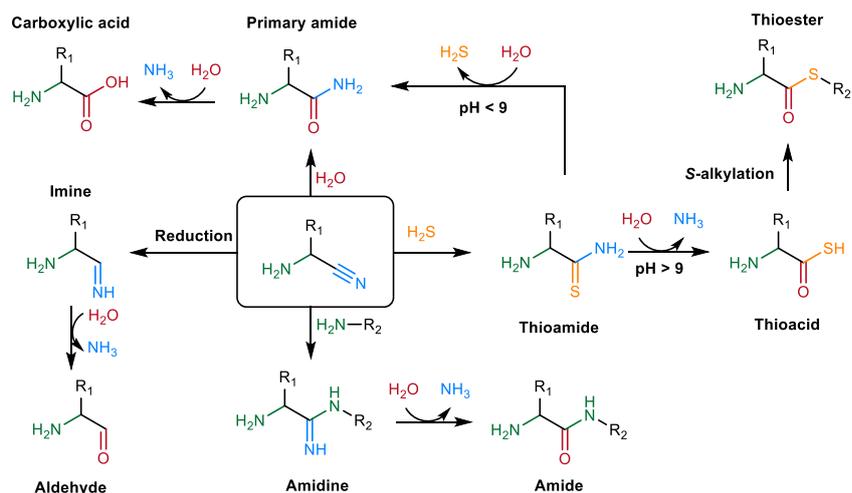
The ability of nitriles to undergo nucleophilic addition led to the hypothesis that aminonitriles could self-condense into amidine-linked peptides, which could in turn hydrolyse to form canonical peptides (**Figure 31**). Although attractive due to its simplicity and to the availability of aminonitriles in the prebiotic environment, this pathway proved extremely inefficient, and mostly results in the formation of diketopiperazines and dipeptides<sup>[117]</sup>.



**Figure 31:** Hypothesised elongation pathway from the self-condensation of aminonitriles.

### 1. 3. Project aims

Despite the inefficiency of their self-condensation, the rich reactivity of nitriles opens the way to a variety of yet-unexplored prebiotic reactional pathways. Aside from yielding  $\alpha$ -aminoamides and  $\alpha$ -amino acids upon hydrolysis<sup>[118]</sup>, nitriles have been shown to afford imines and aldehydes upon aqueous reduction<sup>[119]</sup>; amidines and amides upon direct amine addition<sup>[120]</sup>; and thioamides<sup>[97]</sup> and thioacids<sup>[121 – 122]</sup> through aqueous thiolysis (**Figure 32**).

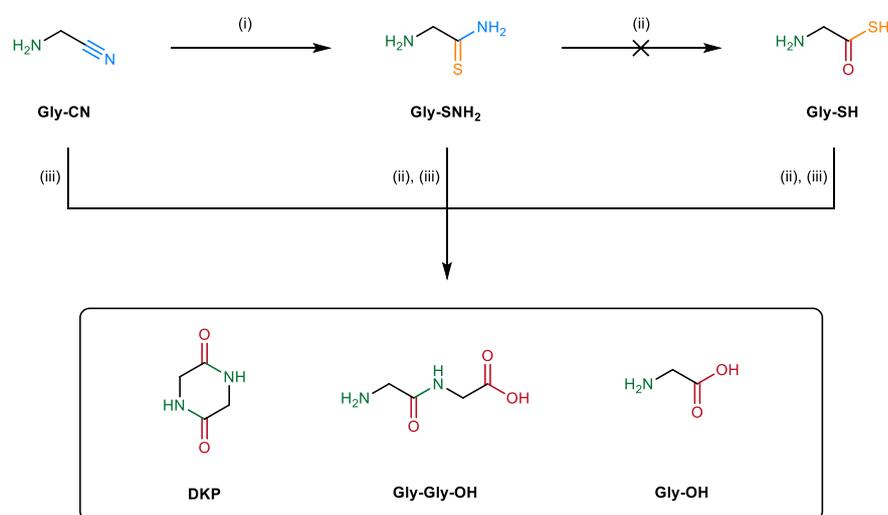


**Figure 32:** Reactivity network of  $\alpha$ -aminonitriles.

The aim of this research project is to explore the reactivity of aminonitriles and their potential applications to the prebiotic synthesis of peptides. Owing to their high reactivity towards activation and amidation, the possibility of transitioning from aminonitrile to aminothioacid derivatives became central to our investigation. This work presents the first high-yielding, prebiotic route to peptide thioacids and amidothioacids reported to date. Then, the controlled, chemoselective, and highly-efficient oxidative ligation of amidothioacids and aminonitriles is demonstrated. This methodology serves as the foundation for a continuous, one-pot peptide elongation cycle, which we show to afford polypeptides in unprecedentedly high yields.

## 2. Attempted oligomerisation of aminonitriles

In order to both confirm the aforescribed results and explore alternative routes towards aminothioacids, we set out to investigate the prebiotic synthesis of aminothiocarbonyls from aminonitriles (**Figure 33**). To this end, we attempted to prepare glycine thioamide **Gly-SNH<sub>2</sub>** through the thiolysis of glycine nitrile **Gly-CN** using sodium hydrosulfide as thionating agent. To our satisfaction, the reaction afforded **Gly-SNH<sub>2</sub>** in quantitative yields after 24 h of incubation at room temperature. Encouraged by these results, we attempted to hydrolyse **Gly-SNH<sub>2</sub>** into glycine thioacid **Gly-SH**. Unfortunately, all attempts resulted in complex mixtures of products, whose main components were identified as glycine **Gly-OH**, diglycine **Gly-Gly-OH**, and 2,5-diketopiperazine **DKP**. No thioacid **Gly-SH** was observed.

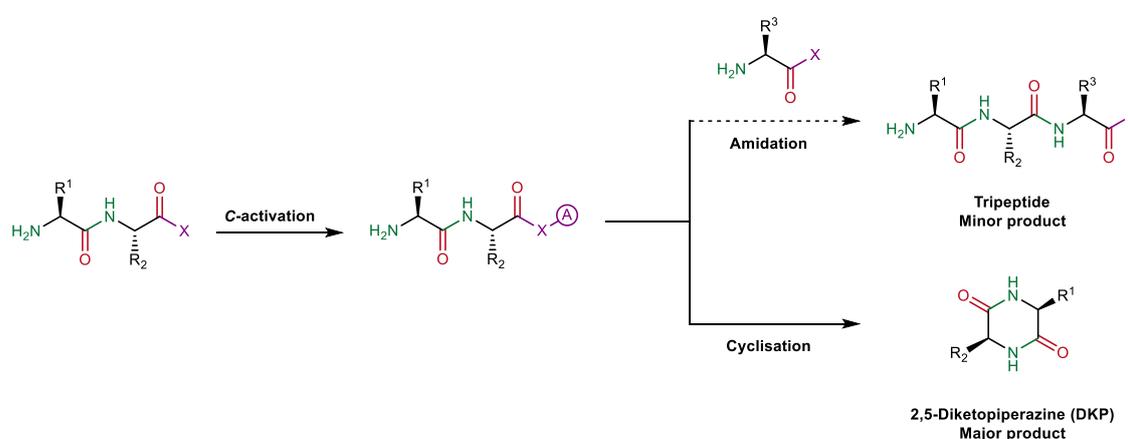


**Figure 33:** Hydrogen sulfide mediated reaction network of glycine nitrile **Gly-CN**; (i) NaSH (10 eq.), pH 9.0, RT, 24 h; (ii) pH 9.0, 60 °C, 24 h; (iii) K<sub>3</sub>[Fe(CN)<sub>6</sub>] (3 eq.), pH 9.0, RT, 30 min.

In the event that an efficient route to free aminothioacids would emerge, we deemed it necessary to reassess the results reported by Orgel regarding the polymerisation of free aminothioacids<sup>[109]</sup>. To this end, the polymerisation of **Gly-SH** was attempted either through high-temperature incubation, or oxidative activation mimicking Orgel's conditions. However, these reactions also descended into complex mixtures, with varying proportions of **Gly-OH**, **Gly-Gly-OH**, **DKP** as their main components. Similar conditions were applied to **Gly-CN** and **Gly-SNH<sub>2</sub>**, whose polymerisation had previously been investigated, albeit with extremely poor yields<sup>[117, 123]</sup>. As expected, those experiments yielded similar mixtures as those obtained from the polymerisation of **Gly-SH**.

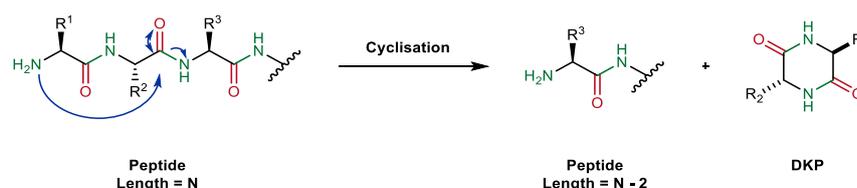
The repeated failures met by all attempts towards polymerising amino acid precursors and their derivatives has highlighted a series of roadblocks, which hinder all polymerisation strategies relying on C-activation, independently of the free monomer used:

- **The formation of DKP** consistently outcompetes oligomerisation (**Figure 34**). Although several free elongation strategies boast high to quantitative yields for the formation of dimers<sup>[124]</sup>, those yields generally plummet upon transitioning to the formation of tripeptides and longer oligomers. Unfortunate as it may be, this outcome was predictable: at prebiotically relevant concentrations, the cyclisation of C-activated free dipeptides is kinetically favoured over the ligation into tripeptides.



**Figure 34:** DKP formation from C-activated dipeptides.

- **The N-terminal self-degradation** of peptides compounds this issue (**Figure 35**). Bada has shown the half-life of N-uncapped peptides to be 35 days under physiological conditions<sup>[125]</sup>. This point is of the utmost significance in a prebiotic framework, which heavily relies on the notion that biogenic materials may have accumulated over thousands of years before the first living organisms emerged. In this context – and given the sluggish speed at which the proposed polymerisations proceed – it is highly unlikely that peptides would reach lengths allowing for catalytic activity, and would accumulate in sufficient concentrations to play any role in abiogenesis.

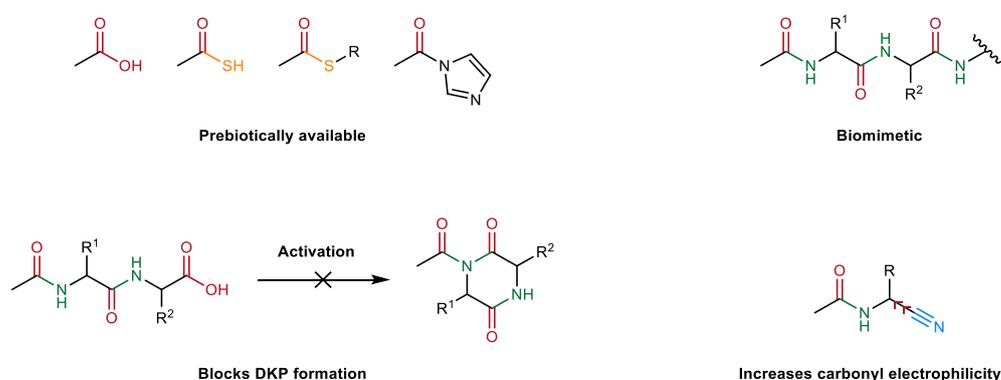


**Figure 35:** DKP formation N-terminal self-immolation of peptides.

### 3. Prebiotic synthesis of *N*-capped peptide thioacids

The severe limitations faced by polymerisation strategies led us to consider alternative approaches to prebiotic peptide elongation. As all of the aforementioned issues arise from the free *N*-terminal amine of the elongating peptide, we turned our attention towards prebiotically plausible *N*-capping strategies. Among those strategies, the *N*-acetylation of aminonitriles seemed the most promising, as it offers a range of advantages (**Figure 36**):

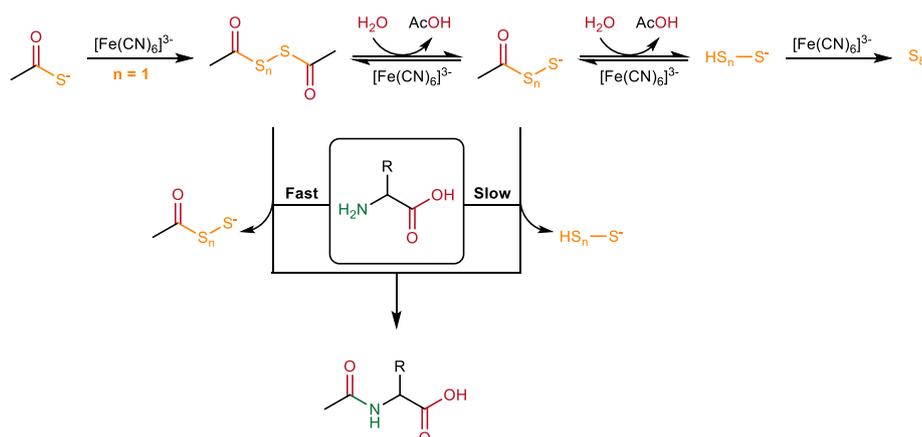
- **Prebiotic plausibility:** The acetylation of amines and amino acids by prebiotically abundant acetyl donors has been extensively demonstrated<sup>[126]</sup>. Those acetyl donors are generally simple and plausible, and proceed in near-quantitative yields in water.
- **No DKP formation:** The ability of *N*-acetylated amines to engage in nucleophilic attacks is greatly reduced. Consequently, the formation of DKP, either from dipeptide cyclisation or *N*-terminal self-immolation, is suppressed by *N*-acetylation.
- **Biomimetism:** *N*-Capping strategies are the most common solution used by living organisms to combat DKP formation and protein self-immolation. The ubiquity of such strategies is demonstrated by the fact that *ca.* 80% of all proteins and peptides in eukaryotes are *N*-acetylated<sup>[127]</sup>. Therefore, unravelling a prebiotic peptide elongation pathway based on *N*-acetylated peptides would bring to light a direct connection between abiogenesis and extant biology.
- **Enhanced C-vicinal electrophilicity:** The electron-withdrawing effect of the resulting  $\alpha$ -amides greatly increases the electrophilicity of neighbouring nitrile and carbonyl groups. This property was hypothesised, and subsequently shown to both accelerate reactions involving said nitriles and carbonyls, which in turn allows for kinetic selectivity in the thiolysis of *N*-capped *versus* free aminonitriles.



**Figure 36:** Key benefits of *N*-acetylated aminonitriles and derivatives.

### 3. 1. From Strecker mixtures to *N*-acetyl aminonitriles

In 1997, Orgel described the oxidative acetylation of amino acids using thioacetate and ferricyanide in water<sup>[126]</sup> (**Figure 37**). The reaction proceeds through the oxidative dimerization of thioacetate into highly reactive acetic dithioperoxyanhydride, which rapidly ligates with amino acids **AA-OH** to form *N*-acetyl amino acids **Ac-AA-OH**. The ligation releases ethane(dithioperoxoic)-*S*-acid, which also ligates with amino acids, albeit at speeds 100 times lower than in the case of acetic dithioperoxyanhydride. Alternatively, excess oxidant can be used to reoxidise *n*-thioperoxy acids and with deprotonated sulfur-bearing species (thioacetate, other *n*-thioperoxy acids, polysulfides) to regenerate a collection of highly electrophilic *n*-peroxyanhydrides and *n*-thioperoxy acids. In turn, these electrophiles rapidly ligate with the remaining amino acids to form **Ac-AA-OH**, and with water to form acetic acid. If the starting stoichiometry of ferricyanide is sufficiently large (typically, 3 eq. with regards to thioacetate), the only products are *N*-acetyl amino acids, acetic acid, ferrocyanide, and elemental sulfur.

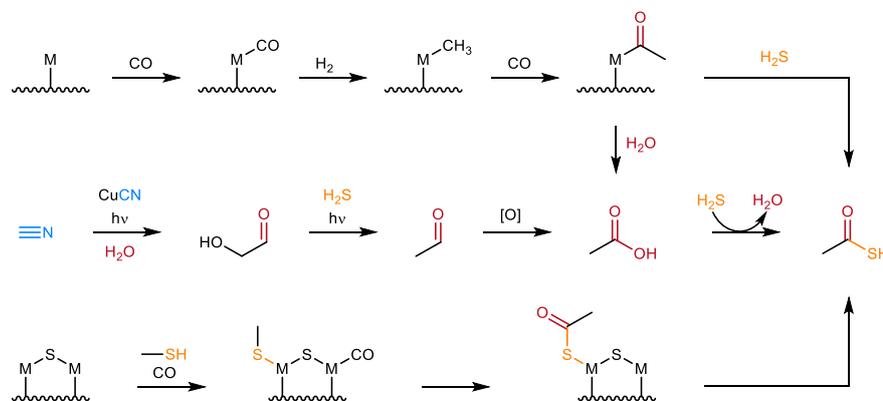


**Figure 37:** Proposed mechanism for the oxidative acetylation of amino acids.

Those conditions are extremely appealing in a prebiotic context due to several factors:

- **Simplicity and high availability of the reagents:** The reaction only requires two simple reagents to proceed: thioacetate, and ferricyanide. Several pathways leading to thioacetate under prebiotic conditions have been demonstrated. Most of those pathways involve the metal-catalysed carbonylation of  $\text{C}_1$  units by carbon monoxide (**Figure 38**). The carbonylation of methanethiol was shown to directly produce thioacetic acid<sup>[128]</sup>. Similarly, the carbonylation of methane on metal-rich rock surfaces was observed to generate both thioacetic and acetic acid<sup>[129]</sup>, which can itself be converted back to thioacetate by hydrogen sulfide. More recently, a photochemical pathway was shown to yield acetaldehyde<sup>[119]</sup>, which is readily converted to acetic acid prebiotic oxidants, such

as chlorates <sup>[130-131]</sup>, and offers another route to thioacetate. In addition, ferricyanide is readily formed from two of the simplest and most naturally ubiquitous chemical building blocks: ferric iron, and cyanide.



**Figure 38:** Prebiotic pathways to the synthesis of thioacetic acid.

- **Continuity with the cyanosulfidic protometabolism:** In a landmark study, Sutherland *et al.* demonstrated the feasibility of a prebiotic network producing both activated ribonucleotides and aminonitriles. This network arises from a restricted set of components, which includes hydrogen sulfide, iron, and cyanide. Therefore, exploiting the same reactants that produced aminonitriles to selectively elongate them into peptides would constitute a direct link between cyanosulfidic protometabolism and canonical protein structures.
- **Rapid, mild, quantitative conversions:** Orgel's oxidative acetylation affords quantitative yields within minutes at room temperature, and its only byproducts are entirely unreactive towards *N*-acetyl aminonitriles. Knowing that the formation of biogenic substrates is at perpetual odds with degradation pathways, including thermal hydrolysis, these mild conditions set a favourable stage for the accumulation of peptides.

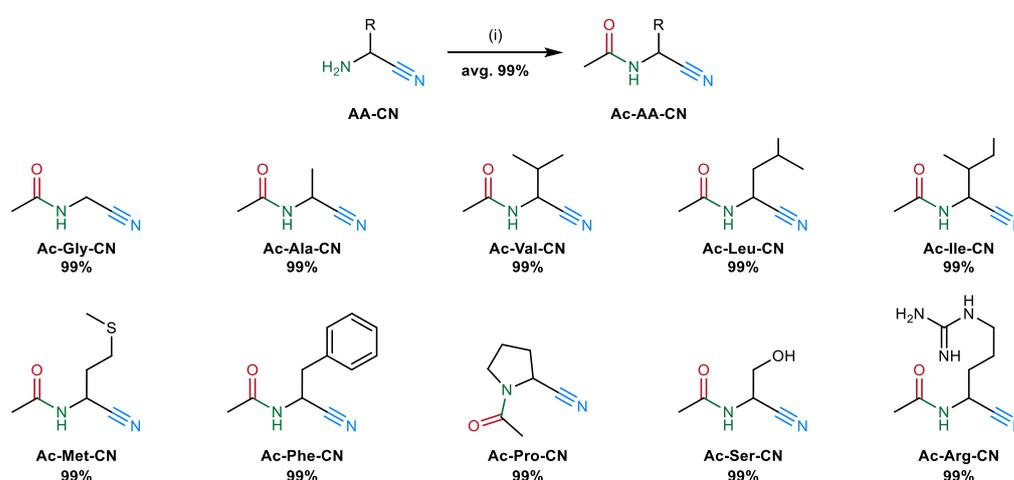
### 3. 1. 1. Strecker synthesis of aminonitriles

In light of these remarkable properties, we set out to apply Orgel's methodology to the acetylation of aminonitriles. To do so, we began by preparing a range of biologically relevant aminonitriles from commercially available aldehyde precursors **1a-g** *via* Strecker synthesis (**Figure 39**). The reaction afforded moderate to high yields across a range of 8 aminonitriles (31 – 95%; 67% avg.). These aminonitriles were chosen due to their demonstrated emergence from prebiotic networks<sup>[97]</sup>, and to their wide variety of sidechains, which would allow the generality of our elongation strategy to be asserted.



### 3. 1. 2. Oxidative acetylation of aminonitriles

With a range of aminonitriles in hand, we went on to adapt Orgel's methodology to the synthesis of *N*-acetyl aminonitriles **Ac-AA-CN**. Pleasingly, quantitative yields were obtained with all residues tested, even in the presence of nucleophilic (*e.g.* serine), oxidisable (*e.g.* methionine) or bulky (*e.g.* phenylalanine) sidechains (**Figure 41**). In addition, the reaction proceeds in under 20 minutes<sup>2</sup> at room temperature, using only a threefold excess of thioacetate, and maintains near-quantitative yields at aminonitrile concentrations ranging from 1.0 M to 1.0  $\mu$ M.

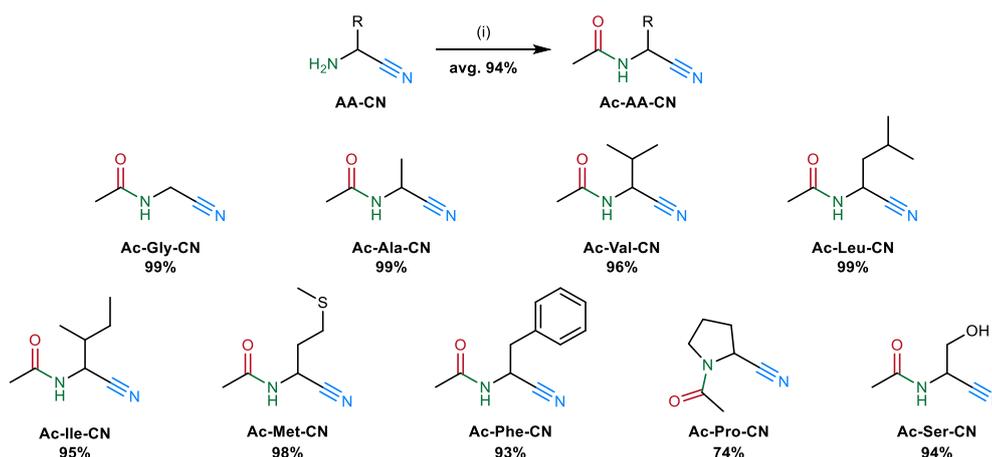


**Figure 41:** Oxidative acetylation of aminonitriles; (i) optimised conditions: **AA-CN** (1 eq.) AcSK (3 eq.),  $K_3[Fe(CN)_6]$  (9 eq.),  $H_2O$  (50 mM **AA-CN**), pH 9.0, RT, 20 min.

### 3. 1. 3. Preparative synthesis of *N*-acetyl aminonitriles

In spite of extremely high <sup>1</sup>H NMR yields, preparing *N*-acetyl aminonitriles **Ac-AA-CN** on a gram scale using Orgel's method proved remarkably challenging in general, and infeasible in the presence of polar sidechains (*e.g.* arginine). These difficulties chiefly arise from the large amounts of polysulfides and potassium ferrocyanide produced by the reaction, which could not be readily separated from organic materials by extraction, recrystallisation, or chromatography. For this reason, we devised a rapid, quantitative acetylation strategy, allowing for the preparation of pure *N*-acetyl aminonitriles on a multigram scale. This procedure, which uses acetic anhydride as single reagent in water pH 9.0, afforded *N*-acetyl aminonitriles in high to quantitative yields (74 – 99%; 94% avg.), which were easily purified by flash column chromatography (**Figure 42**).

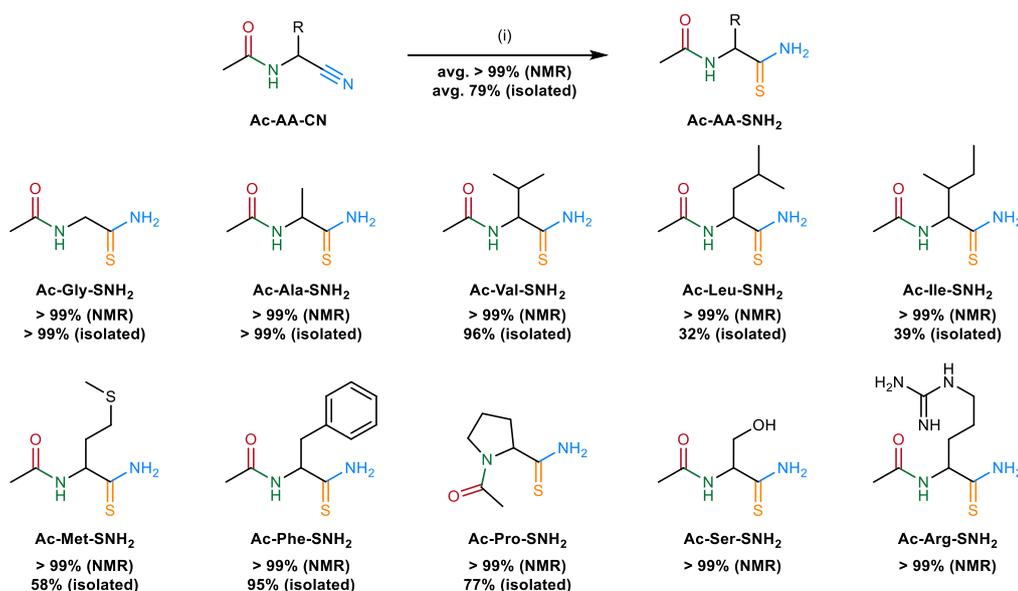
<sup>2</sup> The reaction proceeds faster than <sup>1</sup>H NMR acquisition speed (*ca.* 1 min) under the conditions described in **Figure 40** for **Gly-CN**, **Ala-CN**, **Val-CN**, and **Phe-CN**; 20 minutes was set as a standard reaction time to homogenise the experimental conditions across substrates. **Gly-CN** and **Pro-CN** were purchased as their hydrochloride salts, while the one-pot synthesis and reaction of **Arg-CN** is detailed in section 7.



**Figure 42:** Preparation of *N*-acetyl aminonitriles; (i) **AA-CN** (1 eq.), Ac<sub>2</sub>O (3 eq.), H<sub>2</sub>O (0.2 M **AA-CN**), pH 9.0, RT, 10 min.

### 3. 2. Thiolysis of *N*-acetyl aminonitriles

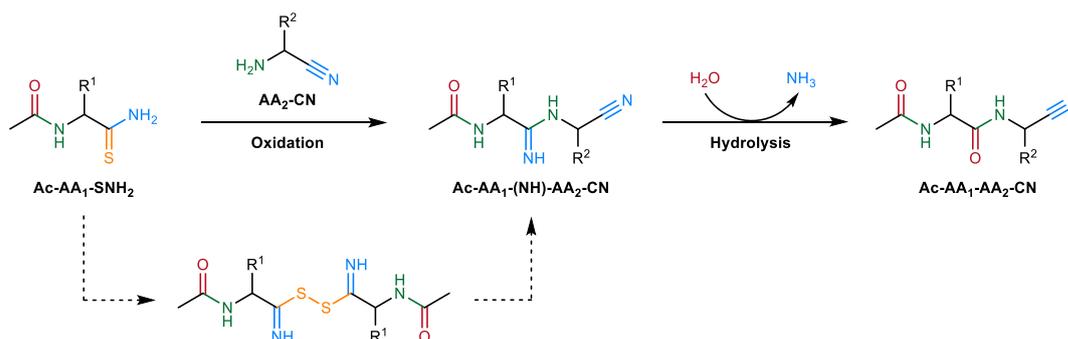
Having demonstrated the prebiotic synthesis of a wide range of *N*-acetyl aminonitriles, we set out to convert those into *N*-acetyl thiocarbonyls. Satisfyingly, we found that the aqueous thiolysis of *N*-acetyl aminonitriles, using sodium hydrosulfide as sole reagent, afforded *N*-acetyl aminothioamides in quantitative yields. The full conversion of most aminonitriles was observed after 24 to 48 h at room temperature, while some substrates required incubation times extended up to 92 h (**Figure 43**). Furthermore, the absence of organic impurities made the isolation of most thioamide products by flash chromatography straightforward, although substantial degradation of to their nitrile and amide analogues was observed for certain sidechains (**Leu, Ile, Met**).



**Figure 43:** Thiolysis of *N*-acetyl aminonitriles; (i) **Ac-AA-CN** (1 eq.), NaSH · (H<sub>2</sub>O)<sub>n</sub> (10 eq.), H<sub>2</sub>O (50 mM **Ac-AA-CN**), pH 9.0, RT, 24 – 96 h.

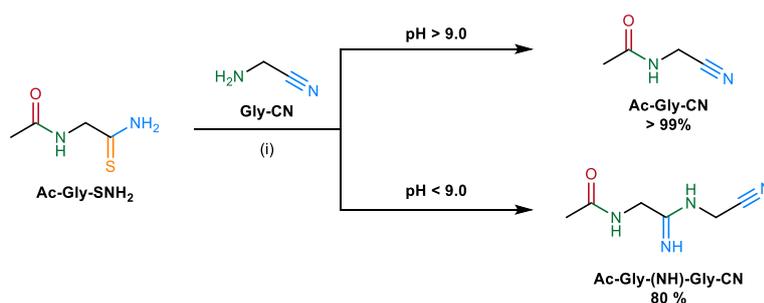
### 3. 3. Attempted amidation of *N*-acetyl thioamides

From this point, we considered the possibility of applying Orgel's oxidative activation strategy to the ligation of *N*-acetyl aminothioamides with aminonitriles. We hypothesised that if thioamides could be activated using conditions related to those used on thioacetate, the resulting dimer could react in a similar fashion with aminonitriles to form amidine-linked propeptides **Ac-AA<sub>1</sub>-(NH)-AA<sub>2</sub>-CN**. Those amidines could then be hydrolysed to afford *N*-acetyldipeptide nitriles (**Figure 44**).



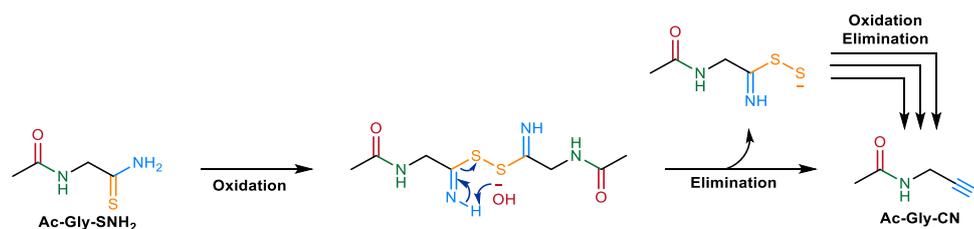
**Figure 44:** Proposed oxidation-ligation-hydrolysis pathway to *N*-acetyl peptide nitriles.

To test our hypothesis, we exposed *N*-acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>** and **Gly-CN** to the oxidative conditions described in **Figure 40**. To our surprise, little to no amidine **Ac-Gly-(NH)-Gly-CN** was observed. Instead, the reaction afforded quantitative yields of **Ac-Gly-CN**. Interestingly, performing the reaction at neutral and acidic pH, as opposed to the mildly alkaline acetylation conditions (pH 9.0), afforded the expected amidine in yields as high as 80% (**Figure 45**).



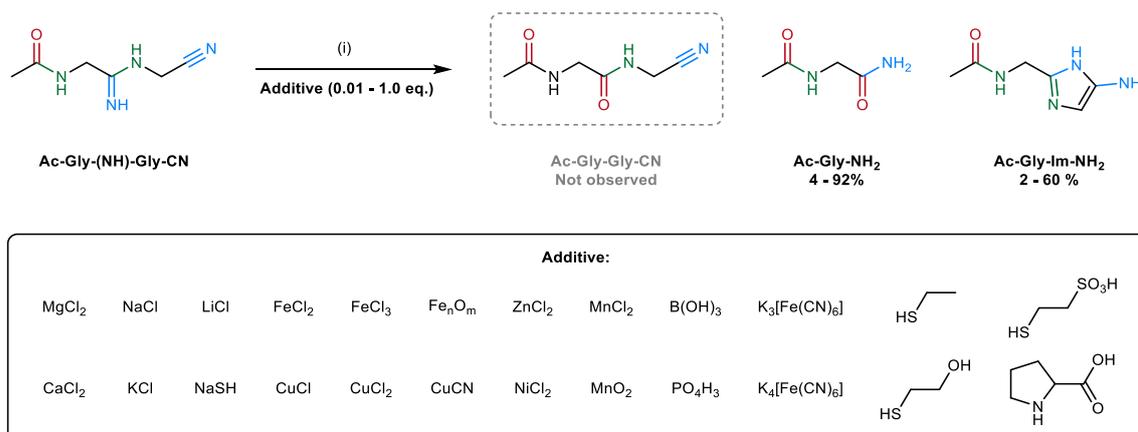
**Figure 45:** Oxidative ligation of *N*-acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>**; (i) **Ac-Gly-SNH<sub>2</sub>** (1 eq.), **Gly-CN** (2 eq.),  $K_3[Fe(CN)_6]$  (3 eq.),  $H_2O$  (50 mM **Ac-Gly-SNH<sub>2</sub>**), RT, 20 min.

Both the structure of the expected intermediate and the dependency on pH of the regeneration of **Ac-Gly-CN** suggest that the reaction proceeds by deprotonating the transient oxidised dimer, leading to the elimination of sulfur-containing leaving groups. The higher the pH, the more this elimination process would be expected to outcompete nucleophilic addition due to the higher concentration of hydroxyl ions in the solution (**Figure 46**).



**Figure 46:** Proposed mechanism for the oxidative elimination of *N*-aminothioamides.

Despite this unforeseen reactivity, a set of conditions was found that could afford amidine propeptides. Unfortunately, all attempts at hydrolysing those amidines into dipeptides proved unfruitful. In spite of an extensive screening for additives, both organic and inorganic, and reaction conditions, no significant amount of the desired peptide could be observed. Those additives include naturally-occurring buffers (*e.g.* phosphate, borate); halide equivalents of common oxidation states of geochemically abundant metals (*e.g.* copper(I) and (II), iron(II) and (III)); and other prebiotic additives (*e.g.* H<sub>2</sub>S, mercaptoethanol). In addition, the effects of potassium ferricyanide and the reduced byproduct of its oxidative coupling, potassium ferrocyanide, were studied (**Figure 47**).

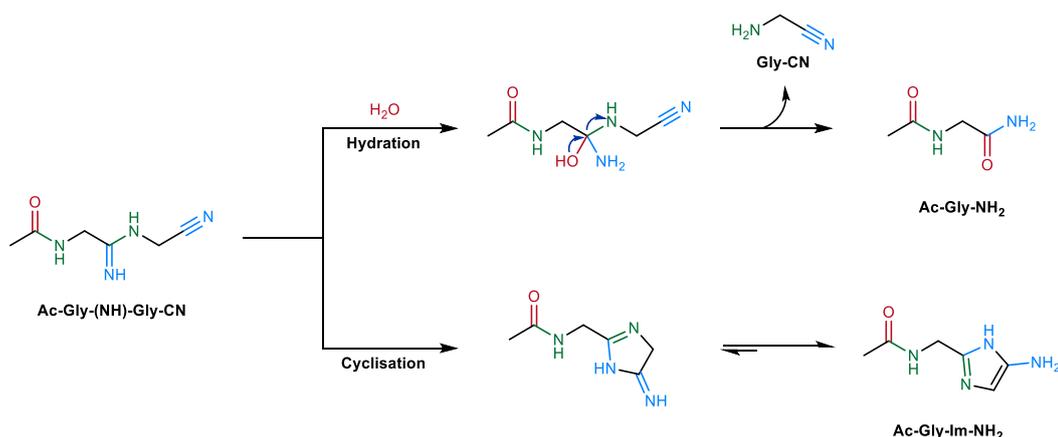


**Figure 47:** Attempted hydrolysis of **Ac-Gly-(NH)-Gly-CN** to **Ac-Gly-Gly-CN**; (i) **Ac-Gly-(NH)-Gly-CN** (1 eq.), **catalyst** (0.01 to 1.0 eq.), H<sub>2</sub>O (50 mM **Ac-Gly-(NH)-Gly-CN**), pH 3.0 to 12.0, RT to 80 °C, 30 min to 24 h.

Regardless of the additive, **Ac-Gly-(NH)-Gly-CN** proved unreactive at room temperature. Oppositely, samples incubated at 80 °C and above or in the presence of organic thiols rapidly devolved into intractable mixtures. However, interpretable results were obtained for most additives after 24 h at 60 °C and pH 9.0. No trace of **Ac-Gly-Gly-CN** was observed. Instead, the reaction afforded two main products in variable proportions: *N*-acetylglycinamide **Ac-Gly-NH<sub>2</sub>**, and *N*-acetylglycine aminoimidazole **Ac-Gly-Im-NH<sub>2</sub>** (**Table 1**).

Table 1: Attempted conversion of Ac-Gly-(NH)-Gly-CN to Ac-Gly-Gly-CN at pH 9.0, 60 °C.			
Entry	Additive	Yields (%)	
		Ac-Gly-NH <sub>2</sub>	Ac-Gly-Im-NH <sub>2</sub>
1	-	81	3
2	Pi 0.5 M	11	60
2	BBS 0.5 M	17	8
4	MgCl <sub>2</sub> (1 eq.)	92	-
5	CaCl <sub>2</sub> (1 eq.)	42	-
7	FeCl <sub>2</sub> (1 eq.)	7	11
8	FeCl <sub>3</sub> (1 eq.)	-	25
9	NiCl <sub>2</sub> (1 eq.)	31	10
11	CuCl <sub>2</sub> (1 eq.)	60	-
12	ZnCl <sub>2</sub> (1 eq.)	49	-
13	MnCl <sub>2</sub> (1 eq.)	63	12
18	NaSH · (H <sub>2</sub> O) <sub>n</sub> (10 eq.)	4	-

This unfortunate outcome finds a rationale in the reactivity of the amidine linkage. On the one hand, hydration of the N=C double bond leads to the formation of a tetrahedral intermediate, which can realistically collapse into three different species: the desired dipeptide **Ac-Gly-Gly-CN**, the observed byproduct **Ac-Gly-NH<sub>2</sub>**, or the starting material **Ac-Gly-(NH)-Gly-CN**. The strong electron-withdrawing effect of its nitrile moiety is likely to make the aminonitrile the best leaving group. This elimination pathway leads to the irreversible formation of **Ac-Gly-NH<sub>2</sub>**. On the other hand, the N=C double bond is ideally positioned to engage in an intramolecular cyclisation onto the neighbouring nitrile. Tautomerisation of the resulting intermediate affords the highly stable, aromatic **Ac-Gly-Im-NH<sub>2</sub>**. This tentative mechanism was subsequently confirmed by studies into the reactivity of amidine-linked peptides by other members of the Powner group (**Figure 48**).

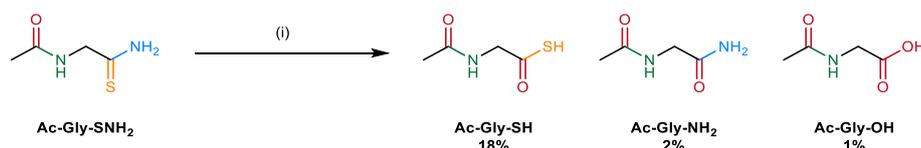


**Figure 48:** Proposed mechanism for the formation of **Ac-Gly-NH<sub>2</sub>** and **Ac-Gly-Im-NH<sub>2</sub>** from **Ac-Gly-(NH)-Gly-CN**.

### 3. 4. From *N*-acetyl thioamides to *N*-acetyl thioacids

#### 3. 4. 1. Hydrolysis of *N*-acetyl thioamides

Undeterred by these results, we proceeded to investigate the possibility of hydrolysing *N*-acetyl aminothioamides **Ac-AA-SNH<sub>2</sub>** into *N*-acetyl aminothioacids **Ac-AA-SH** using **Ac-Gly-SNH<sub>2</sub>** as a model. To our surprise, **Ac-Gly-SNH<sub>2</sub>** proved remarkably resistant to hydrolysis. Room temperature experiments led to no observable conversion after 48 h of incubation at pH 9.0. Even at 60 °C, only 18% of the starting material was hydrolysed, and converted to a 8.5:1:0.5 distribution of **Ac-Gly-Ac-Gly-SNH<sub>2</sub>**, **Ac-Gly-NH<sub>2</sub>**, and **Ac-Gly-OH** (**Figure 49**).



**Figure 49:** Hydrolysis of **Ac-Gly-SNH<sub>2</sub>**; **Ac-Gly-SNH<sub>2</sub>** (1.0 eq.),  $\text{H}_2\text{O}$  (50 mM **Ac-Gly-SNH<sub>2</sub>**), pH 9.0, 21 – 60 °C, 48 h.

The pH at which the reaction is carried out appears to be the main factor controlling the distribution of products, as more alkaline pHs tend to yield a higher ratio of thioacid over amides and carboxylic acids. This observation can be rationalised by examining the mechanism of the hydrolysis: upon addition of water or hydroxide, a tetrahedral intermediate forms, which can collapse either into a thioacid or an amide, depending on whether ammonia or hydrogen sulfide is expelled. The elimination of hydrogen sulfide being dependent on the protonation state of the sulfide leaving group, strongly alkaline pH increases the proportion of thioacid products by favouring the formation of thiolates, which are less likely to collapse into primary amides. Over

time, both products will eventually hydrolyse to their carboxylic acid counterpart. In the case of thioacids, the hydrolysis is dependent on the protonation of the sulfur atom, and is virtually unobservable in oxidant-free conditions above pH 3.0, while the hydrolysis of primary amides requires incubation under strongly acidic or alkaline conditions to proceed at room temperature (Figure 50).

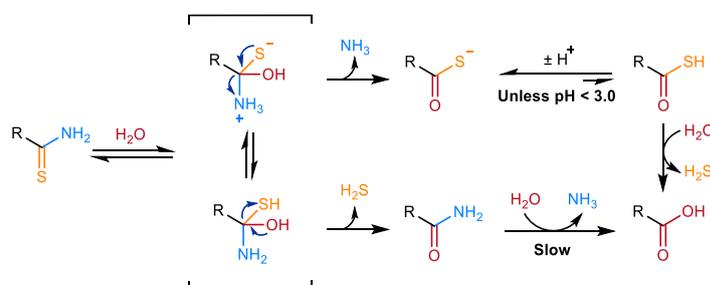


Figure 50: Proposed mechanism for the hydrolysis of thioamides.

In order to both speed up the process and increase the proportion of thioacid produced by the reaction, a range of prebiotically plausible additives and catalysts similar to those described in Figure 47 were investigated (Figure 51)

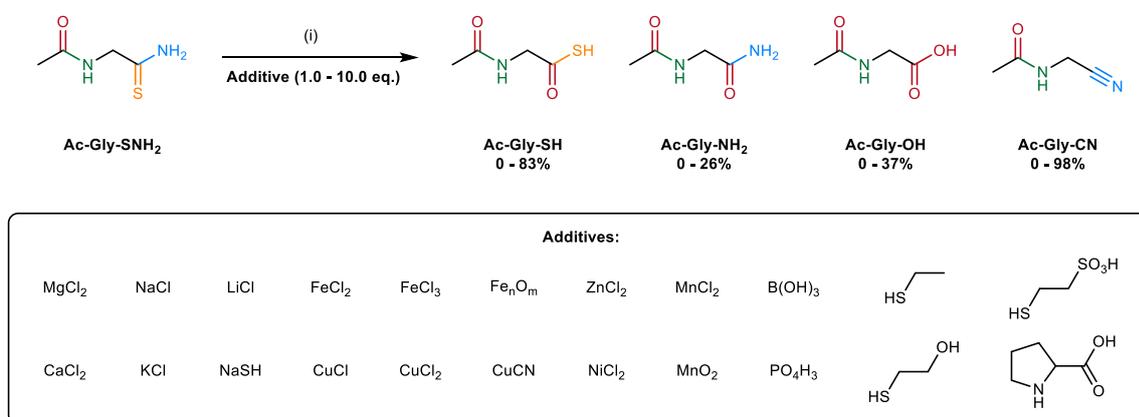


Figure 51: Optimisation of the hydrolysis of **Ac-Gly-SNH<sub>2</sub>**; (i) **Ac-Gly-SNH<sub>2</sub>** (1.0 eq.), additive (1.0 to 10.0 eq.), H<sub>2</sub>O (50 mM **Ac-Gly-SNH<sub>2</sub>**), pH 9.0, 60 °C, 24 h.

To our great satisfaction, we found that sodium hydrosulfide, which is involved in the previous thiolysis step, turned out to be the best catalyst. A secondary screening for conditions established that in the presence of 10 eq. of hydrogen sulfide, yields of 83% could be consistently achieved by incubating **Ac-Gly-SNH<sub>2</sub>** at 60 °C and pH 9.0 for 24 h. Remarkably, we also observed that performing the reaction in the presence of several metal additives (FeCl<sub>n</sub>, CuCl<sub>n</sub>, MnCl<sub>2</sub>, ZnCl<sub>2</sub>) led to near-quantitative regeneration of **Ac-Gly-CN** from **Ac-Gly-SNH<sub>2</sub>** (Table 2).

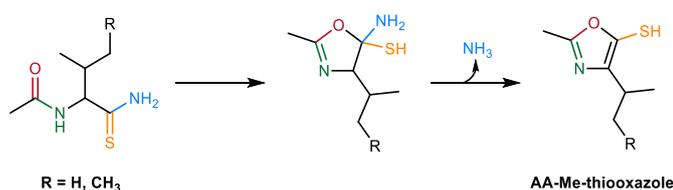
Table 2: optimised hydrolysis of <i>N</i> -acetyl glycine thioamide at pH 9.0, 60 °C for 24 h.						
Entry	Additive	Conversion	Yields (%)			
			Ac-Gly-SH	Ac-Gly-NH <sub>2</sub>	Ac-Gly	Ac-Gly-CN
1	-	18%	15	2	1	-
2	Pi 0.5 M	60%	24	36	-	-
3	BBS 0.5 M	75%	49	26	-	-
4	MgCl <sub>2</sub> (1 eq.)	26%	23	3	-	-
5	CaCl <sub>2</sub> (1 eq.)	2%	-	-	-	-
6	AlCl <sub>3</sub> (1 eq.)	72%	55	17	-	-
7	FeCl <sub>2</sub> (1 eq.)	> 99%	33	77	-	-
8	FeCl <sub>3</sub> (1 eq.)	49%	16	33	-	-
9	NiCl <sub>2</sub> (1 eq.)	65%	54	8	3	-
10	CuCl (1 eq.)	> 99%	-	2	-	98
11	CuCl <sub>2</sub> (1 eq.)	> 99%	-	3	-	97
12	ZnCl <sub>2</sub> (1 eq.)	> 99%	-	15	4	81
13	MnCl <sub>2</sub> (1 eq.)	> 99%	-	18	2	80
14	K <sub>3</sub> [Fe(CN) <sub>6</sub> ] (1 eq.)	-	-	6	-	94
15	K <sub>4</sub> [Fe(CN) <sub>6</sub> ] (1 eq.)	25	11	10	4	-
16	Mercaptoethanol (1 eq.)	> 99	40	23	37	-
17	Na <sub>2</sub> S · 9 H <sub>2</sub> O (10 eq.)	58	33	22	3	-
18	NaSH · (H <sub>2</sub> O) <sub>n</sub> (10 eq.)	> 99	83	12	5	-

The formation of **Ac-Gly-CN** from **Ac-Gly-SNH<sub>2</sub>** can be explained by a simple activation-elimination mechanism akin to that described in **Figure 46**. All investigated metal chloride salts are known to disproportionate and/or hydrolyse into metal hydroxide complexes of the form M(OH)<sub>2</sub> upon



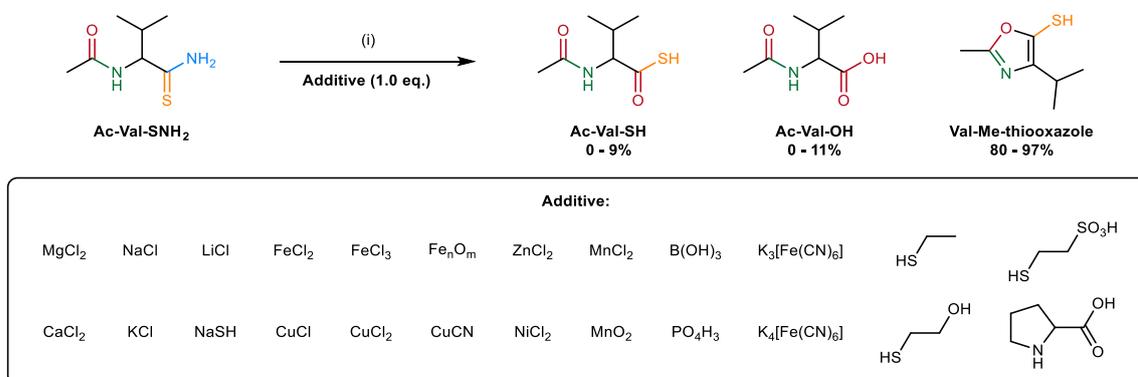
### 3. 4. 1. 1. Attempted suppression of the cyclisation of *N*-acetyl aminothioacids

The formation of methyl thiooxazole species can be explained by an unfortunate combination of factors. On the one hand, bulky sidechains are known to promote ring closure reactions through the Thorpe-Ingold effect<sup>[134]</sup>. On the other, large hydrophobic sidechains can shield the thioamide from the nucleophilic addition of negatively charged hydroxide ions, further favouring intramolecular cyclisation over hydrolysis. These combined factors promote the ring closure of *N*-acetyl aminothioamides *via* *O*-nucleophilic attack of their amide moiety onto the vicinal thiocarbonyl. Elimination of ammonia from the resulting intermediate leads to the irreversible formation of aromatic, highly stable methyl thiooxazoles (**Figure 54**).



**Figure 54:** Proposed mechanism for the formation of methyl amino acid thiooxazoles.

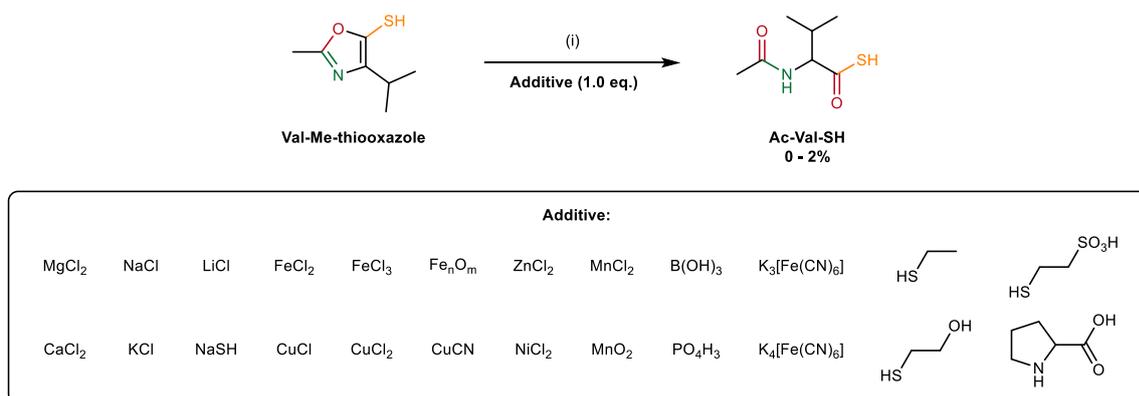
These results are problematic, as far from being rare and biologically irrelevant amino acids, valine and isoleucine represent 6.8 and 7.6% of all codons in vertebrates, respectively<sup>[4]</sup>. Due to their biological ubiquity, the impossibility of incorporating those amino acids into peptides would represent a major deficiency in our strategy. For this reason, we sought to circumvent the problem posed by the formation of methyl thiooxazoles. Our first efforts were directed towards preventing their formation. Using *N*-acetyl valine thioamide **Ac-Val-SNH<sub>2</sub>** as a model, an exhaustive screening was performed in the hope of discovering conditions that would suppress the cyclisation of **Ac-Val-SNH<sub>2</sub>**. Unfortunately, no combination of pH, temperature, reaction time, concentration and additive was found to significantly increase thioacid yields (**Figure 55**).



**Figure 55:** Hydrolysis of **Ac-Val-SNH<sub>2</sub>**; (i) **Ac-Val-SNH<sub>2</sub>** (1 eq.), **additive** (1.0 eq.), H<sub>2</sub>O (5 – 500 mM **Ac-Val-SNH<sub>2</sub>**), pH 1.0 – 12.0, RT to 90 °C, 30 min to 72 h.

### 3. 4. 1. 2. Attempted hydrolysis of methyl thiooxazoles

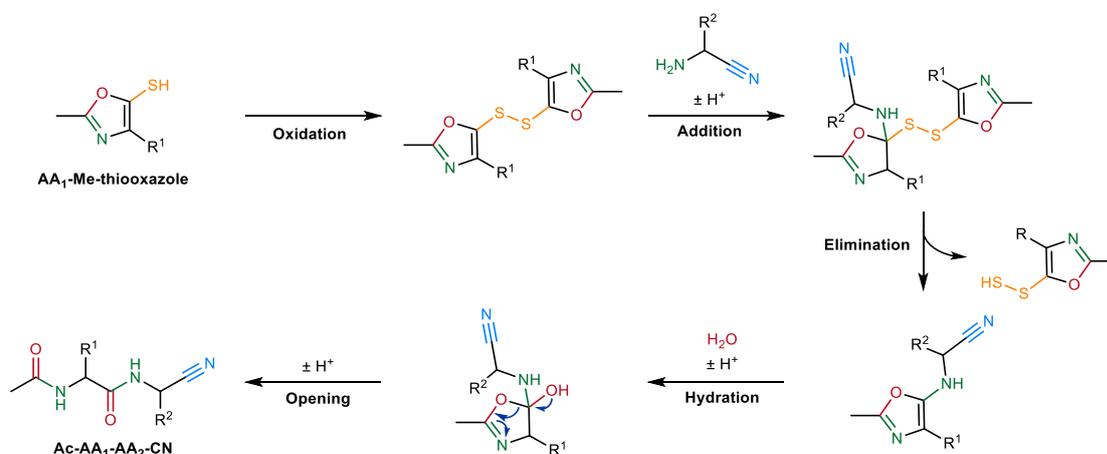
As no way of preventing the formation of **Val-Me-thiooxazole** was found, we next investigated the possibility of converting methyl thiooxazoles into thioacids through hydrolysis. To this end, a second screening was performed. Unfortunately, **Val-Me-thiooxazole** proved impervious to hydrolysis, and to nucleophilic additions in general (**Figure 56**). As a result, only trace amounts of thioacid **Ac-Val-SH** were observed, even after 7 days of incubation at 90 °C.



**Figure 56:** Attempted hydrolysis of **Val-Me-thiooxazole**; (i) **Val-Me-thiooxazole** (1 eq.), **additive** (1.0 eq.), H<sub>2</sub>O (50 mM **Val-Me-thiooxazole**), pH 1.0 – 12.0, RT to 90 °C, 1 to 7 days.

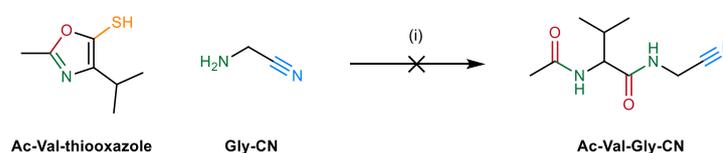
### 3. 4. 1. 3. Attempted oxidative ligation of methyl thiooxazoles

Faced with those difficulties, we went on to explore the possibility of performing oxidative ligations directly on the thiooxazole species. Examination of their structure led us to hypothesise a mechanistic pathway through which methyl thiooxazoles could be oxidised into activated dimers. If the addition of aminonitriles onto their sulfur-bearing carbon could be achieved, then a reasonable hydration-ring opening sequence could deliver *N*-acetyl dipeptide nitriles (**Figure 57**).



**Figure 57:** Proposed mechanism for the oxidative ligation of methyl amino acid thiooxazoles **AA-Me-thiooxazole** and aminonitriles **AA-CN**.

To explore this hypothesis, we exposed **Val-Me-thiooxazole** and **Gly-CN** to the aforescribed oxidative activation conditions at a range of pH, concentrations and temperatures. Unfortunately, all reactions descended into intractable mixtures from which no trace of the desired product **Ac-Val-Gly-CN** could be identified (**Figure 58**).

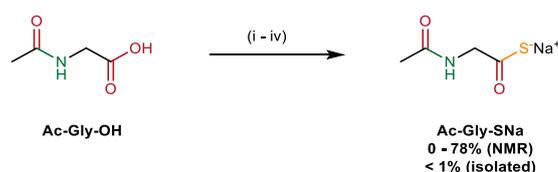


**Figure 58:** Attempted oxidative ligation of **Val-Me-thiooxazole** and **Gly-CN**; (i) **Val-Me-thiooxazole** (1 eq.), **Gly-CN** (2.0 – 10.0 eq.), H<sub>2</sub>O (50-500 mM **Val-Me-thiooxazole**), pH 5.0 – 9.0, RT to 60 °C, 1 to 24 h.

### 3. 4. 1. 4. Preparative synthesis of *N*-acetyl aminothioacids

In a somewhat predictable manner, a similar problem arose with the large-scale preparation of *N*-acetyl aminothioacids **Ac-AA-SH** as in the case of *N*-acetyl aminonitriles **Ac-AA-CN**. While the prebiotic synthesis afforded remarkably high in-solution yields, the isolation of the resulting thioacids proved an impossible challenge. While *N*-acetyl thioacids display exceptional stability in degassed, alkaline water, their sensitivity towards aerobic oxidation, and the extreme reactivity of the resulting dithioperoxyanhydrides, make their purification remarkably difficult<sup>[135]</sup>. This problem is compounded by the fact that most documented isolation procedures were designed to accommodate large, hydrophobic thioacid-bearing substrates<sup>[136]</sup>, which are readily separated from inorganic impurities by either simple workups or flash column chromatography. However, the small size of most proteinogenic *N*-acetyl aminothioacids, and the general absence of sidechains large and hydrophobic enough to counterbalance the presence of highly polar amide and thioacid functions made most such procedures unfruitful. Even when performed using strictly anhydrous, degassed solvents, and in the presence of both antioxidant additives (*e.g.* EDTA) and bases (*e.g.* triethylamine), all attempts made at purifying *N*-acetyl aminothioacids *via* normal or reverse phase flash column chromatography resulted in the near-complete degradation of thioacid products. Conversely, the commonly-recommended alternative to chromatography consisting in rapidly extracting protonated thioacids from acidified aqueous layers proved unsuitable, as most *N*-acetyl aminothioacids are too polar to extract into organic solvents. As a consequence, devising an expedient, gram-scale synthetic route to pure *N*-acetyl peptide thioacids became a necessary step in our inquiry. An extensive investigation of the multitude of thioacid preparation methods described in the literature was undertaken. Our efforts were first focused on the direct activation-thionation of *N*-acetyl amino acids, using conventional C-activating agents and strategies and several sources of hydrosulfide. Unfortunately, none of those

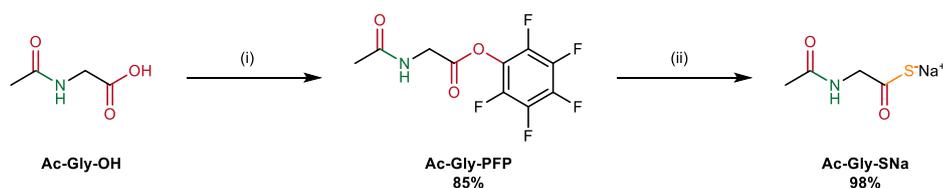
procedures were found to be suitable. While moderate to high conversions of **Ac-Gly-OH** to **Ac-Gly-SH** were observed in several cases, all reactions produced a collection of byproducts which proved inseparable from the target thioacid. These difficulties were compounded by the fact that most commonly prescribed solvents (*e.g.* acetonitrile, DMF) were found to react with hydrosulfide, which further accumulated impurities in the reaction. Consequently, only trace amounts of thioacids could be isolated from the resulting mixtures (**Figure 59**).



**Figure 59:** Attempted preparation of **Ac-Gly-SH** from **Ac-Gly-OH**; **Ac-Gly-OH** (1 eq., 500 mM), and (i) DCC (1.2 eq.), Na<sub>2</sub>S (1.2 – 10.0 eq.), DMF, 0 °C, 2 h; (ii) DIC (1.2 eq.), Na<sub>2</sub>S (5.0 – 10.0 eq.), 0 °C, 2 h; (iii) CDI (1.2 – 2.4 eq.), Na<sub>2</sub>S (5.0 eq.) or NaSH (5.0 eq.), DMF, 0 °C, 2 h; (iv) EDC · HCl DCC (2.0 eq.), Na<sub>2</sub>S (5.0 eq.) or NaSH (5.0 eq.), DMF or MeCN, 0 °C, 2 h; (v) EDC · HCl (2.0 eq.), HATU (0.2 – 2.0 eq.), Na<sub>2</sub>S (5.0 eq.) or NaSH (5.0 eq.), DMF or MeCN, 0 °C, 2 h; (iv) EDC · HCl (2.0 eq.), HOBt (2.0 eq.), Na<sub>2</sub>S (5.0 eq.), DMF, 0 °C, 2 h.

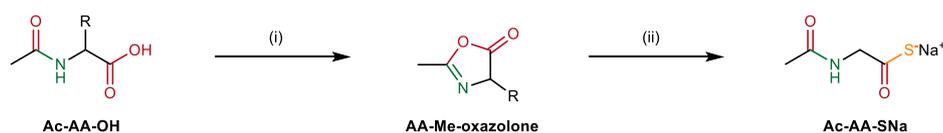
These difficulties, and the incompatibility of *N*-acetyl aminothioacids with most purification procedures, highlighted the need for the thioacid-forming step to produce as few byproducts as possible. This constraint directed our effort towards designing a two-step synthesis, whose first step would afford stable, readily purified intermediates. The conversion of those intermediates to *N*-acetyl aminothioacids should be quantitative, and involve as few reagents as possible – ideally, a hydrosulfide source alone. With these criteria in mind, we elected to derivatise *N*-acetyl amino acids **Ac-AA-OH** into perfluoryl esters **Ac-AA-PFP** due to their high reactivity towards nucleophilic addition, hydrolytic stability, and ease of purification. Using *N*-acetyl glycine **Ac-Gly-OH** as a model system, we designed conditions that successfully afforded *N*-acetyl glycine perfluoryl ester **Ac-Gly-PFP** as a highly pure, crystalline solid (**Figure 60**). Activation of **Ac-Gly-OH** using *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide in DMF in the presence of **PFP** leads to the quantitative conversion of **Ac-Gly-OH** to **Ac-Gly-PFP** in under 2 h. An added advantage of this strategy comes from the lipophilicity of the pentafluorophenol moiety, which facilitates the purification of perfluoryl esters by standard flash chromatography. Subsequently, we proceeded to find optional conditions for the conversion of *N*-acetyl perfluoryl esters into *N*-acetyl aminothioacids. Pleasingly, we found that exposing **Ac-Gly-PFP** to an aqueous solution of sodium hydrosulfide led to the near-quantitative formation of *N*-acetyl glycine thioacid **Ac-Gly-SH**. Concentration of the reaction mixture and rapid trituration of the resulting solid afforded the thioacid sodium salt **Ac-Gly-SNa**, along with minute amounts of **Ac-Gly-OH** (< 1%) and sodium

chloride salts. Small quantities of these purities, which did not interfere with the reactivity of thioacids, were deemed inconsequential and the thioacids were used without further purification.



**Figure 60:** PFP-mediated preparation of **Ac-Gly-SNa** from **Ac-Gly-OH**; (i) **Ac-Gly-OH** (1 eq.), EDC · HCl (1.2 eq.), pentafluorophenol (1.5 eq.), DMF (500 mM **Ac-Gly-OH**), RT, 2 h; (ii) **Ac-Gly-PFP** (1 eq.), NaSH (1.2 eq.), H<sub>2</sub>O (100 mM **Ac-Gly-PFP**), pH 9.0, 1 h.

A second, similar procedure was developed in parallel by other members of the Powner group. This route relies on the well-characterised self-cyclisation of *C*-activated *N*-acetyl amino acids **Ac-AA-OH** to produce stable methyl oxazolones **AA-Me-oxazolone**. The resulting oxazolones were subsequently thionated in dry acetonitrile to afford thioacid sodium salts **Ac-AA-SNa**. While this procedure proved more efficient for larger, hydrophobic sidechains (*e.g.* phenylalanine), the pentafluoroyl route was preferred in the case of smaller residues (*e.g.* glycine) or non-cyclisable amino acids (*e.g.* proline). These two procedures were alternatively used to prepare gram-scale quantities of synthetic *N*-acetyl aminothioacids, which were used both to identify their counterparts from prebiotic mixtures described in **Figure 46**, and to perform extensive investigations into the reactivity of aminothioacids using pure starting materials (**Figure 61**).



**Figure 61:** Oxazolone-mediated preparation of **Ac-AA-SNa** from **Ac-AA-OH**; (i) **Ac-AA-OH** (1 eq.), EDC · HCl (1.4 eq.), DCM (100 mM **Ac-AA-OH**), 0 °C, 1 h; (ii) **AA-Me-oxazolone** (1 eq.), NaSH (2.0 eq.), acetonitrile (100 mM **AA-Me-oxazolone**), RT, 16 h.

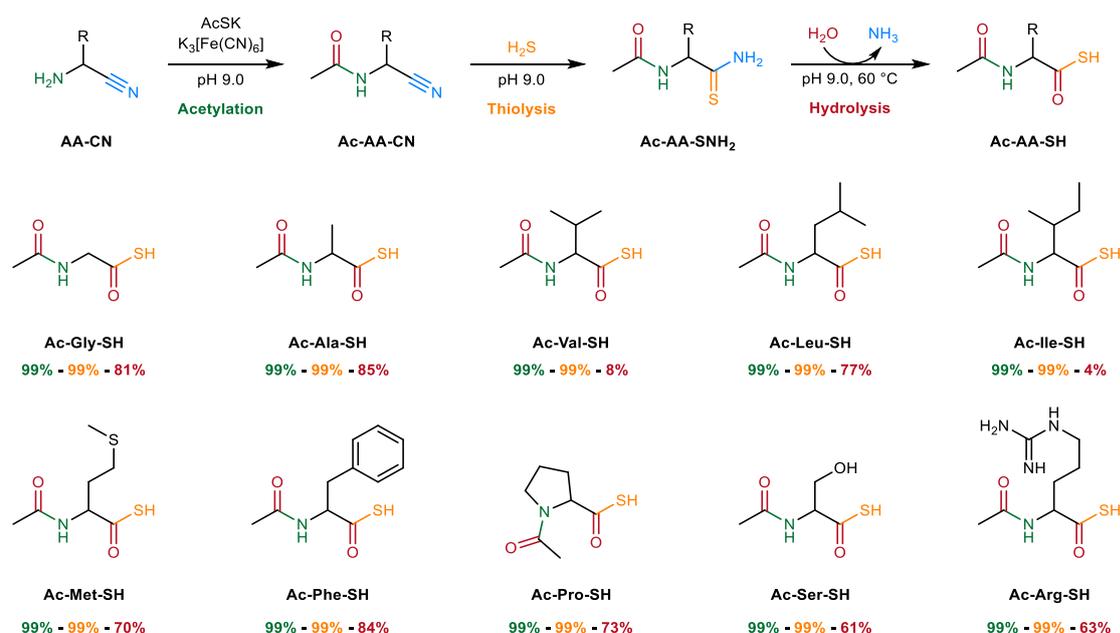
### 3. 5. Conclusions

Our investigation has successfully uncovered a three-steps, prebiotic pathway connecting aminonitriles to *N*-acetyl aminothioacids. This route distinguishes itself from previously-reported attempts at preparing aminothioacids in a prebiotic context by a range of advantageous characteristics (**Figure 62**):

- **High to quantitative yields:** Two out of three steps proceed in uniformly quantitative yields. While the final hydrolysis step presents more variability depending on the

sidechain, *N*-acetyl aminothioacids yields up to 85% were achieved. Those results are to be compared with recent strategies, which form aminothioacids in 1.4% yield <sup>[111]</sup>.

- **Large sidechain tolerance:** With the exception of the unexpected cyclisation of  $\beta$ -branched residues (valine, isoleucine), our strategy has proved fully compatible with all prebiotically relevant sidechains investigated, including notoriously difficult residues, such as serine and arginine. This remarkable generality stands in sharp contrast with the aforementioned routes, whose substrate scope is limited to simple, unreactive sidechains. Furthermore, it is worth noting that even in the case of branched residues, aminothioacid yields outcompete all previously published methods.
- **Rapid, mild conversions:** A recurrent flaw in prebiotic aminothioacid-yielding pathways stems from the required use of reagents and conditions which have been shown to expedite the destruction of the thioacid products, or of the peptides they aim to produce <sup>[111]</sup>. Oppositely, our route relies entirely on reagents which are unreactive toward amide bonds. In addition, the moderate temperatures and near-neutral pH used throughout further guarantee the preservation of both thioacid and peptide products.
- **Unified conditions and reagents:** Our strategy relies solely on simple, plausible prebiotic reagents, and maintains constant pH and concentration across all three steps. Furthermore, hydrosulfide is used both as sulfur source and nucleophilic catalyst for the hydrolysis step. These unified conditions offer a far simpler, more plausible alternative to previously proposed models involving dry/wet and pH cycles and unstable reagents<sup>[137]</sup>.

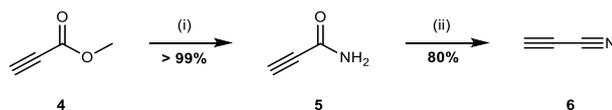


**Figure 62:** Summary of the aminonitrile-to-*N*-acetyl aminothioacids route.

## 4. Ligations of *N*-capped peptide thioacids

### 4. 1. Optimisation studies

Having described a prebiotic route towards *N*-acetyl aminothioacids and prepared pure synthetic samples for prebiotically relevant sidechains, we went on to investigate the ligation of said aminothioacids with aminonitriles. Using the *N*-acetyl glycine thioacid **Ac-Gly-SH** – glycine nitrile **Gly-CN** ligation as a model system, we first set out to compare the reactivity of prebiotically plausible *S*-activating agents. These reagents include previously-described potassium ferricyanide and copper(II) chloride, as well as prebiotic alkylating agents such as cyanoacetylene<sup>[138]</sup>, cyanamide<sup>[139]</sup>, and *N*-cyanoimidazole<sup>[138]</sup>. While both *N*-cyanoimidazole and cyanamide are commercially available, cyanoacetylene **6** was prepared following Eschenmoser's procedure<sup>[140]</sup>. Methyl propionate **4** was reacted with liquid ammonia to afford propiolamide **5** in quantitative yield. Amide **5** was then dehydrated using phosphorous pentoxide on sand to afford cyanoacetylene **6** in 80% yield, which was stored as a 1 M aqueous solution at -80 °C (**Figure 63**).



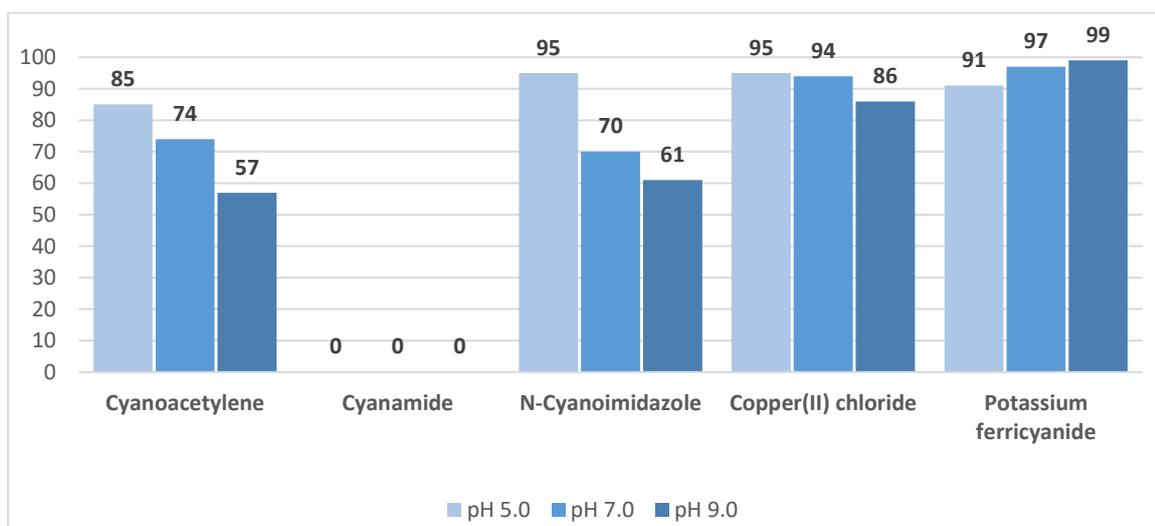
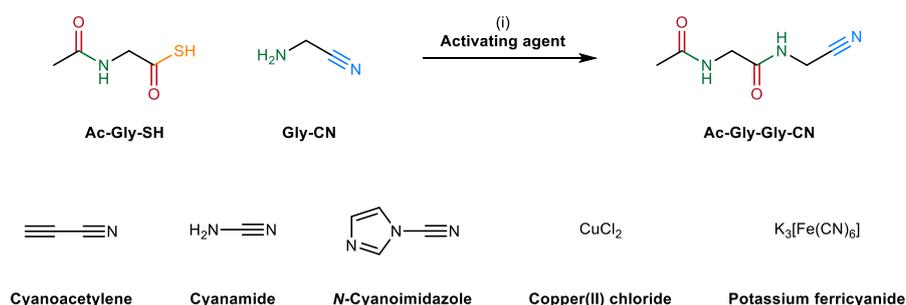
**Figure 63:** Preparation of cyanoacetylene **6**; (i) **4** (1.0 eq.), NH<sub>3</sub>(l) (1.2 M **4**), -80 °C, 4 h; (ii) **5** (1.0 eq.), P<sub>2</sub>O<sub>5</sub> (1.2 eq.), sand, 130 °C, 4 h.

With cyanoacetylene **6** in hand, we proceeded to study the reactivity of each activating agent under acidic, neutral, and alkaline conditions (**Figure 64**). To our satisfaction, most activating agents tested activated *N*-acetyl glycine thioacid **Ac-Gly-SH** and afforded **Ac-Gly-Gly-CN** in high to quantitative yields. Upon closer examination of the data, several informative trends were uncovered:

- Metal-based activating agents afforded better yields than *S*-alkylating agents. A secondary advantage stems from the absence of formation of organic byproducts from metallic activators, which renders NMR analysis simpler and more reliable.
- Overall, potassium ferricyanide proved the best thioacid activating agent, and afforded near-quantitative yields at pH 5.0 and 7.0, and consistently quantitative ligation yields at pH 9.0.
- The different activating agents follow drastically divergent pH sensitivity trends. While alkylating agents afford high yields at pH 5.0, which precipitously drop as the pH increases, copper(II) chloride shows only a weak response to pH variations. Contrarily, potassium

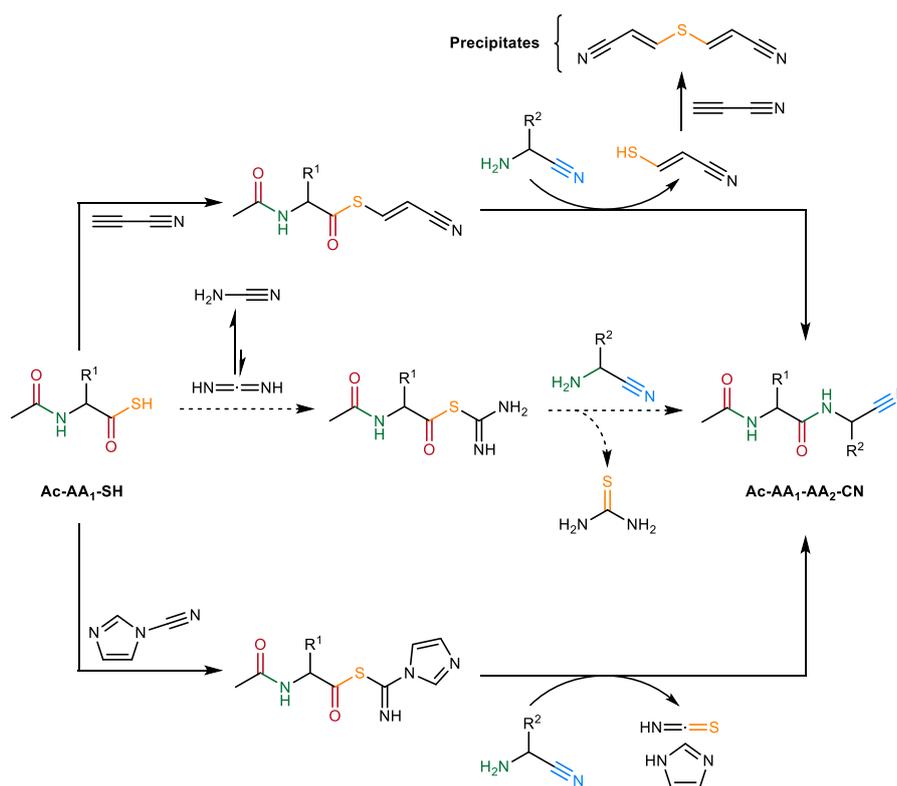
ferricyanide presents a modest increase in yields as the pH increases, with the best yields obtained under alkaline conditions.

- In spite of repeated claims in the literature asserting the significance of cyanamide as a prebiotic carbonyl activating agent<sup>[139, 141]</sup>, no reaction could be observed between cyanamide and **Ac-Gly-SH**, even with incubation times extended up to 7 days and reaction temperatures increased up to 60 °C. It is not implausible that the discrepancy between the alleged and observed reactivity of cyanamide stems from most cyanamide-based studies actually use synthetic substitutes (EDC, most commonly), while failing to provide experimental evidence that similar results could be achieved using cyanamide itself.



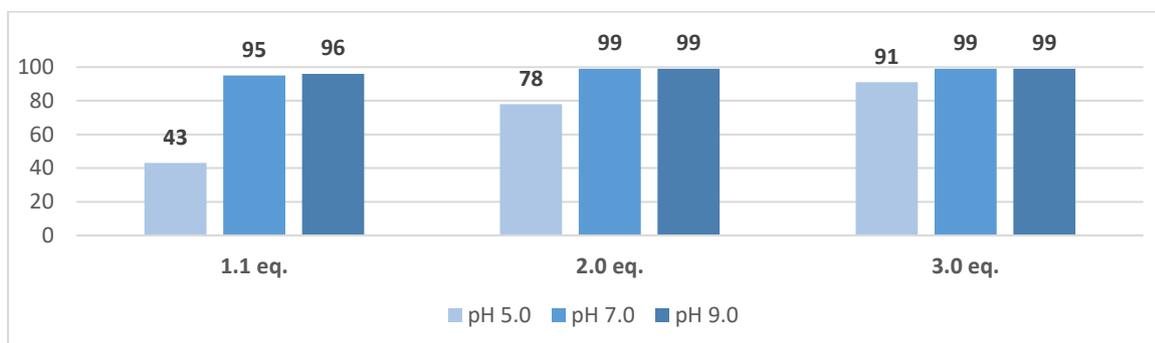
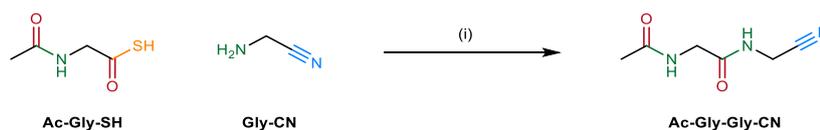
**Figure 64:** Activation and ligation of **Ac-Gly-SH** and **Gly-CN** using alkylating and oxidative activating agents; (i) **Ac-Gly-SH** (1.0 eq.), **Gly-CN** (2.0 eq.), **Activating agent** (3.0 eq.), H<sub>2</sub>O pH 5.0 – 9.0 (50 mM **Ac-Gly-SH**), RT, 20 min.

From a mechanistic point of view, the activation strategies investigated fall into two distinct categories: *S*-alkylation, or *S*-oxidation. Metal-based activations and their postulated mechanisms were described in **Figure 37** and **51**. On the other hand, alkylation-based strategies rely on the high nucleophilicity of thiocarbonyls, which allows their sulfur atom to turn into a highly reactive leaving group by addition to soft electrophiles (**Figure 65**).



**Figure 65:** Proposed mechanisms for the alkylation-based activation of thioacids by cyanoacetylene, cyanamide and *N*-cyanoimidazole.

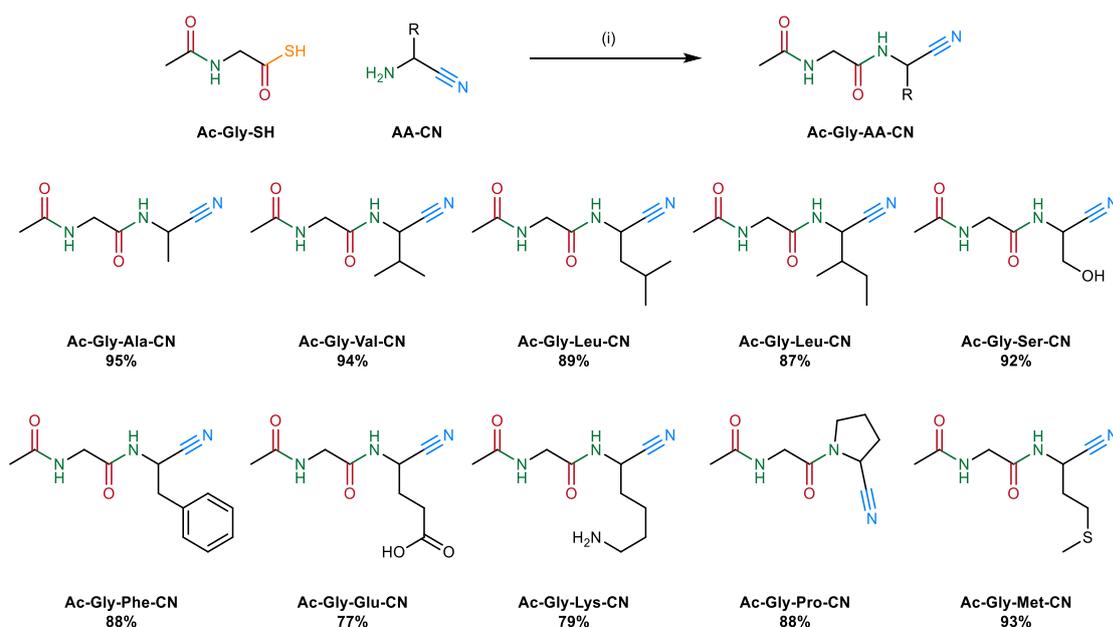
Owing to its high efficiency and continuity with acetylation conditions employed in our first step, potassium ferricyanide was selected as our general thioacid activating agent. An additional step in our investigation consisted in studying the impact of the number of equivalents of aminonitriles on the ligation yields. For this purpose, the ferricyanide-mediated ligation of **Ac-Gly-SH** and **Gly-CN** was repeated using equivalents of **Gly-CN** ranging from 1.1 to 3.0. To our great satisfaction, we found that even coupling performed at near-stoichiometric ratios maintained near-quantitative yields for neutral and alkaline pH. Alternatively, using a threefold excess of aminonitrile brought the pH 5.0 ligation yields from 43 to 91%, thus demonstrating a wide range of conditions under which the ligation reaches near-quantitative dipeptide nitrile yields. Unsurprisingly, the reaction exhibits a very similar kinetic profile to the ferricyanide-mediated acetylation previously described: the ligation was found to proceed in under 1 minute at room temperature, and to maintain near-quantitative yields at aminonitrile concentrations ranging from 1.0 M to 1.0  $\mu$ M. All previous steps have been shown to proceed most efficiently at pH 9.0, and 2.0 equivalents of **Gly-CN** proved sufficient to achieved consistently quantitative yields. Therefore, a thioacid concentration of 50 mM, pH 9.0 and 2.0 equivalents of aminonitrile were set as standard ligation conditions (**Figure 66**).



**Figure 66:** Optimisation of the ligation of **Ac-Gly-SH** and **Gly-CN**; (i) **Ac-Gly-SH** (1.0 eq.), **Gly-CN** (1.1 – 3.0 eq.), potassium ferricyanide (3.0 eq.), H<sub>2</sub>O pH 5.0 – 9.0 (50 mM **Ac-Gly-SH**), RT, 20 min.

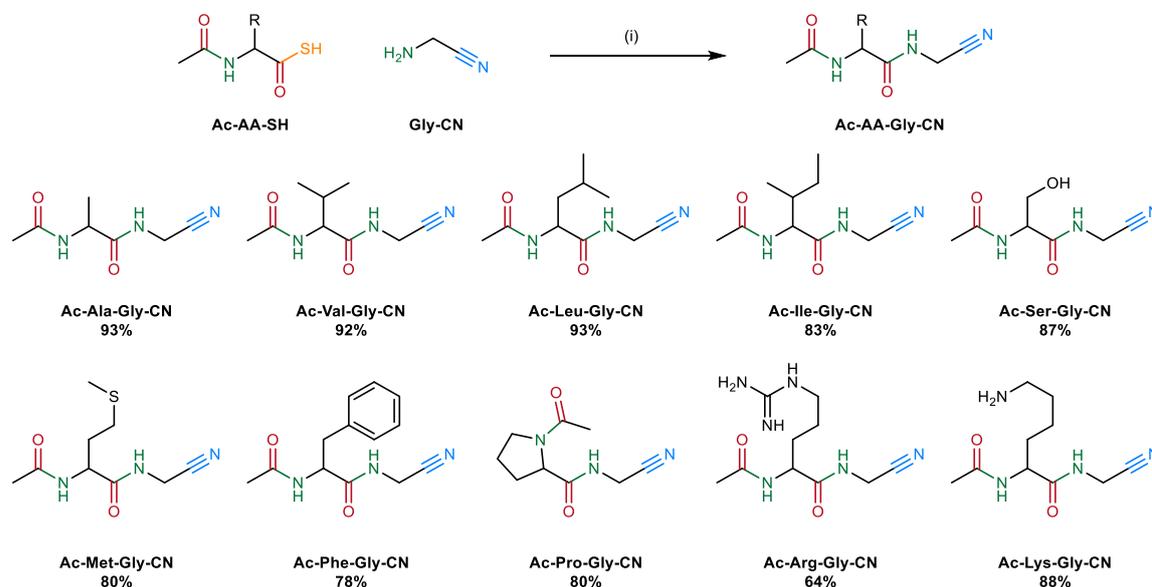
## 4. 2. Scope of the ligation

With optimal conditions established, we went on to assert the generality of our ligation strategy. To do so, a set of 10 different aminonitriles were ligated to *N*-acetyl glycine thioacid **Ac-Gly-SH**. Gratifyingly, high to near-quantitative yields were obtained in all cases (77 – 95%, avg. 88%). Of particular note is the fact that even highly constrained amines (*e.g.* proline) and residues presenting reactive sidechains (*e.g.* lysine, glutamate) proved compatible with our strategy, and afforded yields in the high 70%'s (**Figure 67**).



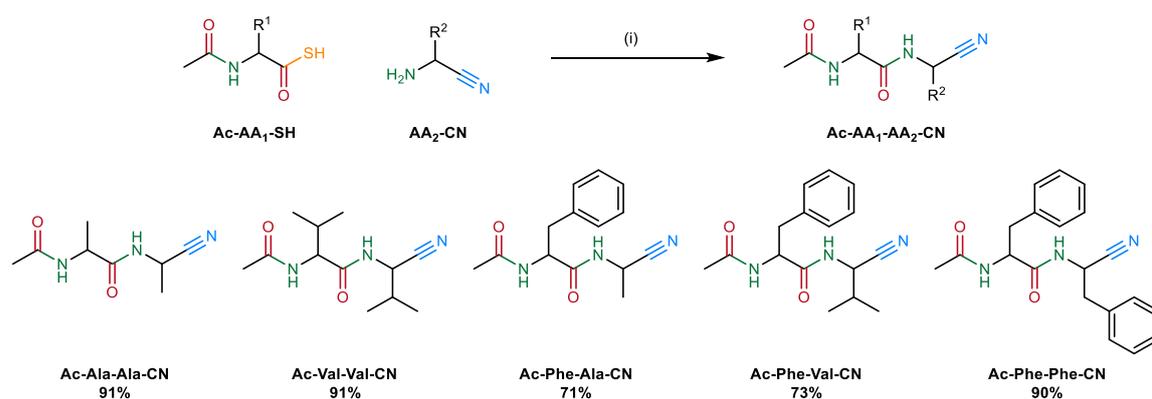
**Figure 67:** Ligation of **Ac-Gly-SH** and aminonitriles **AA-CN**; (i) **Ac-Gly-SH** (1.0 eq.), **AA-CN** (2.0 eq.), potassium ferricyanide (3.0 eq.), H<sub>2</sub>O pH 9.0 (50 mM **Ac-Gly-SH**), RT, 20 min.

In order to demonstrate the C-terminus sidechain tolerance in our ligation, a second series of couplings was performed, this time by varying the thioacid sidechain while setting **Gly-CN** as coupling partner (**Figure 68**). Again, high to near-quantitative yields were consistently achieved (64 – 93%, avg. 84%).



**Figure 68:** Ligation of amino thioacids **Ac-AA-SH** and **Gly-CN**; (i) **Ac-AA-SH** (1.0 eq.), **Gly-CN** (2.0 eq.), potassium ferricyanide (3.0 eq.), H<sub>2</sub>O pH 9.0 (50 mM **Ac-AA-SH**), RT, 20 min.

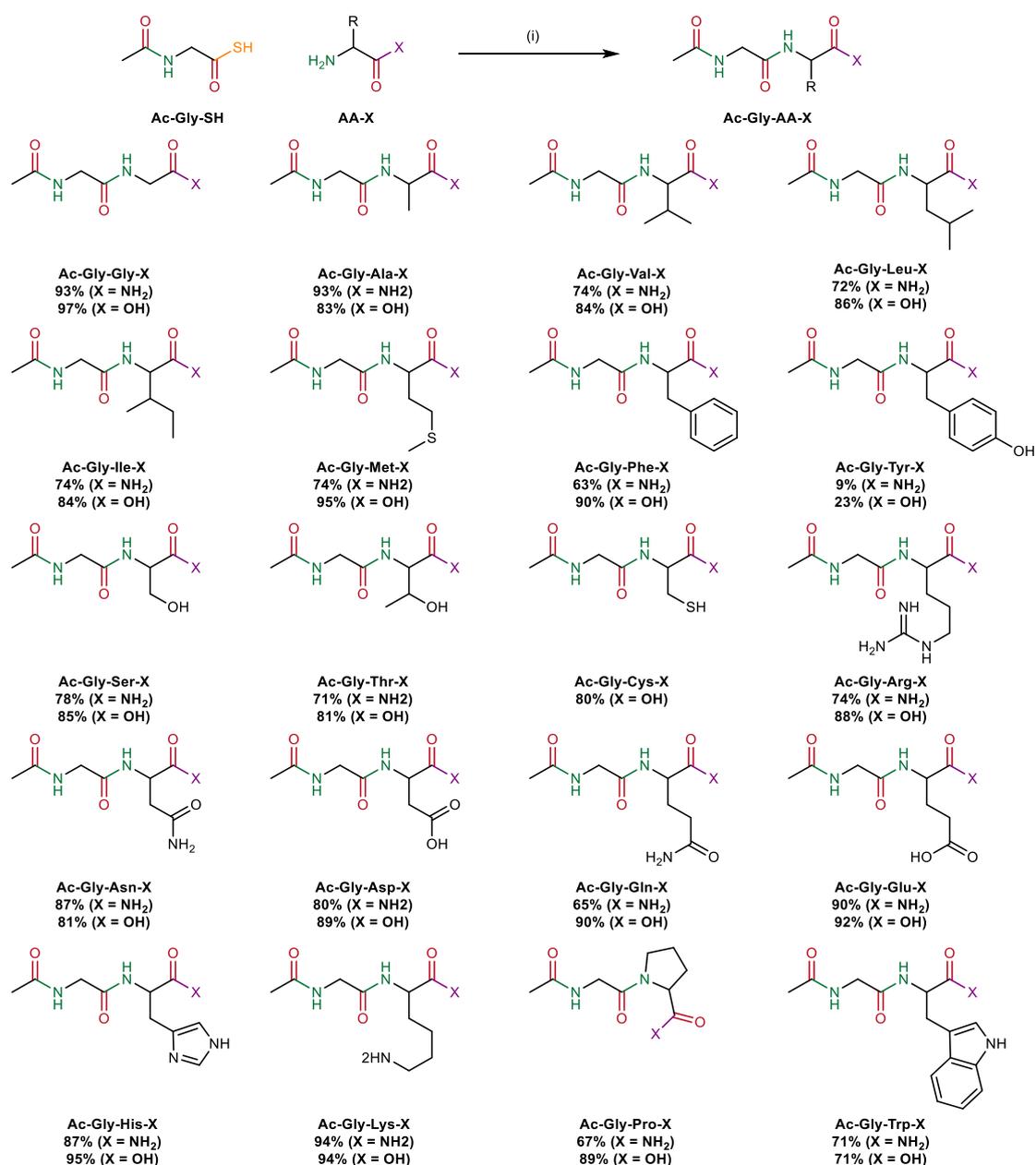
Finally, we concluded that in order to fully demonstrate the generality of our ligation, worst-case scenario couplings should also be examined. To do so, we performed a third series of coupling targeting sterically hindered dimers (**Figure 69**). Satisfactorily, even highly congested dimers were obtained in high to near-quantitative yields<sup>3</sup> (71 – 91%, avg. 83%).



**Figure 69:** Ligation of sterically hindered amino thioacids **Ac-AA<sub>1</sub>-SH** and amino nitriles **AA<sub>2</sub>-CN**; (i) **Ac-AA<sub>1</sub>-SH** (1.0 eq.), **AA<sub>2</sub>-CN** (2.0 eq.), potassium ferricyanide (3.0 eq.), H<sub>2</sub>O pH 9.0 (50 mM **Ac-AA<sub>1</sub>-SH**), RT, 20 min.

<sup>3</sup> These experiments, as well as those described in **Figure 70**, were performed in collaboration with Dr S. Islam, a PDRA in the Powner group.

While our ligation strategy chiefly targets aminonitriles as coupling partners, the ubiquitous presence of amino amides and amino acids in the prebiotic milieu led us to investigate their ligation. Therefore, we performed an extensive series of ligations, involving all 20 proteinogenic amino acids 19 amino amides (excluding Cys-NH<sub>2</sub>). Both amino amides (9 – 94%, avg. 75%) and amino acids (23 – 97%, avg. 84%) ligated in high yields (**Figure 70**). Overall, aminonitriles ligate more efficiently than both amino acids and amino mides. This can be explained by the fact that the  $\alpha$ -amine of both aminoamides and amino acids mostly exist as a protonated ammonium moiety under ligation conditions, while that of aminonitriles is mostly deprotonated.

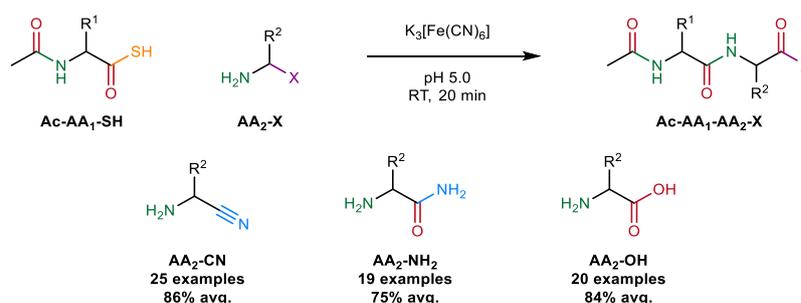


**Figure 70:** Ligation of **Ac-Gly-SH** with amino amides **AA<sub>2</sub>-NH<sub>2</sub>** and amino acids **AA<sub>2</sub>-OH**; (i) **Ac-AA<sub>1</sub>-SH** (1.0 eq.), **AA<sub>2</sub>-X** (2.0 eq.), potassium ferricyanide (3.0 eq.), H<sub>2</sub>O (50 mM **Ac-Gly-SH**), pH 9.0 (**AA<sub>2</sub>-OH**) or pH 9.5 (**AA<sub>2</sub>-NH<sub>2</sub>**), RT, 20 min.

### 4. 3. Conclusions

Exploring the reactivity of *N*-acetyl aminothioacids allowed us to uncover an expedient activation-amidation strategy, affording *N*-acetyl dipeptide nitriles from the ligation of aminothioacids and aminonitriles. This strategy produced exceptional results, and offers a multitude of remarkable characteristics which puts it above all previous literature in the field (**Figure 71**):

- **Unprecedentedly high yields:** Our ligation afforded dipeptides in up to quantitative yields, with an average of 86% over 25 examples. Those yields outcompete even the most efficient prebiotic amidation strategies published to date, and put our strategy on par with state-of-the-art synthetic amidation methodologies.
- **Generality:** While the overwhelming majority of prebiotic peptide elongation strategies either solely focus on the simplest amino acids (*e.g.* glycine, alanine), or were shown to be incompatible with more complex sidechains<sup>[141, 142]</sup>, our method has proved fully compatible with every investigated combination of all 20 natural sidechains across aminonitriles, aminoamides, and amino acid residues. Even in the case of particularly challenging targets, such as highly congested dipeptides or reactive sidechains, high to quantitative yields were consistently obtained with all prebiotically relevant residues.
- **Low equivalents:** Only two equivalents of aminonitrile were found to be sufficient to maintain high to quantitative ligation yields, as opposed to the hundredfold excesses frequently required by previously described systems.
- **Fast, mild reaction:** The reaction is complete within minutes at room temperature, without involving any reactant, byproduct or condition which would endanger the resulting peptides.
- **Robustness and flexibility:** The ligation proceeds using a variety of prebiotically abundant activating agents, both metal-based and *S*-alkylating, to maintain high yields under acidic, neutral and alkaline pH, and concentrations ranging from the molar to micromolar scale.
- **Unified conditions:** Optimal ligation yields were obtained using conditions identical to those used in the oxidative acetylation of aminonitriles.



**Figure 71:** Summary of the *N*-acetyl aminothioacid ligation.

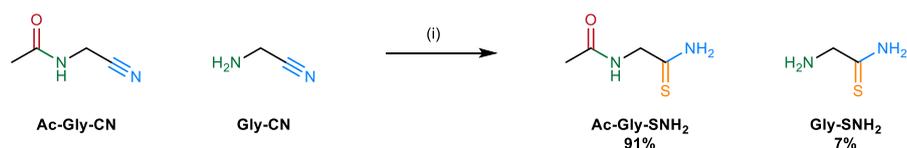
## 5. Chemoselectivity of the elongation process

While stepwise, supervised pathways are a necessary stage in understanding each separate reaction underlying prebiotic systems, the end goal of the field remains continuous, autonomous networks. Although some level of compartmentation of the prebiotic environment can be invoked<sup>[143]</sup>, the predominance of complex mixtures of mutually reactive products remains a safe assumption. It is therefore crucial for any proposed prebiotic system to show substantial robustness towards non-biogenic cross-reactions, both with members of their own networks, and with plausible components of their milieu. With this requirement in mind, we set out to investigate the selectivity of the key steps of our elongation strategy.

### 5. 1. Chemoselectivity of the thiolysis

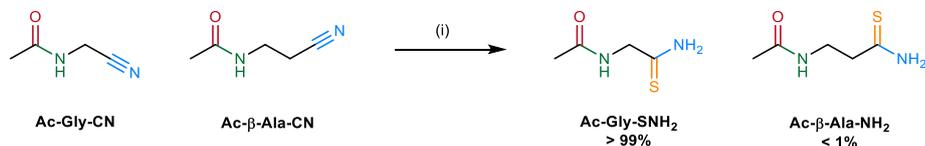
The first critical step of our strategy is the thiolysis of *N*-acetyl aminonitriles **Ac-AA-CN** into *N*-acetyl aminothioamides **Ac-AA-SNH<sub>2</sub>**. While hydrosulfide has proved entirely unreactive towards amide bonds, and thus compatible with the continued growth of peptides, its selectivity against other prebiotic nitriles remains unproven. Such selectivity is, however, paramount to the viability of our route, as an excess of free aminonitriles can be expected to be found in solution alongside their acetylated counterpart. Therefore, a lack of selectivity in the thiolysis step could deplete the reaction milieu of the free aminonitrile monomers required to both begin and progress the elongation process. To make matters worse, non-proteinogenic, nitrile-bearing species are common in prebiotic networks<sup>[97]</sup>. Successful thiolysis of such species would produce non-proteinogenic thioamides and, eventually, thioacids, which could both intercept aminonitriles during the ligation step, irreversibly locking them away into non-biogenic products.

In order to assay the selectivity of the thiolysis step, we thiolysed *N*-acetyl glycine nitrile **Ac-Gly-CN** in the presence of equimolar amounts of prebiotically relevant competitors. Our first experiment examined the selectivity of the thiolysis against glycine nitrile **Gly-CN**. Gratifyingly, the thiolysis of **Gly-CN** proved 13 times slower than that of **Ac-Gly-CN**, with only 7% of the former converting to **Gly-SNH<sub>2</sub>** after 24 h of incubation with 10 eq. of hydrosulfide, whereas 91% of **Ac-Gly-CN** converted to **Ac-Gly-SNH<sub>2</sub>** (**Figure 72**). This remarkable kinetic selectivity can be explained by the strong electron-withdrawing effect stemming from the  $\alpha$ -amide present in both *N*-acetyl aminonitriles and peptide nitriles, which exalts the reactivity of the nitrile group towards nucleophilic addition. Those results back the notion that even if the thiolysis was to be performed in an environment containing both free and *C*-terminal aminonitriles, the reaction would leave aminonitrile monomers mostly untouched.



**Figure 72:** Thiolytic conversion of **Ac-Gly-CN** in the presence of **Gly-CN**; (i) **Ac-Gly-CN** (1.0 eq.), **Gly-CN** (1.0 eq.), NaSH (5.0 eq.), H<sub>2</sub>O (50 mM **Ac-Gly-CN**), pH 9.0, RT, 24 h.

Encouraged by those results, we pursued by performing a second thiolytic conversion, this time involving *N*-acetyl glycine nitrile **Ac-Gly-CN** and *N*-acetyl β-alanine nitrile **Ac-β-Ala-CN**. The choice of **Ac-β-Ala-CN** as competitor arises from its plausible presence in the prebiotic environment<sup>[48, 144]</sup>. Additionally, **Ac-β-Ala-CN** constitutes an ideal model for evaluating the impact of the amide electron-withdrawing effect, as its structure only differs from that of **Ac-Gly-CN** by an extra -CH<sub>2</sub>- group intercalated between its amide and nitrile moieties. To our great satisfaction, only trace amounts of **Ac-β-Ala-CN** were converted into **Ac-β-Ala-SNH<sub>2</sub>**, while quantitative conversion of **Ac-Gly-SNH<sub>2</sub>** was observed (**Figure 73**). These results demonstrate the exquisite selectivity of the thiolytic conversion step for α-amidocyanides over other alkylnitriles, which creates a first chemoselective filter funneling the necessary α-amidocyanides into our route, while excluding non-proteinogenic substrates.



**Figure 73:** Thiolytic conversion of **Ac-Gly-CN** in the presence of **Ac-β-Ala-CN**; (i) **Ac-Gly-CN** (1.0 eq.), **Ac-β-Ala-CN** (1.0 eq.), NaSH (5.0 eq.), H<sub>2</sub>O (50 mM **Ac-Gly-CN**), pH 9.0, RT, 24 h.

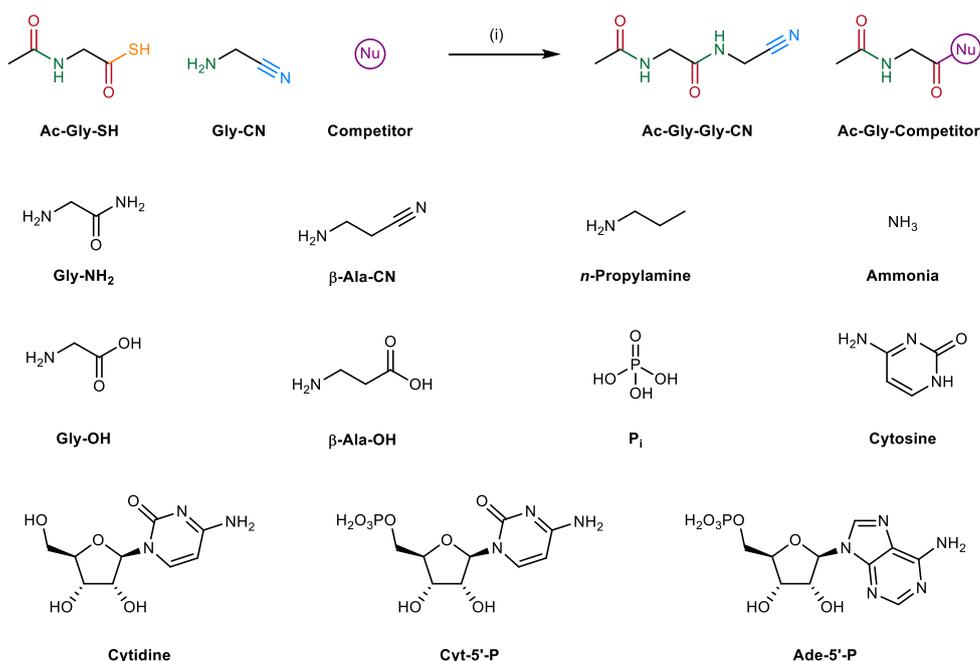
## 5. 2. Chemoselectivity of the ligation

The second key stage of our strategy where chemoselectivity is a crucial requirement is the thioacids-aminonitrile oxidative ligation step. While the reaction was shown to proceed in remarkably high yields in isolation, a multitude of nucleophilic competitors are likely to have been abundant in the prebiotic milieu. The competitors include:

- **Hydrolysis products of aminonitriles:** the incubation of aminonitriles in water will eventually lead to their hydrolysis into aminoamides and amino acids. Although their ligation would result in proteinogenic products, the absence of a C-terminal nitrile moiety would irreversibly terminate the growth of the ligated peptide.

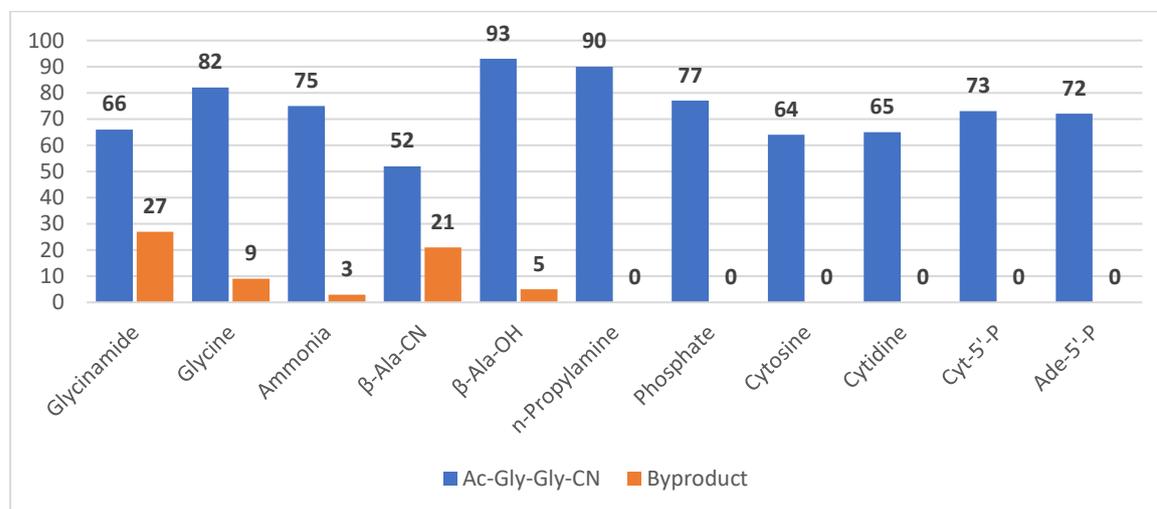
- Prebiotically plausible alkylamines:** Aside from  $\alpha$ -aminocarboxyls, a collection of non-proteinogenic amines have been identified as likely constituents of the prebiotic milieu. As an example,  $\beta$ -alanine  **$\beta$ -Ala-OH** is the third most abundant product of the original Miller experiment and its volcanic variants<sup>[144]</sup>. For this reason, both  **$\beta$ -Ala-OH** and its nitrile precursor  **$\beta$ -Ala-CN** were included in our study. In addition, *n*-propylamine was used as a generic model for simple alkylamines.
- Inorganic nucleophiles:** Inorganic phosphate **P<sub>i</sub>** and ammonia are ubiquitous in prebiotic networks, and are necessary elements of the synthesis of RNA nucleotides as well as the Strecker synthesis of aminonitriles.
- RNA nucleotides and nucleobases:** Considering that one of our main motivation for developing a prebiotic peptide elongation pathway was to provide a route towards catalytically active structures able to supplement the RNA world theory, the possibility of performing such elongation in the presence of RNA components is an obvious necessity. For this reason, we included representative examples of potentially nucleophilic RNA components, such as cytosine, cytidine, and phosphorylated nucleotides.

With the list of competitors established, we embarked on an extensive series of experiments to investigate the chemoselectivity of our oxidative ligation. To this end, we repeated the ligation of **Ac-Gly-SH** and **Gly-CN** under acidic, neutral, and alkaline pH conditions, in the presence of equimolar amounts of nucleophilic competitors (**Figure 74**).



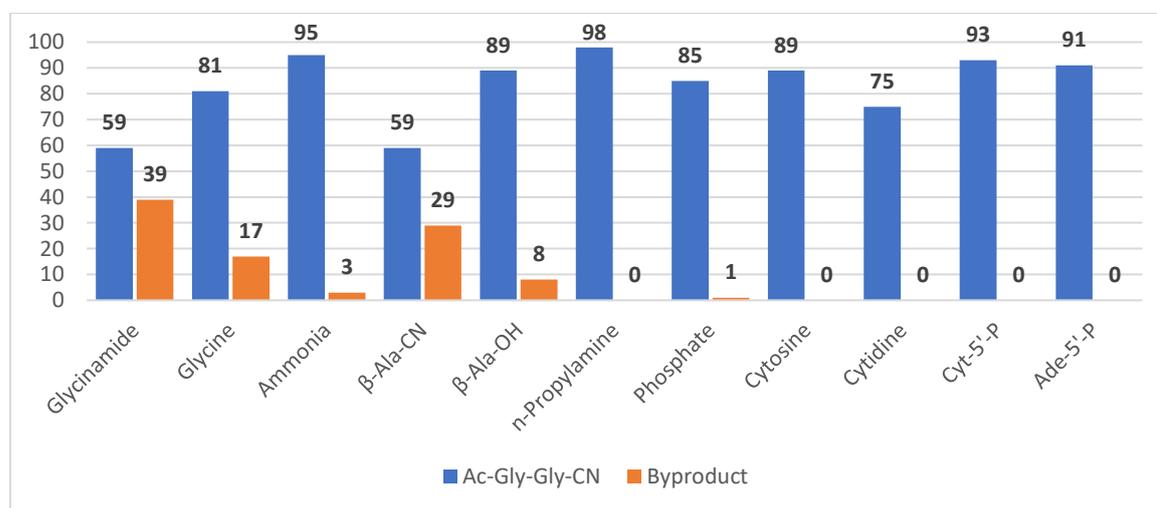
**Figure 74:** Oxidative ligation of **Ac-Gly-SH** and **Gly-CN** in the presence of competitors; (i) **Ac-Gly-SH** (1.0 eq.), **Gly-CN** (2.0 eq.), **competitor** (2.0 eq.), H<sub>2</sub>O (50 mM **Ac-Gly-SH**), pH 5.0 – 9.0, RT, 20 min.

As suspected, the relative yields of the target dipeptide **Ac-Gly-Gly-CN** to competitor ligation byproduct proved highly dependent on the pH at which the reaction is performed. At pH 5.0, most competitors achieved little to no ligation, and **Ac-Gly-Gly-CN** was, by far, the dominant product in all of these reactions (**Figure 75**).



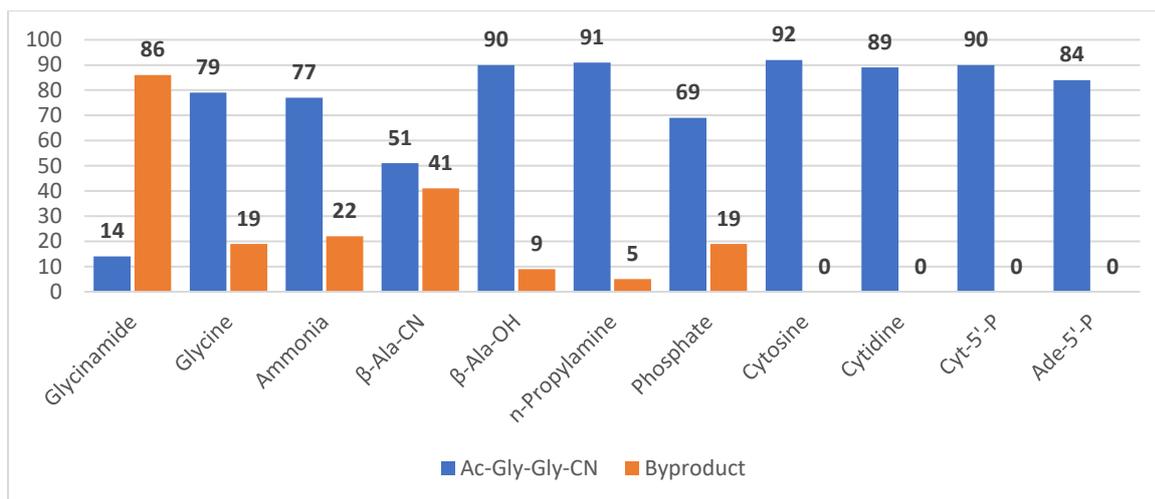
**Figure 75:** Ligation yields of **Ac-Gly-SH** and **Gly-CN** in the presence of competitors at pH 5.0 (%).

As the pH increases to 7.0, nucleophiles nearing their  $pK_{aH}$  (e.g. **Gly-NH<sub>2</sub>**, **β-Ala-OH**) begin to form significant amounts of their respective ligation products, although **Ac-Gly-Gly-CN** remains the main product against all competitors (**Figure 76**).



**Figure 76:** Ligation yields of **Ac-Gly-SH** and **Gly-CN** in the presence of competitors at pH 7.0 (%).

At pH 9.0, more nuanced results were obtained, but once again following the expect profile based on the competing nucleophiles  $pK_{aH}$ . While in most cases, **Ac-Gly-Gly-CN** remains the main product by a large margin, **Gly-CN** was vastly outcompeted by **Gly-NH<sub>2</sub>**, and the product ratio for **β-Ala-CN** approached 1:1 (**Figure 77**).

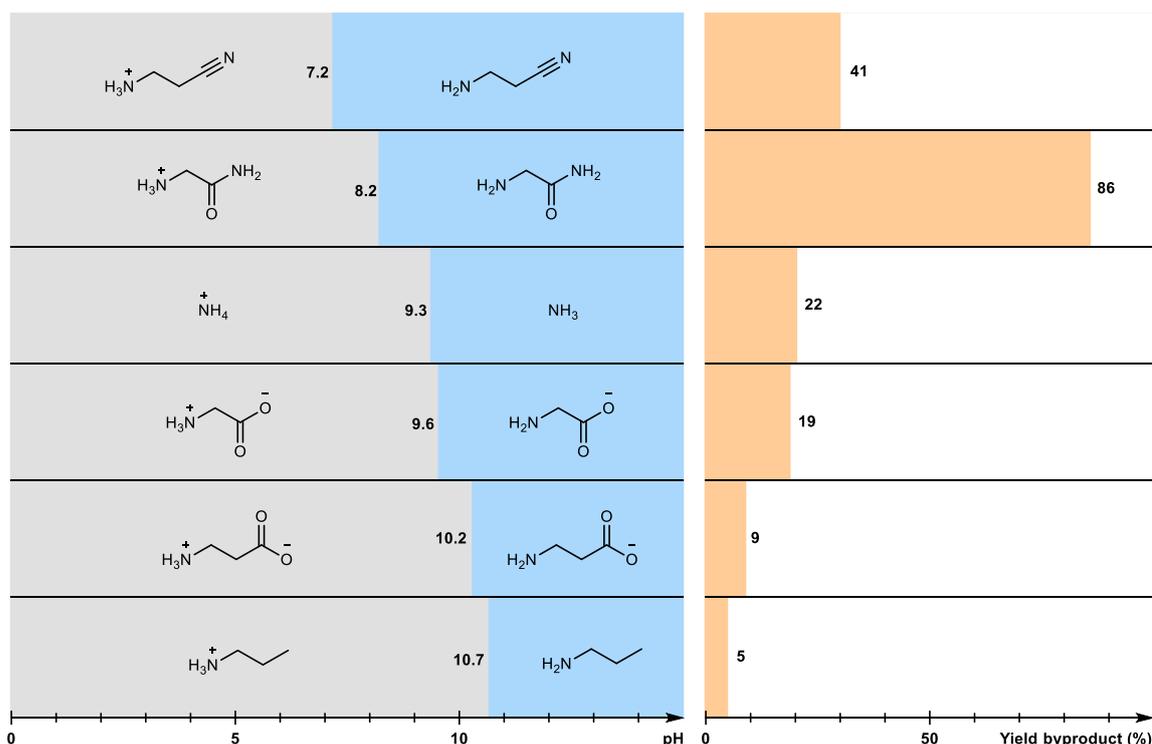


**Figure 77:** Ligation yields of **Ac-Gly-SH** and **Gly-CN** in the presence of competitors at pH 9.0 (%).

Although the results remain largely substrate-specific, separating the competitors into three main groups allows to draw some general conclusions from the data. Overall, the ligation:

- Is highly orthogonal to nucleobases and nucleotides, regardless of the pH;
- shows near-perfect selectivity against alkylamines ( **$\beta$ -Ala**, *n*-propylamine);
- presents high to moderate selectivity against proteinogenic nucleophiles (**Gly-OH**, **Gly-NH<sub>2</sub>**, ammonia).

A rationale for those observations can be put forth by examining the  $pK_{aH}$  of the competitors and the strong pH-dependency of the coupling yields. Indeed, the proportion of competitor ligation products is strongly correlated to the acidity of the nucleophilic moiety. This trend is especially salient when comparing the addition of similar functional groups. **Figure 78** puts in parallel the  $pK_{aH}$  of all amine competitors which achieved ligation with the yields of their respective byproducts at pH 9.0. With the notable exception of  **$\beta$ -Ala-CN**, all substrates show a strong anticorrelation between  $pK_{aH}$  and ligation yields. This relationship can be understood by considering the remarkably low basicity of  $\alpha$ -aminonitriles: with a typical  $pK_{aH}$  of 5.4, aminonitriles mostly exist as neutral, nucleophilic species, even at neutral pH. Oppositely, most other prebiotically relevant amines predominantly exist as unreactive ammonium groups under similar conditions. Owing to this drastic difference in basicity, aminonitriles are able to outcompete most nucleophiles tested across all investigated pH values. This idea is further supported by the fact that the only competition reactions in which **Gly-CN** was outcompeted was by **Gly-NH<sub>2</sub>** at pH 9.0. **Gly-NH<sub>2</sub>** has a  $pK_{aH}$  of 8.2, and is therefore almost 90% free-base at pH 9.0. This fact, coupled to the stronger nucleophilicity of **Gly-NH<sub>2</sub>** free base (relative to **Gly-CN** free base) leads to the high ligation yields observed for **Gly-NH<sub>2</sub>**.



**Figure 78:** Parallel between competitor  $pK_{aH}$  and ligation yields at pH 9.0.

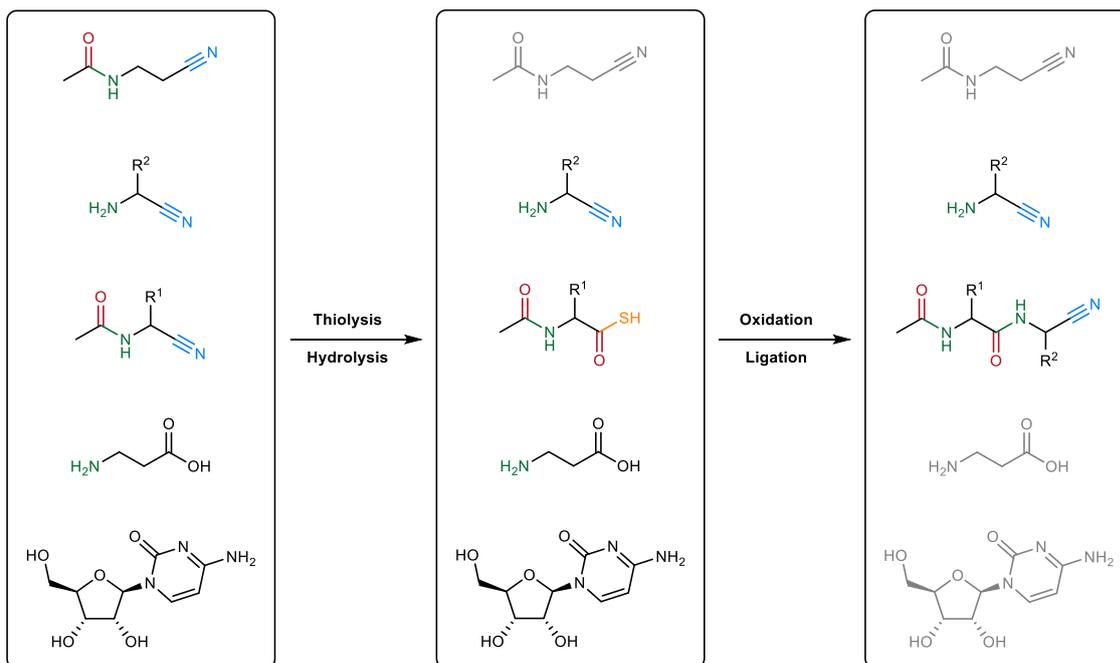
Selectivity studies brought to light the remarkable chemoselectivity of our elongation strategy, with each step building an additional layer of protection against the main classes of prebiotic species that could interfere with peptide elongation. On the one hand, the thiolysis exhibits near-perfect selectivity against both free aminonitrile monomers and non-proteinogenic nitriles. On the other hand, the oxidative ligation has proved highly selective against chain-terminating nucleophiles, including amino acids and nucleotides. Taken together, these two layers of selectivity allow to filter out most non-proteinogenic species, while preserving the free aminonitrile monomers required for the continued elongation of the peptide. This selectivity is enabled by the mutual activation provided by the nitrile and amide moieties of  $\alpha$ -amino- and  $\alpha$ -amidonitriles. The electron-withdrawing effect of the nitrile moiety of aminonitriles dramatically lowers the  $pK_a$  of the amine group, thus allowing aminonitriles to engage in nucleophilic addition even under acidic pH. Conversely, the formation of an  $\alpha$ -amide moiety resulting from the oxidative acylation of aminonitriles enhances the otherwise low electrophilicity of the vicinal nitrile, thus favouring its selective thiolysis (**Figure 79**).



**Figure 79:** Mutual activation of the nitrile and amide moieties.

### 5. 3. Conclusions

Such selectivity is entirely unprecedented among peptide synthesis methodologies, prebiotic or otherwise, which have to operate in the complete absence of competing species to proceed at any appreciable rate. This unique advantage over previous strategies further increases the plausibility of a prebiotic scenario involving our thiolysis-hydrolysis-ligation cycle, and points towards the possibility of performing our cycle as a continuous, one-pot process (**Figure 80**).

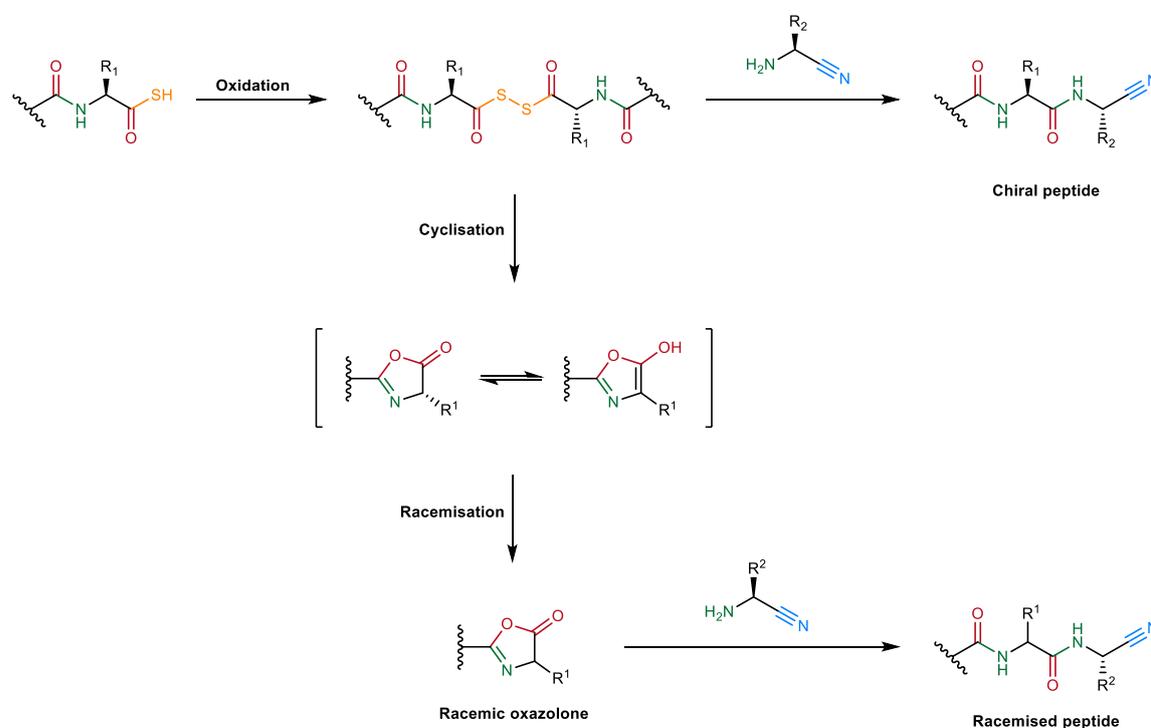


**Figure 80:** Chemoselectivity of the sulfur-mediated oxidative ligation pathway.

## 6. Enantioretention of the elongation process

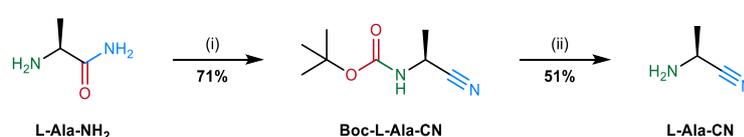
The emergence of homochirality in living organisms remains one of the major unsolved questions in prebiotic chemistry. In this context, prebiotic pathways that could induce homochirality in existing or elongating peptides and their monomers have been a long sought-after goal<sup>[145]</sup>. For this reason, we decided to examine the enantioselectivity and enantioretention properties of our route.

All aminonitrile ligation experiments performed so far have shown no significant enantio- or diastereoselectivity, as all dimer products were obtained as racemic mixtures. This outcome is in accordance with the proposed mechanisms for the thiolysis, hydrolysis, and ligation steps, neither of which involves any obvious source of chirality. In addition, no prebiotic route towards chiral, proteinogenic aminonitriles has been proposed to date. Consequently, devising enantioselective variants of our pathways sits far outside of the scope of present work. However, in the event that such a route was to be uncovered in the future, it nonetheless appeared necessary to ascertain that our process would preserve any enantiomeric excess present in the aminonitrile monomers. Indeed, a major impediment faced by *C*-terminal peptide activation strategies stems from the formation of oxazolone intermediates, whose rapid racemisation leads to the degradation of any enantiomeric excess present in the *C*-terminal residue (**Figure 81**).



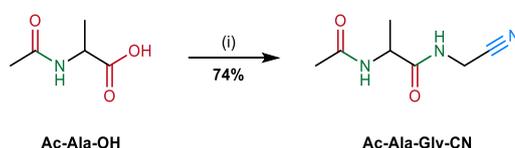
**Figure 81:** Possible cyclisation and racemisation pathways of activated peptide thioacids.

As disulfide intermediates do not typically tautomerize, and no oxazolone derivatives were ever observed from ligation experiments, we hypothesised that the addition of aminonitriles was rapid enough as to avoid the racemisation of activated peptide thioacids. To test this hypothesis, we set out to prepare enantiopure L-alanine nitrile **L-Ala-CN** and follow the evolution of its enantiomeric excess while proceeding to its acetylation, thiolysis, hydrolysis, and ligation into an *N*-acetyl dipeptide nitrile. To do so, a synthetic standard of enantiopure **L-Ala-CN** was prepared through an expedient, two-step process (**Figure 82**). This synthesis proceeds through the one-pot Boc-protection and dehydration of commercial, enantiopure L-alaninamide **L-Ala-NH<sub>2</sub>** to form Boc-protected aminonitrile **Boc-L-Ala-CN**, which was subsequently deprotected to afford L-alanine nitrile **L-Ala-CN**.



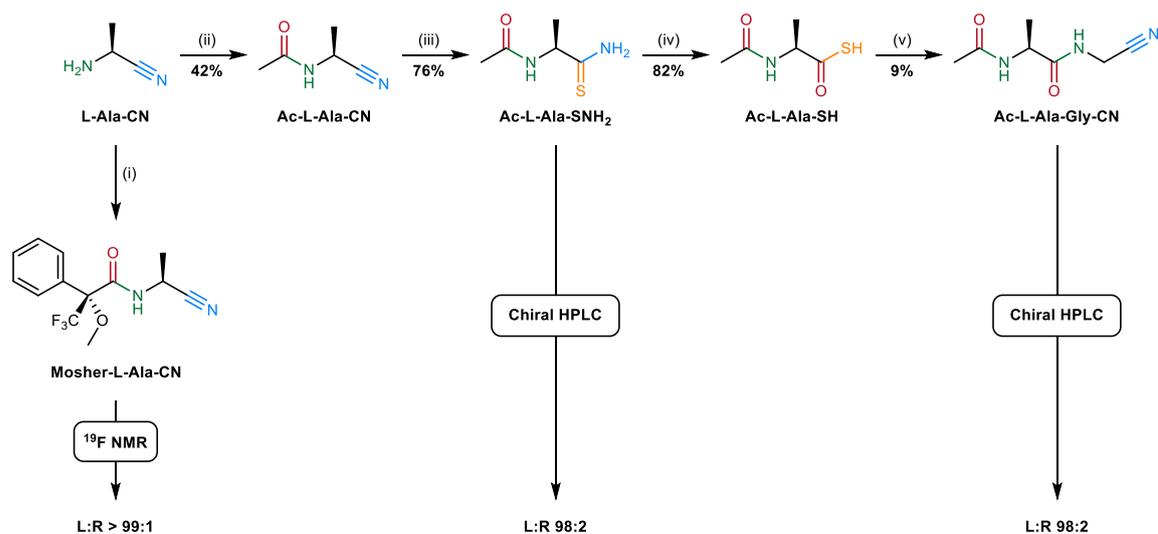
**Figure 82:** Preparation of **L-Ala-CN**; (i) **L-Ala-NH<sub>2</sub>**·HCl (1.0 eq.), Boc<sub>2</sub>O (2.0 eq.), NEt<sub>3</sub> (5.0 eq.), TFA anhydride (3.0 eq.), THF (100 mM **L-Ala-NH<sub>2</sub>**·HCl), 0 °C to RT, 5 h; (ii) **L-Ala-CN** (1.0 eq.), formic acid (235 mM **L-Ala-CN**), RT, 1 h.

In parallel, a synthetic standard for racemic *N*-acetyl alanyl glycine nitrile **Ac-Ala-Gly-CN** was prepared *via* conventional amide synthesis (**Figure 83**).



**Figure 83:** Preparation of **Ac-Ala-Gly-CN** (i) **Ac-Ala-OH** (1.0 eq.), **Gly-CN**·HCl (2.5 eq.), EDC·HCl (2.5 eq.), NEt<sub>3</sub> (5.0 eq.), DCM (150 mM **Ac-Ala-OH**), 0 °C to RT, 3 h.

With synthetic materials in hand, we proceeded to investigate the enantioselectivity of our route on **L-Ala-CN**. First, the enantiomeric ratio of **L-Ala-CN** was asserted by quantitative derivatisation with Mosher's acyl chloride to form **Mosher-L-Ala-CN** (**Figure 84**). <sup>19</sup>F NMR analysis confirmed the diastereomeric ratio of **Mosher-L-Ala-CN** to be above 99:1. With this value as a starting point, we went on to take **L-Ala-CN** through our acetylation-thiolysis-hydrolysis-ligation route to afford **Ac-L-Ala-Gly-CN**. Chiral HPLC analysis performed on **Ac-L-Ala-SNH<sub>2</sub>** and **Ac-L-Ala-Gly-CN** revealed that the diastereomeric ratio of the stereogenic carbon remains constant throughout the elongation process. These findings confirm that should a prebiotic route to chiral aminonitriles be uncovered, our strategy would remain suitable to form canonical, L-homochiral peptides.

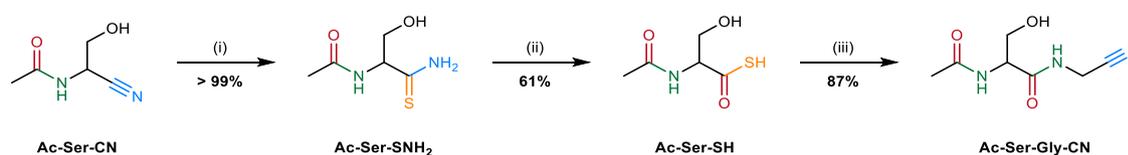


**Figure 84:** Enantioretention studies on **L-Ala-CN**; (i) **L-Ala-CN** (1.0 eq.),  $\text{NEt}_3$  (2.4 eq.), (R)-(+)- $\alpha$ -methoxy- $\alpha$ -(trifluoromethyl)phenylacetyl chloride (5.0 eq.), DMSO (100 mM **L-Ala-CN**), RT, 10 min; (ii) **L-Ala-CN** (1.0 eq.), AcSK (3.0 eq.), potassium ferricyanide (9.0 eq.),  $\text{H}_2\text{O}$  (50 mM **L-Ala-CN**), pH 9.0, RT, 20 min; (iii) **Ac-L-Ala-CN** (1.0 eq.), NaSH (10.0 eq.),  $\text{H}_2\text{O}$  (50 mM **Ac-L-Ala-CN**), pH 9.0, RT, 24 h; (iv) **Ac-L-Ala-SNH<sub>2</sub>** (1.0 eq.), NaSH (10.0 eq.),  $\text{H}_2\text{O}$  (50 mM **Ac-L-Ala-SNH<sub>2</sub>**), pH 9.0, 60 °C, 24 h; (v) **Ac-L-Ala-SH** (1.0 eq.), **Gly-CN** (5.0 eq.), potassium ferricyanide (3.0 eq.),  $\text{H}_2\text{O}$  (50 mM **Ac-L-Ala-SH**), pH 9.0, RT, 20 min.

## 7. One-pot experiments

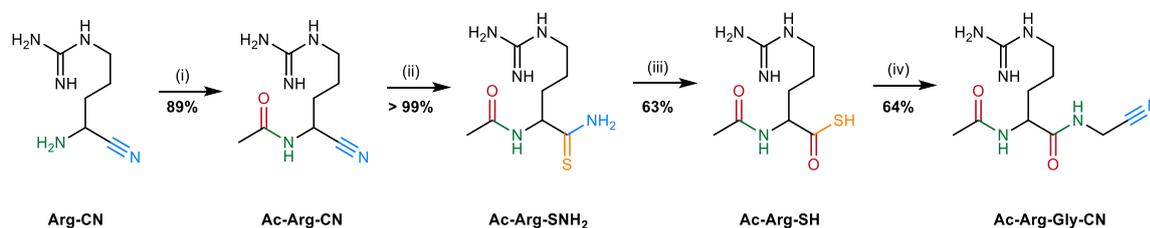
### 7. 1. Prebiotic, one-pot synthesis of *N*-acetyl dipeptide nitriles

Building upon the remarkable properties of our system, we set out to put its robustness to the test by turning our strategy into a one-pot, continuous process. As an additional challenge, we elected to perform this transition using particularly difficult sidechains, such as serine and arginine, which proved so reactive as to make the isolation of their *N*-acetyl aminothioamides counterparts **Ac-Ser-SNH<sub>2</sub>** and **Ac-Arg-SNH<sub>2</sub>** infeasible. To this end, **Ac-Ser-CN** was thiolysed into **Ac-Ser-SNH<sub>2</sub>** following conditions described in **Figure 42**. The resulting thioamide was immediately hydrolysed to form **Ac-Ser-SH** in 61% yield, along with **Ac-Ser-NH<sub>2</sub>** and **Ac-Ser-OH**. Finally, addition of **Gly-CN** and potassium ferricyanide led to the formation of **Ac-Ser-Gly-CN** in 87% yield (**Figure 85**).



**Figure 85:** One-pot thiolysis, hydrolysis and ligation of **Ac-Ser-CN** into **Ac-Ser-Gly-CN**; (i) **Ac-Ser-CN** (1.0 eq.), NaSH (10.0 eq.), H<sub>2</sub>O (50 mM **Ac-Ser-CN**), pH 9.0, RT, 12 h; (ii) **Ac-Ser-SNH<sub>2</sub>** (1.0 eq.), NaSH (10.0 eq.), H<sub>2</sub>O (50 mM **Ac-Ser-SNH<sub>2</sub>**), pH 9.0, 60 °C, 18 h; (iii) **Ac-Ser-SH** (1.0 eq.), **Gly-CN** (2.0 eq.), potassium ferricyanide (3.0 eq.), H<sub>2</sub>O (30.5 mM **Ac-Ser-SH**), pH 9.0, RT, 20 min.

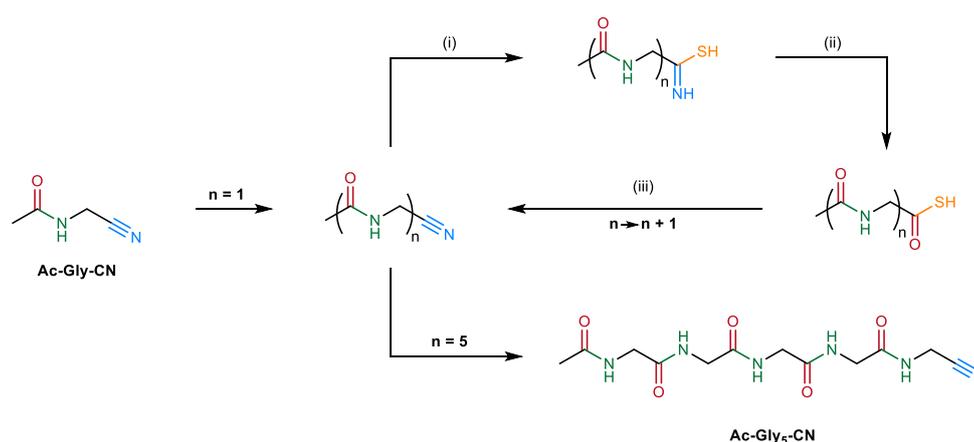
Encouraged by those results, we went on to perform a second series of one-pot experiments, this time starting from **Arg-CN**. In spite of the supplementary acetylation step, similarly high yields were obtained, affording **Ac-Arg-Gly-CN** in 36% yield from **Arg-CN** after a total of 4 steps (**Figure 86**).



**Figure 86:** One-pot acetylation, thiolysis, hydrolysis and ligation of **Arg-CN** into **Ac-Arg-Gly-CN**; (i) **Arg-CN** (1.0 eq.), AcSK (3.0 eq.), potassium ferricyanide (9.0 eq.), H<sub>2</sub>O (50 mM **Arg-CN**), pH 9.0, RT, 20 min; (ii) **Ac-Arg-CN** (1.0 eq.), NaSH (10.0 eq.), H<sub>2</sub>O (45 mM **Ac-Arg-CN**), pH 9.0, RT, 24 h; (iii) **Ac-Arg-SNH<sub>2</sub>** (1.0 eq.), NaSH (10.0 eq.), H<sub>2</sub>O (44.5 mM **Ac-Ser-SNH<sub>2</sub>**), pH 9.0, 60 °C, 24 h; (iv) **Ac-Arg-SH** (1.0 eq.), **Gly-CN** (2.0 eq.), potassium ferricyanide (3.0 eq.), H<sub>2</sub>O (28 mM **Ac-Ser-SH**), pH 9.0, RT, 20 min.

## 7. 2. Iterative elongation of glycine nitrile

Having demonstrated the possibility of performing one-pot sequences converting aminonitriles directly into *N*-acetyl dipeptide nitriles, we set out to apply our strategy to the continuous, iterative elongation of peptides. Starting from **Ac-Gly-CN**, the thiolysis, hydrolysis and ligation steps were performed in sequence to form **Ac-Gly-Gly-CN**. The process was repeated without purification to afford *N*-acetyl pentaglycine nitrile **Ac-Gly<sub>5</sub>-CN** in 13% yield, after a cumulative three weeks of reaction time (**Figure 87**).

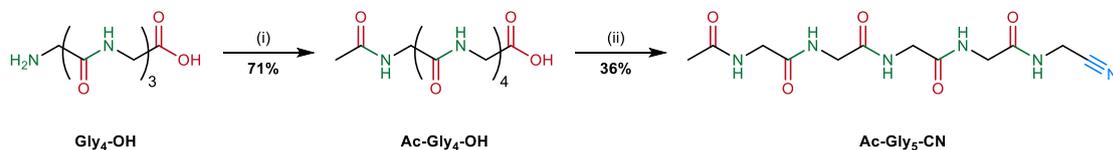


**Figure 87:** Iterative elongation of **Ac-Gly-CN**; (i) **Ac-Gly<sub>n</sub>-CN** (1.0 eq.), H<sub>2</sub>S<sub>(g)</sub>, Dowex® (SH<sup>-</sup> form), H<sub>2</sub>O (100 mM **Ac-Gly-CN**), pH 9.0, RT, 8 h; (ii) **Ac-Gly<sub>n</sub>-SNH<sub>2</sub>** (1.0 eq.), H<sub>2</sub>O pH 9.0, 60 °C, 24 h; (iii) **Ac-Gly<sub>n</sub>-SH** (1.0 eq.), **Gly-CN** (1.2 eq.), potassium ferricyanide (3.0 eq.), H<sub>2</sub>O, pH 9.0, RT, 30 min.

In spite of the considerable number of byproducts that could possibly form and interfere with the elongation process, consistently high yields were maintained up until the last round of ligation (**Table 3**). The water solubility of pentapeptide nitrile **Ac-Gly<sub>5</sub>-CN** proved a limiting factor, as spontaneous precipitation put an end the iteration process, and is likely underpinning the lower ligation yields observed during the last step of the iteration. Nevertheless, unprecedentedly high pentamer yields of 13% were observed by <sup>1</sup>H NMR after a total of 12 successive steps.

Round	Starting material	Product	Yield (step)	Yield (from <b>Ac-Gly-CN</b> )
1	<b>Ac-Gly-CN</b>	<b>Ac-Gly<sub>2</sub>-CN</b>	71%	71%
2	<b>Ac-Gly<sub>2</sub>-CN</b>	<b>Ac-Gly<sub>3</sub>-CN</b>	71%	51%
3	<b>Ac-Gly<sub>3</sub>-CN</b>	<b>Ac-Gly<sub>4</sub>-CN</b>	63%	32%
4	<b>Ac-Gly<sub>4</sub>-CN</b>	<b>Ac-Gly<sub>5</sub>-CN</b>	41%	13%

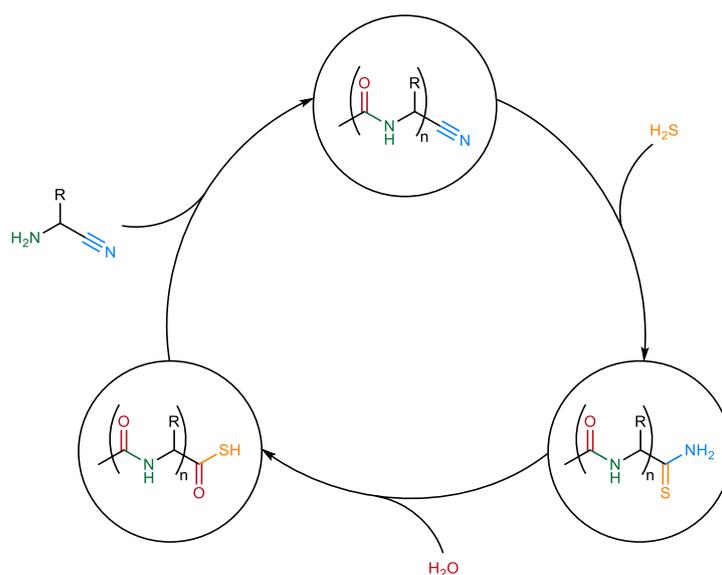
As could have been expected in view of the difficulty of purifying peptide nitriles resulting from ferricyanide-mediated ligation, all attempts at isolating **Ac-Gly<sub>5</sub>-CN** from the final iteration mixture failed. As an alternative, a pure synthetic standard for **Ac-Gly<sub>5</sub>-CN** was prepared through the sequential ligation of tetraglycine **Gly<sub>4</sub>-OH** with **Ac-Gly-PFP** and **Gly-CN** (**Figure 88**).



**Figure 88:** Preparation of synthetic **Ac-Gly<sub>5</sub>-CN** from **Gly<sub>3</sub>-OH**; (i) **Gly<sub>3</sub>-OH** (1.0 eq.), **Ac-Gly-PFP** (1.0 eq.), **NEt<sub>3</sub>** (1.2 eq.), **DMF** (125 mM **Gly<sub>3</sub>-OH**), RT, 2 h; (ii) **Ac-Gly<sub>4</sub>-OH** (1.0 eq.), **Gly-CN·HCl** (3.0 eq.), **EDC·HCl** (1.5 eq.), **NEt<sub>3</sub>** (5.0 eq.), **DMF** (70 mM **Gly<sub>3</sub>-OH**), RT, 4 h.

## 8. Conclusions and future work

In a clear departure from paradigms which have dominated the field for the past 60 years, this work describes the highest-yielding, non-amino acid based prebiotic peptide elongation strategy to date. This strategy is founded on the unique reactivity of amidothioacids, for which the first high-yielding prebiotic synthesis is proposed. The activation and ligation of amidothioacids is demonstrated to proceed in high to quantitative yields, which makes it most efficient amidation reaction documented in prebiotic chemistry. Furthermore, the ligation is fully compatible with all 20 proteinogenic sidechains, making it the most general amidation strategy in all the prebiotic literature. Crucially, our route displays exquisite selectivity against non proteinogenic substrates, including RNA nucleotides. In addition, this strategy offers complete enantioretention of its substrates. Moreover, our route relies entirely on simple, plausible reagents, and on mild conditions guaranteeing the preservation of its peptide products. These advantageous conditions do not require thermal, pH, or dry-wet cycles, and are built on similar reagents as the cyanosulfidic protometabolism producing both activated nucleotides and aminonitrile monomers (**Figure 89**).



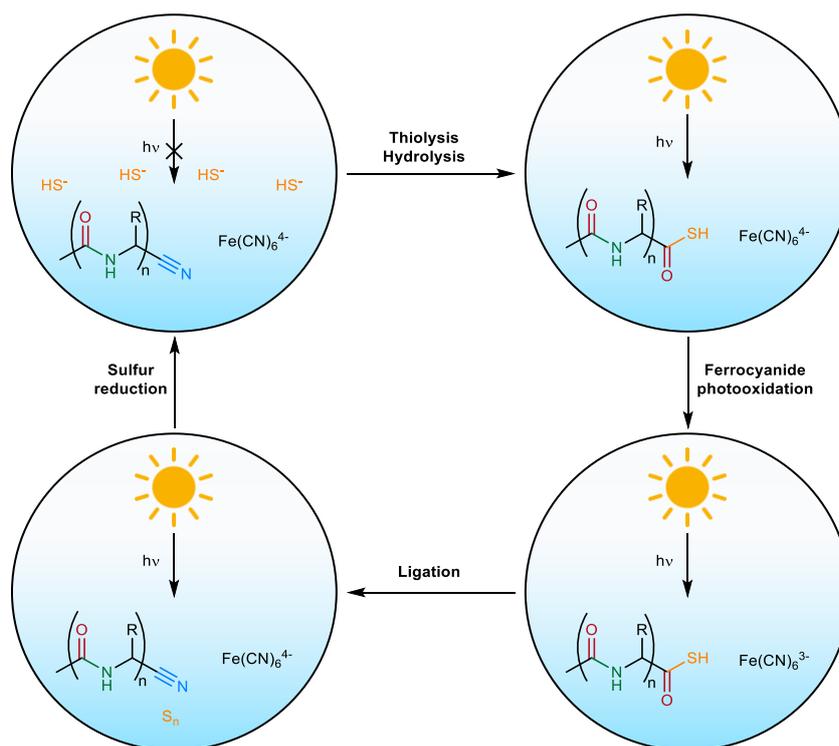
**Figure 89:** Iterative *N*-acetyl peptide elongation by sequential thiolysis, hydrolysis and  $\alpha$ -aminonitrile ligation.

Despite these achievements, our strategy presents a number of limitations. First, the hydrolysis of amidothioamides to amidothioacids has proved extremely inefficient in the case of  $\beta$ -branched residues (*e.g.* valine, isoleucine). While this limitation only affects two out of the 20 natural sidechains, the relatively high frequencies of those residues in extant organisms and the central role played by large hydrophobic sidechains in the folding of peptides and proteins make their incorporation into prebiotic peptides highly desirable. In addition, two key reagents in our cycle –

hydrogen sulfide and ferricyanide – are mutually destructive, and react to form elemental sulfur and ferrocyanide. This reactivity imposes a separation between the thiolysis and ligation steps – either temporal, through sequential delivery of the reagents; or spatial, through physically segregated reaction environments. While a variety of scenarios involving membrane compartmentation or flow chemistry have been invoked to resolve similar limitations in prebiotic systems, the increased complexity of such models renders their prebiotic realisation less likely than single-environment settings. This issue is compounded by the relatively large excess of reagents employed by our strategy, which implies that elongating substantial amounts of peptides would require implausible quantities of hydrogen sulfide and ferricyanide in the absence of a recycling strategy allowing the regeneration of spent reagents. In order to both overcome those limitations and expand the range of applications of the amidonitriles and amidothioacids, several extensions to our strategy can be envisaged.

### **8. 1. Photoredox regeneration of inorganic reagents**

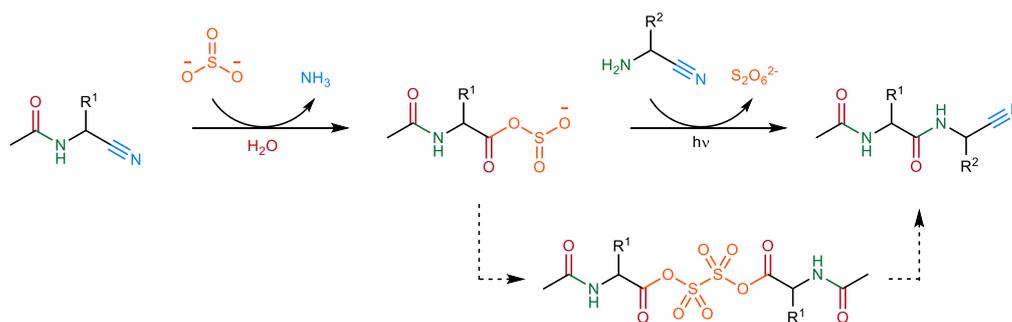
Aqueous ferrocyanide has been shown to be photo-oxidised into ferricyanide under UV irradiation<sup>[146]</sup>. The reduction expels a solvated electron, which could in turn initiate the reduction of elemental sulfur. Interestingly, solvated hydrogen sulfide is known to block a large portion of UV radiation out of its environment<sup>[147]</sup>. Taken together, these properties allow to draw a prebiotically plausible, closed-loop photoredox cycle, which could potentially solve both the mutual reactivity issue and reagent regeneration issues outlined above (**Figure 90**). Assuming an initial state in which sulfur and iron cyanide complexes exist as hydrogen cyanide and ferrocyanide, respectively, amidonitriles could be converted to amidothioacids without interference from metal oxidants. Sequestration of hydrogen sulfide as thiocarbonyles would allow for the photooxidation of ferrocyanide to progressively generate ferricyanide, which would in turn react with amidothioacids to form elongated amidonitriles and oxidised sulfur species. Finally, both UV irradiation and the solvated electrons released by the photooxidation of ferrocyanide would convert polysulfides back to hydrogen sulfide. If experimentally verified, this self-sustained process could eliminate the need for the supervised, sequential delivery of reagents. By taking advantage of natural day – night cycles as a source of UV irradiation, this process would also allow to regenerate both hydrogen sulfide and ferricyanide from spent materials. In addition, such cycles would couple peptide synthesis to solar radiation, thus harnessing a virtually endless energy source driving forward the elongation process.



**Figure 90:** Tentative photoredox cycle for the hydrogen sulfide – ferricyanide redox tandem.

## 8. 2. Sulfites as amidonitrile activating agents

The inclusion of photoredox cycles opens the way to alternative aminonitrile activation strategies. Sulfite ions have been proposed as plausible prebiotic forms of sulfur<sup>[147]</sup>, and have been shown to readily add onto aminonitriles<sup>[148]</sup>. Building upon those properties, a potential research avenue would consist of investigating the possibility of forming sulfurous anhydrides from the addition of sulfite dianions onto amidonitriles (**Figure 91**). In a similar fashion to amidothioacids, these negatively-charged sulfurous anhydrides can be expected to be hydrolytically stable under non-acidic pH. Importantly, sulfite anions have been shown to dimerise under UV illumination to form dithionate species. Such reactivity could be leveraged to oxidise sulfurous anhydrides into uncharged dimers<sup>[148]</sup>, which could in turn ligate with aminonitriles to afford *N*-acetylated dipeptide nitriles. This alternative elongation strategy would present several advantages over the hydrogen sulfide – ferricyanide system. First, it would do away with the need for a metal oxidant and replace it with UV-driven photoactivation, thus eliminating the issue of mutual reactivity, and reducing the chemical complexity of the peptide elongation network. Second, both the sulfurous anhydrides and their dimerised counterpart would be unable to cyclise into methyl thiooxazoles, which could allow for the incorporation of  $\beta$ -branched residues into the elongating peptide. Finally, the addition of a light-driven step would create an energy entry point providing the peptide elongation with a driving force.

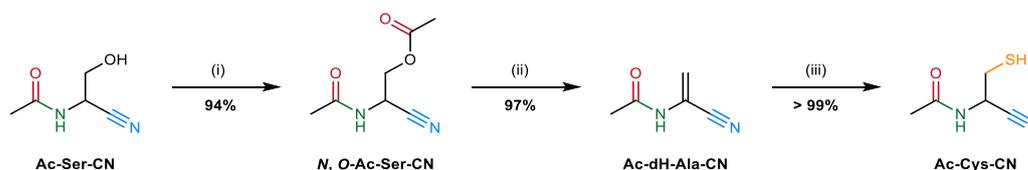


**Figure 91:** Tentative sulfite-based amidonitrile elongation strategy.

### 8. 3. Dehydroalanine and post-ligation sidechain synthesis

Taking inspiration from protein modification strategies pioneered by B. Davis<sup>[149]</sup>, the author proposed that serine nitrile derivatives could undergo a simple, acetylation-driven activation-elimination to form dehydroalanine nitrile species, which are known to readily undergo nucleophilic functionalisation. From there on, addition of prebiotically available building blocs onto dehydroalanine would allow access to key amino acid sidechains. Among those, cysteine is of prime interest. The thiol sidechain of cysteine is included in several catalytic triads<sup>[150]</sup>, which play pivotal roles in biocatalysis, and would open the way towards NCL-like ligation strategies. Moreover, it has recently been shown to assemble with ferrous ions to form iron-sulfur cluster under prebiotic conditions<sup>[151]</sup>, which could also have played a crucial role in abiotic catalysis. However, a prebiotic pathway to cysteine has yet to be uncovered.

In a series of preliminary experiments, we discovered that applying oxidative acetylation conditions to *N*-acetyl serine nitrile **Ac-Ser-CN** afforded its bis-acetylated counterpart, ***N, O*-Ac-Ser-CN**. Incubation of ***N, O*-Ac-Ser-CN** at pH > 9.0 led to the rapid, quantitative elimination of the acetyl group to afford dehydroalanine **Ac-dH-Ala-CN**. The subsequent addition of NaSH led to the quantitative conversion of **Ac-dH-Ala-CN** to *N*-acetyl cysteine nitrile **Ac-Cys-CN** in under 10 minutes. These mild, high-yielding reactions were carried out as a one-pot sequence, and demonstrate the first prebiotic route towards cysteine sidechains (**Figure 92**).



**Figure 92:** Prebiotic one-pot synthesis of **Ac-Cys-CN** from **Ac-Ser-CN**; (i) **Ac-Ser-CN** (1.0 eq.), AcSK (3.0 eq.), K<sub>3</sub>[Fe(CN)<sub>6</sub>] (9.0 eq.), H<sub>2</sub>O (50 mM **Ac-Ser-CN**), pH 9.0, RT, 20 min; (ii) H<sub>2</sub>O (50 mM ***N, O*-Ac-Ser-CN**), pH 9.0, RT, 10 min; (iii) ***N, O*-Ac-Ser-CN** (1.0 eq.), NaSH (5.0 eq.), H<sub>2</sub>O (50 mM ***N, O*-Ac-Ser-CN**), pH 9.0, RT, 20 min.

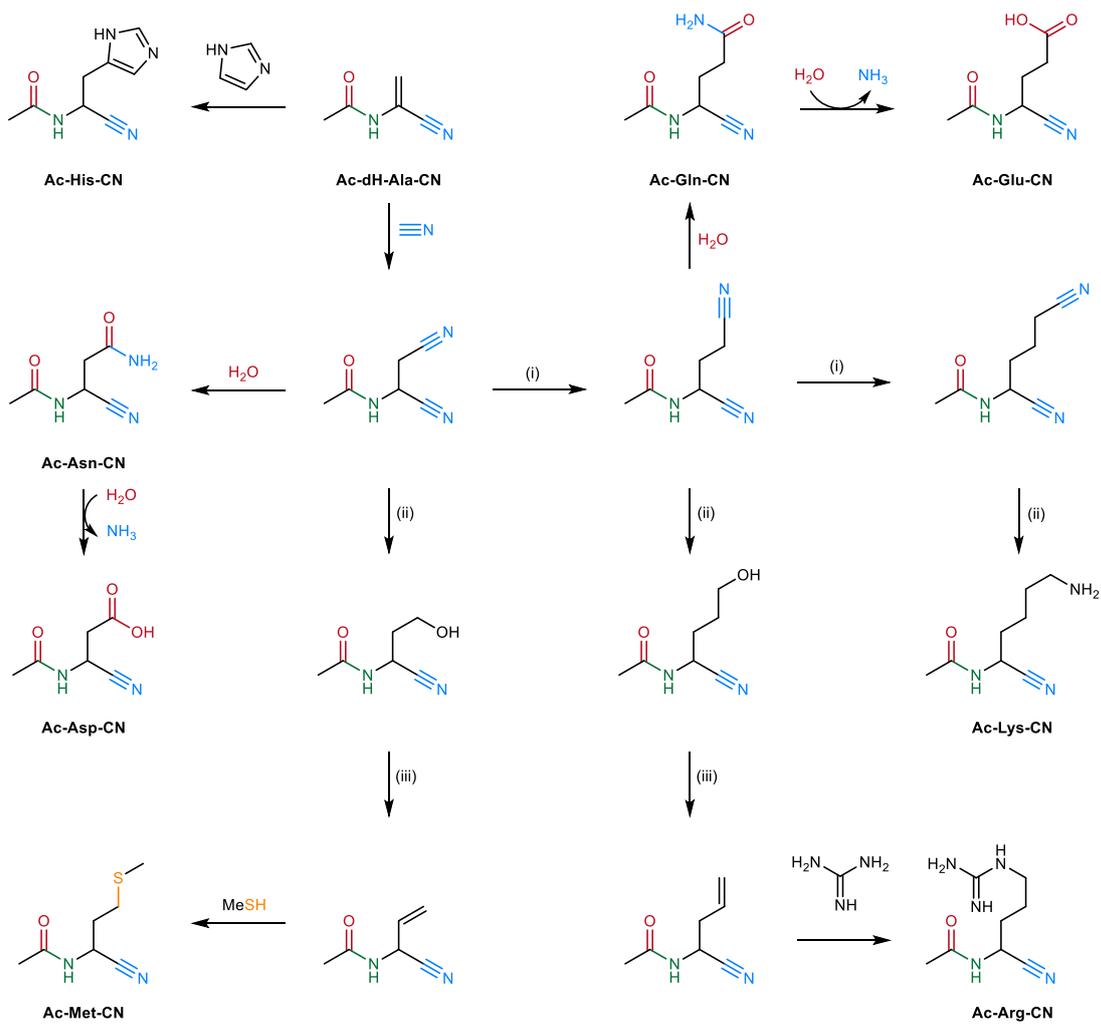
Encouraged by those results, the author proposed a reaction network stemming from **Ac-dH-Ala-CN** which would allow to access biologically relevant, yet elusive sidechains, including histidine and lysine (**Figure 93**). This simple network relies on a set of three reactions:

- The photoredox elongation of nitrile-terminated alkyl sidechains, described by Sutherland as part of a cyanosulfidic protometabolism<sup>[97]</sup>.
- The reduction of nitriles into primary alcohols and amines, which can be achieved by interrupting the photoredox sequence described above, or by using prebiotically abundant metal catalysts, including FeS and Ni/Fe alloys<sup>[152]</sup>.
- The acetylation-elimination sequence described above.

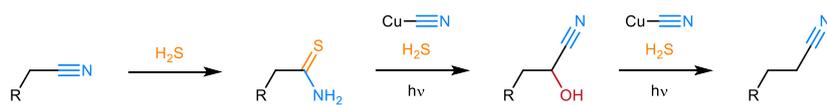
If experimentally realised, this network would bring to light the first route to catalytically crucial sidechains, such as histidine. In addition, it would further strengthen the connections between protein synthesis and cyanosulfidic protometabolism. Finally, this network would represent the first successful biomimetic, post-ligation sidechain synthesis, thus establishing a direct connection between prebiotic peptide synthesis and biological, post-translational protein modification.

Although promising, the exploration of this network may prove experimentally intensive. First, all three reactions listed above can be expected to become increasingly challenging as the sidechain grows larger and the electron-withdrawing effect of the amidonitrile group diminishes. Furthermore, while all substrates shown in **Figure 93** bear aminonitrile handles, the impact of variations of the C-terminal moiety should be examined. For this purpose, the pathway described in **Figure 91** should be repeated on substrates including both free and acetylated variants of **Ser-OH**, **Ser-NH<sub>2</sub>**. Crucially, the selectivity of the acetylation – elimination process for serine monomers *versus* amide-linked, serine-containing oligomers should be asserted. In addition, other prebiotic O-activating groups, including phosphates, could offer interesting and biomimetic alternatives to O-acylation.

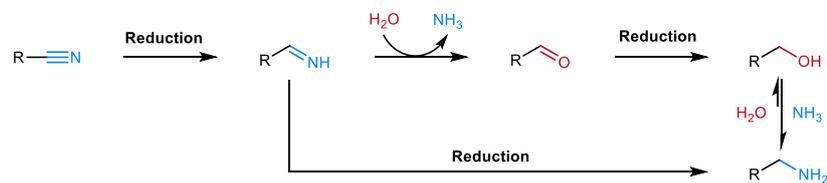
A supplementary application of dehydroalanine functionalisation resides in the possibility of adding nucleobases to poly-dehydroalanine oligomers to form peptide nucleic acids (PNAs)<sup>[153]</sup>. While PNAs have long been hypothesised as a possible abiotic alternative to ribonucleic polymers, no prebiotic synthesis leading to either PNA oligomers or monomers has been proposed to date<sup>[154]</sup>. Considering that serine nitrile and its hydrolysis products are readily formed through the Strecker reaction of glycolaldehyde, one of the most abundant prebiotic building blocks, and that a multitude of pathways to free nucleobases have been demonstrated, assembling poly-dehydroalanine oligomers and nucleobases into base-pairing PNAs remains a promising perspective.



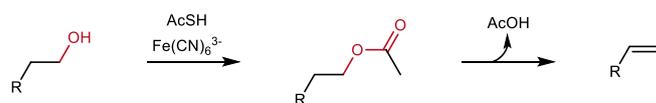
(i)



(ii)



(iii)



**Figure 93:** Tentative reaction network stemming from **Ac-dH-Ala-CN**.

## 9. Experimental

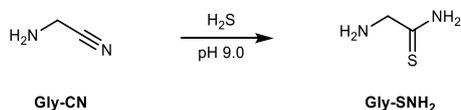
### 9. 1. General experimental

Reagents and solvents were obtained and used without further purification, unless specified, from the following commercial sources: Sigma Aldrich, Alfa Aesar, Fluorochem, Acros Organics, Merck, Fisher Scientific, VWR International, Carbosynth, Manchester Organics, BDH, Lancaster, Apollo Scientific, Molekula, TCI and Santa Cruz Biotechnology. Dowex® 50W × 8 ion-exchange resin (200-400 mesh) ion exchange resin was purchased from Acros Organics and was washed with methanol and sodium hydroxide solution before being regenerated with hydrochloric acid solution. Deionized water was obtained from an Elga Option 3 purification system. <sup>1</sup>H and <sup>13</sup>C NMR spectra were recorded on Bruker NMR spectrometers AVANCE III 600, AVANCE III 400 and AVANCE 300 equipped with a Bruker 5 mm cryoprobe (600 MHz) and a gradient probe (400 and 300 MHz). All chemical shifts (δ) are reported in parts per million (ppm) relative to residual solvent peaks, and <sup>1</sup>H and <sup>13</sup>C chemical shifts relative to TMS were calibrated using the residual solvent peak. 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), or a combination of these. NMR data are reported as follows: chemical shift (multiplicity, coupling constants (J), number of protons, nuclear assignment). Spectra were recorded at 298 K. Methylsulfonylmethane (MSM) was used as an internal standard for all prebiotically relevant <sup>1</sup>H NMR experiments (characteristic signal at 3.17 ppm, s, 6H). Infrared spectra (IR) were recorded on a Shimadzu IR Tracer 100 FT-IR spectrometer. Absorption maxima are reported in wavenumber (cm<sup>-1</sup>). Mass spectra and accurate mass measurements were recorded on a VG70-SE, Waters LCT Premier XE or Thermo Finnigan MAT 900XP instrument at the Department of Chemistry, University College London. Solution pH values were measured using a Mettler Toledo Seven Compact pH meter with a Mettler Toledo InLab semi-micro pH probe, or a Corning pH meter 430 with a Fisherbrand FB68801 semi-micro pH probe. Unless stated otherwise, all experiments were carried out under argon atmosphere. Aqueous solvent systems (H<sub>2</sub>O, D<sub>2</sub>O and their mixtures) were systematically degassed by boiling and recondensation under continuous argon flow. All aqueous buffers were dissolved in degassed aqueous solvents. The pH of aqueous solutions was adjusted using degassed NaOH/HCl or NaOD/DCl aqueous solutions. PE refers to petroleum ether (b. p. 40 – 60 °C). NaSH·xH<sub>2</sub>O purchased from Sigma-Aldrich was used as a source of hydrosulfide and was assumed to contain 50% of NaSH by weight.

## 9. 2. Prebiotic experiments

### 9. 2. 1. Attempted polymerisation of glycine nitrile

#### Thiolysis of glycine nitrile



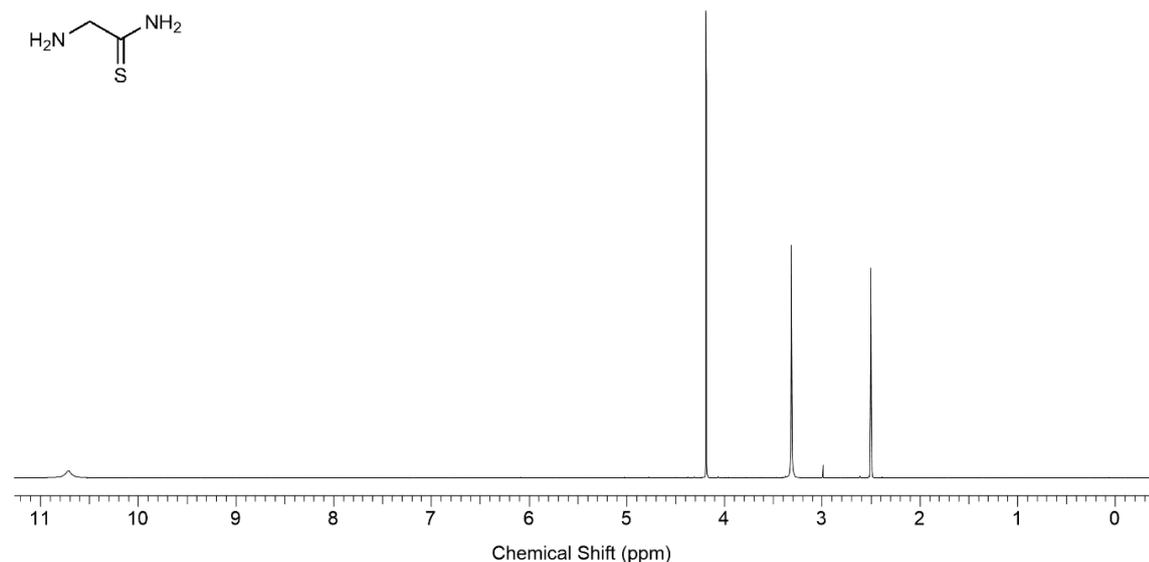
Glycine nitrile (**Gly-CN**, 0.50 mmol), NaSH·xH<sub>2</sub>O (561 mg, 5.00 mmol) and methylsulfonylmethane (0.50 mmol) were dissolved in degassed H<sub>2</sub>O/D<sub>2</sub>O (98:2, 10 mL). The solution was adjusted to pH 9.0 with NaOH/HCl, and NMR spectra were periodically acquired. A white precipitate appeared after 16 h, which started crystallising after 5 d. After 7 d, NMR analysis showed complete conversion of **Gly-CN** to glycine thioamide **Gly-SNH<sub>2</sub>**. The crystals were isolated and dissolved in DMSO-*d*<sub>6</sub>. NMR and HRMS analyses identified the solid as glycine thioamide **Gly-SNH<sub>2</sub>** (45.1 mg, 0.50 mmol, > 99%).

<sup>1</sup>H NMR (700 MHz, DMSO-*d*<sub>6</sub>): δ 10.71 (br. s., 1H, NH) 4.18 (s, 2H, (C2)–H<sub>2</sub>).

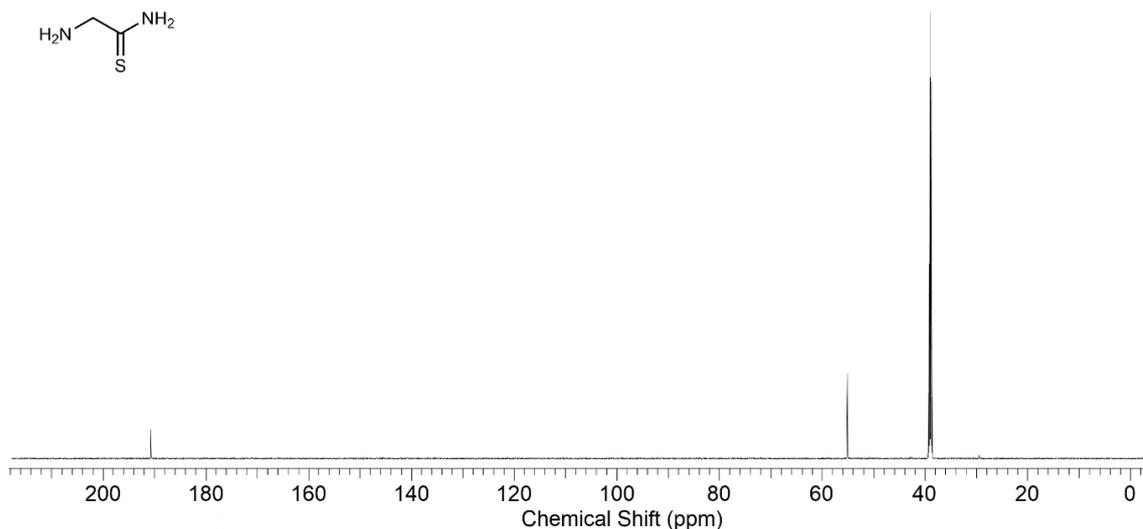
<sup>13</sup>C NMR (176 MHz, DMSO-*d*<sub>6</sub>): δ 190.7 (C1), 55.2 (C2).

IR (solid): 3181, 3085, 2985, 2880, 2179, 1577 cm<sup>-1</sup>.

M.p. 71.2–71.6 °C (decomp.)



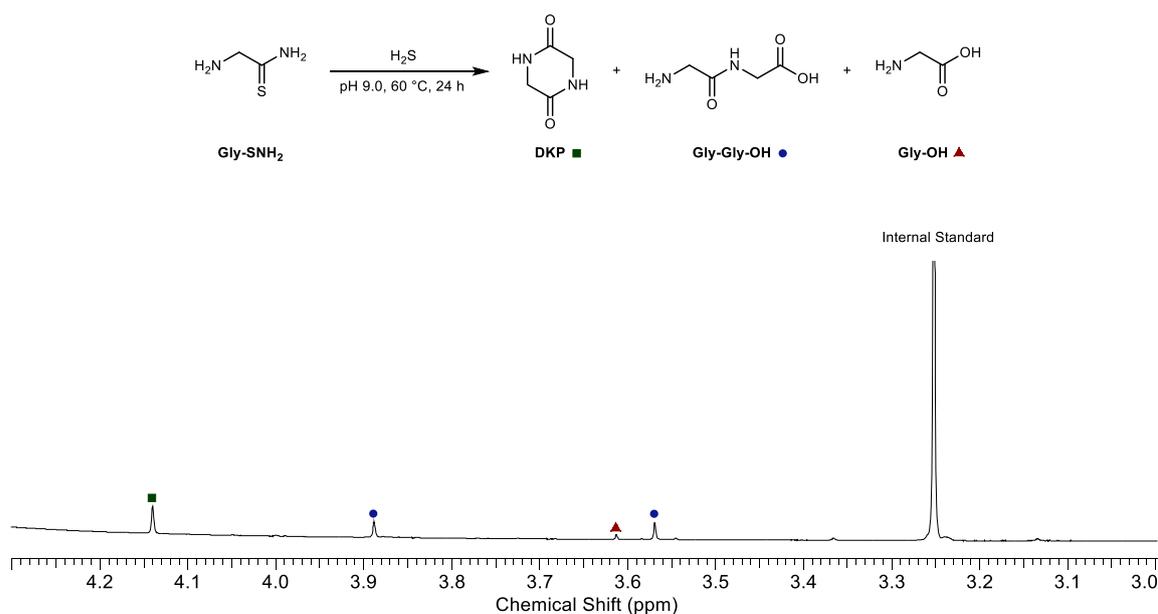
**Experimental figure 1:** <sup>1</sup>H NMR (700 MHz, DMSO-*d*<sub>6</sub>, -0.40 – 11.20 ppm) spectrum of glycine thioamide **Gly-SNH<sub>2</sub>**.



**Experimental figure 2:**  $^{13}\text{C}$  NMR (176 MHz,  $\text{DMSO-}d_6$ , 0 – 220 ppm) spectrum of glycine thioamide **Gly-SNH<sub>2</sub>**.

#### Attempted hydrolysis of glycine thioamide to glycine thioacid

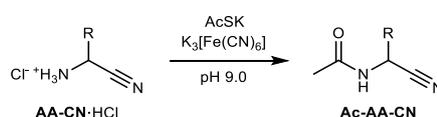
Glycine thioamide **Gly-SNH<sub>2</sub>** (9.0 mg, 0.10 mmol) was added to a solution of methylsulfonylmethane (0.10 mmol) in degassed  $\text{H}_2\text{O}/\text{D}_2\text{O}$  (98:2, 2 mL). The solution was adjusted to pH 9.0 with NaOH/HCl, and incubated at 60 °C for 24 h. After that time, NMR spectra were acquired. The presence of piperazine-2,5-dione **DKP** (25%), glycyglycine **Gly-Gly-OH** (14%) and glycine **Gly** (3%) was confirmed by  $^1\text{H} - ^{13}\text{C}$  HMBC NMR analysis and spiking with pure synthetic standards. The observable products were quantified using MSM as internal standard.



**Experimental figure 3:**  $^1\text{H}$  NMR (600 MHz,  $\text{D}_2\text{O}/\text{H}_2\text{O}$  2:98, 3.00 – 4.30 ppm) spectrum to show the incubation of glycine thioamide **Gly-SNH<sub>2</sub>** (50 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 2 mL) at 60 °C for 24 h.

## 9. 2. 2. Oxidative acetylation of aminonitriles

### General procedure A: Oxidative acetylation of aminonitriles

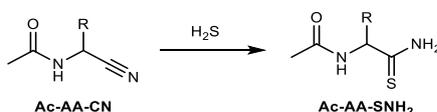


Aminonitrile hydrochloride **AA-CN·HCl** (0.1 mmol) and potassium thioacetate (0.3 mmol) were dissolved in H<sub>2</sub>O (2 mL) and the solution was adjusted to pH 9.0 with NaOH. Potassium hexacyanoferrate(III) (296 mg, 0.9 mmol) was added, and the solution was stirred at room temperature for 20 min. The solution was adjusted to pH 9.0, centrifuged, and NMR spectra of the supernatant were acquired. All identified products were confirmed by signal amplification upon addition of pure synthetic standards.

Experimental table 1: Oxidative acetylation of aminonitriles		
Entry	Residue	<sup>1</sup> H NMR yield (%)
1	Gly	> 99
2	Ala	> 99
3	Arg	> 99
4	Leu	> 99
5	Ile	> 99
6	Lys	> 99
7	Met	> 99
8	Phe	> 99
9	Pro	> 99
10	Ser	> 99
11	Val	> 99

### 9. 2. 3. Thiolysis of *N*-acetylaminonitriles

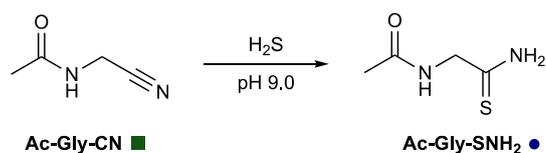
#### **General procedure B: Thiolysis of *N*-acetylaminonitriles**



*N*-Acetylaminonitrile **Ac-AA-CN** (2.50 mmol) and NaSH·xH<sub>2</sub>O (2.80 g, 25.0 mmol) were dissolved in degassed H<sub>2</sub>O/D<sub>2</sub>O (98:2, 50 mL, 50 mM). The solution was adjusted to pH 9.0 and stirred at room temperature for 24 h. NMR spectra were periodically acquired, until complete conversion of *N*-acetylaminonitrile **Ac-AA-CN** to *N*-acetylaminothioamide **Ac-AA-SNH<sub>2</sub>** was observed. The solution was then sparged with argon for 15 min at pH 5.0 and concentrated *in vacuo*. The residue was purified by flash column chromatography (SiO<sub>2</sub>; eluting with a gradient of MeOH/CH<sub>2</sub>Cl<sub>2</sub> 0:100 → 10:90) to afford *N*-acetylaminothioamide **Ac-AA-SNH<sub>2</sub>**.

Experimental table 2: Thiolysis of <i>N</i> -acetylaminonitriles				
Entry	Residue	NMR yield (%)	Isolated yield (%)	Reaction time (h)
1	Gly	> 99	> 99	24
2	Ala	> 99	> 99	48
3	Arg	> 99	-	36
4	Ile	> 99	39	48
5	Leu	> 99	32	48
6	Lys	> 99	-	36
7	Met	> 99	58	48
8	Phe	> 99	95	24
9	Pro	> 99	77	96
10	Ser	> 99	-	24
11	Val	> 99	96	72

### Thiolysis of *N*-Acetylglycine nitrile **Ac-Gly-CN**

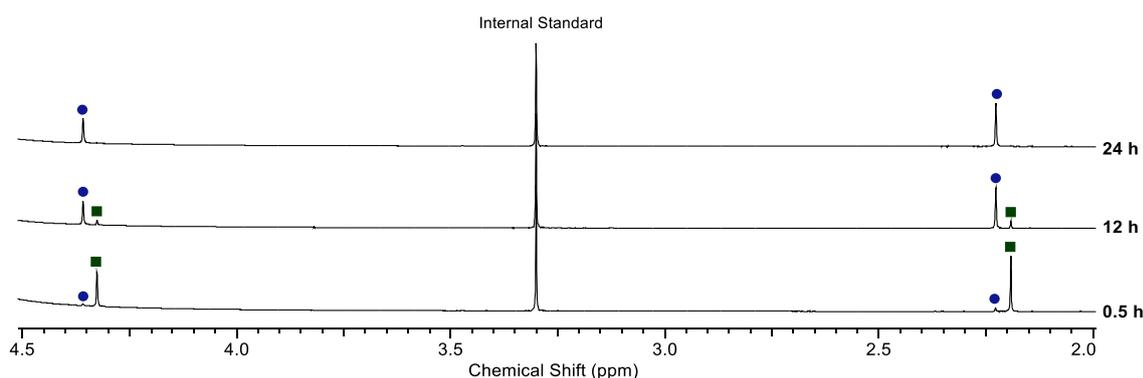


*N*-Acetylglycine nitrile **Ac-Gly-CN** (24.5 mg, 0.25 mmol) and NaSH·xH<sub>2</sub>O (280 mg, 2.50 mmol) were dissolved in degassed H<sub>2</sub>O/D<sub>2</sub>O (98:2, 5 mL). The solution was adjusted to pH 9.0 with HCl/NaOH, and stirred at room temperature. NMR spectra were periodically acquired until the quantitative conversion of *N*-acetylglycine nitrile **Ac-Gly-CN** to *N*-acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>** was observed after 24 h.

<sup>1</sup>H NMR (600 MHz, H<sub>2</sub>O/D<sub>2</sub>O 98:2, pH 9.0, 2.0 – 4.5 ppm):

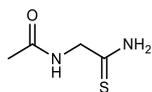
*N*-Acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>** (●): δ 4.36 (s, 2H, (C2)–H<sub>2</sub>), 2.23 (s, 3H, (CO)CH<sub>3</sub>).

*N*-Acetylglycine nitrile **Ac-Gly-CN** (■): δ 4.33 (s, 2H, (C2)–H<sub>2</sub>), 2.19 (s, 3H, (CO)CH<sub>3</sub>).



**Experimental figure 4:** <sup>1</sup>H NMR (600 MHz, H<sub>2</sub>O/D<sub>2</sub>O 98:2, pH 9.0, 2.0 – 4.5 ppm) spectra to show the reaction of *N*-acetylglycine nitrile (**Ac-Gly-CN**; 50 mM) with NaSH (500 mM) to yield *N*-acetylglycine thioamide (**Ac-Gly-SNH<sub>2</sub>**) at pH 9.0 and room temperature after 0.5 h, 12 h and 24 h.

*N*-Acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>**



*N*-Acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>** was prepared following *general procedure B* using *N*-acetylglycine nitrile **Gly-CN** (245 mg, 2.5 mmol). The reaction afforded *N*-acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>** as a white solid (329 mg, 2.49 mmol, > 99%).

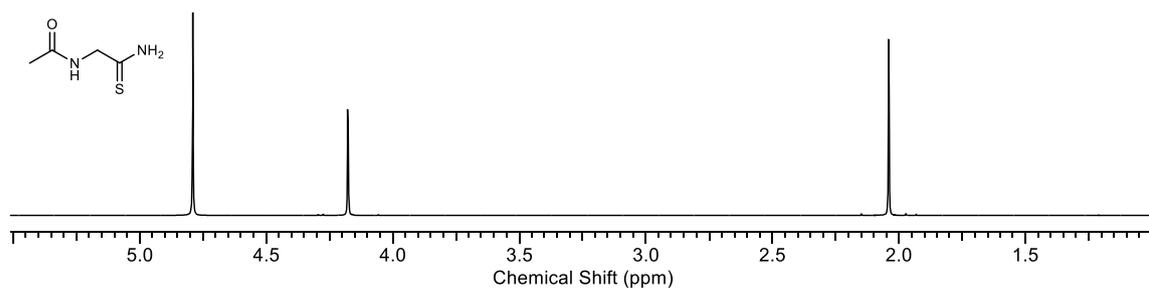
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.18 (s, 2H, (C2)-H<sub>2</sub>), 2.04 (s, 3H, (CO)CH<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 204.9 (C1), 175.6 (COCH<sub>3</sub>), 50.0 (C2), 22.5 (COCH<sub>3</sub>).

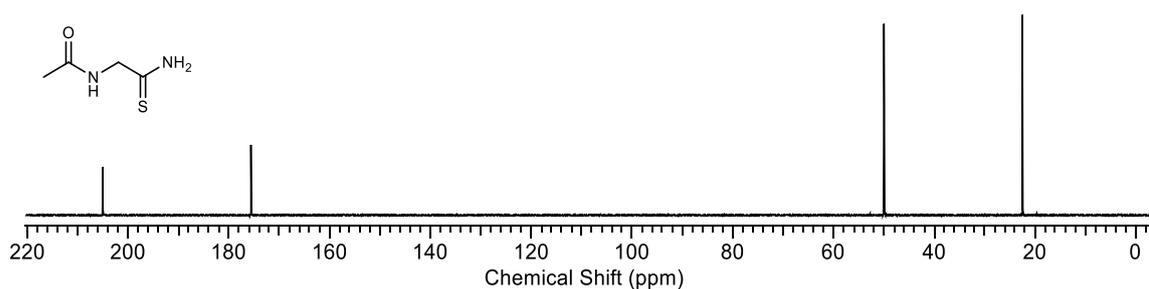
HRMS (ESI): [M + H]<sup>+</sup> calcd. for [C<sub>4</sub>H<sub>7</sub>N<sub>2</sub>OS-H]<sup>+</sup> 131.0279; observed: 131.0278.

IR (solid): 3341, 3287, 3077, 2931, 2143, 1640, 1522 cm<sup>-1</sup>.

M.p: 125.2 – 127.8 °C; lit. m.p. 124 – 126 °C.<sup>4</sup>



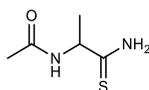
**Experimental figure 5:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 1.00 – 5.50 ppm) spectrum to show *N*-acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>**.



**Experimental figure 6:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show *N*-acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>**.

<sup>4</sup> G. Lowe, Y. Yuthavong, *Biochemical Journal*, 1971, 124 (1) 107-115;

*N*-Acetylalanine thioamide **Ac-Ala-SNH<sub>2</sub>**



*N*-Acetylalanine thioamide **Ac-Ala-SNH<sub>2</sub>** was prepared following *general procedure B* using *N*-acetylalanine nitrile **Ac-Ala-CN** (280 mg, 2.5 mmol). The reaction afforded *N*-acetylalanine thioamide **Ac-Ala-SNH<sub>2</sub>** as a white solid (362 mg, 2.48 mmol, > 99%).

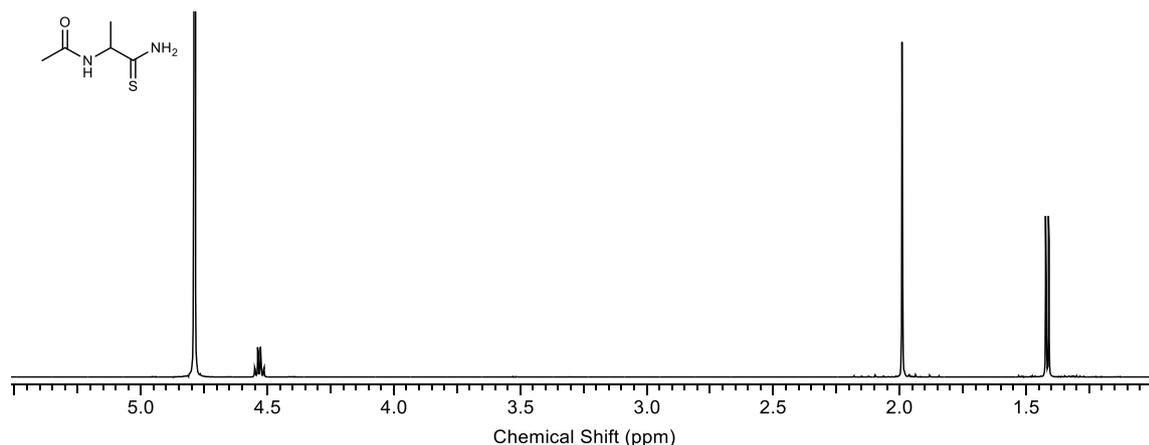
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.53 (q, *J* = 7.2 Hz, 1H, (C2)-H), 1.99 (s, 3H, (CO)CH<sub>3</sub>), 1.41 (d, *J* = 7.2 Hz, 3H, (C3)).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 210.7 (C1), 174.7 (CO), 56.5 (C2), 22.3 (CO)CH<sub>3</sub>, 20.3 (C3).

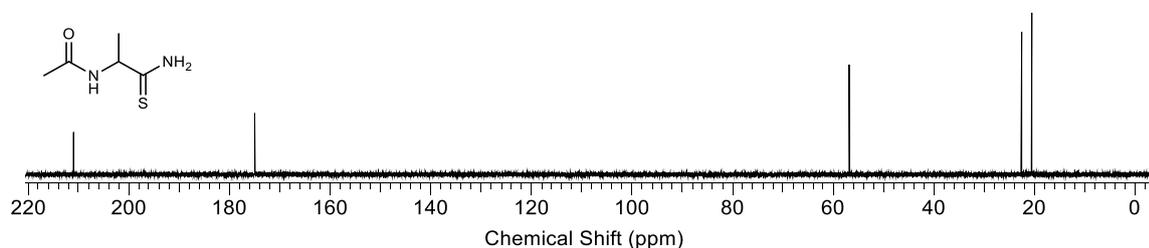
HRMS (ESI): [M + H]<sup>+</sup> calcd. for [C<sub>5</sub>H<sub>10</sub>N<sub>2</sub>OS+H]<sup>+</sup>: 147.0592; observed: 147.0594.

IR (solid): 3292, 3233, 3097, 1623, 1541, 1524 cm<sup>-1</sup>.

M.p. 152.7 –154.8 °C; lit. m.p. 152 – 154 °C.<sup>5</sup>



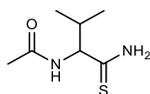
**Experimental figure 7:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 1.00 – 5.50 ppm) spectrum to show *N*-acetylalanine thioamide **Ac-Ala-SNH<sub>2</sub>**.



**Experimental figure 8:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show *N*-acetylalanine thioamide **Ac-Ala-SNH<sub>2</sub>**.

<sup>5</sup> T. P. Culbertson, J. M. Domagala, P. Peterson *et al.*, JOC, 1987 vol. 24, 6 p. 1509 – 1520.

*N*-Acetylvaline thioamide **Ac-Val-SNH<sub>2</sub>**



*N*-Acetylvaline thioamide **Ac-Val-SNH<sub>2</sub>** was prepared following *general procedure B* using *N*-acetylvaline nitrile **Ac-Val-CN** (351 mg, 2.5 mmol). The reaction afforded *N*-acetylvaline thioamide **Ac-Val-SNH<sub>2</sub>** as a white solid (419 mg, 2.41 mmol, > 96%).

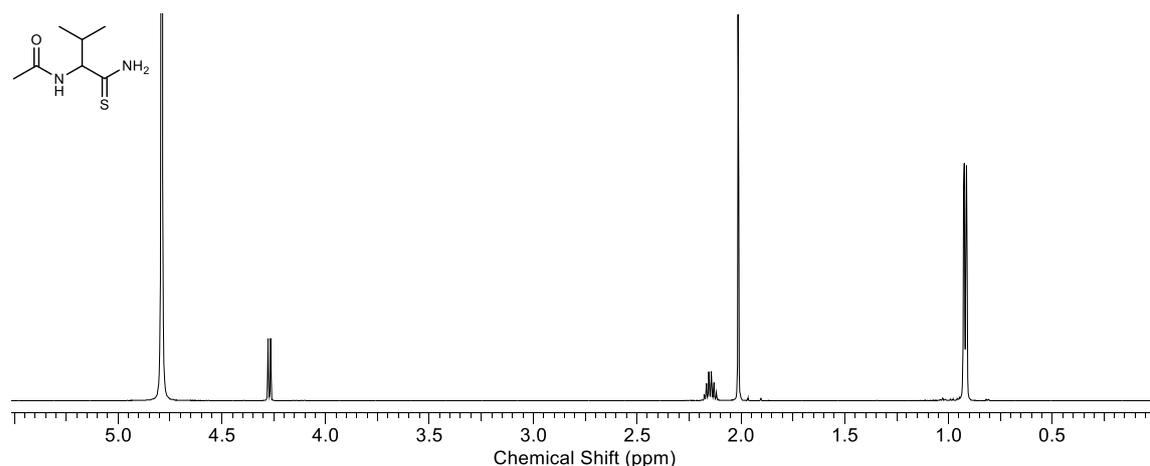
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.27 (d, *J* = 7.5 Hz, 1H, (C2)–H), 2.19 – 2.11 (m, 1H, (C3)–H), 2.01 (s, 3H, (CO)CH<sub>3</sub>), 0.92 (d, *J* = 6.8 Hz, 6H, (C4)–H<sub>3</sub> + (C4')–H<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 208.7 (C1), 175.0 (CO), 65.9 (C2), 32.3 (C3), 22.4 ((CO)CH<sub>3</sub>), 19.2 (C4) 17.7 (C4').

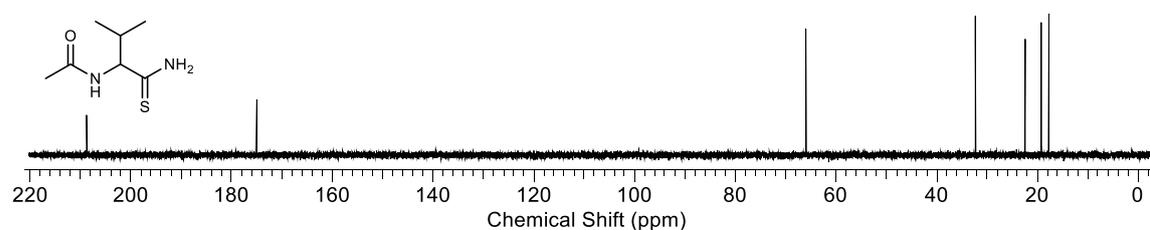
HRMS (ESI): [M+H]<sup>+</sup> calcd. for [C<sub>7</sub>H<sub>13</sub>N<sub>2</sub>OS+H]<sup>+</sup>: 174.0821; observed: 174.0822.

IR (solid): 3269, 3091, 2966, 2869, 1656, 1535 cm<sup>-1</sup>.

M.p. 161.9 – 162.3 °C.

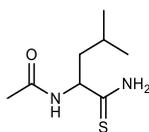


**Experimental figure 9:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 0.00 – 5.50 ppm) spectrum to show *N*-acetylvaline thioamide **Ac-Val-SNH<sub>2</sub>**.



**Experimental figure 10:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show *N*-acetylvaline thioamide **Ac-Val-SNH<sub>2</sub>**.

*N*-Acetylleucine thioamide **Ac-Leu-SNH<sub>2</sub>**



*N*-Acetylleucine thioamide **Ac-Leu-SNH<sub>2</sub>** was prepared following *general procedure B* using *N*-acetylleucine nitrile **Ac-Leu-CN** (386 mg, 2.5 mmol). The reaction afforded *N*-acetylleucine thioamide **Ac-Leu-SNH<sub>2</sub>** as a white solid (150 mg, 0.79 mmol, 32%).

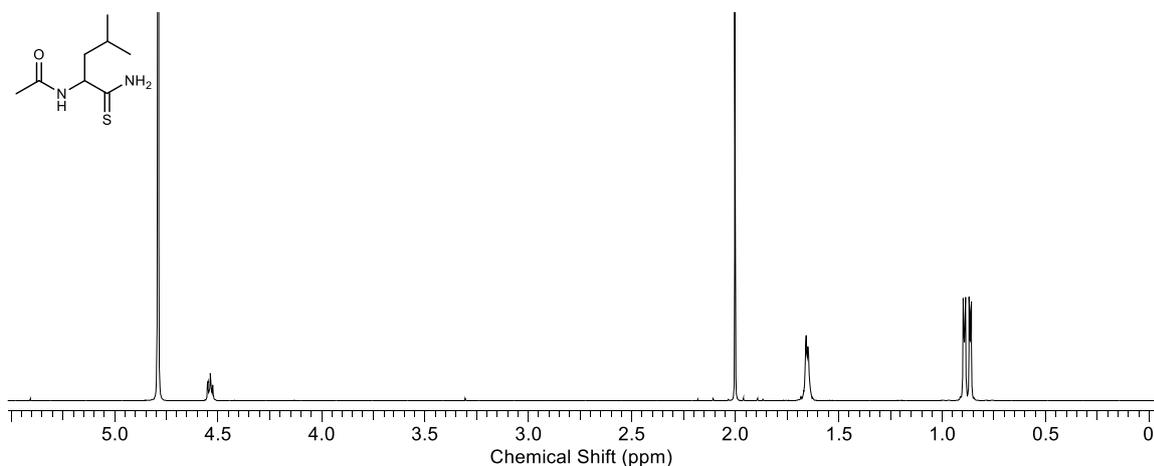
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.54 (app. t, *J* = 7.1 Hz, 1H, (C2)–H), 2.00 (s, 3H, (CO)CH<sub>3</sub>), 1.68 – 1.63 (m, 3H, (C3)–H<sub>2</sub> + (C4)–H), 0.92 (d, *J* = 6.1 Hz, 3H, (C5)–H<sub>3</sub>), 0.86 (d, *J* = 6.1 Hz, 3H, (C5')–H<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 210.4 (C1), 174.9 (CO), 59.1 (C2), 43.5 (C3), 25.2 (C4), 22.4 (C5') 22.4 (C5), 21.1 (COCH<sub>3</sub>).

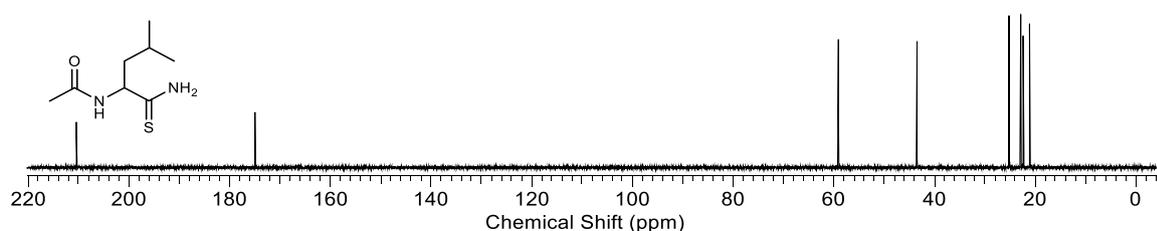
HRMS (EI): calcd. for [C<sub>8</sub>H<sub>16</sub>N<sub>2</sub>OS]<sup>+</sup>: 188.0978; observed: 188.0978.

IR (solid): 3255, 3086, 2956, 2930, 2915, 2869, 1652, 1532 cm<sup>-1</sup>.

M.p. 127.5 – 128.0 °C.

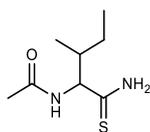


**Experimental figure 11:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 0.0 – 5.50 ppm) spectrum to show *N*-acetylleucine thioamide **Ac-Leu-SNH<sub>2</sub>**.



**Experimental figure 12:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show *N*-acetylleucine thioamide **Ac-Leu-SNH<sub>2</sub>**.

*N*-Acetylisoleucine thioamide **Ac-Ile-SNH<sub>2</sub>**



*N*-Acetylisoleucine thioamide **Ac-Ile-SNH<sub>2</sub>** was prepared following *general procedure B* using *N*-acetylisoleucine nitrile **Ac-Ile-CN** (386 mg, 2.5 mmol). The reaction afforded *N*-acetylisoleucine thioamide **Ac-Ile-SNH<sub>2</sub>** as a white solid (185 mg, 0.98 mmol, 32%).

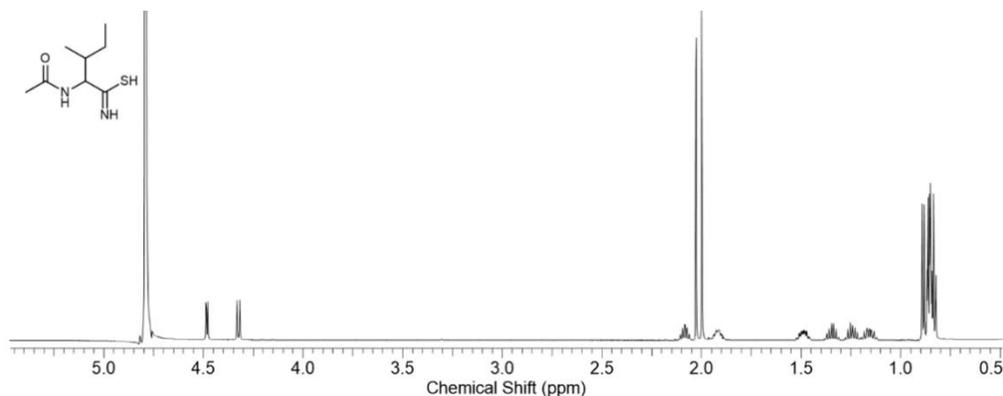
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 52:48 mixture of diastereoisomers a/b): diastereoisomer a: δ 4.32 (d, *J* = 8.0 Hz, 1H, (C2)-H), 2.00 (s, 3H, (CO)-CH<sub>3</sub>), 1.96 – 1.88 (m, 1H, (C3)-H), 1.53 – 1.44 (m, 1H, (C4)-H), 1.20 – 1.11 (m, 1H, (C4')-H), 0.91 – 0.81 (m, 6H, (C3)-H<sub>3</sub>, (C5)-H<sub>3</sub>); diastereoisomer b: δ 4.48 (d, *J* = 5.8 Hz, 1H, (C2)-H), 2.11 – 2.06 (m, 1H, (C3)-H), 2.03 (s, 3H, (CO)-CH<sub>3</sub>), 1.39 – 1.31 (m, 1H, (C4)-H), 1.28 – 1.21 (m, 1H, (C4')-H), 0.91 – 0.81 (m, 6H, (C3)-H<sub>3</sub>, (C5)-H<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 52:48 mixture of diastereoisomers): diastereoisomer a: δ 208.8 (C1), 174.9 (CO), 64.0 (C2), 38.3 (C3), 24.7 (C4), 22.4 ((CO)-CH<sub>3</sub>), 15.4 ((C3)-CH<sub>3</sub>), 10.7 (C5); diastereoisomer b: δ 208.9 (C1), 175.3 (CO), 64.7 (C2), 38.6 (C3), 26.4 (C4), 22.3 ((CO)-CH<sub>3</sub>), 13.9 ((C3)-CH<sub>3</sub>), 11.5 (C5).

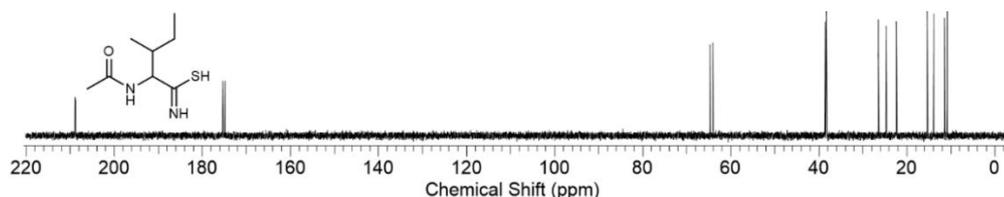
HRMS (ES<sup>+</sup>): calcd. for [C<sub>8</sub>H<sub>16</sub>N<sub>2</sub>OS]<sup>+</sup>: 188.0978; observed: 188.0978.

IR (solid): 3243, 3099, 2961, 2927, 2875, 1656, 1526, 1502 cm<sup>-1</sup>.

M.p. 152.0 – 151.9 °C.

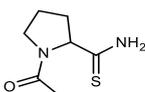


**Experimental figure 13:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 0.5 – 5.50 ppm) spectrum to show *N*-acetylisoleucine thioamide **Ac-Ile-SNH<sub>2</sub>**.



**Experimental figure 14:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show *N*-acetylisoleucine thioamide **Ac-Ile-SNH<sub>2</sub>**.

1-Acetylproline thioamide **Ac-Pro-SNH<sub>2</sub>**



1-Acetylproline thioamide **Ac-Pro-SNH<sub>2</sub>** was prepared following *general procedure B* using *N*-acetylproline nitrile **Ac-Pro-CN** (345 mg, 2.5 mmol). The reaction afforded 1-acetylproline thioamide **Ac-Pro-SNH<sub>2</sub>** as a white solid (334 mg, 1.93 mmol, 77%).

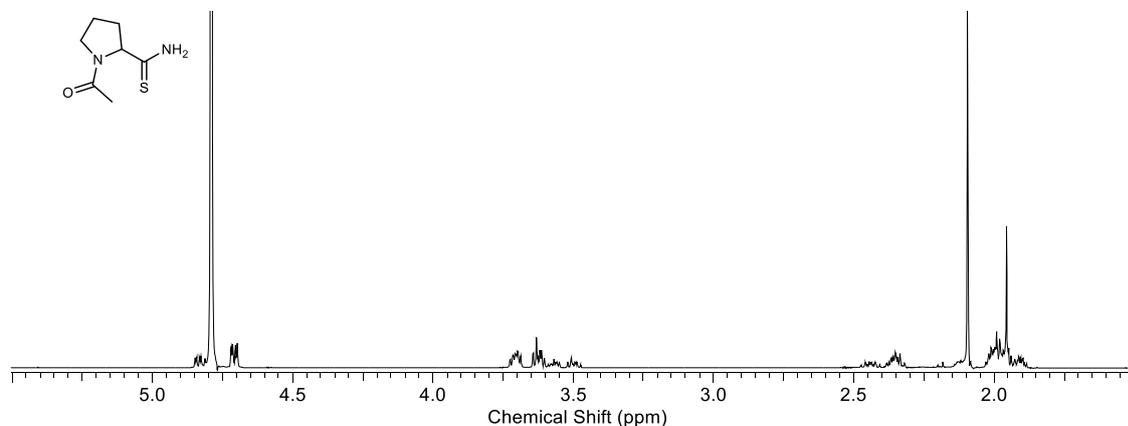
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 65:35 mixture of rotamers a/b): a: δ 4.71 (dd, *J* = 3.6, 9.0 Hz, 1H, (C2)-H), 3.73 – 3.68 (m, 1H, (C3)-H), 3.65 – 3.60 (m, 1H, (C3')-H), 2.39 – 2.31 (m, 1H, (C4)-H), 2.09 (s, 3H, (CO)CH<sub>3</sub>), 2.03 – 1.97 (m, 2H, (C4')-H, (C5)-H), 1.94 – 1.88 (m, 1H, (C5')-H); b: δ 4.84 (dd, *J* = 3.1, 9.0 Hz, 1H, (C2)-H), 3.59 – 3.54 (m, 1H, (C3)-H), 3.53 – 3.46 (m, 1H, (C3')-H), 2.48 – 2.40 (m, 1H, (C4)-H), 2.15 – 2.08 (m, 2H, (C4')-H, (C5)-H), 1.95 (s, 3H, (CO)CH<sub>3</sub>), 1.98 – 1.93 (m, 1H, (C5')-H).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, mixture of rotamers a/b): rotamer a: δ 209.1 (C1), 174.3 (CO), 67.5 (C2), 49.7 (C3), 33.5 (C5), 24.5 (C4), 22.2 (CO)CH<sub>3</sub>; rotamer b: δ 209.2 (C1), 174.7 (CO), 68.9 (C2), 48.0 (C3), 35.0 (C5), 22.9 (C4), 22.1 (CO)CH<sub>3</sub>.

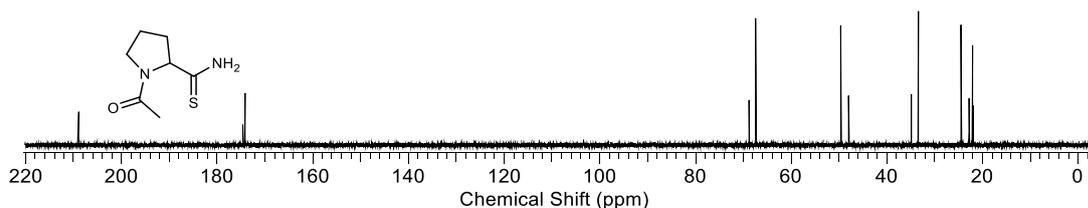
HRMS (EI): calcd. for [C<sub>7</sub>H<sub>12</sub>N<sub>2</sub>OS]<sup>+</sup>: 172.0665; observed: 172.0664.

IR (solid): 3381, 3297, 3079, 2974, 2951, 2870, 1645, 1619 cm<sup>-1</sup>.

M.p. 159.0 – 161.2 °C.

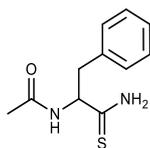


**Experimental figure 15:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 1.50 – 5.50 ppm) spectrum to show 1-acetylproline thioamide **Ac-Pro-SNH<sub>2</sub>**.



**Experimental figure 16:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show 1-acetylproline thioamide **Ac-Pro-SNH<sub>2</sub>**.

*N*-Acetylphenylalanine thioamide **Ac-Phe-SNH<sub>2</sub>**



*N*-Acetylphenylalanine thioamide **Ac-Phe-SNH<sub>2</sub>** was prepared following *general procedure B* using *N*-acetylphenylalanine nitrile **Ac-Phe-CN** (471 mg, 2.5 mmol). The reaction afforded *N*-acetylphenylalanine thioamide **Ac-Phe-SNH<sub>2</sub>** as a white solid (529 mg, 2.38 mmol, 95%).

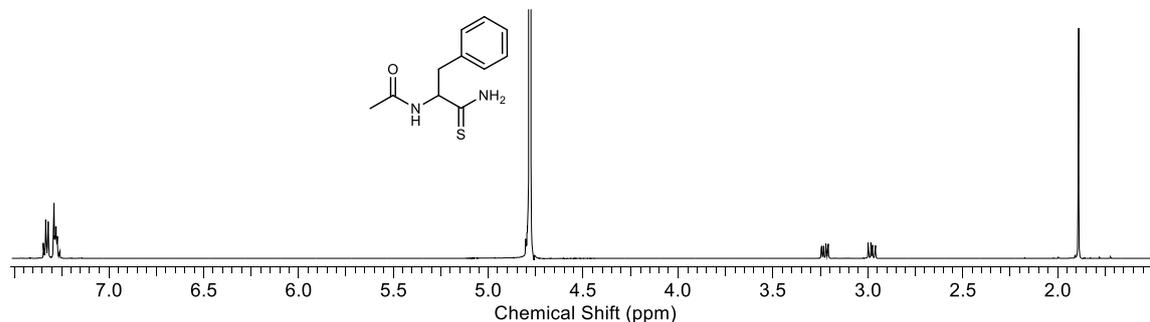
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O) δ 7.36 – 7.25 (m, 5H, Ar), 4.80 (obs., 1H, (C2)–H), 3.23 (ABX, *J* = 6.1, 13.9 Hz, 1H, (C3)–H), 2.99 (ABX, *J* = 9.0, 13.9 Hz, 1H, (C3')–H), 1.90 (s, (CO)CH<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 208.2 (C1), 174.6 (CO), 137.2 (Ar), 129.8 (Ar × 2), 129.3 (Ar × 2), 127.8 (Ar), 61.3 (C2), 40.5 (C3), 22.3 (CO)CH<sub>3</sub>).

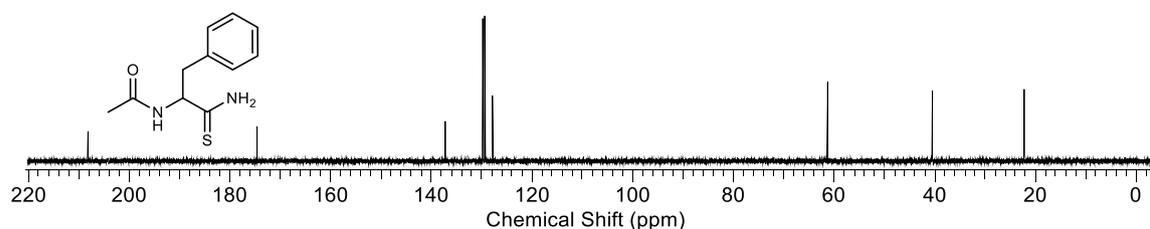
HRMS (EI): [M+H]<sup>+</sup>: calcd. for [C<sub>11</sub>H<sub>14</sub>N<sub>2</sub>OS]<sup>+</sup>: 222.0821; observed: 222.0822.

IR (solid): 3362, 3293, 3136, 3025, 2954, 1653, 1628, 1519, 1500 cm<sup>-1</sup>.

M.p.: 129.9 – 130.2 °C.

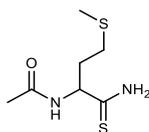


**Experimental figure 17:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 1.50 – 7.50 ppm) spectrum to show *N*-acetylphenylalanine thioamide **Ac-Phe-SNH<sub>2</sub>**.



**Experimental figure 18:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show *N*-acetylphenylalanine thioamide **Ac-Phe-SNH<sub>2</sub>**.

*N*-Acetylmethionine thioamide **Ac-Met-SNH<sub>2</sub>**



*N*-Acetylmethionine thioamide **Ac-Met-SNH<sub>2</sub>** was prepared following *general procedure B* using *N*-acetylmethionine nitrile **Ac-Met-CN** (431 mg, 2.5 mmol). The reaction afforded *N*-acetylmethionine thioamide **Ac-Met-SNH<sub>2</sub>** as a white solid (298 mg, 1.44 mmol, 58%).

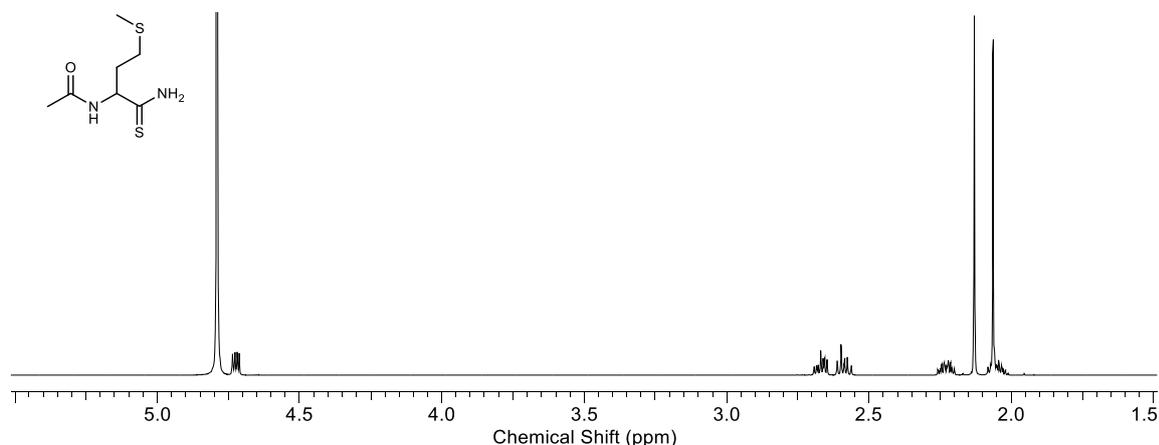
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.72 (dd, *J* = 4.8, 9.4 Hz, 1H, (C2)-H), 2.70 – 2.64 (m, 1H, (C3)-H), 2.62 – 2.55 (m, 1H, (C3')-H), 2.26 – 2.19 (m, 1H, (C4)-H), 2.13 (s, 3H, SCH<sub>3</sub>), 2.08 – 1.99 (obs. m, 1H, (C4')-H), 2.05 (s, 3H, (CO)CH<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 208.8 (C1), 175.0 (CO), 52.2 (C1), 33.7 (C2), 30.1 (C4), 22.4 (CO)CH<sub>3</sub>, 14.8 (SCH<sub>3</sub>).

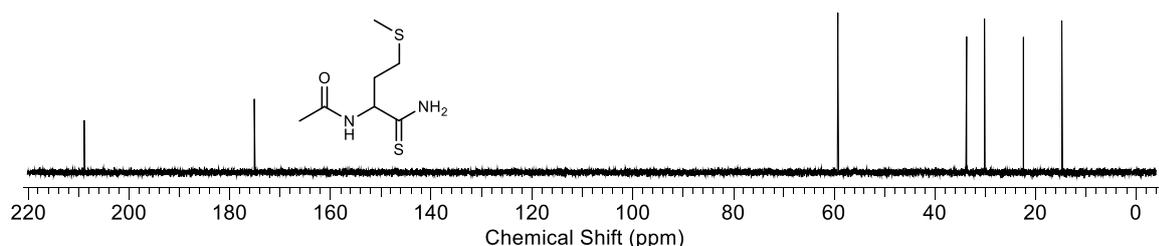
HRMS (EI): calcd. for [C<sub>7</sub>H<sub>14</sub>N<sub>2</sub>OS<sub>2</sub>]<sup>+</sup>: 206.0542; observed: 206.0543.

IR (solid): 3339, 3240, 3067, 2964, 2916, 1641, 1527 cm<sup>-1</sup>.

M.p.: 103.1 – 105.2 °C.



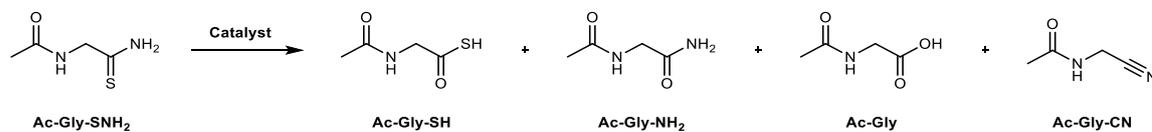
**Experimental figure 19:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 1.50 – 5.50 ppm) spectrum to show *N*-acetylmethionine thioamide **Ac-Met-SNH<sub>2</sub>**.



**Experimental figure 20:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show *N*-acetylmethionine thioamide **Ac-Met-SNH<sub>2</sub>**.

## 9. 2. 4. Prebiotic *N*-acetylaminoacyl thioacid synthesis

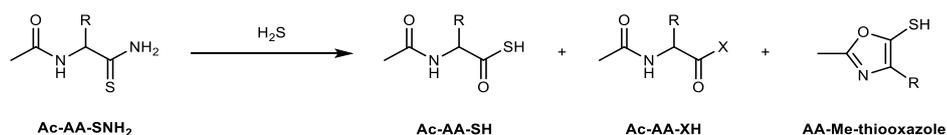
### Optimisation of the hydrolysis of *N*-acetyl glycine nitrile



*N*-Acetyl glycine thioamide **Ac-Gly-SNH<sub>2</sub>** (0.1 mmol) was dissolved in a 2% D<sub>2</sub>O:H<sub>2</sub>O solution (1 mL). The appropriate catalyst was added, and the pH was adjusted to 9.0 (final volume: 2 mL). The reaction mixture was incubated at 60 °C for 24 h. After that time, the reaction was analysed by <sup>1</sup>H NMR and comparison with a synthetic standard for *N*-acetyl glycine thioacid **Ac-Gly-SH**, *N*-acetyl glycineamide **Ac-Gly-NH<sub>2</sub>**, *N*-acetyl glycine **Ac-Gly** and *N*-acetyl glycine **Ac-Gly-CN**.

Experimental table 3: optimisation of the hydrolysis of <i>N</i> -acetyl glycine nitrile						
Entry	Additive	Conversion	Yields (%)			
			Ac-Gly-SH	Ac-Gly-NH <sub>2</sub>	Ac-Gly	Ac-Gly-CN
1	-	18%	15	2	1	-
2	Pi 0.5 M	60%	24	36	-	-
3	BBS 0.5 M	75%	49	26	-	-
4	MgCl <sub>2</sub> (1 eq.)	26%	23	3	-	-
5	CaCl <sub>2</sub> (1 eq.)	2%	-	-	-	-
6	AlCl <sub>3</sub> (1 eq.)	72%	55	17	-	-
7	FeCl <sub>2</sub> (1 eq.)	> 99%	33	77	-	-
8	FeCl <sub>3</sub> (1 eq.)	49%	16	33	-	-
9	NiCl <sub>2</sub> (1 eq.)	65%	54	8	3	-
10	CuCl (1 eq.)	> 99%	-	2	-	98
11	CuCl <sub>2</sub> (1 eq.)	> 99%	-	3	-	97
12	ZnCl <sub>2</sub> (1 eq.)	> 99%	-	15	4	81
13	MnCl <sub>2</sub> (1 eq.)	> 99%	-	18	2	80
14	K <sub>3</sub> [Fe(CN) <sub>6</sub> ] (1 eq.)	-	-	6	-	94
15	K <sub>4</sub> [Fe(CN) <sub>6</sub> ] (1 eq.)	25	11	10	4	-
16	Mercaptoethanol (1 eq.)	> 99	40	23	37	-
17	Na <sub>2</sub> S · 9 H <sub>2</sub> O (10 eq.)	58	33	22	3	-
18	NaSH · (H <sub>2</sub> O) <sub>n</sub> (10 eq.)	> 99	83	12	5	-

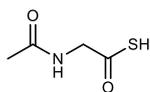
### General procedure C: Hydrolysis of $\alpha$ -amidothioamides



$\alpha$ -Amidothioamide **Ac-AA-SNH<sub>2</sub>** (0.05 mmol), NaSH·xH<sub>2</sub>O (56 mg, 0.50 mmol) and methylsulfonylmethane (0.05 mmol) were dissolved in degassed H<sub>2</sub>O/D<sub>2</sub>O (98:2, 1 mL), and the solution was adjusted to pH 9.0 with NaOH/HCl. The solution was incubated at 60 °C, whilst maintaining the solution at pH 9.0 with NaOH/HCl, and NMR spectra were periodically acquired until complete conversion of  $\alpha$ -amidothioamide **Ac-AA-SNH<sub>2</sub>** was observed. The presence of  $\alpha$ -amidothioacid **Ac-AA-SH** and *N*-acetylamino acid **Ac-AA-OH** were confirmed by <sup>1</sup>H – <sup>13</sup>C HMBC NMR analysis, spiking or comparison of NMR data with pure synthetic standards. The reaction mixture was quantified using methylsulfonylmethane (MSM) an internal standard. A second by-product of  $\alpha$ -amidothioamide **Ac-AA-SNH<sub>2</sub>** hydrolysis was also observed in varying quantities, depending on the  $\alpha$ -amidothioamide side chain residue. This was tentatively assigned as the aromatic species **AA-Me-thiooxazole** based on NMR spectral data.

Entry	Residue	Time (h)	Ac-AA-SH (%)	Ac-AA-OH (%)	AA-Me-thiooxazole (%)
1	Gly	24	81	19	0
2	Ala	72	85	8	6
3	Arg	36	51	18	24
4	Leu	72	77	1	21
5	Met	72	70	6	22
6	Phe	48	84	0	10
7	Pro	168	72	26	0
8	Ser	24	61	16	0
9	Val	72	8	9	82

*N*-Acetylglycine thioacid **Ac-Gly-SH**

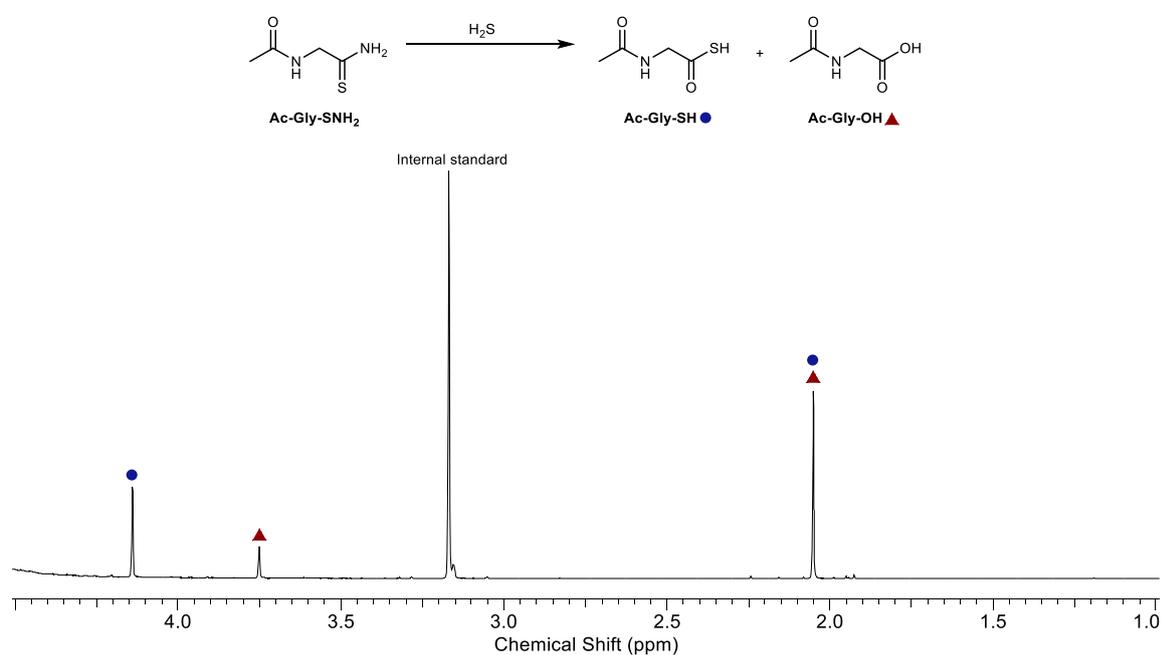


*N*-Acetylglycine thioacid **Ac-Gly-SH** (81%, 24 h) was prepared following *general procedure C* using *N*-acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>** (6.61 mg, 0.05 mmol).

<sup>1</sup>H NMR (600 MHz, H<sub>2</sub>O/D<sub>2</sub>O 98:2):

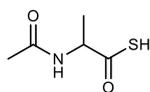
*N*-acetylglycine thioacid **Ac-Gly-SH** (●): δ 4.14 (s, 2H, (C2)–CH<sub>2</sub>), 2.05 (s, 3H, (CO)CH<sub>3</sub>).

*N*-Acetylglycine **Ac-Gly-OH** (▲): δ 3.75 (s, 2H, (C2)–CH<sub>2</sub>), 2.05 (s, 3H, (CO)CH<sub>3</sub>).



**Experimental figure 21:** <sup>1</sup>H NMR (600 MHz, H<sub>2</sub>O/D<sub>2</sub>O 98:2, 1.0 – 4.5 ppm) spectrum for the reaction of *N*-acetylglycine thioamide (**Ac-Gly-SNH<sub>2</sub>**; 50 mM) with H<sub>2</sub>S (500 mM) at 60 °C and pH 9.0 after 24 h.

*N*-Acetylanine thioacid **Ac-Ala-SH**



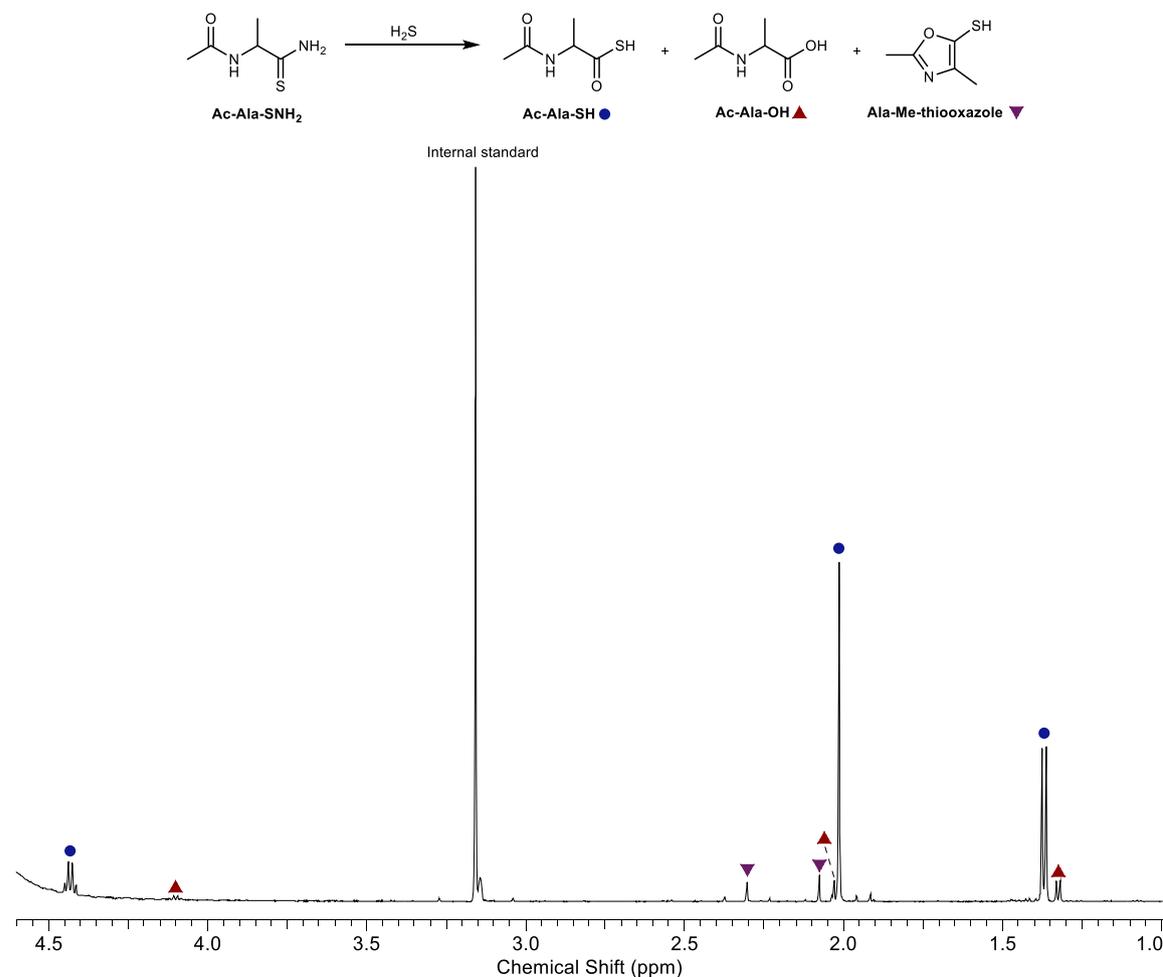
*N*-Acetylanine thioacid **Ac-Ala-SH** (85%, 72 h) was prepared following *general procedure C* using *N*-acetylanine thioamide **Ac-Ala-SNH<sub>2</sub>** (7.3 mg, 0.05 mmol).

<sup>1</sup>H NMR (600 MHz, H<sub>2</sub>O/D<sub>2</sub>O 98:2):

*N*-Acetylanine thioacid **Ac-Ala-SH** (●): δ 4.43 (q, *J* = 7.2 Hz, 1H, (C2)-H), 2.02 (s, 3H, (CO)CH<sub>3</sub>), 1.37 (d, *J* = 7.2 Hz, 3H, (C3)-H<sub>3</sub>).

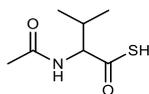
*N*-Acetylglycine **Ac-Gly-OH** (▲): δ 4.11 (q, *J* = 7.3 Hz, 1H, (C2)-H), 2.08 (s, 3H, (CO)CH<sub>3</sub>), 1.32 (d, *J* = 7.3 Hz, 3H, (C3)-H<sub>3</sub>).

2,4-Dimethyloxazole-5-thiol **Ala-Me-thiooxazole** (▼): δ 2.31 (s, 3H, (CO)-CH<sub>3</sub>), 2.07 (s, (C3)-H<sub>3</sub>).



**Experimental figure 22:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O/H<sub>2</sub>O 98:2, 1.0 – 4.6 ppm) spectrum for the reaction of *N*-acetylanine thioamide (**Ac-Ala-SNH<sub>2</sub>**; 50 mM) with H<sub>2</sub>S (500 mM) at 60 °C and pH 9.0 after 72 h.

*N*-Acetylvaline thioacid **Ac-Val-SH**



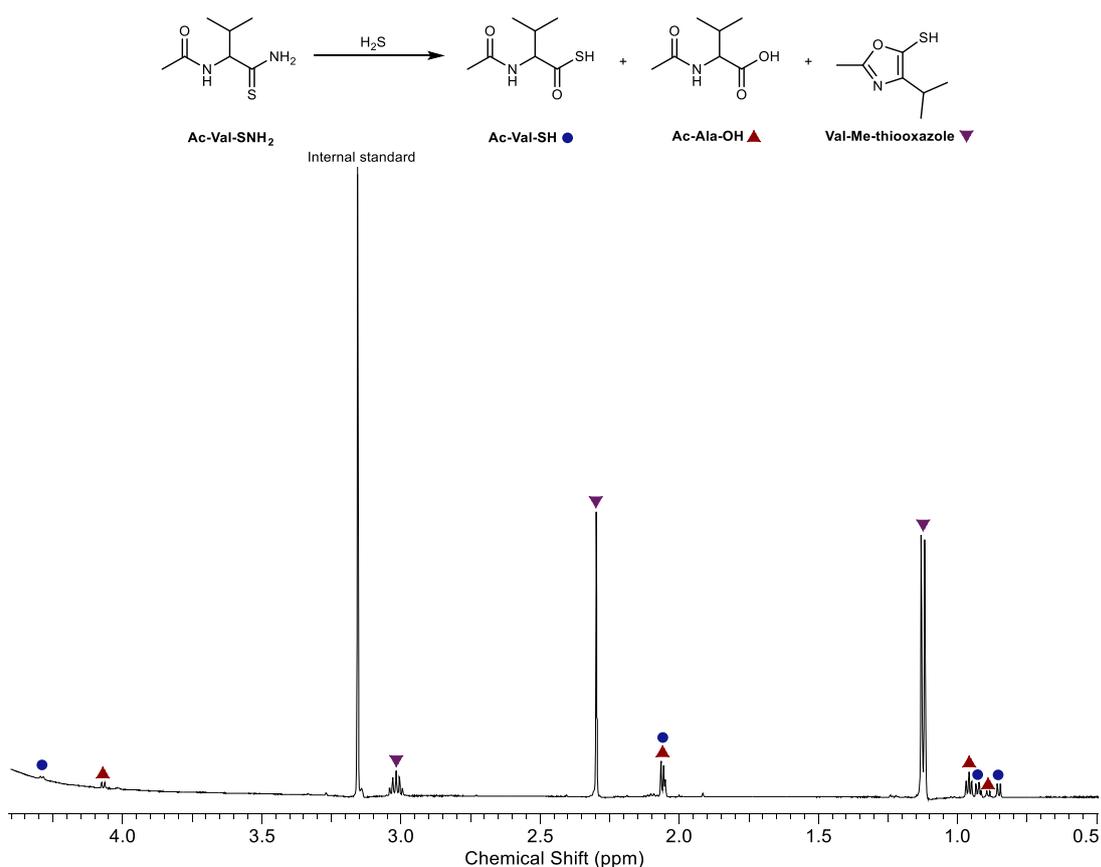
*N*-Acetylvaline thioacid **Ac-Val-SH** (8%, 72 h) was prepared following *general procedure C* using *N*-acetylvaline thioamide **Ac-Val-SNH<sub>2</sub>** (8.6 mg, 0.05 mmol).

<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O/H<sub>2</sub>O 98:2, partial assignments):

*N*-Acetylvaline thioacid **Ac-Val-SH** (●): δ 4.29 (d, *J* = 5.9 Hz, 1H, (C2)–H), 2.07 (s, 3H, (CO)CH<sub>3</sub>), 0.93 (d, *J* = 6.9 Hz, 3H, (C4)–H<sub>3</sub>), 0.85 (d, *J* = 6.9 Hz, 3H, (C4')–H<sub>3</sub>).

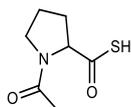
*N*-Acetylvaline **Ac-Val-OH** (▲): δ 4.06 (d, *J* = 7.3 Hz, 1H, (C2)–H), 2.06 (s, 3H, (CO)CH<sub>3</sub>), 0.96 (d, *J* = 6.9 Hz, 3H, (C4)–H<sub>3</sub>), 0.89 (d, *J* = 6.9 Hz, 3H, (C4')–H<sub>3</sub>).

4-Isopropyl-2-methyloxazole-5-thiol **Val-Me-thiooxazole** (▼): δ 3.06 – 2.98 (m, 1H, (C3)–H), 2.30 (s, 3H, (CO)CH<sub>3</sub>), 1.13 (d, *J* = 6.7 Hz, 6H, (C4)–H<sub>3</sub>, (C4')–H<sub>3</sub>).



**Experimental figure 23:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O/H<sub>2</sub>O 98:2, 1.00 – 4.60 ppm) spectrum for the reaction of *N*-acetylvaline thioamide (**Ac-Val-SNH<sub>2</sub>**; 50 mM) with H<sub>2</sub>S (500 mM) at 60 °C and pH 9.0 after 72 h.

*N*-Acetylproline thioacid **Ac-Pro-SH**

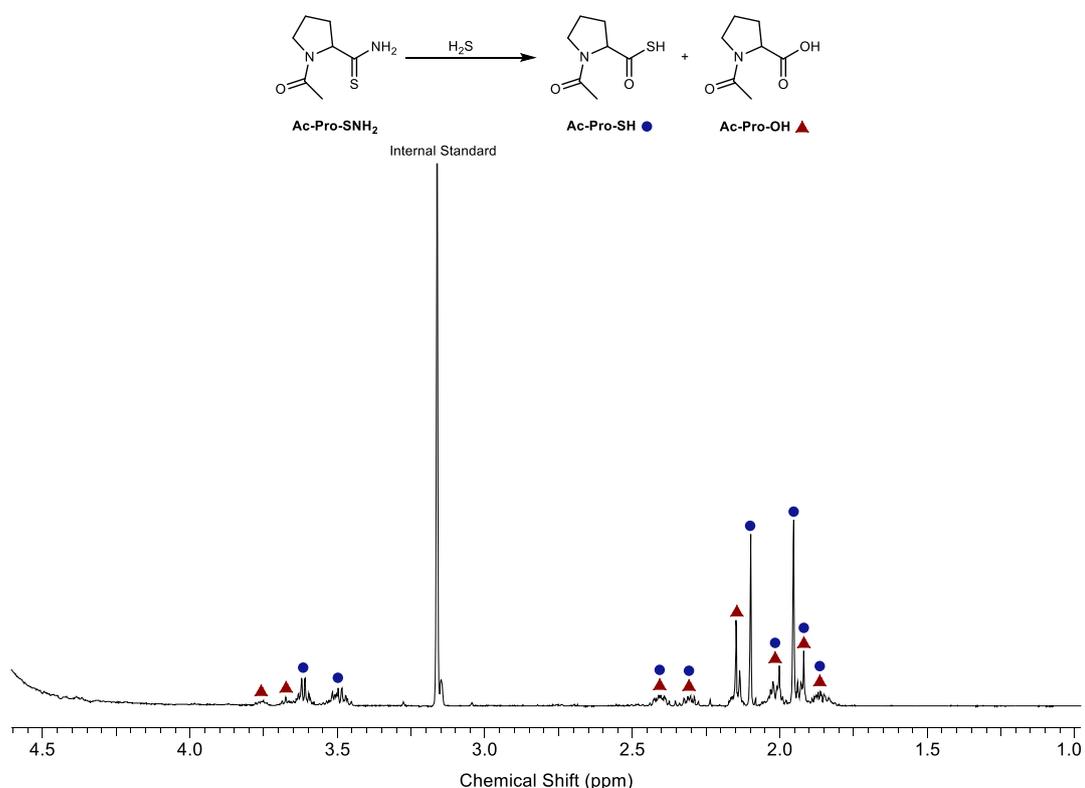


*N*-Acetylproline thioacid **Ac-Pro-SH** (72%, 168 h) was prepared following *general procedure C* using *N*-acetylproline thioamide **Ac-Pro-SNH<sub>2</sub>** (8.6 mg, 0.05 mmol).

<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O/H<sub>2</sub>O 98:2):

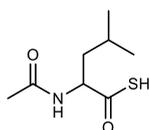
*N*-Acetylproline thioacid **Ac-Pro-SH** (●, 56:46 mixture of rotamers a/b, partial assignment):  
rotamer a: δ 3.66-3.57 (m, 1H, (C5)-H<sub>2</sub>), 2.43-2.35 (m, 1H, (C3)-H), 2.34-2.26 (m, 1H, (C3')-H), 1.96 (s, 3H, (CO)CH<sub>3</sub>), 1.89-1.83 (m, 2H, (C4)-H<sub>2</sub>); rotamer b: δ 3.55-3.43 (m, 1H, (C5)-H<sub>2</sub>), 2.33-2.27 (m, 1H, (C3)-H), 2.07 (s, 3H, (CO)CH<sub>3</sub>), 2.07-1.98 (m, 1H, (C3')-H), 1.93-1.84 (m, 2H, (C4)-H<sub>2</sub>).

*N*-Acetylproline **Ac-Pro-OH** (▲, 55:45 mixture of rotamers a/b, partial assignment) rotamer a: δ 3.78-3.74 (m, 2H, (C5)-H<sub>2</sub>), 2.34-2.23 (m, 2H, (C3)-H<sub>2</sub>), 2.15 (s, 3H, (CO)CH<sub>3</sub>), 2.05-1.98 (m, 2H, (C4)-H<sub>2</sub>); rotamer b: δ 3.70-3.65 (m, 2H, (C5)-H<sub>2</sub>), 2.44-2.37 (m, 2H, (C3)-H<sub>2</sub>), 1.93 (s, 3H, (CO)CH<sub>3</sub>), 1.90-1.83 (m, 2H, (C4)-H<sub>2</sub>).



**Experimental figure 24:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 0.50 – 4.60 ppm) spectrum for the reaction of *N*-acetylproline thioamide (**Ac-Pro-SNH<sub>2</sub>**; 50 mM) with  $\text{H}_2\text{S}$  (500 mM) at 60 °C and pH 9.0 after 168 h.

*N*-Acetylleucine thioacid **Ac-Leu-SH**



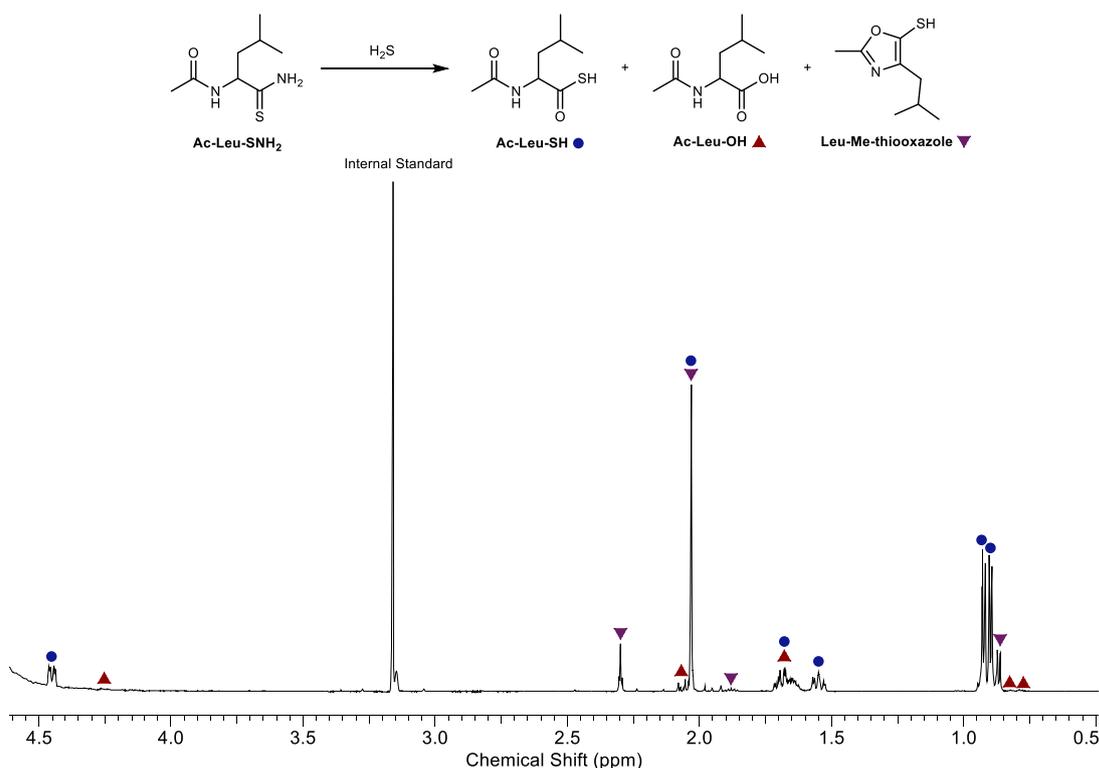
*N*-Acetylleucine thioacid **Ac-Leu-SH** (77%, 72 h) was prepared following *general procedure C* using *N*-acetylleucine thioamide **Ac-Leu-SNH<sub>2</sub>** (9.4 mg, 0.05 mmol).

$^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2):

*N*-Acetylleucine thioacid **Ac-Leu-SH** (•):  $\delta$  4.45 (dd,  $J = 3.6, 10.7$  Hz, 1H, (C2)–H), 2.03 (s, 3H, (CO)CH<sub>3</sub>), 1.72-1.61 (m, 2H, (C3)–H<sub>2</sub>), 1.58-1.52 (m, 1H, (C4)–H), 0.92 (d,  $J = 6.4$  Hz, 3H, (C5)–H<sub>3</sub>), 0.90 (d,  $J = 6.4$  Hz, 3H, (C5')–H<sub>3</sub>).

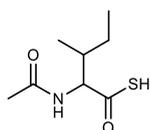
*N*-Acetylleucine **Ac-Leu-OH** (▲, partial assignment):  $\delta$  4.28-4.23 (obs. m, 1H, (C2)–H), 2.06 (s, 3H, (CO)CH<sub>3</sub>), 1.72-1.61 (obs. m, (C3)–H<sub>2</sub>) 1.32 (d,  $J = 7.3$  Hz, 3H).

4-Isobutyl-2-methyloxazole-5-thiol **Leu-Me-thiooxazole** (▼):  $\delta$  2.30 (t,  $J = 3.5$  Hz, 2H, (C3)–H<sub>2</sub>), 2.03 (s, 3H, (CO)–CH<sub>3</sub>), 1.91 – 1.84 (m, 1H, (C4)–H), 0.87 (d,  $J = 6.6$  Hz, 6H, (C5)–H<sub>3</sub>, (C5')–H<sub>3</sub>).



**Experimental figure 25:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 0.50 – 4.60 ppm) spectrum for the reaction of *N*-acetylleucine thioamide (**Ac-Leu-SNH<sub>2</sub>**; 50 mM) with  $\text{H}_2\text{S}$  (500 mM) at 60 °C and pH 9.0 after 72 h.

*N*-Acetylisoleucine thioacid **Ac-Ile-SH**

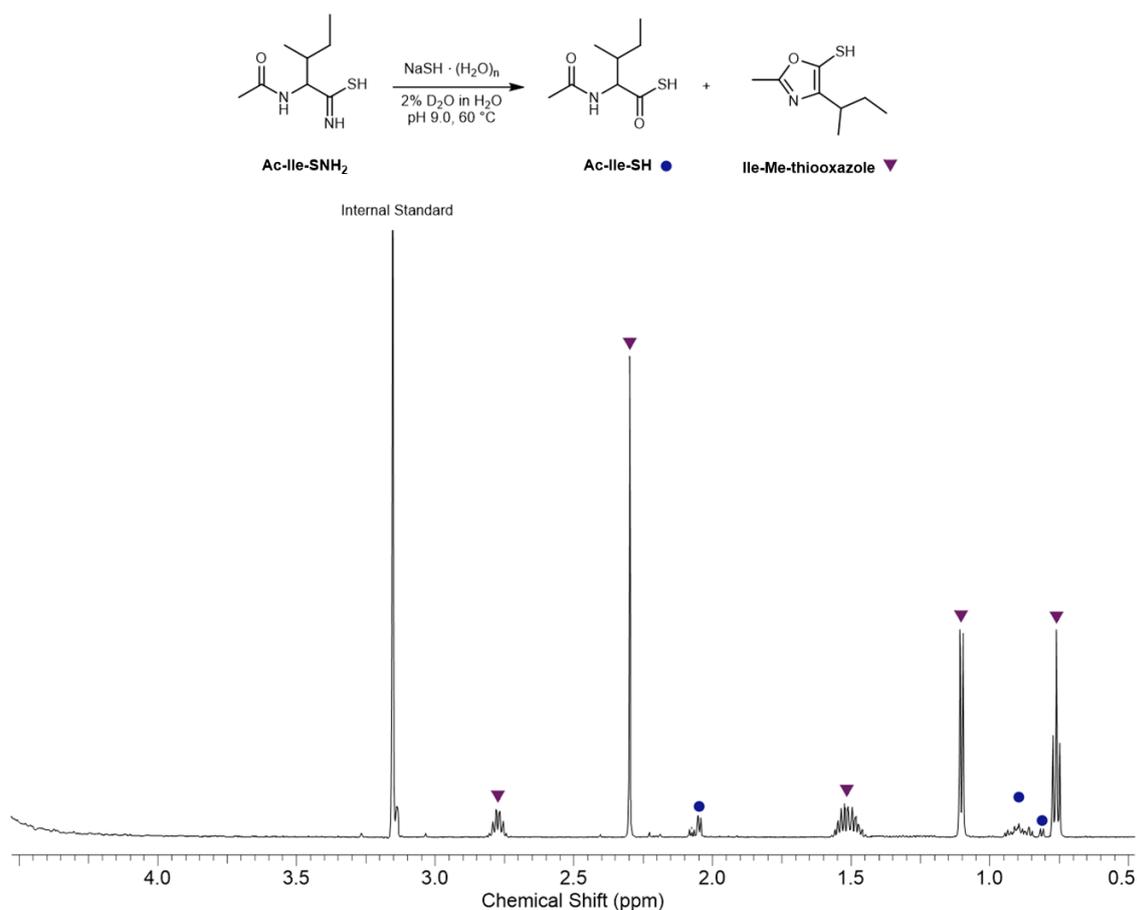


*N*-Acetylisoleucine thioacid **Ac-Ile-SH** (4%, 72 h) was prepared following *general procedure C* using *N*-acetylisoleucine thioamide **Ac-Ile-SNH<sub>2</sub>** (9.4 mg, 0.05 mmol).

$^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2):

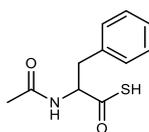
*N*-Acetylisoleucine thioacid **Ac-Ile-SH** (•):  $\delta$  2.05 (s, 3H, (CO)-CH<sub>3</sub>), 0.90 (m, 3H, (C5)-H<sub>3</sub>), 0.81 (d,  $J$  = 7.4 Hz, 3H, (C3)-CH<sub>3</sub>);

4-(*sec*-butyl)-2-methyloxazole-5-thiol **Ile-Me-thiooxazole** (▼):  $\delta$  2.80 – 2.72 (m, 1H, (C3)-H), 2.30 (s, 3H, (CO)-CH<sub>3</sub>), 1.57 – 1.45 (m, 2H, (C4)-H<sub>2</sub>), 1.10 (d,  $J$  = 7.0 Hz, 3H, (C3)-CH<sub>3</sub>), 0.76 (t,  $J$  = 7.4 Hz, 3H, (C5)-H<sub>3</sub>).



**Experimental figure 26:** <sup>1</sup>H NMR (600 MHz, H<sub>2</sub>O/D<sub>2</sub>O 98:2, 0.50 – 4.50 ppm) spectrum for the reaction of *N*-acetylisoleucine thioamide (**Ac-Ile-SNH<sub>2</sub>**; 50 mM) with H<sub>2</sub>S (500 mM) at 60 °C and pH 9.0 after 72 h.

#### *N*-Acetylphenylalanine thioacid **Ac-Phe-SH**

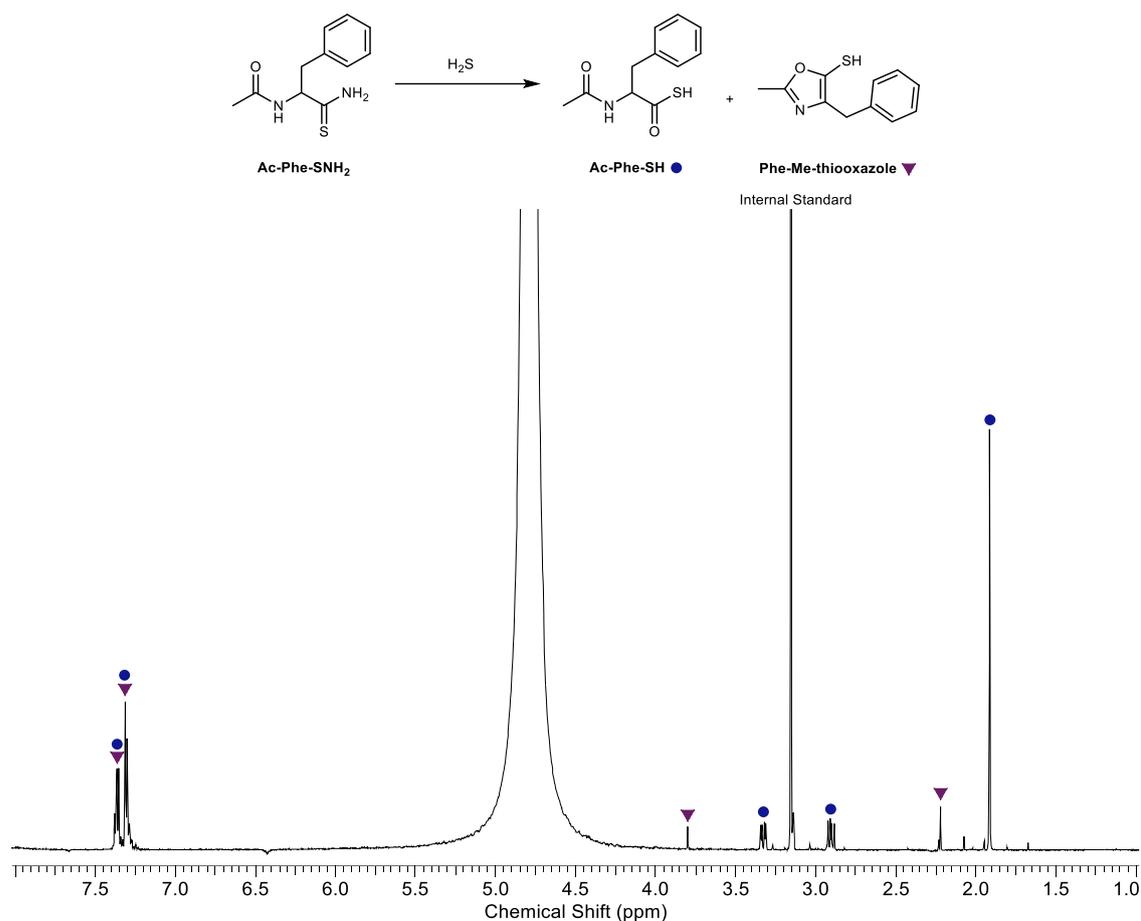


*N*-Acetylphenylalanine thioacid **Ac-Phe-SH** (84%, 48 h) was prepared following *general procedure C* using *N*-acetylphenylalanine thioamide **Ac-Phe-SNH<sub>2</sub>** (9.4 mg, 0.05 mmol).

<sup>1</sup>H NMR (600 MHz, H<sub>2</sub>O/D<sub>2</sub>O 98:2):

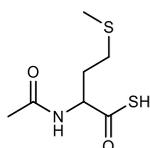
*N*-Acetylphenylalanine thioacid **Ac-Phe-SH** (●, partial assignment): δ 7.40 – 7.22 (m, 5H, Ar), 3.33 (ABX, *J* = 4.8, 14.5 Hz, 1H, (C3)–H), 2.90 (ABX, *J* = 9.6, 14.5 Hz, 1H, (C3')–H), 1.87 (s, 3H, (CO)CH<sub>3</sub>).

*4*-Benzyl-2-methyloxazole-5-thiol **Phe-Me-thiooxazole** (▼): δ 7.40 – 7.22 (m, 5H, Ar), 3.80 (s, 2H, (C3)–H<sub>2</sub>), 2.22 (s, 3H, (CO)-CH<sub>3</sub>).



**Experimental figure 27:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 0.50 – 4.60 ppm) spectrum for the reaction of *N*-acetylphenylalanine thioamide (**Ac-Phe-SNH<sub>2</sub>**; 50 mM) with  $\text{H}_2\text{S}$  (500 mM) at 60 °C and pH 9.0 after 48 h.

*N*-Acetylmethionine thioacid **Ac-Met-SH**



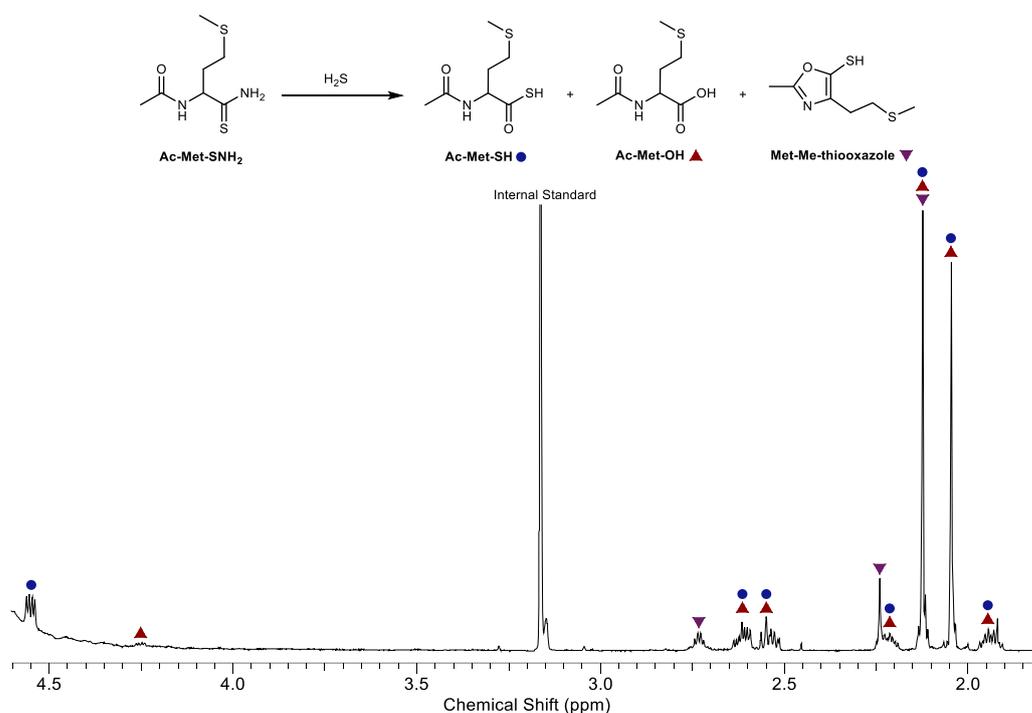
*N*-Acetylmethionine thioacid **Ac-Met-SH** (70%, 72 h) was prepared following *general procedure C* using *N*-acetylmethionine thioamide **Ac-Met-SNH<sub>2</sub>** (10.6 mg, 0.05 mmol).

$^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2):

*N*-acetylmethionine thioacid **Ac-Met-SH** (•):  $\delta$  4.55 (dd,  $J = 4.1, 9.4$  Hz, 1H, (C2)–H), 2.64 – 2.58 (m, 1H, (C3)–H), 2.57 – 2.51 (m, 1H, (C3')–H), 2.25 – 2.18 (m, 1H, (C4)–H), 2.12 (s, 3H,  $\text{SCH}_3$ ), 2.05 (s, 3H,  $(\text{CO})\text{CH}_3$ ), 1.97 – 1.80 (m, 1H, (C4')–H).

*N*-Acetylmethionine **Ac-Met-OH** (▲):  $\delta$  4.24 (dd,  $J = 4.1, 9.4$  Hz, 1H, (C2)-H), 2.64 – 2.58 (m, 1H, (C3)-H), 2.57 – 2.51 (m, 1H, (C3')-H), 2.25 – 2.18 (m, 1H, (C4)-H), 2.12 (s, 3H, SCH<sub>3</sub>), 2.05 (s, 3H, (CO)CH<sub>3</sub>), 1.97 – 1.80 (m, 1H, (C4')-H).

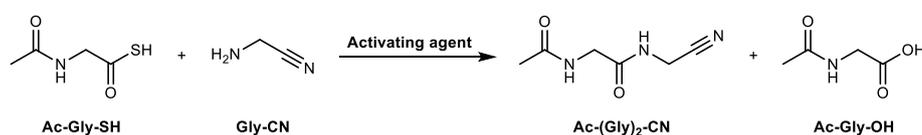
2-Methyl-4-(2-(methylthio)ethyl)oxazole-5-thiol **Met-Me-thiooxazole** (▼):  $\delta$  2.76 – 2.69 (m, 4H, (C3)-H<sub>2</sub>, (C4)-H<sub>2</sub>), 2.24 (s, 3H, (CO)-CH<sub>3</sub>), 2.12 (s, 3H, SCH<sub>3</sub>).



**Experimental figure 28:** <sup>1</sup>H NMR (600 MHz, H<sub>2</sub>O/D<sub>2</sub>O 98:2, 1.80 – 4.60 ppm) spectrum for the reaction of *N*-acetylmethionine thioamide (**Ac-Met-SNH<sub>2</sub>**; 50 mM) with H<sub>2</sub>S (500 mM) at 60 °C and pH 9.0 after 72 h.

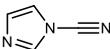
### 9. 2. 5. Ligation of *N*-acetyl thioacids

#### *Prebiotic ligation of N-acetyl glycine thioacid and glycine nitrile*

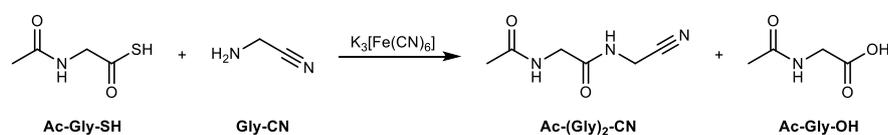


*N*-Acetylglycine thioacid **Ac-Gly-SH** (50 mM, 0.10 mmol), glycine nitrile **Gly-CN** (100 mM, 0.20 mmol) and methylsulfonylmethane (50 mM, 0.10 mmol) were dissolved in degassed H<sub>2</sub>O/D<sub>2</sub>O (98:2, 1 mL), and the solution adjusted to pH 9.0 with NaOH/HCl. The total volume adjusted to 2 mL with H<sub>2</sub>O/D<sub>2</sub>O (98:2), and the specified activating agent (150 mM, 0.30 mmol) was added. The solution was stirred at room temperature for 20 min. The resulting suspension was adjusted to pH

9.0 with NaOH/HCl, centrifuged, and the supernatant was analysed by 1D and 2D NMR spectroscopy in H<sub>2</sub>O/D<sub>2</sub>O (98:2). The presence of **Ac-(Gly)<sub>2</sub>-CN** and **Ac-Gly-OH** was confirmed by <sup>1</sup>H-<sup>13</sup>C HMBC NMR analysis and signal amplification or comparison with pure synthetic standards. The reaction mixture was quantified using methylsulfonylmethane as an internal standard.

Experimental table 5: Prebiotic ligation of <i>N</i> -acetyl glycine thioacid and glycine nitrile			
Entry	Activating agent	pH	Yield Ac-Gly-Gly-CN (%)
1		5.0	85
2		7.0	74
3		9.0	57
4		5.0	95
5		7.0	70
6		9.0	61
7	CuCl <sub>2</sub>	5.0	95
8		7.0	94
9		9.0	86
10	K <sub>3</sub> [Fe(CN) <sub>6</sub> ]	5.0	91
11		7.0	97
12		9.0	99

#### Optimisation of ferricyanide-mediated ligation of *N*-acetylaminoacyl thioacids

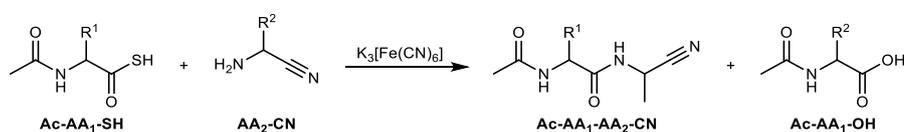


*N*-Acetylglycine thioacid **Ac-Gly-SH** (50 mM, 0.10 mmol), glycine nitrile **Gly-CN** (55/100/150 mM, 0.11/0.20/0.30 mmol) and methylsulfonylmethane (50 mM, 0.10 mmol) were dissolved in H<sub>2</sub>O/D<sub>2</sub>O (98:2, 1 mL). The solution was adjusted to the specified pH value with HCl/NaOH and the total volume was adjusted to 2 mL with H<sub>2</sub>O/D<sub>2</sub>O (98:2). Potassium hexacyanoferrate(III) (150 mM, 98.8 mg, 0.30 mmol) was then added and the resultant solution stirred at room temperature for 20 min. The resulting suspension was then adjusted to pH 9.0, centrifuged and the supernatant was analysed by 1D and 2D NMR spectroscopy in H<sub>2</sub>O/D<sub>2</sub>O (98:2). The presence of **Ac-Gly-Gly-CN**

was confirmed by  $^1\text{H}$ - $^{13}\text{C}$  HMBC NMR analysis and spiking or comparison with pure synthetic standards. The reaction mixture was quantified using methylsulfonylmethane as an internal standard.

Entry	Gly-CN (eq.)	Gly-CN (mM)	pH	Yield Ac-Gly-Gly-CN (%)
1	1.1	55	5.0	43
2	1.1	55	7.0	78
3	1.1	55	9.0	91
4	2	100	5.0	95
5	2	100	7.0	> 99
6	2	100	9.0	> 99
7	3	150	5.0	96
8	3	150	7.0	> 99
9	3	150	9.0	> 99

**General procedure D: oxidative ligation of *N*-acetyl aminothioacids and aminonitriles**

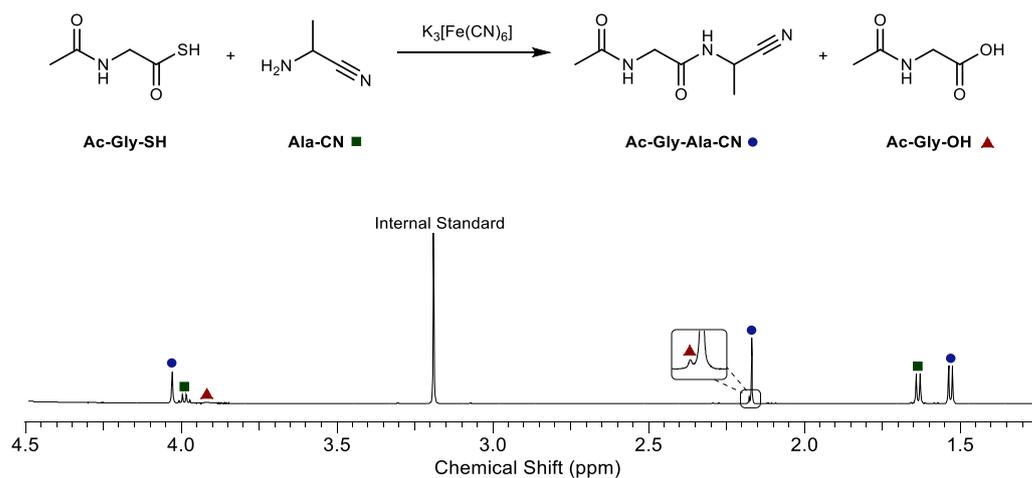


*N*-Acetyl aminothioacid **Ac-AA<sub>1</sub>-SH** (50 mM, 0.05 mmol), aminonitrile **AA<sub>2</sub>-CN** (100 mM, 0.1 mmol) and methylsulfonylmethane (50 mM, 0.05 mmol) were dissolved in H<sub>2</sub>O/D<sub>2</sub>O (98:2, 1 mL), and the solution was adjusted to pH 9.0 with NaOH/HCl. Potassium hexacyanoferrate(III) (49.4 mg, 0.15 mmol) was added, and the solution was stirred at room temperature for 20 minutes. The resulting suspension was then adjusted to pH 9.0 and lyophilised. The resulting crude solid was dissolved in DMSO-*d*<sub>6</sub> (1 mL) and NMR spectra were acquired. The presence of *N*-acetyl dipeptide

nitrile **Ac-AA<sub>1</sub>-AA<sub>2</sub>-CN** was confirmed by <sup>1</sup>H-<sup>13</sup>C HMBC NMR and HRMS. The products were quantified using methylsulfonylmethane as internal standard.

<b>Experimental table 7: oxidative ligation of N-acetyl aminothioacids and aminonitriles</b>			
<b>Entry</b>	<b>AA<sub>1</sub></b>	<b>AA<sub>2</sub></b>	<b>Yield Ac-AA<sub>1</sub>-AA<sub>2</sub>-CN (%)</b>
1	Gly	Ala	95
2	Gly	Val	94
3	Gly	Leu	89
4	Gly	Ile	87
5	Gly	Ser	92
6	Gly	Phe	88
7	Gly	Glu	77
8	Gly	Lys	79
9	Gly	Pro	88
9	Gly	Met	93
10	Ala	Gly	93
11	Val	Gly	92
12	Leu	Gly	93
13	Ile	Gly	83
14	Met	Gly	80
15	Phe	Gly	78
16	Pro	Gly	80
17	Lys	Gly	88
18	Ala	Ala	91

*N*-Acetylglycylalanine nitrile **Ac-Gly-Ala-CN**



**Experimental figure 29:**  $^1H$  NMR (600 MHz,  $H_2O/D_2O$  98:2, 1.25 – 4.50 ppm) spectrum to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), alanine nitrile (**Ala-CN**; 100 mM) and  $K_3[Fe(CN)_6]$  (150 mM) at room temperature and pH 9.0 after 20 min.

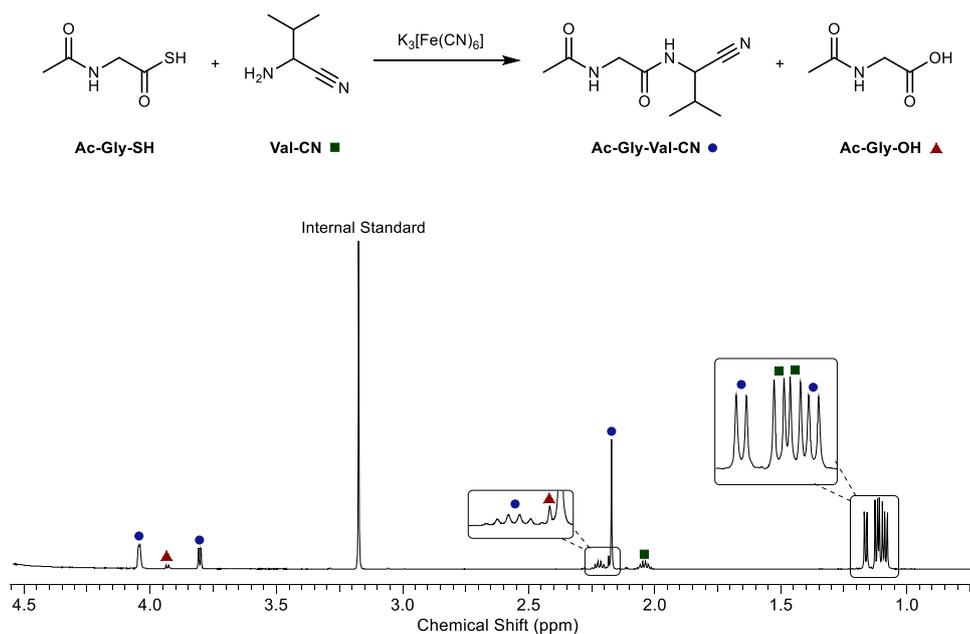
$^1H$  NMR (600 MHz,  $H_2O/D_2O$  (98:2), pH 9.0):

*N*-Acetylglycylalanine nitrile **Ac-Gly-Ala-CN** (•, partial assignment):  $\delta$  4.03 (s, 2H, Gly-(C2)-H<sub>2</sub>), 2.16 (s, 3H, (CO)CH<sub>3</sub>), 1.53 (d,  $J = 7.1$  Hz, 3H, Ala-(C3)-H<sub>3</sub>).

Alanine nitrile **Ala-CN** (■):  $\delta$  3.99 (q,  $J = 7.1$  Hz, 1H, (C2)-H), 1.64 (d,  $J = 7.1$  Hz, 3H, (C3)-H<sub>3</sub>).

*N*-Acetylglycine **Ac-Gly-OH** (▲):  $\delta$  3.92 (s, 2H, (C2)-H<sub>2</sub>), 2.17 (s, 3H, (CO)CH<sub>3</sub>).

*N*-Acetylglycylvaline nitrile **Ac-Gly-Val-CN**



**Experimental figure 30:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 0.70 – 4.55 ppm) spectrum to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), valine nitrile (**Val-CN**; 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and pH 9.0 after 20 min.

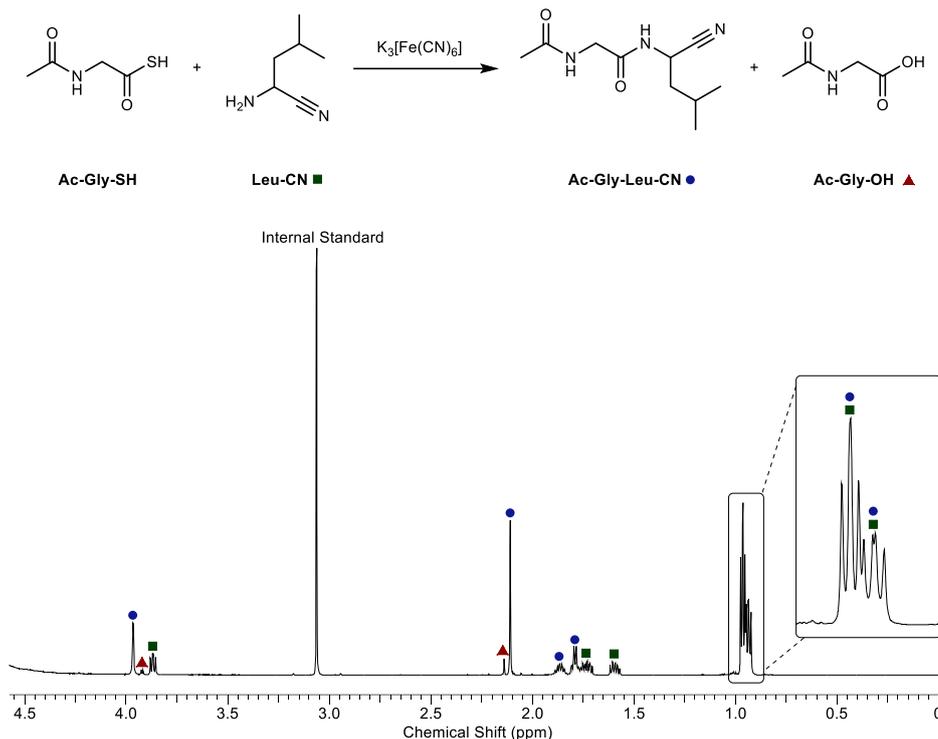
$^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, pH 9.0):

*N*-Acetylglycylvaline nitrile **Ac-Gly-Val-CN** (•):  $\delta$  4.04 (s, 2H, Gly-(C2)–H<sub>2</sub>), 2.25 – 2.20 (m, 1H, Val-(C3)–H), 2.17 (s, 3H, (CO)CH<sub>3</sub>), 1.16 (d,  $J = 6.8$  Hz, 3H, Val-(C4)–H<sub>3</sub>), 1.07 (d,  $J = 6.8$  Hz, 3H, Val-(C4')–H<sub>3</sub>).

Valine nitrile **Val-CN** (■):  $\delta$  3.80 (d,  $J = 7.0$  Hz, 1H, (C2)–H), 2.06 – 2.00 (m, 1H, (C3)–H), 1.13 (d,  $J = 6.8$  Hz, 3H, (C4)–H<sub>3</sub>), 1.10 (d,  $J = 6.8$  Hz, 3H, (C4')–H<sub>3</sub>).

*N*-Acetylglycine **Ac-Gly-OH** (▲):  $\delta$  3.97 (app. d,  $J = 7.4$  Hz, 2H, (C2)–H<sub>2</sub>), 2.18 (s, 3H, (CO)CH<sub>3</sub>).

*N*-Acetylglycylleucine nitrile **Ac-Gly-Leu-CN**



**Experimental figure 31:** <sup>1</sup>H NMR (600 MHz, H<sub>2</sub>O/D<sub>2</sub>O 98:2, 0.00 – 4.55 ppm) spectrum to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), leucine nitrile (**Leu-CN**; 100 mM) and K<sub>3</sub>[Fe(CN)<sub>6</sub>] (150 mM) at room temperature and pH 9.0 after 20 min.

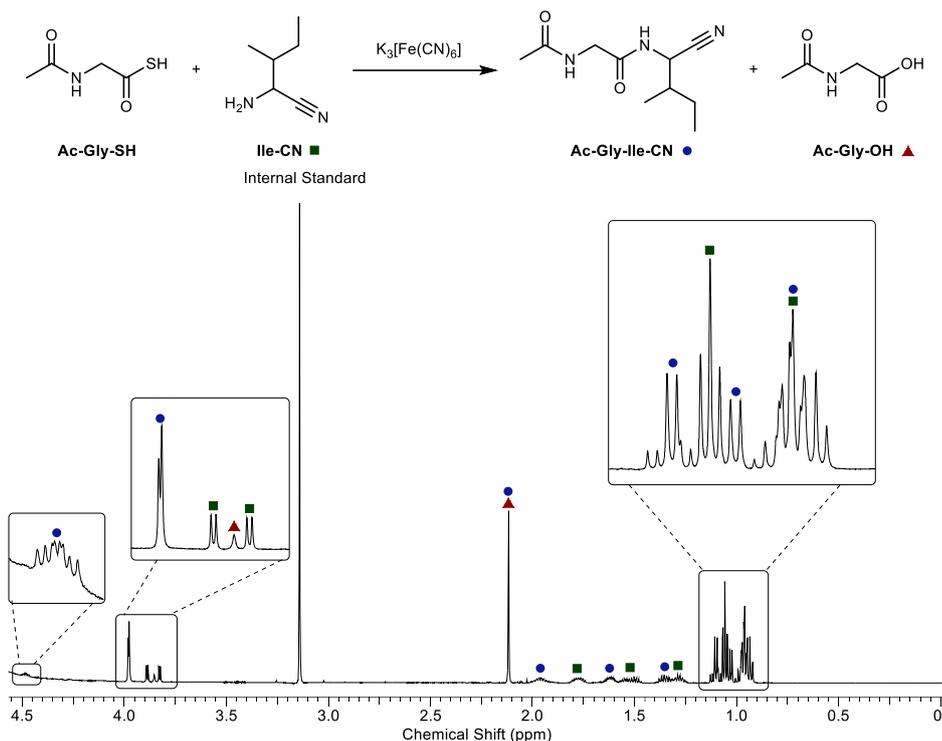
<sup>1</sup>H NMR (600 MHz, H<sub>2</sub>O/D<sub>2</sub>O 98:2, pH 9.0):

*N*-Acetylglycylleucine nitrile **Ac-Gly-Leu-CN** (●): δ 3.96 (s, 2H, Gly-(C2)–H<sub>2</sub>), 2.21 (s, 3H, (CO)CH<sub>3</sub>), 1.90 – 1.82 (m, 2H, Leu-(C3)–H<sub>2</sub>), 1.82 – 1.77 (m, 1H, Leu-(C4)–H), 0.98 – 0.95 (m, 3H, Leu-(C5)–H<sub>3</sub>), 0.94 – 0.91 (m, 3H, Leu-(C5')–H<sub>3</sub>).

Leucine nitrile **Leu-CN** (■): δ 3.87 (dd, *J* = 5.9, 10.2 Hz, 1H, (C2)–H), 1.77 – 1.79 (m, 2H, (C3)–H<sub>2</sub>), 1.62 – 1.56 (m, 1H, (C4)–H), 0.98 – 0.95 (m, 3H, (C5)–H<sub>3</sub>), 0.94 – 0.91 (m, 3H, (C5')–H<sub>3</sub>).

*N*-Acetylglycine **Ac-Gly-OH** (▲): δ 3.92 (app. d, *J* = 7.1 Hz, 2H, (C2)–H<sub>2</sub>), 2.14 (s, 3H, (CO)CH<sub>3</sub>).

*N*-Acetylglycylisoleucine nitrile **Ac-Gly-Ile-CN**



**Experimental figure 32:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 0.00 – 4.55) spectrum to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), isoleucine nitrile (**Ac-Ile-CN**; 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and pH 9.0 after 20 min.

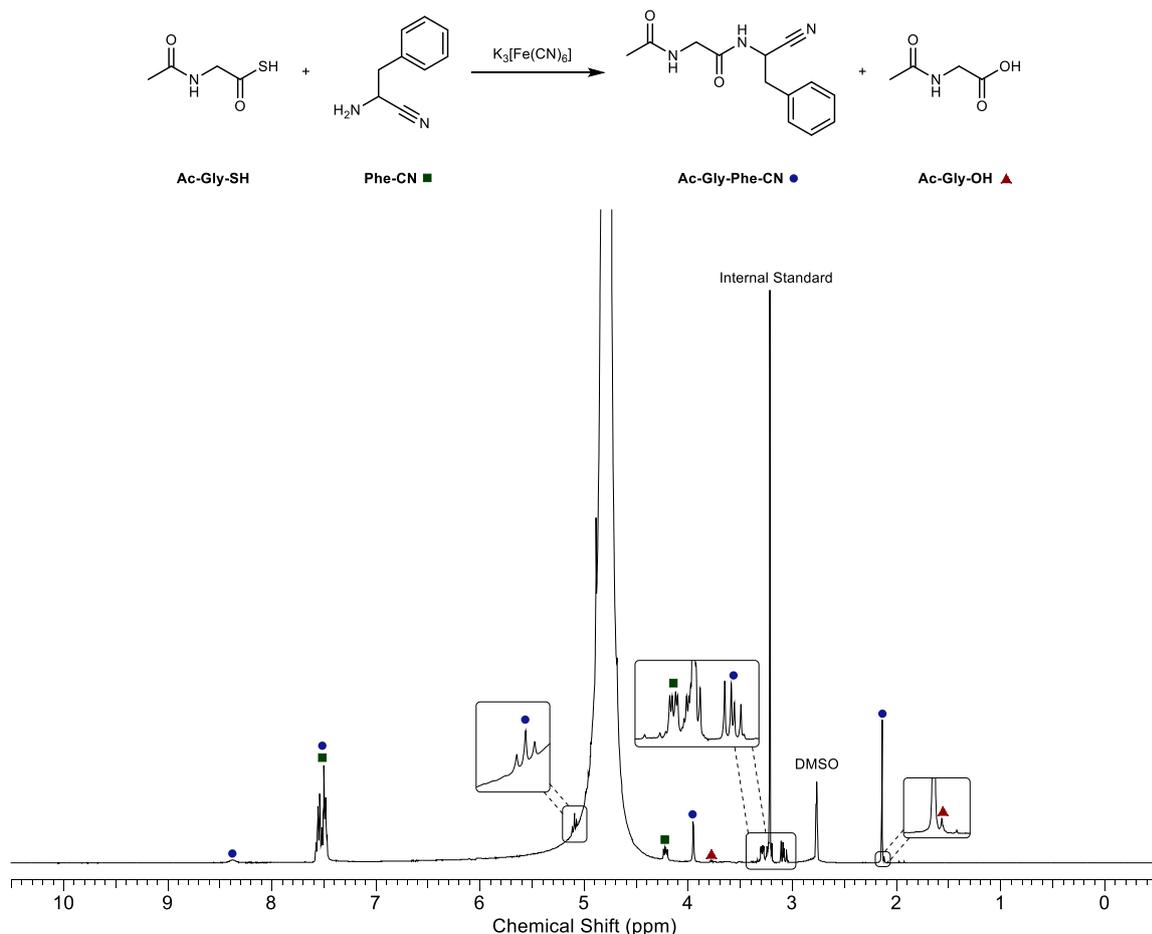
$^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, pH 9.0):

*N*-Acetylglycylisoleucine nitrile **Ac-Gly-Ile-CN** (•, 57:43 mixture of diastereoisomers a/b): diastereoisomer a:  $\delta$  4.50 – 4.46 (m, 1H, Ile-(C2)–H), 3.97 (s, 2H, Gly-(C2)–H<sub>2</sub>), 2.11 (s, 3H, (CO)CH<sub>3</sub>), 1.81 – 1.74 (m, 1H, Ile-(C3)–H), 1.57 – 1.51 (m, 1H, Ile-(C4)–H), 1.34 – 1.26 (m, 1H, Ile-(C4')–H), 1.10 (obs. d, 3H, Ile-(C3)–CH<sub>3</sub>), 0.98 – 0.94 (m, 3H, Ile-(C5)–H<sub>3</sub>); diastereoisomer b:  $\delta$  4.50 – 4.46 (m, 1H, Ile-(C2)–H), 3.97 (s, 2H, Gly-(C2)–H<sub>2</sub>), 2.12 (s, 3H, (CO)CH<sub>3</sub>), 1.81 – 1.74 (m, 1H, Ile-(C3)–H), 1.57 – 1.51 (m, 1H, Ile-(C4)–H), 1.34 – 1.26 (m, 1H, Ile-(C4')–H), 1.03 (obs. d, 3H, Ile-(C3)–CH<sub>3</sub>), 0.98 – 0.94 (m, 3H, Ile-(C5)–H<sub>3</sub>).

*Isoleucine nitrile Ile-CN* (■, 52:48 mixture of diastereoisomer a/b): diastereoisomer a:  $\delta$  3.89 (d,  $J = 5.1$  Hz, 1H, (C2)–H), 2.01 – 1.90 (m, 1H, (C3)–H), 1.57 – 1.48 (m, 1H, (C4)–H), 1.38 – 1.30 (m, 1H, (C4')–H), 0.94 (m, 3H, (C3)–CH<sub>3</sub>), 0.92 (m, 3H, (C5)–H<sub>3</sub>); diastereoisomer b:  $\delta$  3.87 (d,  $J = 6.0$  Hz, 1H, (C2)–H), 2.09 – 2.00 (m, 1H, (C3)–H), 1.65 – 1.57 (m, 1H, (C4)–H), 1.30 – 1.24 (m, 1H, (C4')–H), 0.94 (m, 3H, (C3)–CH<sub>3</sub>), 0.92 (m, 3H, (C5)–H<sub>3</sub>).

*N*-Acetylglycine **Ac-Gly-OH** (▲):  $\delta$  3.85 (s, 2H, (C2)–H<sub>2</sub>), 2.12 (s, 3H, (CO)CH<sub>3</sub>).

*N*-Acetylglycylphenylalanine nitrile **Ac-Gly-Phe-CN**



**Experimental figure 33:**  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-}d_6$  1:49:50, -0.50 – 10.50 ppm) spectrum to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), phenylalanine nitrile (**Phe-CN**; 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and pH 9.0 in  $\text{H}_2\text{O}/\text{D}_2\text{O}$  (98:2, 1 mL). After 20 min, the solution was adjusted to pH 9.0 and diluted with  $\text{DMSO-}d_6$  (1 mL) for analysis.

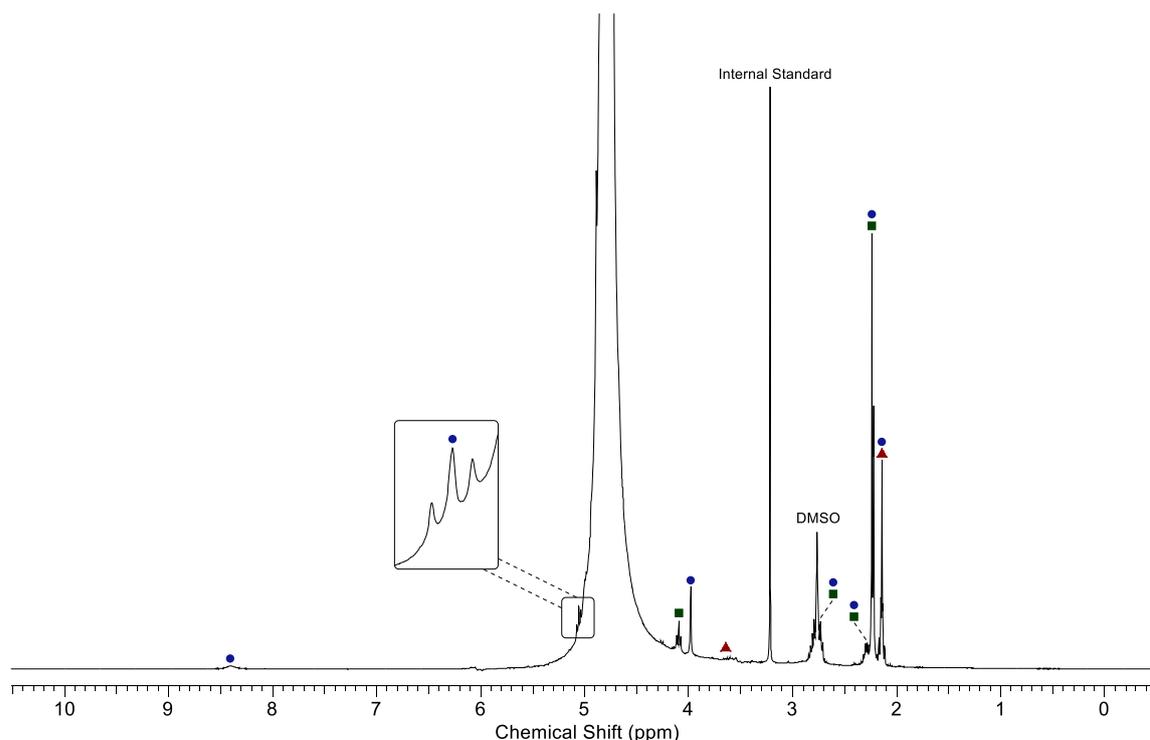
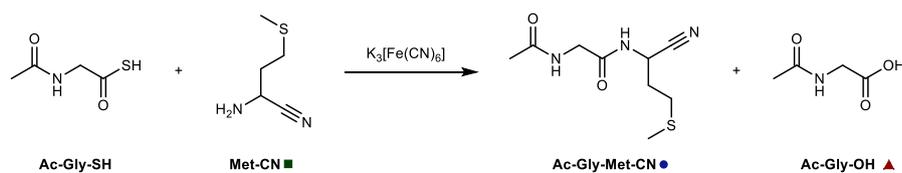
$^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-}d_6$  1:49:50, pH 9.0, partial assignment):

*N*-Acetylglycylphenylalanine nitrile **Ac-Gly-Phe-CN** (•)  $\delta$  8.38 (br s, 1H, Phe-NH), 7.59 – 7.46 (m, 5H, Phe-Ar), 5.10 (app. t,  $J = 8.2$  Hz, 1H, Phe-(C2)-H), 3.95 (s, 2H, Gly-(C2)-H<sub>2</sub>), 3.09 (ABX,  $J = 8.2, 13.8$  Hz, 1H, Phe-(C3)-H), 2.14 (s, 3H, (CO)CH<sub>3</sub>).

Phenylalanine nitrile **Phe-CN** (■)  $\delta$  7.59 – 7.46 (obs. m, 5H, Ar), 4.22 (dd,  $J = 6.1, 9.2$  Hz, 1H, (C2)-H), 3.29 (ABX,  $J = 3.4, 8.2$  Hz, 1H, (C3)-H).

*N*-Acetylglycine **Ac-Gly-OH** (▲):  $\delta$  3.78 (s, 2H, (C2)-H<sub>2</sub>), 2.12 (s, 3H, (CO)CH<sub>3</sub>).

*N*-Acetylglycylmethionine nitrile **Ac-Gly-Met-CN**



**Experimental figure 34:**  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-}d_6$  1:49:50, -0.50 – 10.50 ppm) spectrum to show the reaction of N-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), methionine nitrile (**Met-CN**, 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) in  $\text{H}_2\text{O}/\text{D}_2\text{O}$  (98:2, 1 mL) at pH 9.0 and room temperature. After 20 min, the solution was adjusted to pH 9.0 and diluted with  $\text{DMSO-}d_6$  (1 mL) for analysis.

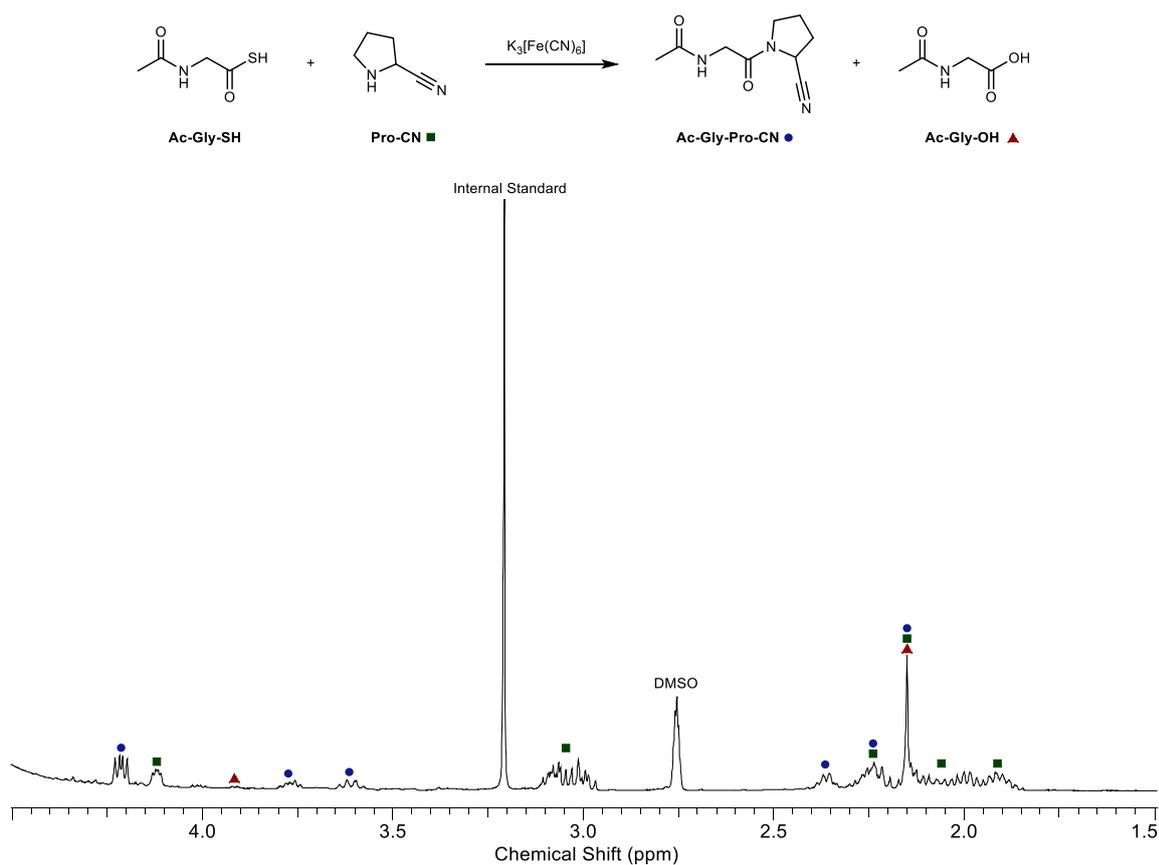
$^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-}d_6$  1:49:50, pH 9.0):

*N*-Acetylglycylmethionine nitrile **Ac-Gly-Met-CN** (●):  $\delta$  8.40 (br s, 1H, Met-NH), 5.08 (obs. t,  $J = 7.8$  Hz, 1H, Met-(C2)-H), 3.98 (s, 2H, Gly-(C2)-H<sub>2</sub>), 2.86 – 2.67 (obs. m, 2H, Met-(C4)-H<sub>2</sub>), 2.34 – 2.18 (obs. m, 2H, Met-(C3)-H<sub>2</sub>), 2.22 (obs. s, 3H,  $\text{SCH}_3$ ); 2.14 (s, 3H,  $(\text{CO})\text{CH}_3$ ).

*Methionine nitrile* **Met-CN** (■):  $\delta$  4.09 (t,  $J = 7.3$  Hz, 1H, (C2)-H), 2.86 – 2.67 (obs. m, 2H, (C4)-H<sub>2</sub>), 2.34 – 2.18 (obs. m, 2H, (C3)-H<sub>2</sub>), 2.24 (obs. s, 3H,  $\text{SCH}_3$ ).

*N*-Acetylglycine **Ac-Gly-OH** (▲):  $\delta$  3.70 (s, 2H, (C2)-H<sub>2</sub>), 2.14 (s, 3H,  $(\text{CO})\text{CH}_3$ ).

*N*-Acetylglycylproline nitrile **Ac-Gly-Pro-CN**



**Experimental figure 35:** <sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O/H<sub>2</sub>O/DMSO-*d*<sub>6</sub> 1:49:50, 4.50 – 1.50 ppm) spectrum to show the reaction of N-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), proline nitrile (**Pro-CN**, 100 mM) and K<sub>3</sub>[Fe(CN)<sub>6</sub>] (150 mM) in D<sub>2</sub>O/H<sub>2</sub>O (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the solution was adjusted to pH 9.0 and diluted with DMSO-*d*<sub>6</sub> (1 mL) for analysis.

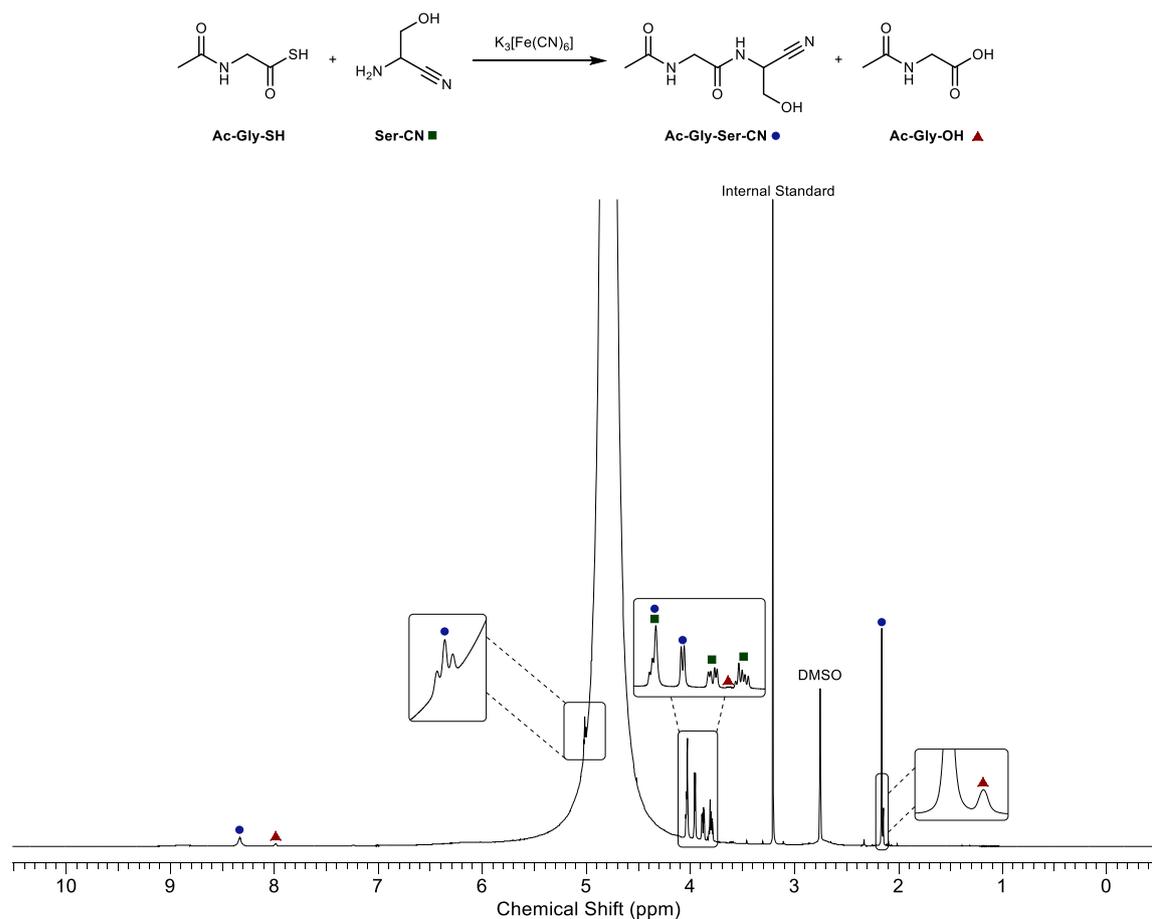
<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O/H<sub>2</sub>O/DMSO-*d*<sub>6</sub> 1:49:50, pH 9.0):

*N*-Acetylglycylproline nitrile **Ac-Gly-Pro-CN** (•, major rotamer, partial assignment): δ 4.21 (dd, *J* = 4.6, 8.0 Hz, Gly-(C2)-H<sub>2</sub>), 3.80 – 3.73 (m, 1H, Pro-(C5)-H), 3.65 – 3.58 (m, 1H, Pro-(C5')-H), 2.41 – 2.32 (m, 2H, Pro-(C3)-H<sub>2</sub>), 2.30 – 2.19 (m, 2H, Pro-(C4)-H<sub>2</sub>), 2.15 (s, 3H, (CO)CH<sub>3</sub>).

*Proline nitrile* **Pro-CN** (■) δ 4.12 (m, 1H, (C2)-H), 3.12 – 2.85 (m, 2H, (C5)-H<sub>2</sub>), 2.31 – 2.17 (obs. m, (C3)-H), 2.14 – 1.82 (obs. m, 3H, (C3')-H, (C4)-H<sub>2</sub>).

*N*-Acetylglycine **Ac-Gly-OH** (▲): δ 3.90 (s, 2H, (C2)-H<sub>2</sub>), 2.15 (s, 3H, (CO)CH<sub>3</sub>).

*N*-Acetylglycylserine nitrile **Ac-Gly-Ser-CN**



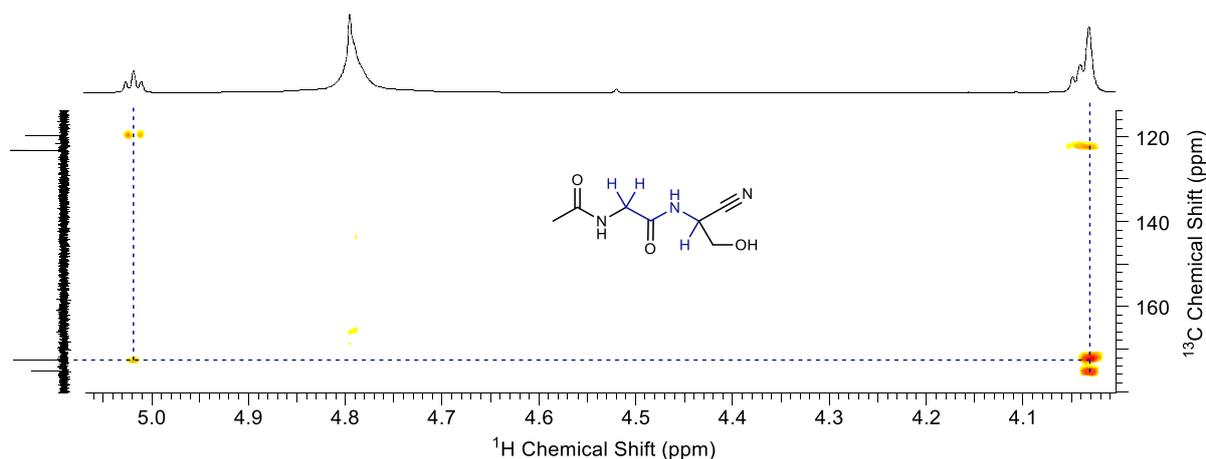
**Experimental figure 36:**  $^1\text{H}$  NMR (700 MHz,  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-}d_6$  1:49:50, 10.50 – -0.50 ppm) spectrum to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), serine nitrile (**Ser-CN**, 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and pH 9.0 after 20 min.

$^1\text{H}$  NMR (700 MHz,  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-}d_6$  1:49:50):

*N*-Acetylglycylserine nitrile **Ac-Gly-Ser-CN** (●):  $\delta$  8.33 (br s, 1H, Ser-NH), 5.02, (obs. t,  $J = 5.6$  Hz, 1H, Ser-(C2)-H), 4.03 (s, 2H, Gly-(C2)-H<sub>2</sub>), 3.96 (d,  $J = 5.6$  Hz, 2H, Ser-(C3)-H<sub>2</sub>), 2.16 (s, 3H, (CO)CH<sub>3</sub>).

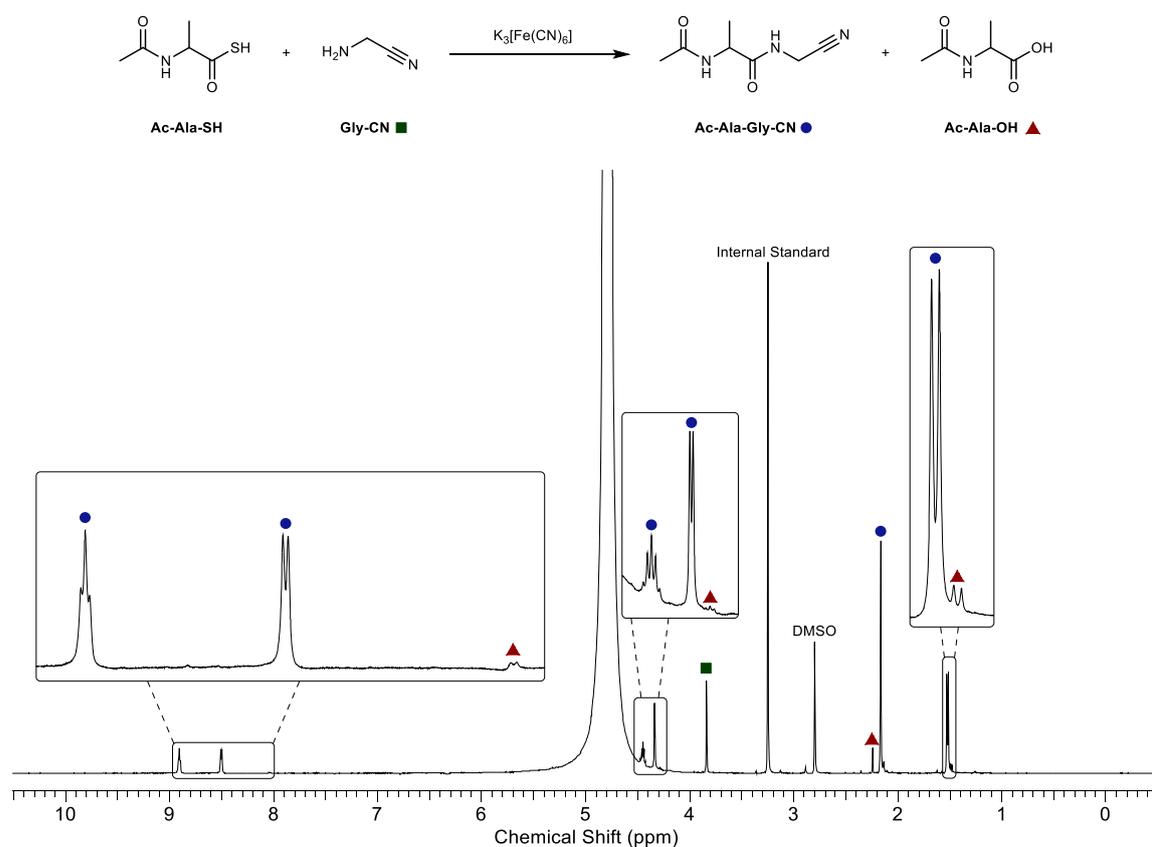
Serine nitrile **Ser-CN** (■):  $\delta$  4.05 – 4.01 (obs. m, 1H, (C2)-H), 3.88 (ABX,  $J = 5.8, 11.6$  Hz, 1H, (C3)-H), 3.80 (app. ABX,  $J = 5.8, 11.6$  Hz, 1H, (C3)-H').

*N*-Acetylglycine **Ac-Gly-OH** (▲):  $\delta$  7.99 (br s, 1H, NH), 3.81 (app. d,  $J = 5.8$  Hz, 2H, (C2)-H<sub>2</sub>), 2.14 (s, 3H, (CO)CH<sub>3</sub>).



**Experimental figure 37:**  $^1\text{H}$  –  $^{13}\text{C}$  HMBC ( $^1\text{H}$  – 700 MHz [5.05 – 4.05 ppm],  $^{13}\text{C}$  – 176 MHz [115 – 180 ppm], water-suppressed,  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-}d_6$  1:49:50) spectrum to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), serine nitrile (**Ser-CN**; 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the solution was adjusted to pH 9.0 and diluted with  $\text{DMSO-}d_6$  (1 mL) for analysis. The characteristic  $^3J_{\text{CH}}$  coupling between serine nitrile proton resonance (5.02 ppm, obs. t,  $J = 5.6$  Hz, 1H, Ser2-(C2)–H), and glycine amide carbon resonance (172.2 ppm, Gly1-(C1)) that demonstrates peptide ligation is highlighted with overlaid blue dotted lines.

*N*-Acetylalanylglycine nitrile **Ac-Ala-Gly-CN**



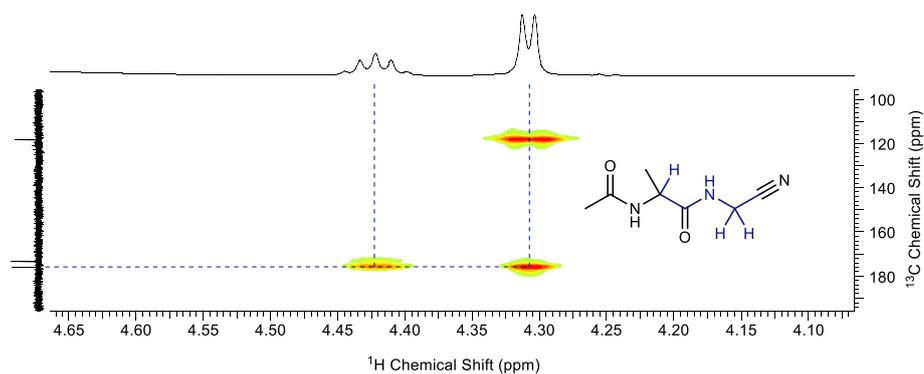
**Experimental figure 38:** <sup>1</sup>H NMR 600 MHz, D<sub>2</sub>O:H<sub>2</sub>O:DMSO-*d*<sub>6</sub> 1:49:50, -0.50 – 10.50 ppm) spectrum to show the reaction of *N*-acetylalanine thioacid (**Ac-Ala-SH**; 50 mM), glycine nitrile (**Gly-CN**; 100 mM) and K<sub>3</sub>[Fe(CN)<sub>6</sub>] (150 mM) in H<sub>2</sub>O/D<sub>2</sub>O (98:2, 1 mL) at pH 9.0 and room temperature. After 20 min, the solution was adjusted to pH 9.0 and diluted with DMSO-*d*<sub>6</sub> (1 mL) for analysis.

<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O/H<sub>2</sub>O/DMSO-*d*<sub>6</sub> 1:49:50):

*N*-Acetylalanylglycine nitrile **Ac-Ala-Gly-CN** (•): δ 8.91 (br t, *J* = 5.4 Hz, 1H, Gly-NH), 8.50 (d, *J* = 6.2 Hz, 1H, Ala-NH), 4.48 – 4.42, (m, 1H, Ala-(C2)-H), 4.34, (d, *J* = 5.4 Hz, 2H, Gly-(C2)-H<sub>2</sub>), 2.16 (s, 3H, (CO)CH<sub>3</sub>), 1.52 (d, *J* = 7.3 Hz, 3H, Ala-(C3)-H<sub>3</sub>).

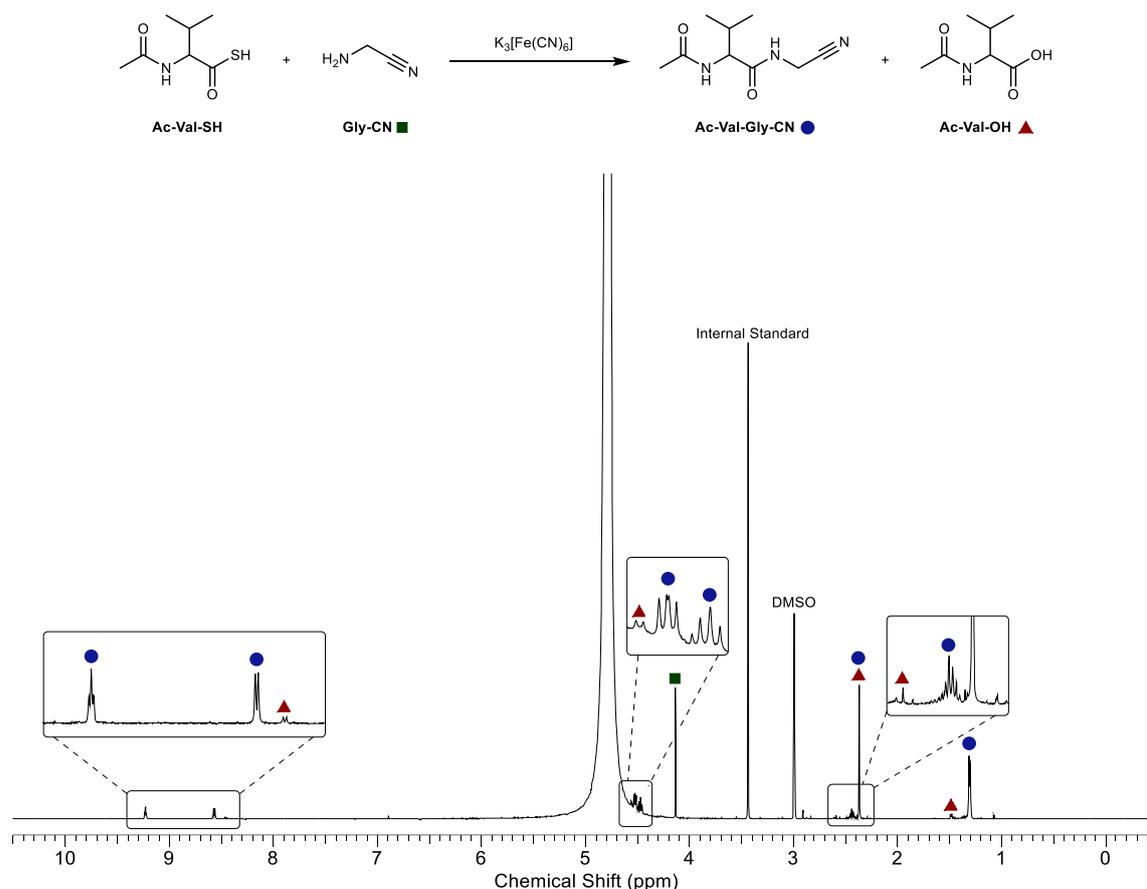
*N*-Acetylalanine **Ac-Ala-OH** (▲): δ 8.04 (br d, *J* = 6.8 Hz, 1H, NH), 4.30 – 4.26, (m, 1H, (C2)-H), 2.24 (s, 3H, (CO)CH<sub>3</sub>), 1.49 (d, *J* = 7.1 Hz, 3H, (C3)-H<sub>3</sub>).

Glycine nitrile **Gly-CN** (■): δ 3.84 (s, 2H, (C2)-H<sub>2</sub>).



**Experimental figure 39:**  $^1\text{H}$  –  $^{13}\text{C}$  HMBC ( $^1\text{H}$  – 600 MHz [4.6 – 4.0 ppm],  $^{13}\text{C}$  – 151 MHz [110 – 185 ppm], water-suppressed,  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-d}_6$  1:49:50) spectrum to show the reaction of *N*-acetylglycine thioacid **Ac-Ala-SH** (50 mM), glycine nitrile **Gly-CN** (100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the solution was adjusted to pH 9.0 and diluted with  $\text{DMSO-d}_6$  (1 mL) for analysis. The characteristic the diagnostic  $^3J_{\text{CH}}$  coupling between glycine nitrile proton resonance (4.34 ppm, d,  $J = 5.5$  Hz, 2H, Gly2-(C2)–H<sub>2</sub>), and alanine amide carbon resonance (179.6 ppm Ala1-(C1)) that demonstrates peptide ligation is highlighted with overlaid blue dotted lines.

*N*-Acetylvalylglycine nitrile **Ac-Val-Gly-CN**



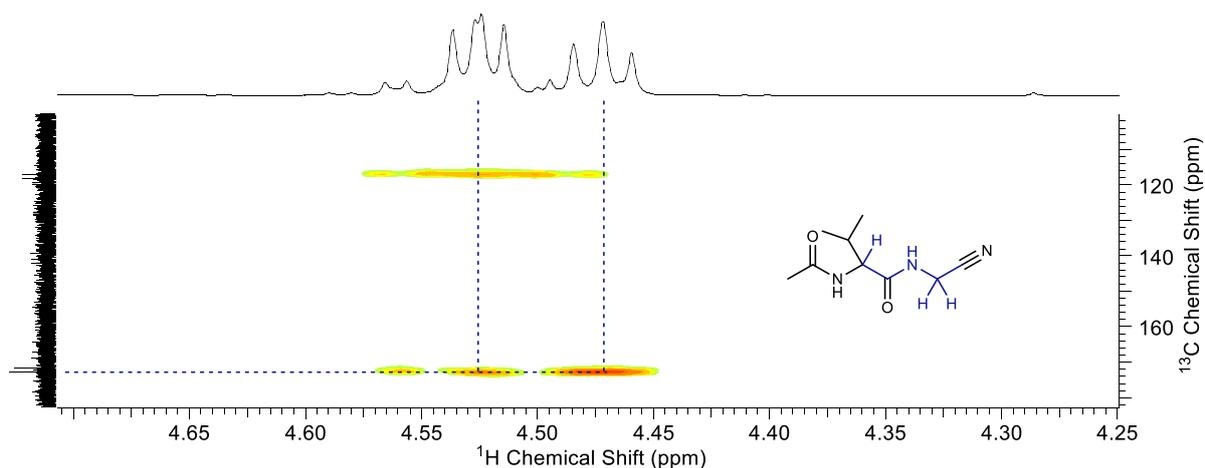
**Experimental figure 40:**  $^1H$  NMR (600 MHz,  $D_2O/H_2O/DMSO-d_6$  1:49:50, -0.50 – 10.50 ppm) spectrum showing the reaction of *N*-acetylvaline thioacid (**Ac-Val-SH**; 50 mM), glycine nitrile (**Gly-CN**; 100 mM) and  $K_3[Fe(CN)_6]$  (150 mM) in  $D_2O/H_2O$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the solution was adjusted to pH 9.0 and diluted with  $DMSO-d_6$  (1 mL) for analysis.

$^1H$  NMR (600 MHz,  $D_2O/H_2O/DMSO-d_6$  1:49:50):

*N*-Acetylvalylglycine nitrile **Ac-Val-Gly-CN** (●):  $\delta$  9.23 (br t,  $J = 5.6$  Hz, 1H, Gly-NH), 8.57 (d,  $J = 7.9$  Hz, 1H, Val-NH), 4.52 (dd,  $J = 5.6, 7.6$  Hz, 1H, Val-(C2)-H), 4.47, (t,  $J = 7.6$  Hz, 2H, Gly-(C2)-H<sub>2</sub>), 2.48 – 2.41 (m, 1H, Val-(C3)-H), 2.37 (s, 3H, (CO)CH<sub>3</sub>), 1.33 – 1.28 (m, 6H, Val1-(C4)-H<sub>3</sub>, Val1-(C4')-H<sub>3</sub>).

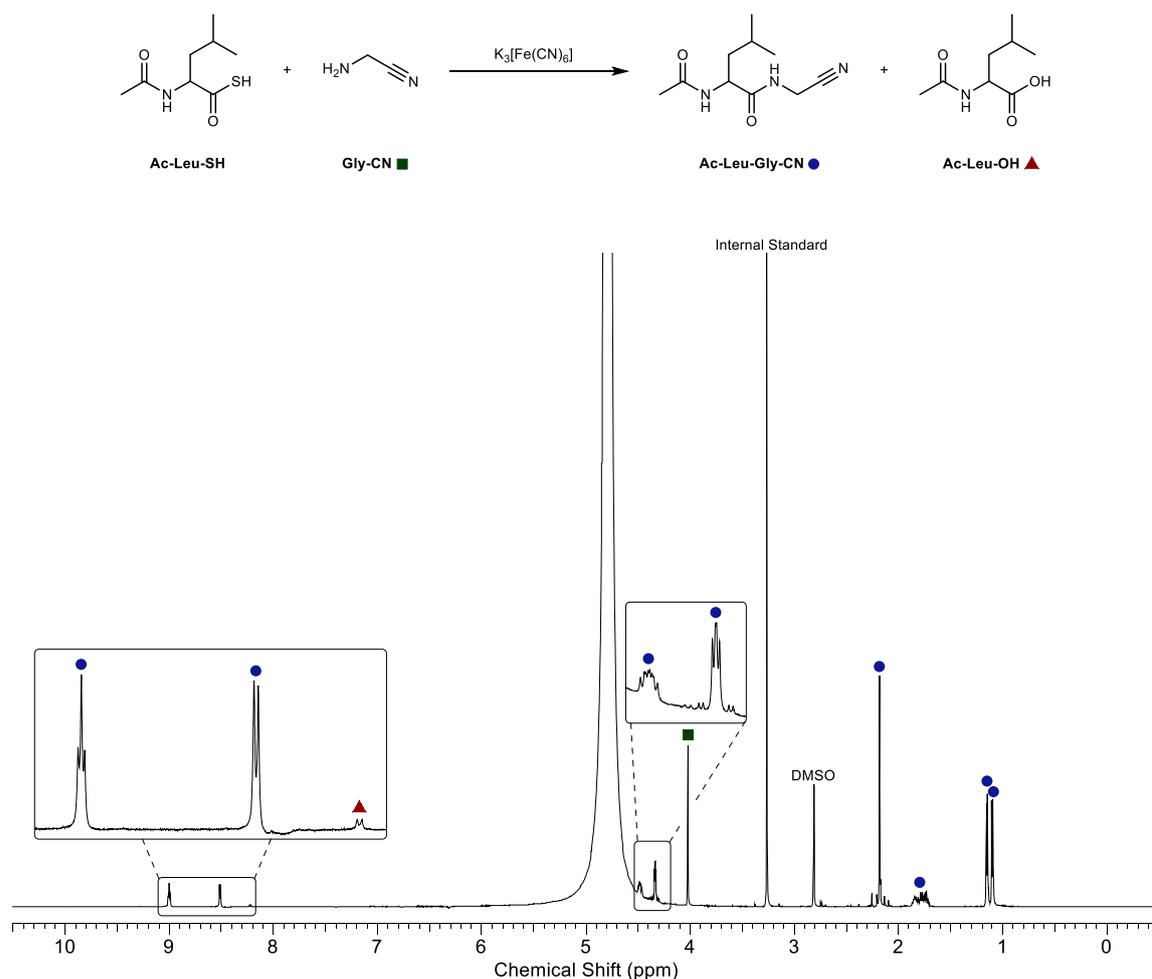
*N*-Acetylvaline **Ac-Val-OH** (▲):  $\delta$  8.46 (br d,  $J = 8.0$  Hz, 1H, NH), 4.56, (app. d,  $J = 6.2$  Hz, 1H, (C2)-H), 2.63 – 2.57 (m, 1H, (C3)-H), 2.37 (s, 3H, (CO)CH<sub>3</sub>), 1.50 – 1.47 (m, 6H, (C4)-H<sub>3</sub>, (C4')-H<sub>3</sub>).

Glycine nitrile **Gly-CN** (■):  $\delta$  4.14 (s, 2H, (C2)-H<sub>2</sub>).



**Experimental figure 41:**  $^1\text{H} - ^{13}\text{C}$  HMBC ( $^1\text{H} - 600$  MHz [4.70 – 4.25 ppm],  $^{13}\text{C} - 151$  MHz [110 – 180 ppm], water-suppressed,  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-d}_6$  1:49:50) spectrum to show the reaction of N-acetylvaline thioacid (50 mM), glycine nitrile (100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the pH of the solution was adjusted to 9.0 and diluted with  $\text{DMSO-d}_6$  (1 mL) for analysis. The diagnostic  $^3J_{\text{CH}}$  coupling between glycine nitrile proton resonance (4.47 ppm, t,  $J = 7.6$  Hz, 2H, Gly2-(C2)–H<sub>2</sub>), and valine amide carbon resonance (174.8 ppm, Val1-C1) that demonstrates peptide ligation is highlighted with overlaid blue dotted lines.

*N*-Acetylleucylglycine nitrile **Ac-Leu-Gly-CN**



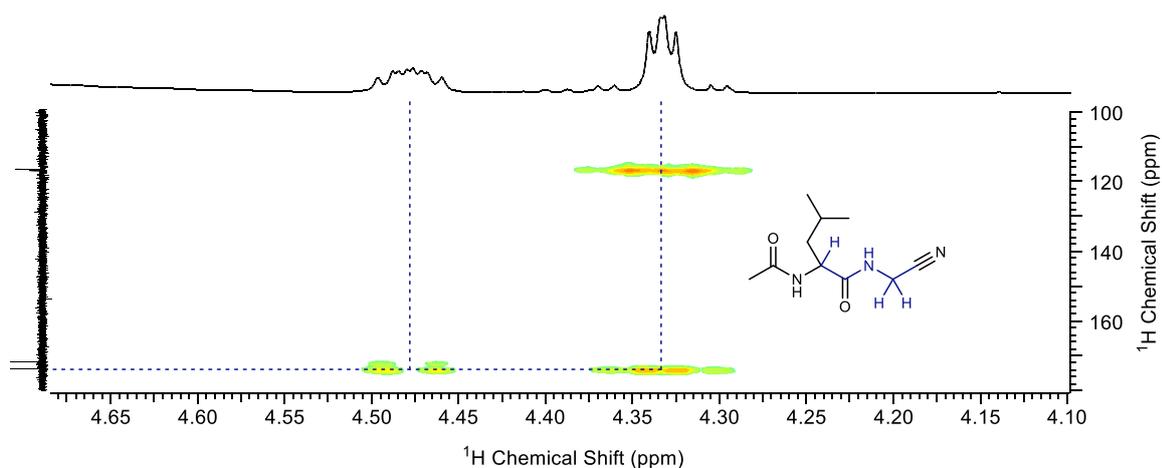
**Experimental figure 42:**  $^1\text{H}$  NMR (600 MHz, 1:49:50  $\text{D}_2\text{O}:\text{H}_2\text{O}:\text{DMSO-}d_6$ , -0.50 – 10.50 ppm) spectrum to show the reaction of *N*-acetylvaline thioacid (**Ac-Val-SH**; 50 mM), glycine nitrile (**Gly-CN**, 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the solution was adjusted to pH 9.0 and diluted with  $\text{DMSO-}d_6$  (1 mL) for analysis.

$^1\text{H}$  NMR 600 MHz, 1:49:50  $\text{D}_2\text{O}:\text{H}_2\text{O}:\text{DMSO-}d_6$ ):

*N*-Acetylleucylglycine nitrile **Ac-Leu-Gly-CN** (●):  $\delta$  9.00 (t,  $J = 6.5$  Hz, 1H, Gly-NH), 8.51 (d,  $J = 7.3$  Hz, 1H, Leu-NH), 4.50 – 4.45, (m, 1H, Leu-(C2)-H), 4.34 – 4.32, (m, 2H, Gly-(C2)-H<sub>2</sub>), 2.18 (s, 3H, (CO)CH<sub>3</sub>), 1.88 – 1.70 (m, 3H, Leu-(C3)-H<sub>2</sub>, Leu-(C4)-H), 1.15 (d,  $J = 6.5$  Hz, 3H, Leu-(C5)-H<sub>3</sub>), 1.10 (d,  $J = 6.5$  Hz, 3H, Leu-(C5')-H<sub>3</sub>).

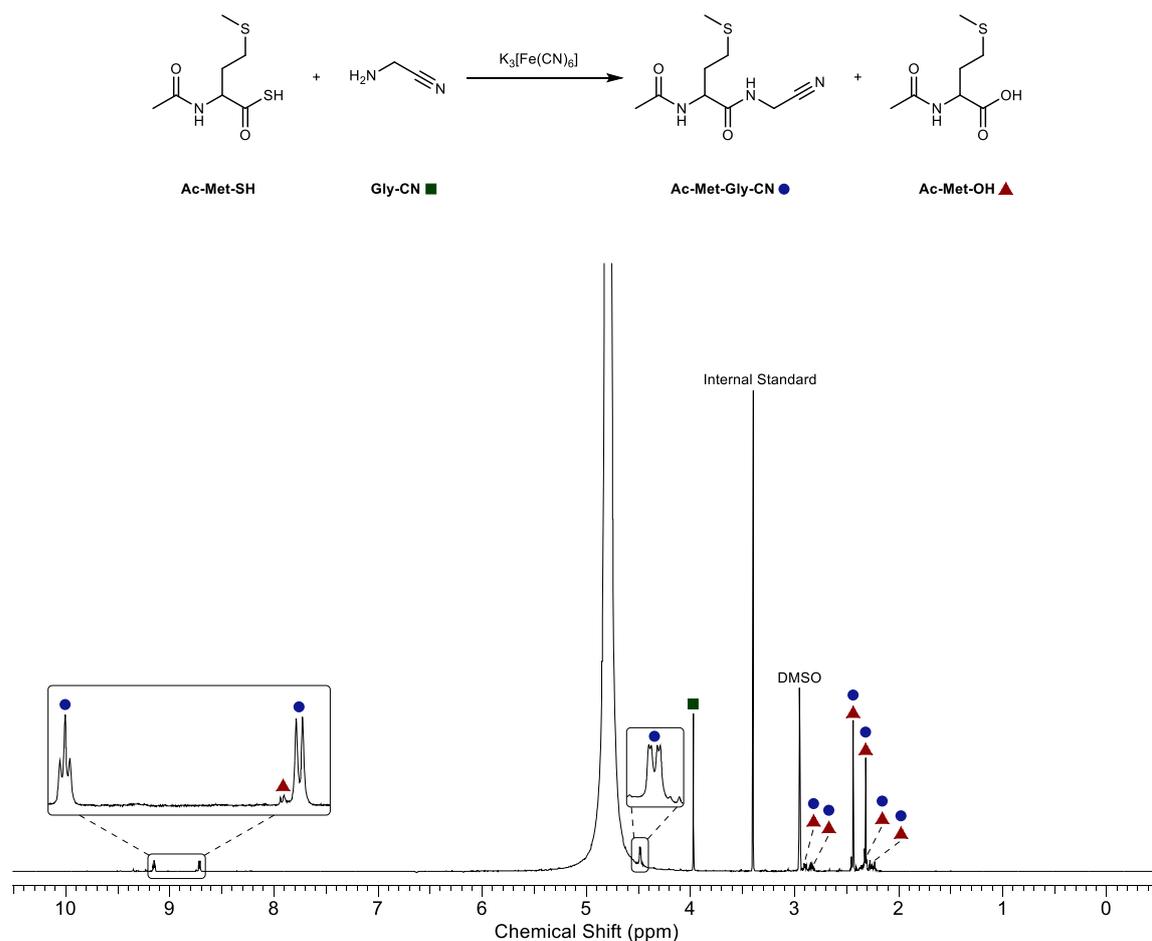
*N*-Acetylleucine **Ac-Leu-OH** (partial assignment) (▲):  $\delta$  8.23 (d,  $J = 7.6$  Hz, 1H, NH).

Glycine nitrile **Gly-CN** (■):  $\delta$  4.03 (s, 2H, (C2)-H<sub>2</sub>).



**Experimental figure 43:**  $^1\text{H}$  –  $^{13}\text{C}$  HMBC ( $^1\text{H}$  – 600 MHz [4.70 – 4.10 ppm],  $^{13}\text{C}$ –151 MHz [100 – 180 ppm], water-suppressed,  $\text{D}_2\text{O}:\text{H}_2\text{O}:\text{DMSO-d}_6$  1:49:50) spectrum to show the reaction of N-acetyl-leucine thioacid (50 mM), glycine nitrile (100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the pH of the solution was adjusted to 9.0 and diluted with  $\text{DMSO-d}_6$  (1 mL) for analysis. The diagnostic  $^3J_{\text{CH}}$  coupling between glycine nitrile proton resonance (4.34 ppm, m, 2H, Gly2-(C2)–H<sub>2</sub>), and leucine amide carbon resonance (174.7 ppm, Leu1-C1) that demonstrates peptide ligation is highlighted with overlaid blue dotted lines.

*N*-Acetylmethionylglycine nitrile **Ac-Met-Gly-CN**



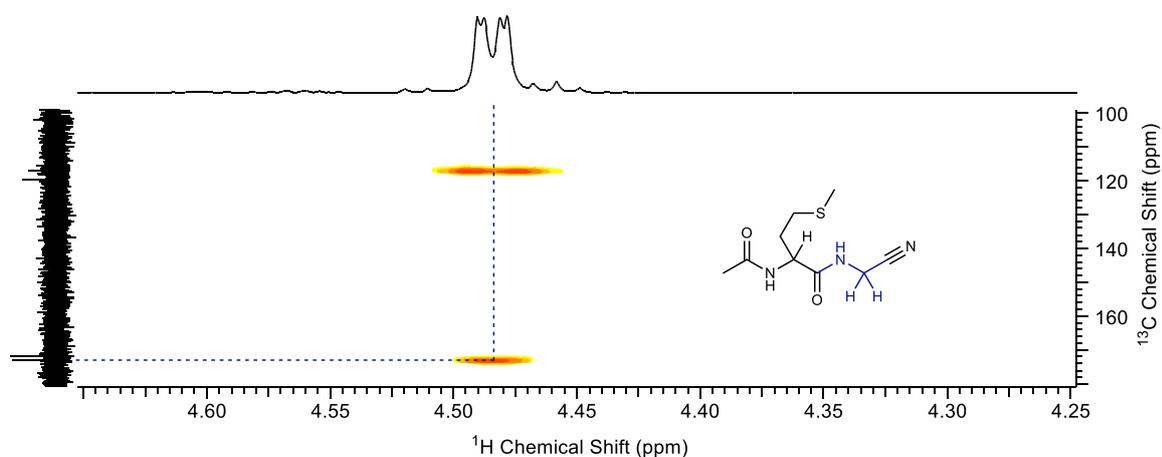
**Experimental figure 44:**  $^1\text{H}$  NMR (600 MHz, 1:49:50  $\text{D}_2\text{O}:\text{H}_2\text{O}:\text{DMSO-}d_6$ , -0.50–10.50 ppm) to show the reaction of *N*-acetylmethionine thioacid (**Ac-Met-SH**; 50 mM), glycine nitrile (**Gly-CN**, 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the solution was adjusted to pH 9.0 and diluted with  $\text{DMSO-}d_6$  (1 mL) for analysis.

$^1\text{H}$  NMR (600 MHz, 1:49:50  $\text{D}_2\text{O}:\text{H}_2\text{O}:\text{DMSO-}d_6$ , partial assignment):

*N*-Acetylmethionylglycine nitrile **Ac-Met-Gly-CN** (●):  $\delta$  9.15 (t,  $J = 6.6$  Hz, 1H, Gly-NH), 8.72 (d,  $J = 7.9$  Hz, 1H, Met-NH), 4.48 (dd,  $J = 1.7, 5.6$  Hz, 2H, Gly-(C2)-H<sub>2</sub>), 2.96 – 2.87 (m, 1H, Met-(C3)-H), 2.86 – 2.80 (m, 1H, Met-(C3')-H), 2.44 (s, 3H, Met1-(C4)-SCH<sub>3</sub>), 2.41 – 2.37 (m, 2H, Met1-(C4)-H<sub>2</sub>), 2.32 (s, 3H, (CO)CH<sub>3</sub>).

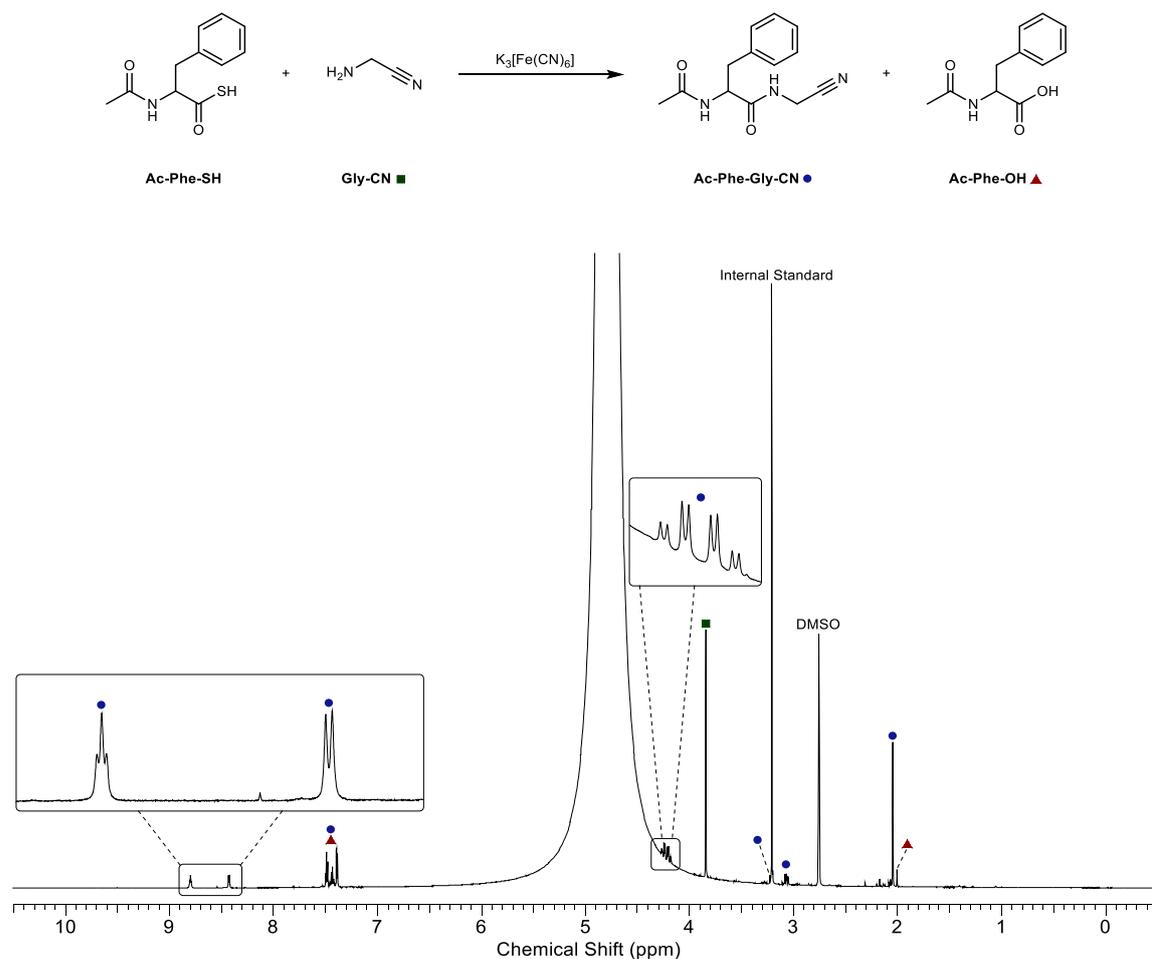
*N*-Acetylmethionine **Ac-Met-OH** (▲):  $\delta$  8.75 (br d,  $J = 7.7$  Hz, 1H, NH), 2.96 – 2.87 (m, 1H, (C3)-H), 2.86 – 2.80 (m, 1H, (C3')-H), 2.43 (s, 3H, (C4)-SCH<sub>3</sub>), 2.33 (s, 3H, (CO)CH<sub>3</sub>), 2.38 – 2.30 (m, 1H, (C4)-H), 2.28 – 2.20 (m, 1H, (C4')-H).

Glycine nitrile **Gly-CN** (■):  $\delta$  3.97 (s, 2H, (C2)-H<sub>2</sub>).



**Experimental figure 45:**  $^1\text{H}$  –  $^{13}\text{C}$  HMBC ( $^1\text{H}$  – 600 MHz [4.65 – 4.25 ppm],  $^{13}\text{C}$  – 151 MHz [100 – 180 ppm], water-suppressed,  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-d}_6$  1:49:50) spectrum to show the reaction of N-acetylmethionine thioacid (50 mM), glycine nitrile (100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the pH of the solution was adjusted to 9.0 and diluted with  $\text{DMSO-d}_6$  (1 mL) for analysis. The diagnostic  $^3J_{\text{CH}}$  coupling between glycine nitrile proton resonance (4.48 ppm, dd,  $J = 1.7, 5.6$  Hz, 2H, Gly2-(C2)–H<sub>2</sub>), and leucine amide carbon resonance (173.3 ppm, Leu1-C1) that demonstrates peptide ligation is highlighted with overlaid blue dotted lines.

*N*-Acetylphenylalanylglycine nitrile **Ac-Phe-Gly-CN**



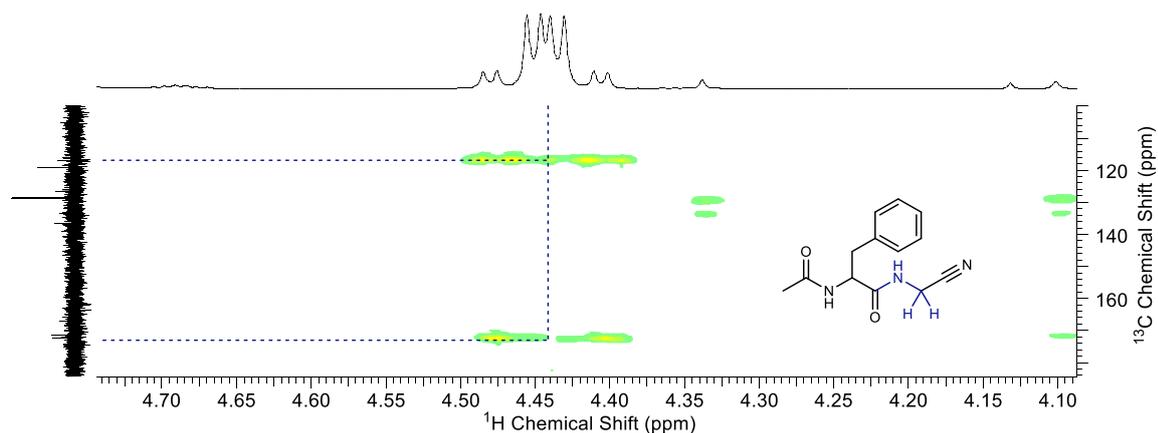
**Experimental figure 46:**  $^1\text{H}$  NMR (600 MHz,  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-}d_6$  1:49:50, -0.50 – 10.50 ppm) spectrum to show the reaction of *N*-acetylphenylalanine thioacid (**Ac-Phe-SH**; 50 mM), glycine nitrile (**Gly-CN**, 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the solution was adjusted to pH 9.0 and diluted with  $\text{DMSO-}d_6$  (1 mL) for analysis.

$^1\text{H}$  NMR (600 MHz, 1:49:50  $\text{D}_2\text{O}:\text{H}_2\text{O}:\text{DMSO-}d_6$ , -0.50–10.50 ppm, partial assignment):

*N*-Acetylphenylalanylglycine nitrile **Ac-Phe-Gly-CN** (●):  $\delta$  8.80 (t,  $J = 5.2$  Hz, 1H, Gly-NH), 8.43 (d,  $J = 6.3$  Hz, 1H, Phe-NH), 7.51 – 7.37 (m, 5H, Phe-Ar), 4.22 (app. qd,  $J = 17.9, 4.7$  Hz, 2H, Gly-(C2)–H<sub>2</sub>), 3.21 (obs. ABX, 1H, Phe-(C3)–H), 3.07 (ABX, 1H,  $J = 9.1, 12.6$  Hz, Phe-(C3')–H), 2.04 (s, 3H, (CO)CH<sub>3</sub>).

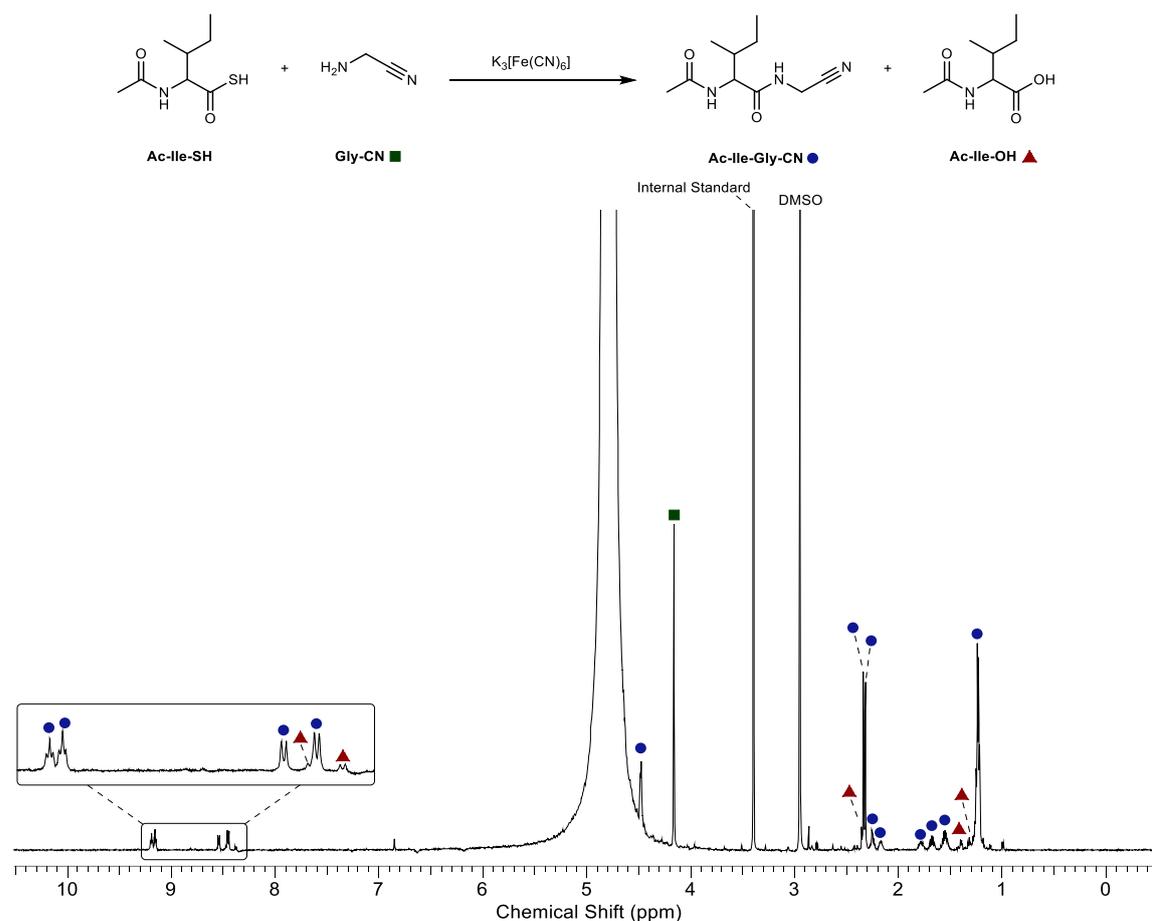
*N*-Acetylphenylalanine **Ac-Phe-OH** (▲):  $\delta$  7.51 – 7.37 (m, 5H, Ar), 2.01 (s, 3H, (CO)-CH<sub>3</sub>).

Glycine nitrile **Gly-CN** (■):  $\delta$  3.84 (s, 2H, (C2)–H<sub>2</sub>).



**Experimental figure 47:**  $^1\text{H}$  –  $^{13}\text{C}$  HMBC ( $^1\text{H}$  – 600 MHz [4.75 – 4.10 ppm],  $^{13}\text{C}$  – 151 MHz [100 – 180 ppm], water-suppressed,  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-d}_6$  1:49:50) spectrum to show the reaction of *N*-acetylphenylalanine thioacid (50 mM), glycine nitrile (100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the pH of the solution was adjusted to 9.0 and diluted with  $\text{DMSO-d}_6$  (1 mL) for analysis. The diagnostic  $^3J_{\text{CH}}$  coupling between glycine nitrile proton resonance (4.22 ppm, app. qd,  $J = 20.1, 4.8$  Hz, 2H, Gly2-(C2)–H<sub>2</sub>), and phenylalanine amide carbon resonance (172.5 ppm, Phe1-(C1)) that demonstrates peptide ligation is highlighted with overlaid blue dotted lines.

*N*-Acetylisoleucylglycine nitrile **Ac-Ile-Gly-CN**



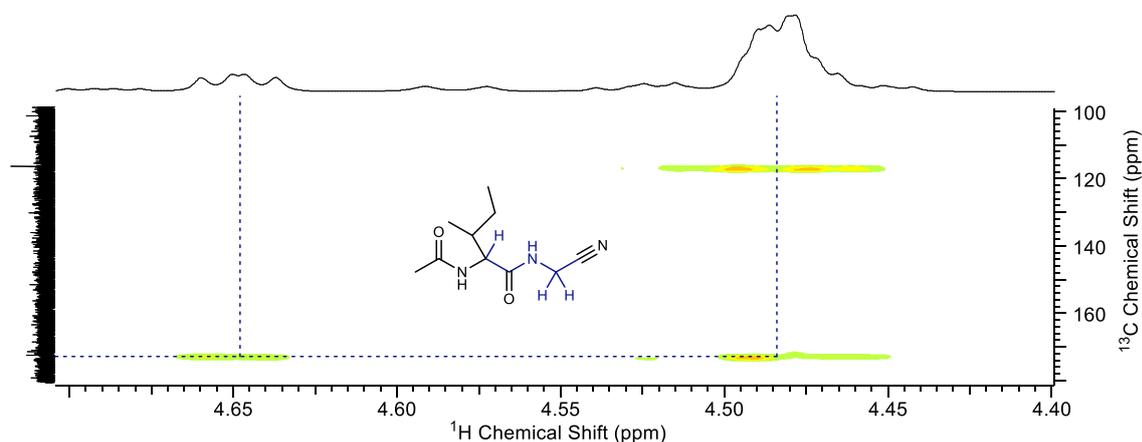
**Experimental figure 48:**  $^1\text{H}$  NMR (600 MHz,  $\text{D}_2\text{O}:\text{H}_2\text{O}:\text{DMSO-}d_6$  1:49:50, -0.50 – 10.50 ppm) spectrum to show the reaction of *N*-acetylisoleucine thioacid (**Ac-Ile-SH**; 50 mM), glycine nitrile (**Gly-CN**, 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the solution was adjusted to pH 9.0 and diluted with  $\text{DMSO-}d_6$  (1 mL) for analysis.

$^1\text{H}$  NMR (600 MHz, 1:49:50  $\text{D}_2\text{O}:\text{H}_2\text{O}:\text{DMSO-}d_6$ , -0.50 – 10.50 ppm):

*N*-Acetylisoleucylglycine nitrile **Ac-Ile-Gly-CN** (•, 54:46 mixture of diastereoisomers a/b): diastereoisomer a (partial assignment):  $\delta$  9.16 (t,  $J = 5.7$  Hz, 1H, Gly-NH), 8.45 (d,  $J = 8.1$  Hz, 1H, Ile-NH), 4.50 – 4.44 (m, 2H, Gly-(C2)-H<sub>2</sub>), 2.32 (s, 3H, (CO)CH<sub>3</sub>), 2.29 – 2.22 (m, 1H, Ile-(C3)-H), 1.72 – 1.63 (m, 1H, Ile-(C4)-H), 1.59 – 1.49 (obs. m, 1H, Ile-(C4')-H), 1.27 – 1.21 (obs. m, 6H, Ile-(C3)-CH<sub>3</sub>, Ile-(C5)-H<sub>3</sub>); diastereoisomer b (partial assignment):  $\delta$  9.19 (t,  $J = 5.6$  Hz, 1H, Gly-NH), 8.54 (d,  $J = 8.1$  Hz, 1H, Ile-NH), 4.50 – 4.44 (m, 2H, Gly-(C2)-H<sub>2</sub>), 2.34 (s, 3H, (CO)CH<sub>3</sub>), 2.20 – 2.13 (m, 1H, Ile-(C3)-H), 1.83 – 1.73 (m, 1H, Ile-(C4)-H), 1.59 – 1.49 (obs. m, 1H, Ile-(C4')-H), 1.27 – 1.21 (obs. m, 6H, Ile-(C3)-CH<sub>3</sub>, Ile-(C5)-H<sub>3</sub>).

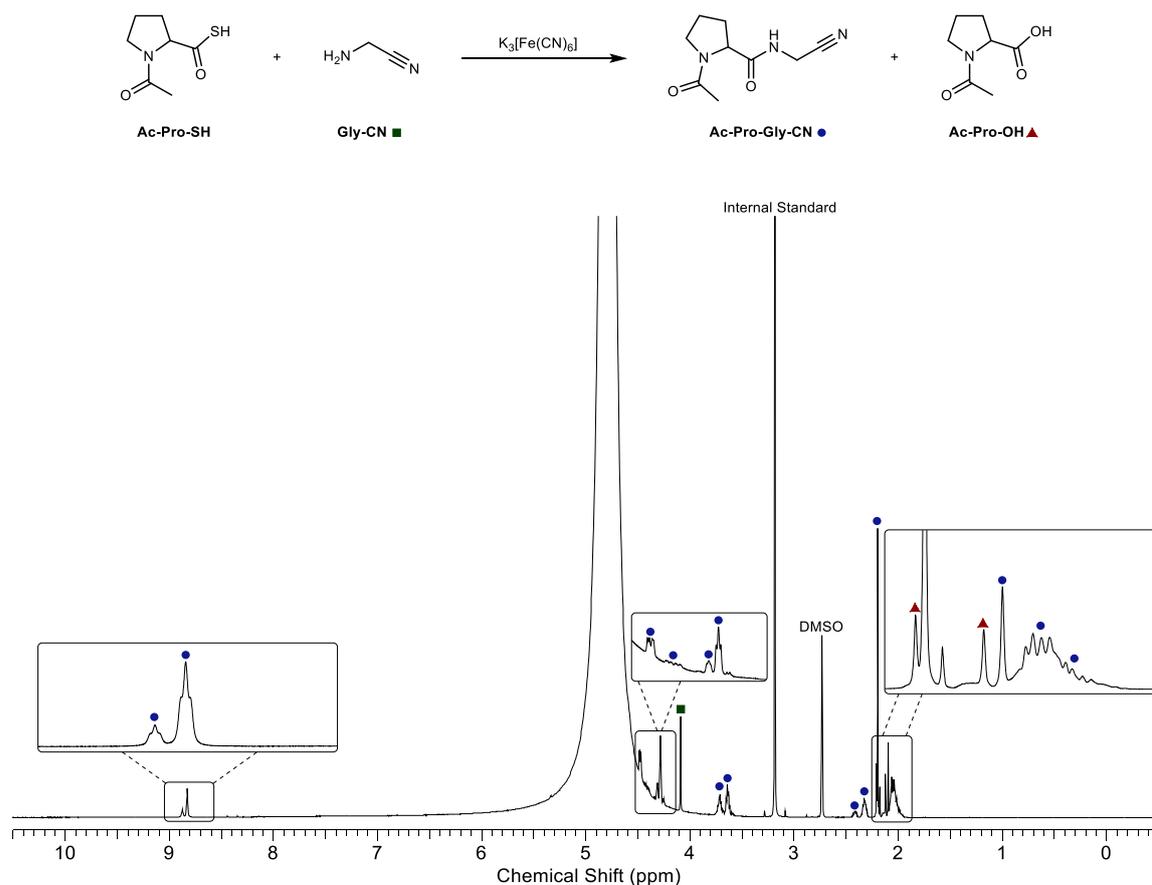
*N*-Acetylisoleucine **Ac-Leu-OH** (▲, 51:49 mixture of diastereoisomers a/b): diastereoisomer a (partial assignment):  $\delta$  8.48 (obs. d, 1H, NH), 2.33 (s, 3H, (CO)CH<sub>3</sub>), 1.34 – 1.29 (obs. m, 3H, (C5)–H<sub>3</sub>); diastereoisomer b (partial assignment):  $\delta$  8.38 (d,  $J$  = 8.4 Hz, 1H, NH), 2.36 (s, 3H, (CO)CH<sub>3</sub>), 1.44 – 1.37 (obs. m, 3H, (C5)–H<sub>3</sub>).

Glycine nitrile **Gly-CN** (■):  $\delta$  4.16 (s, 3H, (C2)–H<sub>2</sub>).



**Experimental figure 49:** <sup>1</sup>H – <sup>13</sup>C HMBC (<sup>1</sup>H – 600 MHz [4.70 – 4.40 ppm], <sup>13</sup>C – 151 MHz [100 – 180 ppm], water-suppressed, D<sub>2</sub>O/H<sub>2</sub>O/DMSO-d<sub>6</sub> 1:49:50) spectrum to show the reaction of *N*-acetylisoleucine thioacid (50 mM), glycine nitrile (100 mM) and K<sub>3</sub>[Fe(CN)<sub>6</sub>] (150 mM) in D<sub>2</sub>O/H<sub>2</sub>O (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the pH of the solution was adjusted to 9.0 and diluted with DMSO-d<sub>6</sub> (1 mL) for analysis. The diagnostic <sup>3</sup>J<sub>CH</sub> coupling between glycine nitrile proton resonance (4.50 – 4.44, m, 2H, Gly2-(C2)–H<sub>2</sub>), and isoleucine amide carbon resonance (173.4 ppm, Ile1-(C1)) that demonstrates peptide ligation is highlighted with overlaid blue dotted lines.

*N*-Acetylprolylglycine nitrile **Ac-Pro-Gly-CN**



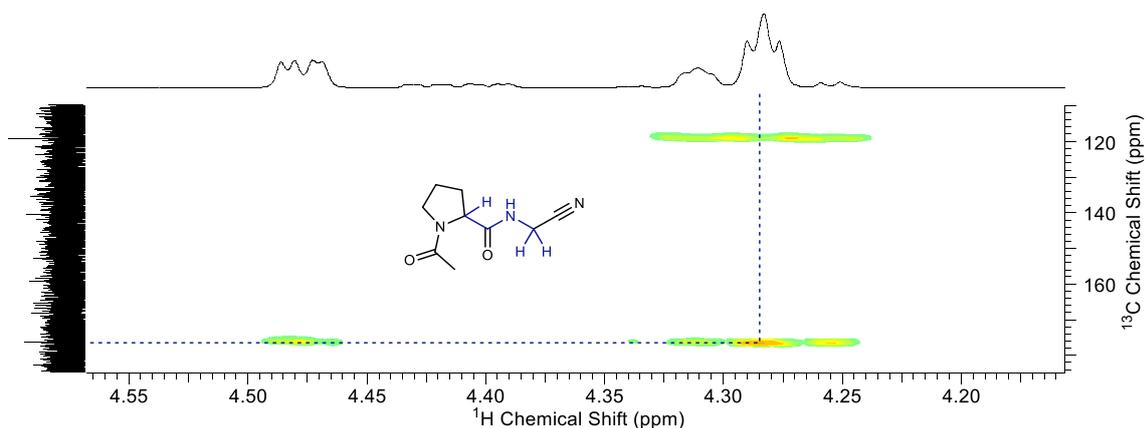
**Experimental figure 50:**  $^1H$  NMR spectrum (700 MHz, 1:49:50  $D_2O:H_2O:DMSO-d_6$ , -0.50 – 10.50 ppm) for the reaction of *N*-acetylproline thioacid (**Ac-Pro-SH**; 50 mM), glycine nitrile (**Gly-CN**; 100 mM) and  $K_3[Fe(CN)_6]$  (150 mM) in  $D_2O/H_2O$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the pH of the solution was adjusted to 9.0 and diluted with  $DMSO-d_6$  (1 mL) for analysis.

$^1H$  NMR (600 MHz, 1:49:50  $D_2O:H_2O:DMSO-d_6$ , -0.50 – 10.50 ppm):

*N*-Acetylprolylglycine nitrile **Ac-Pro-Gly-CN** (•, 78:22 mixture of rotamers): major rotamer  $\delta$  8.83 (br t,  $J = 5.6$  Hz, 1H, Gly-NH), 4.48 (br dd,  $J = 8.4, 5.6$  Hz, 1H, Pro-(C2)-H), 4.29 (m, 2H, Gly-(C2)-H<sub>2</sub>), 3.76 – 3.67 (m, 1H, Pro-(C5)-H), 3.66 – 3.61 (m, 1H, Pro-(C5)-H'), 2.36 – 2.27 (m, 1H, Pro-(C3)-H), 2.20 (s, 3H, (CO)CH<sub>3</sub>), 2.09 – 1.97 (m, 3H, Pro-(C3)-H'); Pro-(C4)-H<sub>2</sub>); minor rotamer  $\delta$  8.88 (br t,  $J = 5.1$  Hz, 1H, Gly-NH), 4.44 – 4.39 (m, 1H, Pro-(C2)-H), 4.31 (m, 2H, Gly-(C2)-H<sub>2</sub>), 3.76 – 3.67 (m, 1H, Pro-(C5)-H), 3.66 – 3.61 (m, 1H, Pro-(C5)-H'), 2.45 – 2.38 (m, 1H, Pro1-(C3)-H), 2.10 (s, 3H, (CO)-CH<sub>3</sub>), 2.09 – 1.97 (m, 3H, Pro-(C3)-H'); (Pro)-(C4)-H<sub>2</sub>).

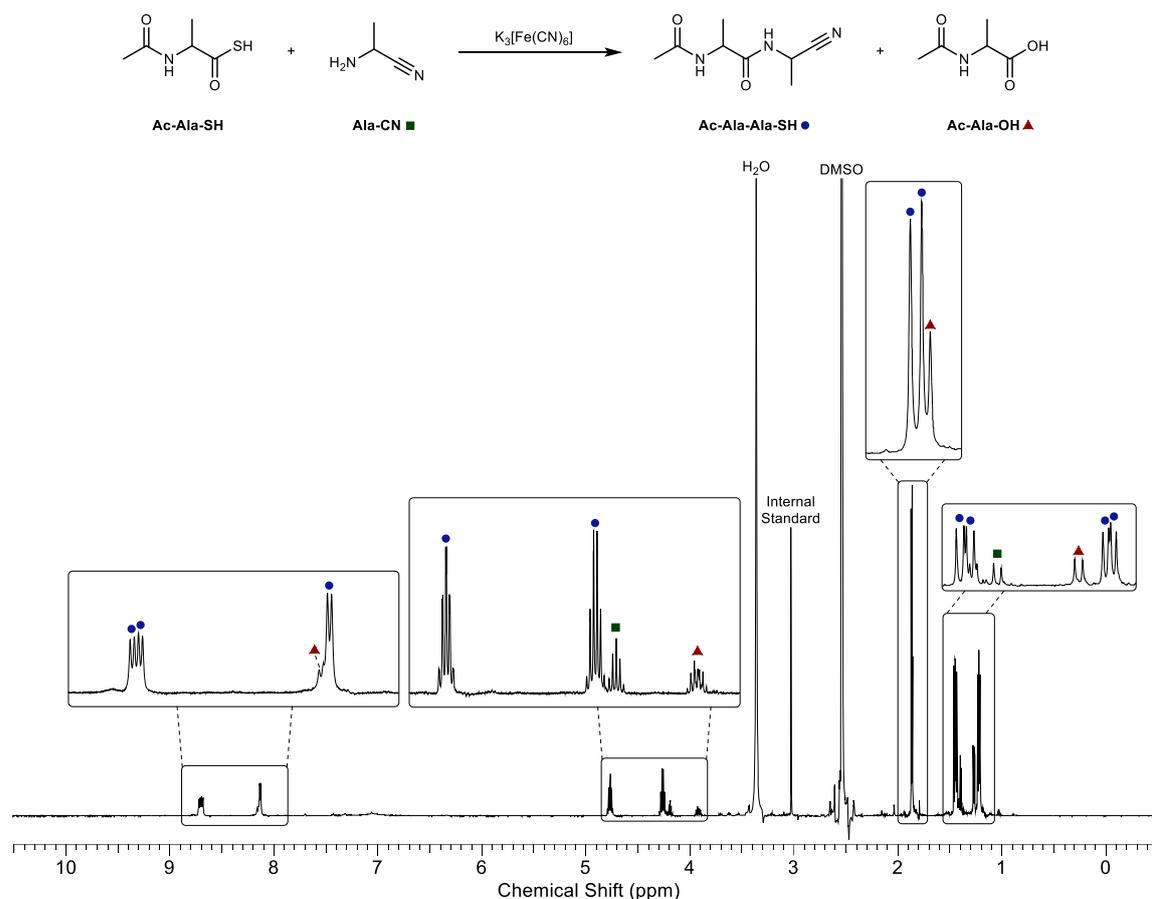
*N*-Acetylproline **Ac-Pro-OH** (▲, 55:45 mixture of rotamers): major rotamer (partial assignment)  $\delta$  2.21 (s, 3H, (CO)CH<sub>3</sub>); minor rotamer (partial assignment)  $\delta$  2.11 (s, 3H, (CO)CH<sub>3</sub>).

Glycine nitrile **Gly-CN** (■):  $\delta$  4.09 (s, 3H, (C2)-H<sub>2</sub>).



**Experimental figure 51:**  $^1\text{H}$  –  $^{13}\text{C}$  HMBC NMR (700 – 176 MHz, water-suppressed, 1:49:50  $\text{D}_2\text{O}:\text{H}_2\text{O}:\text{DMSO-d}_6$ ,  $\delta_{\text{H}}$  4.55 – 4.10 ppm,  $\delta_{\text{C}}$  110 – 180 ppm, major rotamer) spectrum to show the reaction of *N*-acetyl proline thioacid (50 mM), glycine nitrile (100 mM) and ferricyanide (150 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the pH of the solution was adjusted to 9.0 and diluted with  $\text{DMSO-d}_6$  (1 mL) for analysis. The characteristic HMBC cross-coupling between glycine nitrile proton resonance (4.29 ppm, t,  $J = 5.0$  Hz, 2H, Gly2-(C2)–H<sub>2</sub>), and proline amide carbon resonance (176.2 ppm, Pro1-C1) that demonstrates peptide ligation is highlighted with overlaid blue dotted lines.

*N*-Acetylalanylalanine nitrile **Ac-Ala-Ala-CN**



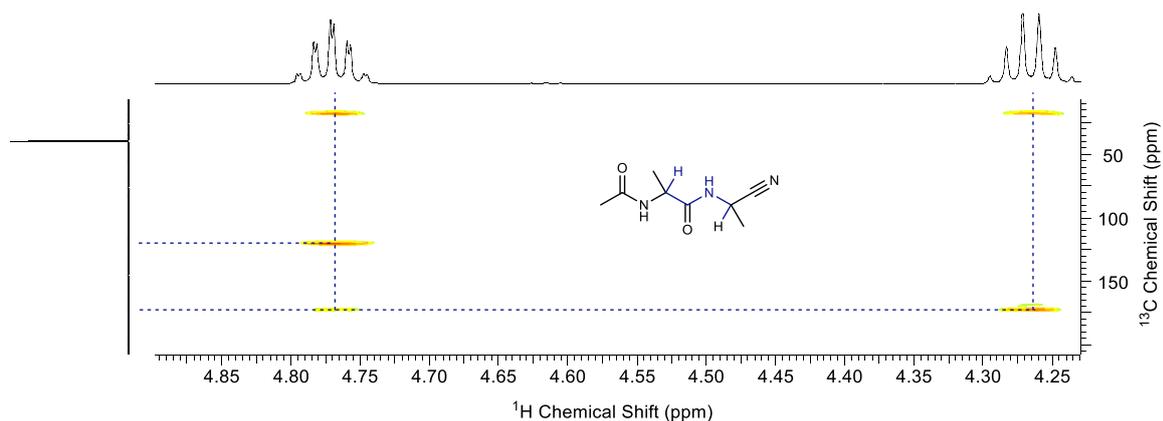
**Experimental figure 52:**  $^1\text{H}$  NMR (600 MHz,  $\text{DMSO-}d_6$ , -0.50 – 10.50 ppm) spectrum to show the reaction of *N*-acetylalanine thioacid (**Ac-Ala-SH**; 50 mM), alanine nitrile (**Ala-CN**, 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature.

$^1\text{H}$  NMR (600 MHz,  $\text{DMSO-}d_6$ ):

*N*-Acetylalanylalanine nitrile **Ac-Ala-Ala-CN** (●, 50:50 mixture of diastereoisomers a/b):  
 diastereoisomer a:  $\delta$  8.72 (d,  $J = 7.2$  Hz, 1H,  $\text{Ala}_2\text{-NH}$ ), 8.14 (d,  $J = 7.2$ , 1H,  $\text{Ala}_1\text{-NH}$ ), 4.77 (qnd,  $J = 7.2$ , 1.3 Hz, 1H,  $\text{Ala}_2\text{-(C2)-H}$ ), 4.27 (qn,  $J = 7.2$  Hz, 1H,  $\text{Ala}_1\text{-(C2)-H}$ ), 1.88 (s, 3H,  $(\text{CO})\text{CH}_3$ ), 1.45 (d,  $J = 7.2$  Hz, 3H,  $\text{Ala}_2\text{-(C3)-H}_3$ ), 1.23 (d,  $J = 7.2$  Hz, 3H,  $\text{Ala}_1\text{-(C3)-H}_3$ ); diastereoisomer b:  $\delta$  8.69 (d,  $J = 7.2$  Hz, 1H,  $\text{Ala}_2\text{-NH}$ ), 8.14 (d,  $J = 7.2$ , 1H,  $\text{Ala}_1\text{-NH}$ ), 4.77 (qnd,  $J = 7.2$ , 1.3 Hz, 1H,  $\text{Ala}_2\text{-(C2)-H}$ ), 4.27 (qn,  $J = 7.2$  Hz, 1H,  $\text{Ala}_1\text{-(C2)-H}$ ), 1.87 (s, 3H,  $(\text{CO})\text{CH}_3$ ), 1.46 (d,  $J = 7.2$  Hz, 3H,  $\text{Ala}_2\text{-(C3)-H}_3$ ), 1.22 (d,  $J = 7.2$  Hz, 3H,  $\text{Ala}_1\text{-(C3)-H}_3$ ).

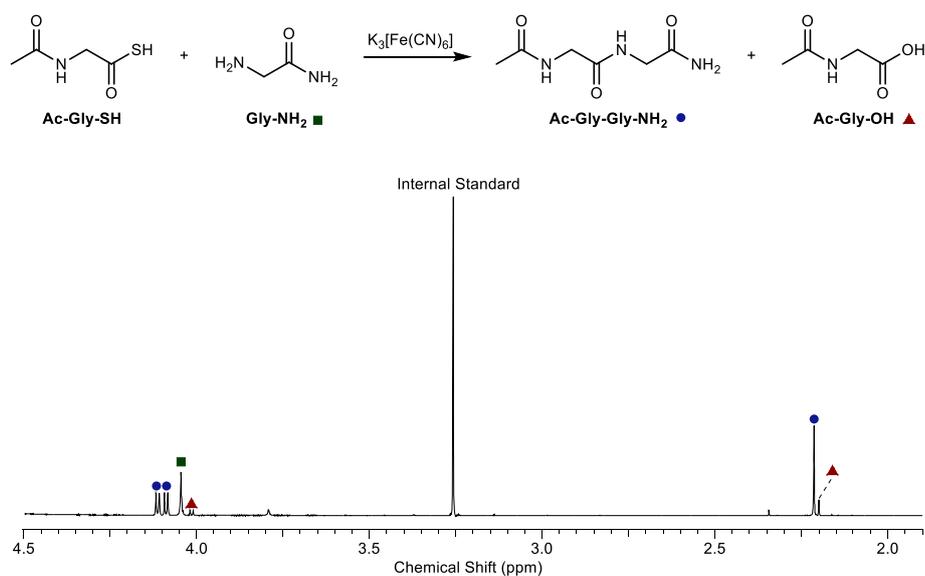
*N*-Acetylalanine **Ac-Ala-OH** (▲):  $\delta$  8.16 (obs. d,  $J = 7.8$  Hz, 1H, NH), 3.95 – 3.89 (m, 1H,  $(\text{C2})\text{-H}$ ), 1.86 (s, 3H,  $(\text{CO})\text{CH}_3$ ), 1.27 (d,  $J = 7.8$  Hz, 3H,  $(\text{C3})\text{-H}_3$ ).

Alanine nitrile **Ala-CN** (■):  $\delta$  4.20 (qn,  $J = 7.3$  Hz, 1H,  $(\text{C2})\text{-H}$ ), 1.37 (d,  $J = 7.3$  Hz, 3H,  $(\text{C3})\text{-H}_3$ ).



**Experimental figure 53:**  $^1\text{H}$  –  $^{13}\text{C}$  HMBC ( $^1\text{H}$  – 600 MHz [4.25 – 4.90 ppm],  $^{13}\text{C}$  – 151 MHz [20 – 200 ppm],  $\text{DMSO-}d_6$ ) spectrum to show the reaction of *N*-acetylcysteine thioacid (**Ac-Ala-SH**; 50 mM), alanine nitrile (**Ala-CN**; 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. The characteristic  $^3J_{\text{CH}}$  coupling between alanine nitrile proton resonance (4.77, qnd,  $J = 7.2, 1.3$  Hz, 1H, (Ala2)-(C2)-H), and alanine amide carbon resonance (174.4 ppm, Ala1-(C1)) that demonstrates peptide ligation is highlighted with overlaid blue dotted lines.

*N*-Acetylglycylglycinamide **Ac-Gly-Gly-NH<sub>2</sub>**



**Experimental figure 54:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O/H<sub>2</sub>O 2:98, 1.90 – 4.50 ppm) spectrum to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), glycinamide (**Gly-NH<sub>2</sub>**, 100 mM) and K<sub>3</sub>[Fe(CN)<sub>6</sub>] (150 mM) in H<sub>2</sub>O/D<sub>2</sub>O (98:2, 1 mL) at pH 9.0 and room temperature.

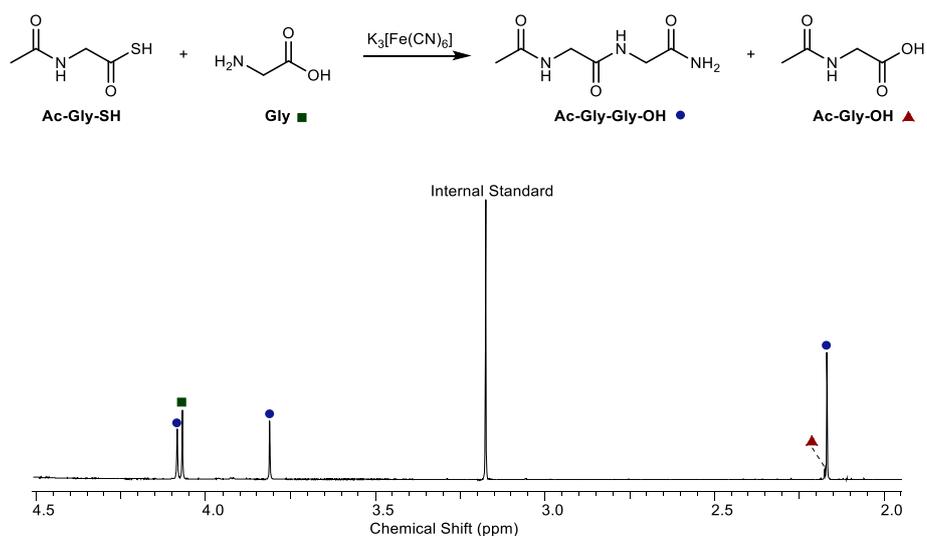
<sup>1</sup>H NMR (600 MHz, 2:98, D<sub>2</sub>O:H<sub>2</sub>O, pH 9.0):

*N*-Acetylglycylglycinamide **Ac-Gly-Gly-NH<sub>2</sub>** (•): δ 4.12 (app. d, *J* = 6.7 Hz, 2H, Gly<sub>2</sub>-(C2)-H<sub>2</sub>), 4.09 (app. d, *J* = 6.4 Hz, 2H, Gly<sub>1</sub>-(C2)-H<sub>2</sub>), 2.21 (s, 3H, (CO)CH<sub>3</sub>).

*N*-Acetylglycine **Ac-Gly-OH** (▲): δ 4.02 (d, *J* = 6.4 Hz, 2H, (C2)-H<sub>2</sub>), 2.20 (s, 3H, (CO)CH<sub>3</sub>).

Glycinamide **Gly-NH<sub>2</sub>** (■): δ 4.04 (s, 2H, (C2)-H<sub>2</sub>).

### *N*-Acetylglycylglycine **Ac-Gly-Gly-OH**



**Experimental figure 55:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O/H<sub>2</sub>O 2:98, 1.90 – 4.50 ppm) spectrum to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), glycine (**Gly**, 100 mM) and K<sub>3</sub>[Fe(CN)<sub>6</sub>] (150 mM) in H<sub>2</sub>O/D<sub>2</sub>O (98:2, 1 mL) at pH 9.0 and room temperature.

<sup>1</sup>H NMR (600 MHz, 2:98 D<sub>2</sub>O:H<sub>2</sub>O, pH 9.0):

*N*-Acetylglycylglycine **Ac-Gly-Gly-OH** (●): δ 4.08 (s, 2H, Gly<sub>2</sub>-(C2)-H<sub>2</sub>), 3.81 (s, 2H, Gly<sub>1</sub>-(C2)-H<sub>2</sub>), 2.16 (s, 3H, (CO)CH<sub>3</sub>).

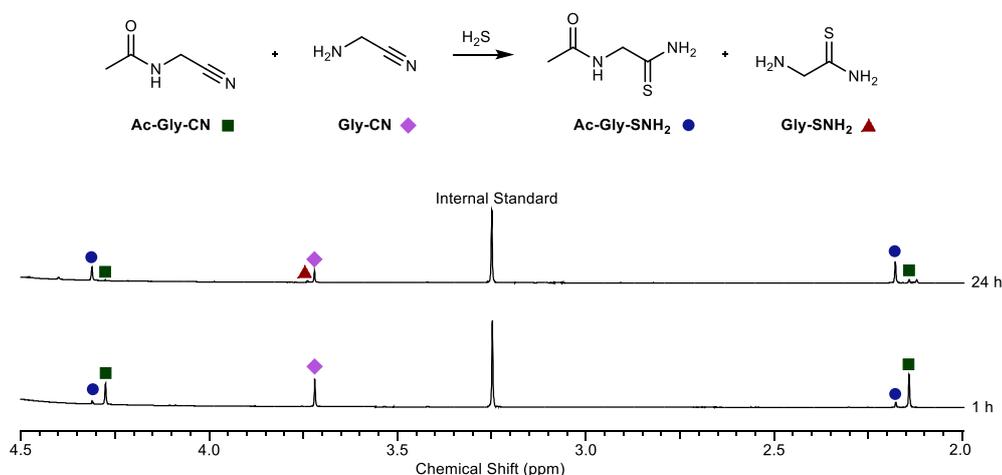
*N*-Acetylglycine **Ac-Gly-OH** (▲): δ 3.98 (br. app. d, *J* = 4.8 Hz, 2H, (C2)-H<sub>2</sub>), 2.16 (s, 3H, (CO)CH<sub>3</sub>).

Glycine **Gly-OH** (■): δ 4.07 (s, 2H, (C2)-H<sub>2</sub>).

## 9. 2. 6. Selectivity studies

### *Chemoselective thiolysis of N-acetylglycine nitrile in the presence of glycine nitrile*

*N*-Acetylglycine nitrile **Ac-Gly-CN** (0.10 mmol), glycine nitrile hydrochloride **Gly-CN**·HCl (9.3 mg, 0.10 mmol) and NaSH·xH<sub>2</sub>O (56 mg, 0.50 mmol) were dissolved in a degassed H<sub>2</sub>O/D<sub>2</sub>O (98:2, 2 mL), and the solution was adjusted to pH 9.0 with NaOH. The solution was stirred at room temperature for 24 h. NMR spectra were periodically acquired. The conversion of *N*-acetylglycine nitrile **Ac-Gly-CN** (■) and glycine nitrile **Gly-CN** (◆) into *N*-acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>** (●, 91%) and glycine thioamide **Gly-SNH<sub>2</sub>** (▲, 7%) was observed.



**Experimental figure 56:** <sup>1</sup>H NMR (600 MHz, H<sub>2</sub>O/D<sub>2</sub>O 98:2, 2.00 – 4.50 ppm) spectra to show the reaction *N*-acetylglycine nitrile (**Ac-Gly-CN**; 50 mM), glycine nitrile (**Gly-CN**; 50 mM) and H<sub>2</sub>S (250 mM) at pH 9.0 and room temperature after 1 h and 24 h.

<sup>1</sup>H NMR (600 MHz H<sub>2</sub>O/D<sub>2</sub>O 98:2, pH 9.0, 4.50 – 2.00 ppm):

*N*-Acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>** (●): δ 4.31 (s, 2H, (C2)–H<sub>2</sub>), 2.17 (s, 3H, (CO)CH<sub>3</sub>).

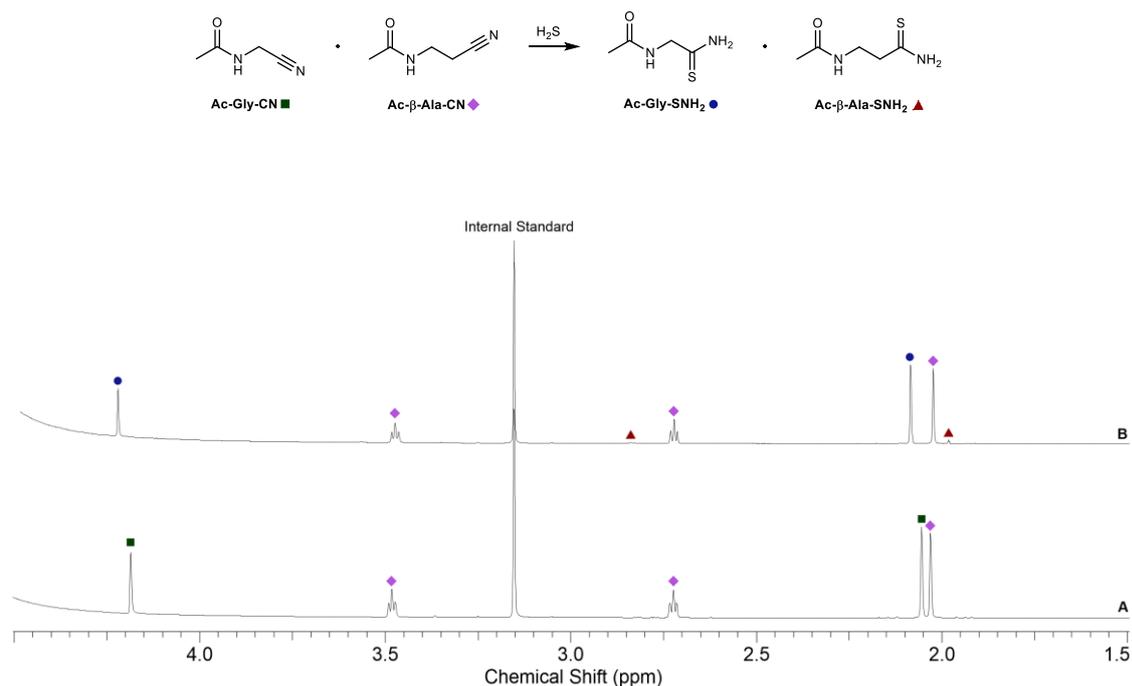
Glycine thioamide **Gly-SNH<sub>2</sub>** (▲): δ 3.73 (s, 2H, (C2)–H<sub>2</sub>).

Glycine nitrile **Gly-CN** (◆): δ 3.72 (s, 2H, (C2)–H<sub>2</sub>).

*N*-Acetylglycine nitrile **Ac-Gly-CN** (■): δ 4.27 (s, 2H, (C2)–H<sub>2</sub>), 2.14 (s, 3H, (CO)CH<sub>3</sub>).

### Chemoselective thiolysis of *N*-acetylglycine nitrile in the presence of *N*-acetyl- $\beta$ -alanine nitrile

*N*-Acetylglycine nitrile **Ac-Gly-CN** (0.10 mmol), *N*-acetyl- $\beta$ -alanine nitrile **Ac- $\beta$ -Ala-CN** (0.10 mmol) and NaSH·xH<sub>2</sub>O (56 mg, 0.50 mmol) were dissolved in a degassed H<sub>2</sub>O/D<sub>2</sub>O (98:2, 2 mL), and the solution was adjusted to pH 9.0 with NaOH. The solution was stirred at room temperature for 24 h. NMR spectra were periodically acquired. The conversion of *N*-acetylglycine nitrile **Ac-Gly-CN** (■) and *N*-acetyl- $\beta$ -alanine nitrile **Ac- $\beta$ -Ala-CN** (◆) into *N*-acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>** (●, > 99%) and *N*-acetyl- $\beta$ -alanine thioamide **Ac- $\beta$ -Ala-SNH<sub>2</sub>** (▲, < 1%) was observed.



**Experimental figure 57:** <sup>1</sup>H NMR (700 MHz, H<sub>2</sub>O/D<sub>2</sub>O 98:2, 1.50 – 4.50 ppm) spectra to show the reaction of *N*-acetylglycine nitrile (**Ac-Gly-CN**; 50 mM), *N*-acetyl- $\beta$ -alanine nitrile (**Ac- $\beta$ -Ala-CN**; 50 mM) and H<sub>2</sub>S (250 mM) at pH 9.0 and room temperature after **A**: 5 min; **B**: 24 h.

<sup>1</sup>H NMR (700 MHz H<sub>2</sub>O/D<sub>2</sub>O 98:2, pH 9.0, partial assignment):

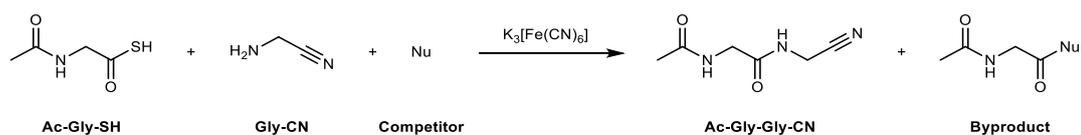
*N*-Acetylglycine thioamide **Ac-Gly-SNH<sub>2</sub>** (●):  $\delta$  4.22 (s, 2H, (C2)–H<sub>2</sub>), 2.08 (s, 3H, (CO)CH<sub>3</sub>).

*N*-Acetyl- $\beta$ -alanine thioamide **Ac- $\beta$ -Ala-SNH<sub>2</sub>** (▲):  $\delta$  2.83 (obs. t,  $J$  = 6.4 Hz, 2H, (C3)–H<sub>2</sub>), 1.98 (s, 3H, (CO)–CH<sub>3</sub>).

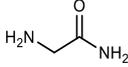
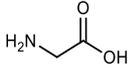
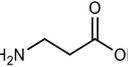
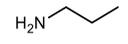
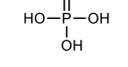
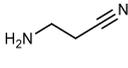
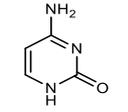
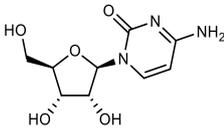
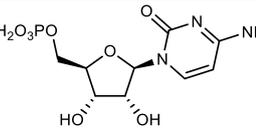
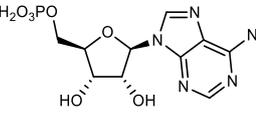
*N*-Acetyl- $\beta$ -alanine nitrile **Ac- $\beta$ -Ala-CN** (◆):  $\delta$  3.48 (t,  $J$  = 6.7 Hz, 2H, (C2)–H<sub>2</sub>), 2.72 (t,  $J$  = 6.7 Hz, 2H, (C3)–H<sub>2</sub>), 2.03 (s, 3H, (CO)CH<sub>3</sub>).

*N*-Acetylglycine nitrile **Ac-Gly-CN** (■):  $\delta$  4.18 (s, 2H, (C2)–H<sub>2</sub>), 2.06 (s, 3H, (CO)–CH<sub>3</sub>).

### Oxidative couplings in the presence of prebiotic competitors



*N*-Acetylglycine thioacid **Ac-Gly-SH** (50 mM, 0.1 mmol) and methylsulfonylmethane (50 mM, 0.1 mmol) were dissolved in H<sub>2</sub>O/D<sub>2</sub>O (98:2, 1 mL). Glycine nitrile **Gly-CN** (100 mM, 0.2 mmol) and the specified competitor (100 mM, 0.2 mmol) were added, and the solution was adjusted to the specified pH value (5.0, 7.0 or 9.0). The volume of the reaction was adjusted to 2 mL with H<sub>2</sub>O/D<sub>2</sub>O (98:2). Potassium hexacyanoferrate(III) (99 mg, 0.3 mmol) was added and the reaction was stirred. After 10 min, the solution was adjusted to pH 9.0, centrifuged, and the supernatant was analysed by NMR spectroscopy. The presence of **Ac-Gly-Gly-CN** was confirmed by <sup>1</sup>H – <sup>13</sup>C HMBC NMR and signal amplification by spiking, or comparison with pure synthetic standards. Yields were determined by integration using methylsulfonylmethane as an internal NMR standard.

Experimental figure 8: oxidative ligations in the presence of competitors			
Competitor	pH	Yield Ac-Gly-Gly-CN (%)	Yield byproduct (%)
	5.0	66	27
	7.0	59	39
	9.0	14	86
	5.0	82	9
	7.0	81	17
	9.0	79	19
NH <sub>3</sub>	5.0	75	3
	7.0	95	3
	9.0	77	22
	5.0	93	5
	7.0	89	8
	9.0	90	9
	5.0	90	-
	7.0	98	-
	9.0	91	5
	5.0	77	-
	7.0	85	< 1
	9.0	69	19
	5.0	52	21
	7.0	59	29
	9.0	51	41
	5.0	64	-
	7.0	89	-
	9.0	92	-
	5.0	65	-
	7.0	75	-
	9.0	89	-
	5.0	73	-
	7.0	83	-
	9.0	90	-
	5.0	72	-
	7.0	91	-
	9.0	84	-

### Competition between glycine nitrile and glycinamide

$^1\text{H}$  NMR (600 MHz, 2%  $\text{D}_2\text{O}/\text{H}_2\text{O}$  pH 9.0, 4.50 – 2.00 ppm):

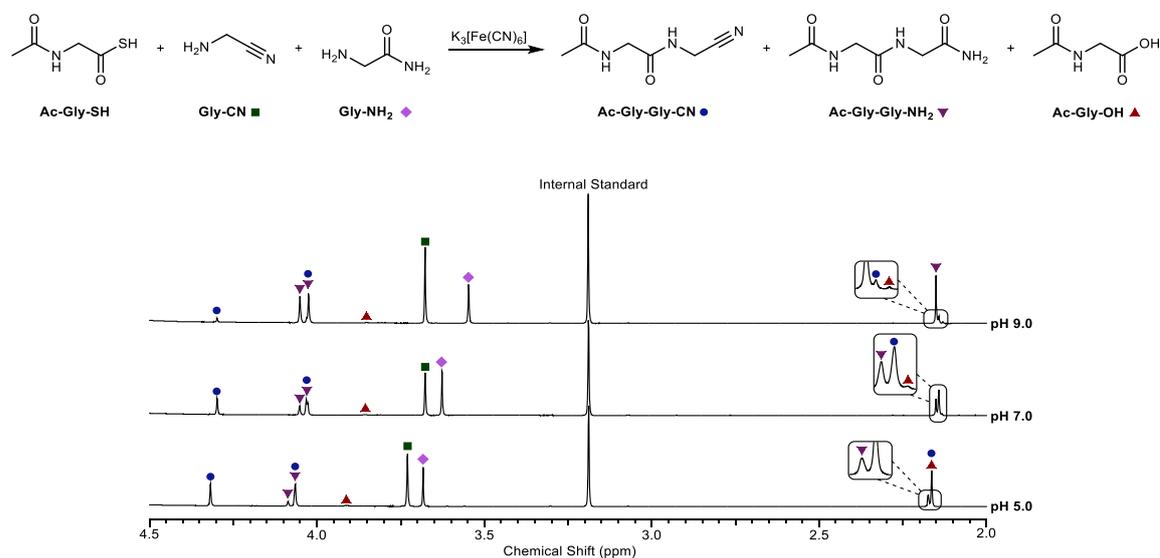
*N*-Acetylglycyl glycine nitrile **Ac-Gly-Gly-CN** (●)  $\delta$  4.32 (s, 2H, (GlyCN2)-(C2)-H<sub>2</sub>), 4.06 (s, 2H, (Gly1)-(C2)-H<sub>2</sub>), 2.16 (s, 3H, (CO)-CH<sub>3</sub>);

*N*-Acetylglycyl glycinamide **Ac-Gly-Gly-NH<sub>2</sub>** (▼)  $\delta$  4.09 (s, 2H, (GlyNH<sub>2</sub>2)-(C2)-H<sub>2</sub>), 4.06 (s, 2H, (Gly1)-(C2)-H<sub>2</sub>), 2.17 (s, 3H, (CO)-CH<sub>3</sub>);

*N*-Acetylglycine **Ac-Gly-OH** (▲)  $\delta$  3.92 (s, 2H, (C2)-H<sub>2</sub>), 2.16 (s, 3H, (CO)-CH<sub>3</sub>);

Glycine nitrile **Gly-CN** (■)  $\delta$  3.73 (s, 2H, (C2)-H<sub>2</sub>);

Glycinamide **Gly-NH<sub>2</sub>** (◆)  $\delta$  3.68 (s, 2H, (C2)-H<sub>2</sub>).



**Experimental figure 58:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 2.00 – 4.50 ppm) spectra to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), glycine nitrile (**Gly-CN**; 100 mM), glycinamide (**Gly-NH<sub>2</sub>**; 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and the specified pH (5.0, 7.0 or 9.0) after 10 min.

### Competition between glycine nitrile and glycine

$^1\text{H}$  NMR (600 MHz, 2%  $\text{D}_2\text{O}:\text{H}_2\text{O}$  pH 9.0, 4.50 – 2.00 ppm):

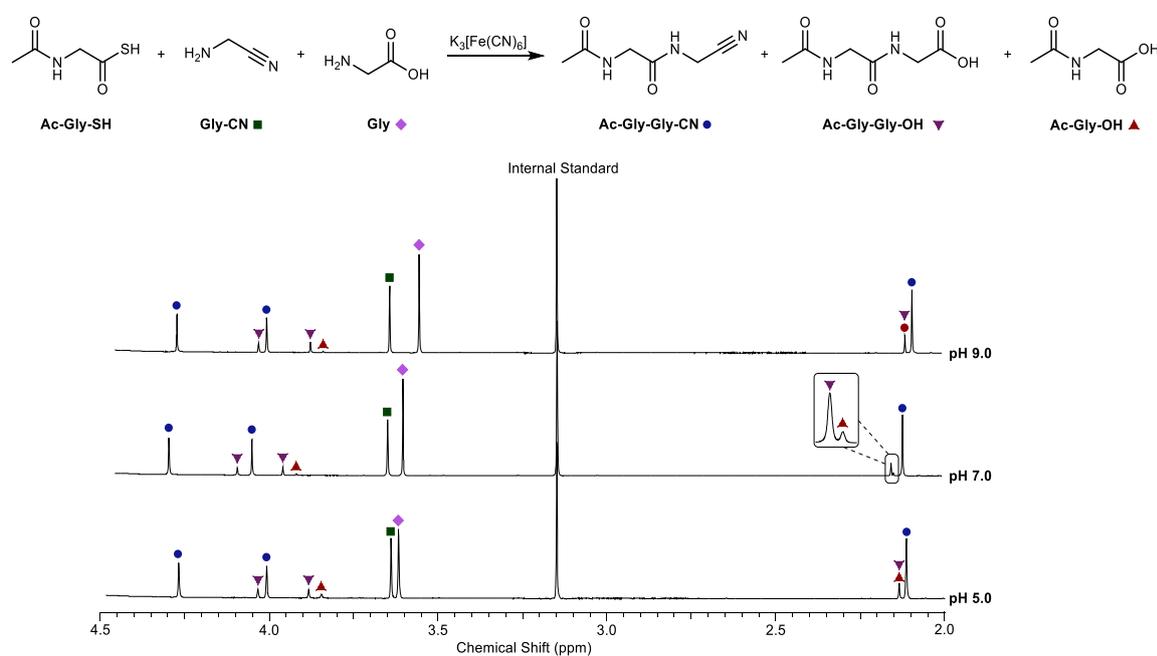
Glycine nitrile **Gly-CN** (■):  $\delta$  3.64 (s, 2H, (C2)-H<sub>2</sub>);

Glycine **Gly-OH** (◆):  $\delta$  3.61 (s, 2H, (C2)-H<sub>2</sub>);

*N*-Acetylglycyl glycine nitrile **Ac-Gly-Gly-CN** (●):  $\delta$  4.27 (s, 2H, (GlyCN2)-(C2)-H<sub>2</sub>), 4.01 (s, 2H, (Gly1)-(C2)-H<sub>2</sub>), 2.11 (s, 3H, (CO)-CH<sub>3</sub>);

; *N*-Acetylglycyl glycine **Ac-Gly-Gly** (▼):  $\delta$  4.04 (s, 2H, (Gly2)-(C2)-H<sub>2</sub>), 3.88 (s, 2H, (Gly1)-(C2)-H<sub>2</sub>), 2.14 (s, 3H, (CO)-CH<sub>3</sub>);

*N*-Acetylglycine **Ac-Gly-OH** (▲):  $\delta$  3.84 (s, 2H, (C2)-H<sub>2</sub>), 2.14 (s, 3H, (CO)-CH<sub>3</sub>);



**Experimental figure 59:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 2.00 – 4.50 ppm) spectra to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), glycine nitrile (**Gly-CN**; 100 mM), glycine (**Gly-OH**, 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and the specified pH (5.0, 7.0 or 9.0) after 10 min.

### Competition between glycine nitrile and ammonia

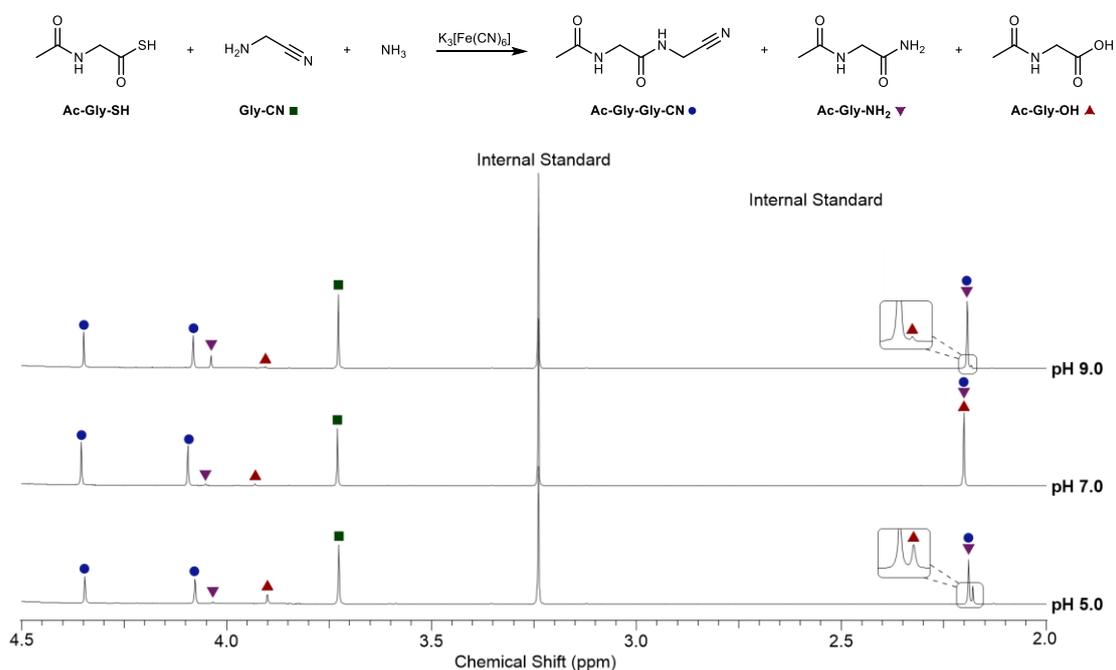
$^1\text{H}$  NMR (600 MHz, 2%  $\text{D}_2\text{O}:\text{H}_2\text{O}$  pH 9.0, 4.50 – 2.00 ppm):

*N*-Acetylglycyl glycine nitrile **Ac-Gly-Gly-CN** (●):  $\delta$  4.34 (s, 2H, (GlyCN2)-(C2)-H<sub>2</sub>), 4.07 (s, 2H, (Gly1)-(C2)-H<sub>2</sub>), 2.18 (s, 3H, (CO)-CH<sub>3</sub>);

*N*-Acetylglycinamide **Ac-Gly-NH<sub>2</sub>** (▼):  $\delta$  4.04 (s, 2H, (C2)-H<sub>2</sub>), 2.18 (s, 3H, (CO)-CH<sub>3</sub>);

*N*-Acetylglycine **Ac-Gly-OH** (▲):  $\delta$  3.90 (s, 2H, (C2)-H<sub>2</sub>), 2.17 (s, 3H, (CO)-CH<sub>3</sub>);

**GGly-CN** (■):  $\delta$  3.73 (s, 2H, (C2)-H<sub>2</sub>).



**Experimental figure 60:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 2.00 – 4.50 ppm) spectra to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), glycine nitrile (**Gly-CN**, 100 mM), ammonia (100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and the specified pH (5.0, 7.0 or 9.0) after 10 min.

### Competition between glycine nitrile and $\beta$ -alanine

$^1\text{H}$  NMR (600 MHz, 2%  $\text{D}_2\text{O}/\text{H}_2\text{O}$  pH 9.0, 4.50 – 2.00 ppm):

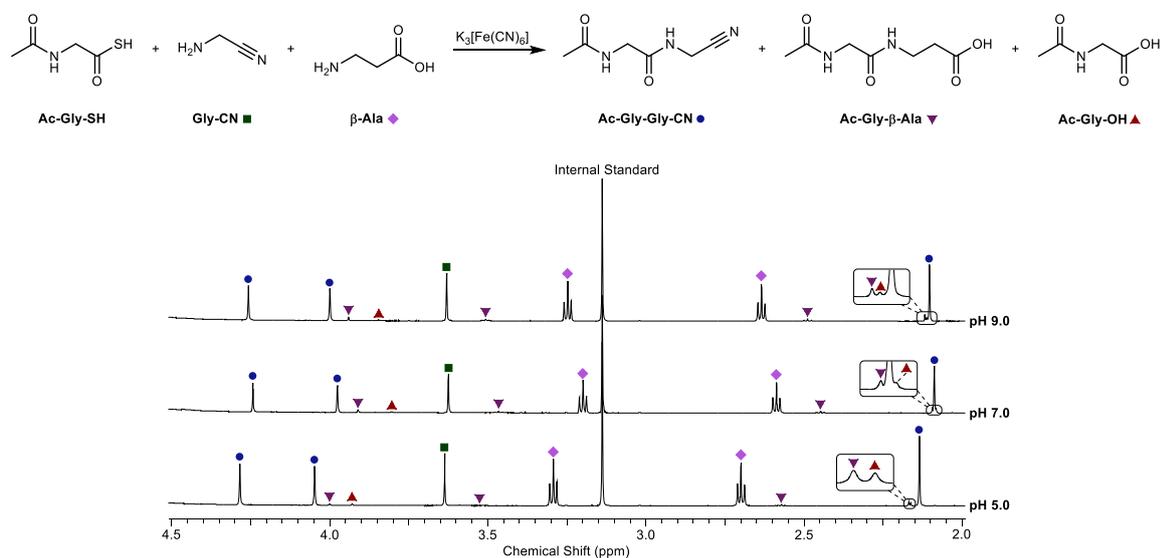
*N*-Acetylglycyl glycine nitrile **Ac-Gly-Gly-CN** (●):  $\delta$  4.28 (s, 2H, (GlyCN2)-(C2)-H<sub>2</sub>), 4.05 (s, 2H, (Gly1)-(C2)-H<sub>2</sub>), 2.14 (s, 3H, (CO)-CH<sub>3</sub>);

*N*-Acetylglycyl  $\beta$ -alanine **Ac-Gly- $\beta$ -Ala** (▼):  $\delta$  4.00 (s, 2H, (Gly1)-(C2)-H<sub>2</sub>), 3.47 (t,  $J = 7.0$  Hz, 2H, ( $\beta$ -Ala2)-(C2)-H<sub>2</sub>), 2.56 (t,  $J = 7.0$  Hz, 2H, ( $\beta$ -Ala2)-(C3)-H<sub>2</sub>), 2.17 (s, 3H, (CO)-CH<sub>3</sub>);

*N*-Acetylglycine **Ac-Gly-OH** (▲):  $\delta$  3.93 (s, 2H, (C2)-H<sub>2</sub>), 2.16 (s, 3H, (CO)-CH<sub>3</sub>);

Glycine nitrile **Gly-CN** (■):  $\delta$  3.64 (s, 2H, (C2)-H<sub>2</sub>);

$\beta$ -Alanine  **$\beta$ -Ala** (◆):  $\delta$  3.74 (t,  $J = 7.0$  Hz, 2H, (C3)-H<sub>2</sub>), 2.69 (t,  $J = 7.0$  Hz, 2H, (C2)-H<sub>2</sub>).



**Experimental figure 61:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 2.00 – 4.50 ppm) spectra to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), glycine nitrile (**Gly-CN**, 100 mM),  $\beta$ -alanine ( **$\beta$ -Ala**, 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and the specified pH (5.0, 7.0 or 9.0) after 10 min.

### Competition between glycine nitrile and *n*-propylamine

$^1\text{H}$  NMR (600 MHz, 2%  $\text{D}_2\text{O}/\text{H}_2\text{O}$  pH 9.0, 4.50 – 0.80 ppm):

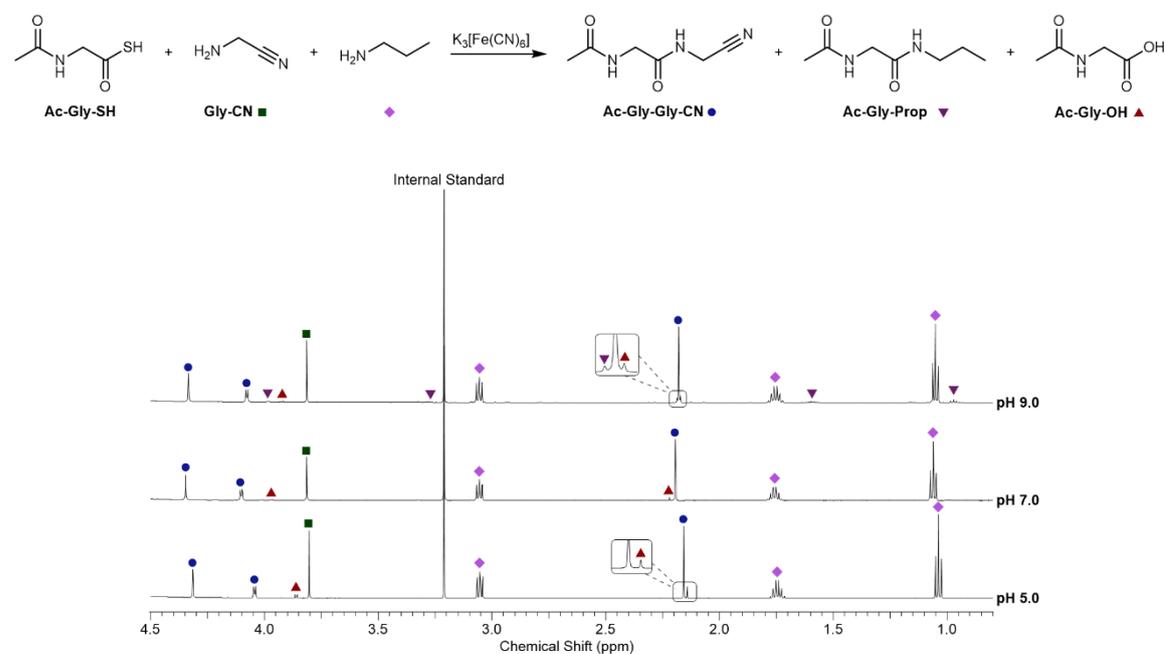
*N*-Acetylglycyl glycine nitrile **Ac-Gly-Gly-CN** (●)  $\delta$  4.31 (s, 2H, (GlyCN2)-(C2)-H<sub>2</sub>), 4.04 (d,  $J$  = 6.1 Hz, 2H, (Gly1)-(C2)-H<sub>2</sub>), 2.15 (s, 3H, (CO)-CH<sub>3</sub>);

*N*-Acetylglycyl *n*-propylamine **Ac-Gly-Prop** (▼)  $\delta$  4.00 (s, 2H, (Gly1)-(C2)-H<sub>2</sub>), 3.28 (t,  $J$  = 6.9 Hz, 2H, CONH-CH<sub>2</sub>), 2.18 (s, 3H, (CO)-CH<sub>3</sub>), 1.61 – 1.54 (m, 2H, CH<sub>2</sub>), 0.98 (t,  $J$  = 7.4 Hz, 3H, CH<sub>3</sub>);

*N*-Acetylglycine **Ac-Gly-OH** (▲)  $\delta$  3.87 (d,  $J$  = 6.3 Hz, 2H, (C2)-H<sub>2</sub>), 2.14 (s, 3H, (CO)-CH<sub>3</sub>);

Glycine nitrile **Gly-CN** (■)  $\delta$  3.60 (s, 2H, (C2)-H<sub>2</sub>);

*n*-Propylamine (◆)  $\delta$  3.05 (t,  $J$  = 6.9 Hz, 2H, (H<sub>2</sub>N)-CH<sub>2</sub>), 1.77 – 1.71 (m, 2H, CH<sub>2</sub>), 1.04 (t,  $J$  = 7.4 Hz, 3H, CH<sub>3</sub>).



**Experimental figure 62:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 0.80 – 4.50 ppm) spectra to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), glycine nitrile (**Gly-CN**, 100 mM), *n*-propylamine (100 mM) and K<sub>3</sub>[Fe(CN)<sub>6</sub>] (150 mM) at room temperature and the specified pH (5.0, 7.0 or 9.0) after 10 min.

### Competition between glycine nitrile and phosphate

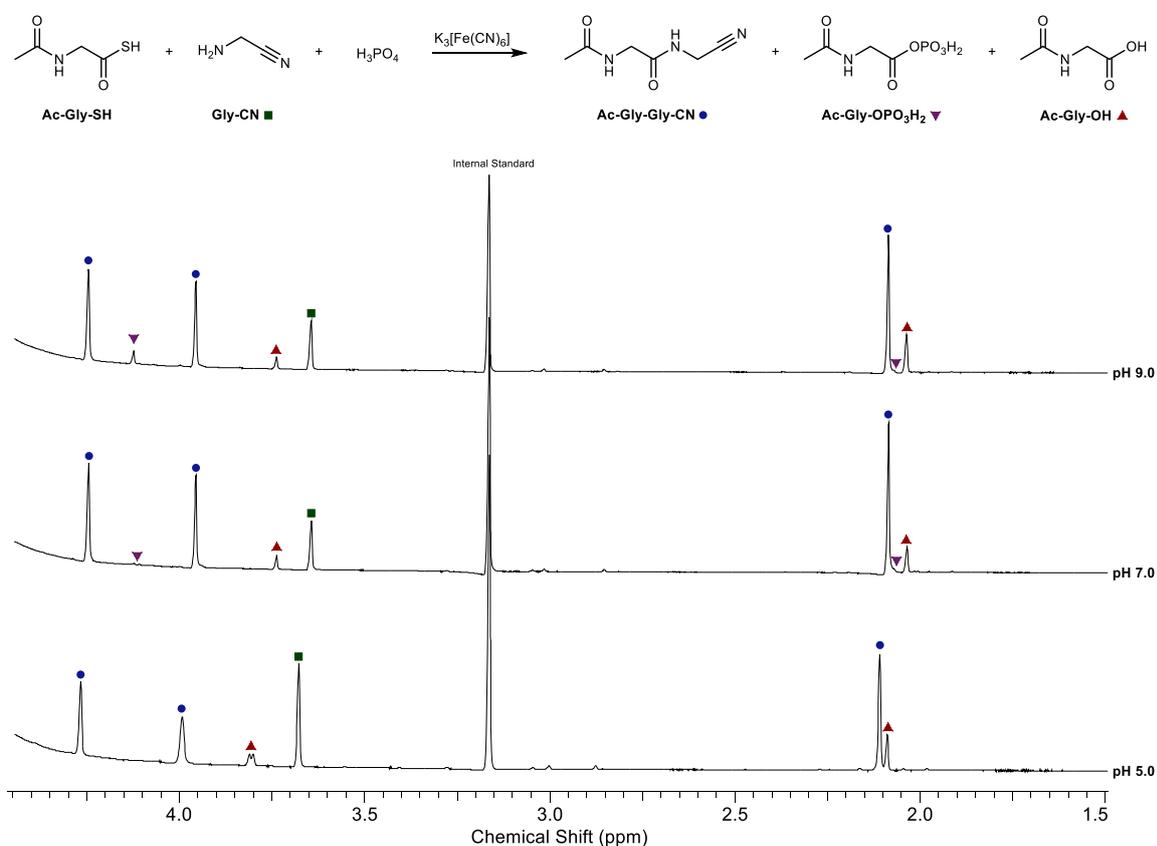
$^1\text{H}$  NMR (600 MHz, 2%  $\text{D}_2\text{O}/\text{H}_2\text{O}$  pH 9.0, 4.40 – 1.50 ppm):

3.69 (s, 2H, (C2)- $\text{H}_2$ ) *N*-Acetylglycyl glycine nitrile **Ac-Gly-Gly-CN** (●)  $\delta$  4.27 (s, 2H, (GlyCN2)-(C2)- $\text{H}_2$ ), 3.99 (br s, 2H, (Gly1)-(C2)- $\text{H}_2$ ), 2.11 (s, 3H, (CO)- $\text{CH}_3$ );

*N*-Acetyl glycyl phosphoric anhydride **Ac-Gly-OPO<sub>3</sub>H<sub>2</sub>** (▼)  $\delta$  4.12 (s, 2H, (C2)- $\text{H}_2$ ), 2.07 (s, 3H, (CO)- $\text{CH}_3$ );

*N*-Acetylglycine **Ac-Gly-OH** (▲)  $\delta$  3.81 (d,  $J = 6.3$  Hz, 2H, (C2)- $\text{H}_2$ ), 2.09 (s, 3H, (CO)- $\text{CH}_3$ );

Glycine nitrile **Gly-CN** (■)  $\delta$  .



**Experimental figure 63:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 1.50 – 4.45 ppm) spectra to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), glycine nitrile (**Gly-CN**, 100 mM), phosphate (100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and the specified pH (5.0, 7.0 or 9.0) after 10 min.

### Competition between glycine nitrile and $\beta$ -alanine nitrile

$^1\text{H}$  NMR (600 MHz, 2%  $\text{D}_2\text{O}/\text{H}_2\text{O}$  pH 9.0, 4.40 – 2.00 ppm):

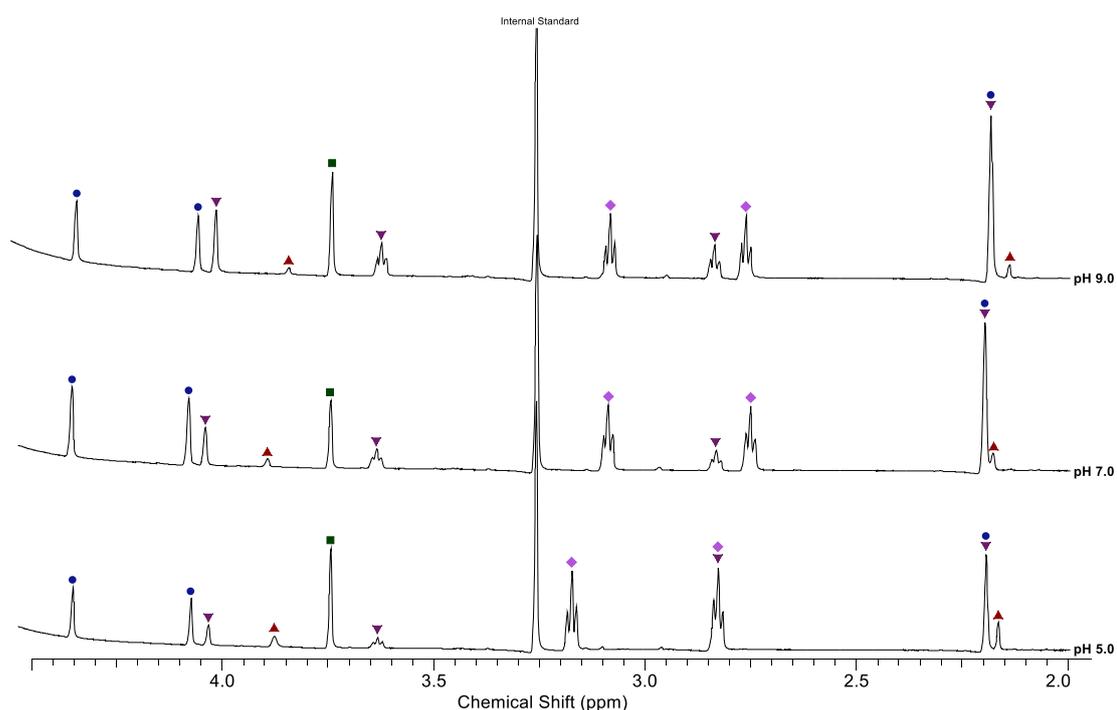
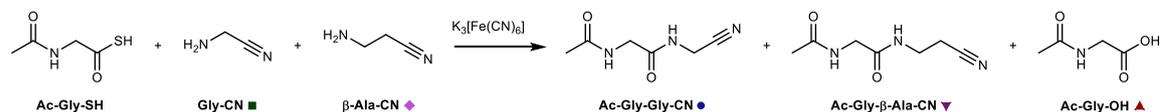
( )  $\delta$  3.67 (s, 2H, (C2)- $\text{H}_2$ );

$\beta$ -Alanine nitrile  **$\beta$ -Ala-CN** (  $\blacklozenge$  ):  $\delta$  3.10 (t,  $J = 6.6$  Hz, 2H, (C2)- $\text{H}_2$ ), 2.76 (t,  $J = 6.6$  Hz, 2H, (C3)- $\text{H}_2$ ).  
*N*-Acetylglycyl glycine nitrile **Ac-Gly-Gly-CN** (  $\bullet$  )  $\delta$  4.28 (s, 2H, (GlyCN2)-(C2)- $\text{H}_2$ ), 4.00 (s, 2H, (Gly1)-(C2)- $\text{H}_2$ ), 2.12 (s, 3H, (CO)- $\text{CH}_3$ );

*N*-Acetylglycyl  $\beta$ -alanine nitrile **Ac-Gly- $\beta$ -Ala-CN** (  $\blacktriangledown$  )  $\delta$  3.98 (s, 2H, (Gly1)-(C2)- $\text{H}_2$ ), 3.56 (t,  $J = 6.3$  Hz, 2H, ( $\beta$ -AlaCN2)-(C3)- $\text{H}_2$ ), 2.76 (t,  $J = 6.3$  Hz, 2H, ( $\beta$ -AlaCN2)-(C2)- $\text{H}_2$ ) 2.12 (s, 3H, (CO)- $\text{CH}_3$ );

*N*-Acetylglycine **Ac-Gly-OH** (  $\blacktriangle$  )  $\delta$  3.81 (d,  $J = 6.3$  Hz, 2H, (C2)- $\text{H}_2$ ), 2.09 (s, 3H, (CO)- $\text{CH}_3$ );

### Glycine nitrile **Gly-CN**



**Experimental figure 64:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 2.00 – 4.50 ppm) spectra to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), glycine nitrile (**Gly-CN**; 100 mM),  $\beta$ -alanine nitrile ( **$\beta$ -Ala-CN**, 100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and the specified pH (5.0, 7.0 or 9.0) after 10 min.

### Competition between glycine nitrile and cytosine

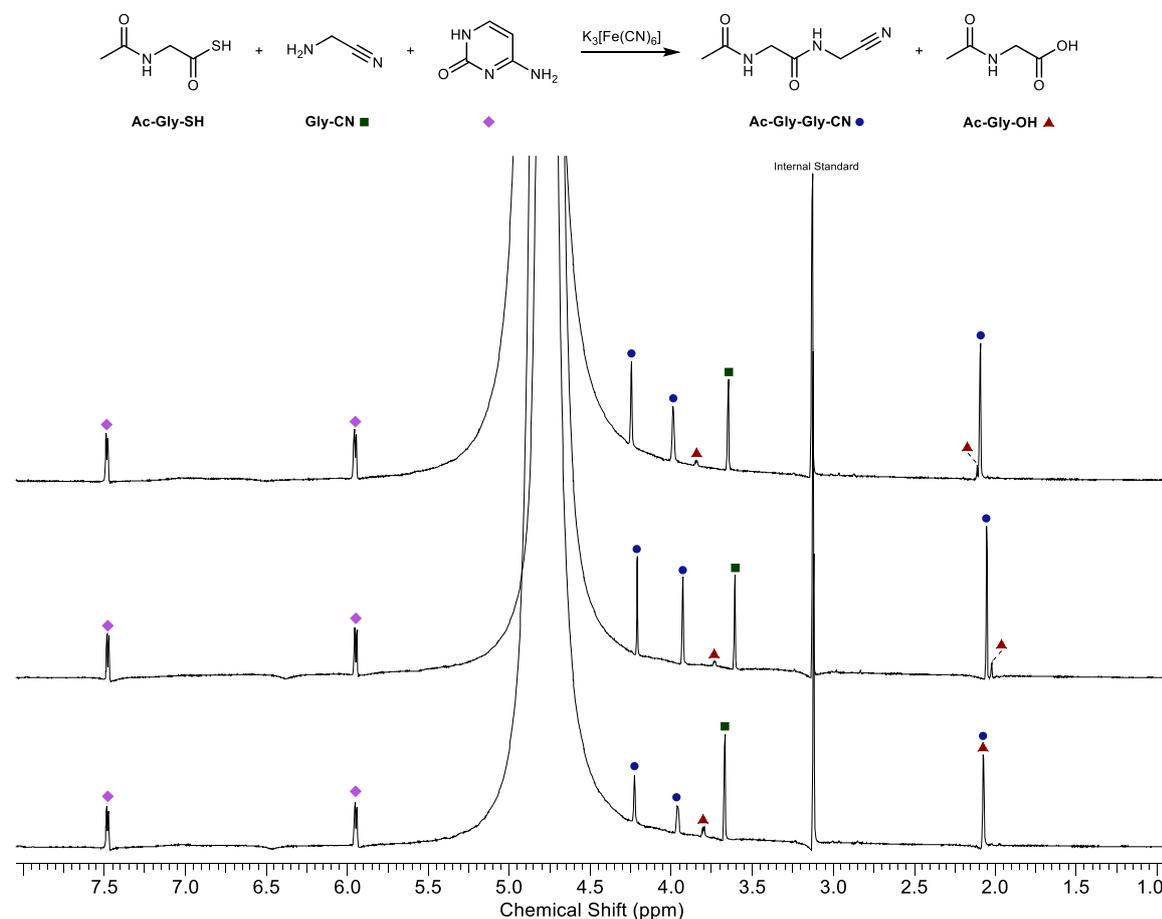
$^1\text{H}$  NMR (600 MHz, 2%  $\text{D}_2\text{O}/\text{H}_2\text{O}$  pH 9.0, 4.40 – 2.00 ppm):

( )  $\delta$  3.67 (s, 2H, (C2)- $\text{H}_2$ ). *N*-Acetylglycyl glycine nitrile **Ac-Gly-Gly-CN** (●)  $\delta$  4.23 (s, 2H, (GlyCN2)-(C2)- $\text{H}_2$ ), 3.96 (s, 2H, (Gly1)-(C2)- $\text{H}_2$ ), 2.07 (s, 3H, (CO)- $\text{CH}_3$ );

*N*-Acetylglycine **Ac-Gly-OH** (▲)  $\delta$  3.80 (d,  $J = 6.3$  Hz, 2H, (C2)- $\text{H}_2$ ), 2.07 (s, 3H, (CO)- $\text{CH}_3$ );

Cytosine (◆):  $\delta$  7.48 (d,  $J = 6.9$  Hz, 1H, (C6)-H), 5.95 (d,  $J = 6.9$  Hz, 1H, (C5)-H);

### Glycine nitrile **Gly-CN**



**Experimental figure 65:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 1.00 – 8.00 ppm) spectra to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), glycine nitrile (**Gly-CN**, 100 mM), cytosine (100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and the specified pH (5.0, 7.0 or 9.0) after 10 min.

### Competition between glycine nitrile and cytidine

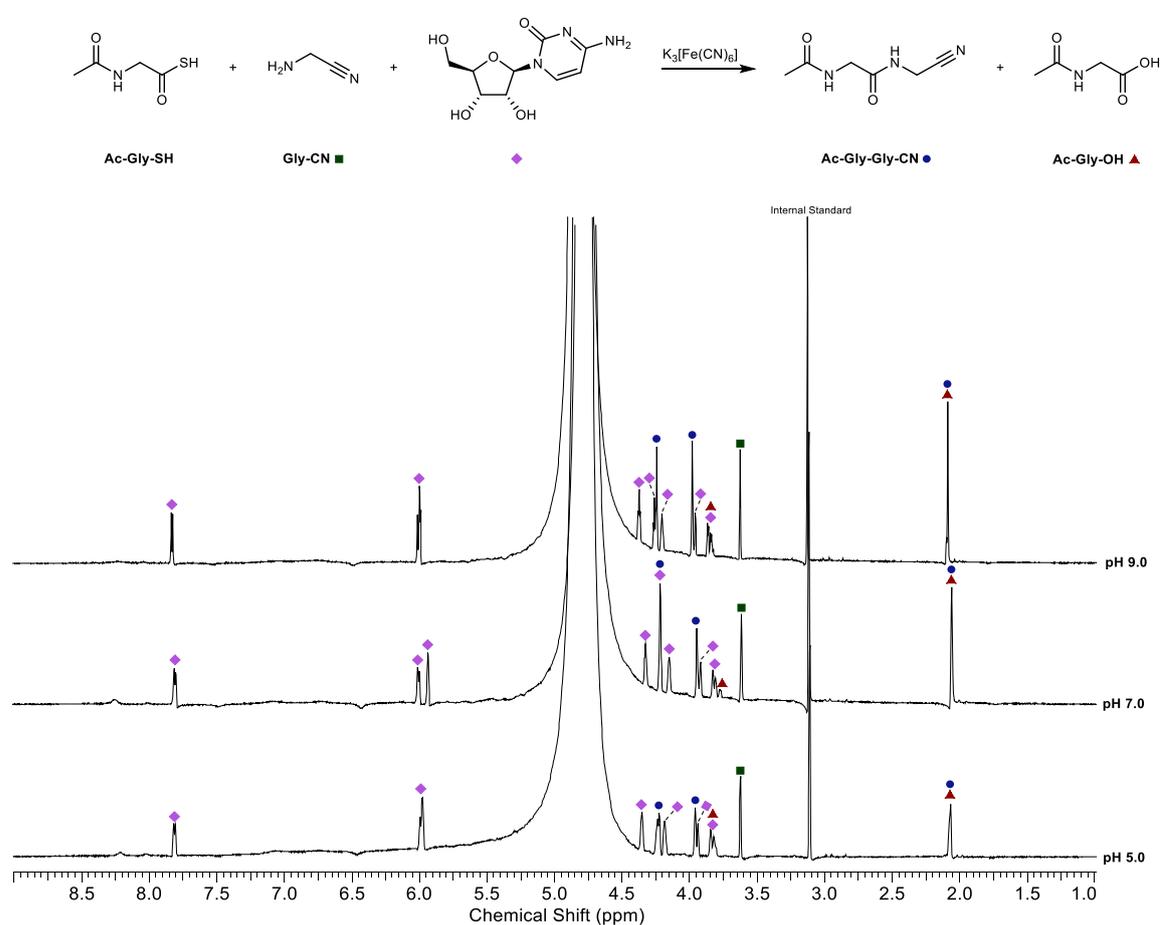
$^1\text{H}$  NMR (600 MHz, 2%  $\text{D}_2\text{O}/\text{H}_2\text{O}$  pH 9.0, 4.40 – 2.00 ppm):

( $\bullet$ )  $\delta$  3.62 (s, 2H, (C2)- $\text{H}_2$ ); *N*-Acetylglycyl glycine nitrile **Ac-Gly-Gly-CN** ( $\bullet$ )  $\delta$  4.22 (s, 2H, (GlyCN2)-(C2)- $\text{H}_2$ ), 3.96 (s, 2H, (Gly1)-(C2)- $\text{H}_2$ ), 2.07 (s, 3H, (CO)- $\text{CH}_3$ );

*N*-Acetylglycine **Ac-Gly-OH** ( $\blacktriangle$ )  $\delta$  3.80 (d,  $J = 6.3$  Hz, 2H, (C2)- $\text{H}_2$ ), 2.07 (s, 3H, (CO)- $\text{CH}_3$ );

### Glycine nitrile **Gly-CN**

*Cytidine* ( $\blacklozenge$ ):  $\delta$  7.48 (d,  $J = 6.3$  Hz, 1H, (C6)-H), 5.95 (d,  $J = 6.3$  Hz, 1H, (C5)-H), 5.94 (br s, 1H, (C1')-H), 4.37 (br t,  $J = 5.1$  Hz, 1H, (C2')-H), 4.26 (obs. t,  $J = 5.7$  Hz, 1H, (C3')-H), 4.20 (br s, 1H, (C4')-H), 3.96 (br s, 1H, (C5')-H), 3.85 (dd,  $J = 14.4, 4.8$  Hz, 1H, (C5')-H').



**Experimental figure 66:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 1.00 – 9.00 ppm) spectra to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), glycine nitrile (**Gly-CN**; 100 mM), cytidine (100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and the specified pH (5.0, 7.0 or 9.0) after 10 min.

### Competition between glycine nitrile and cytidine-5'-monophosphate

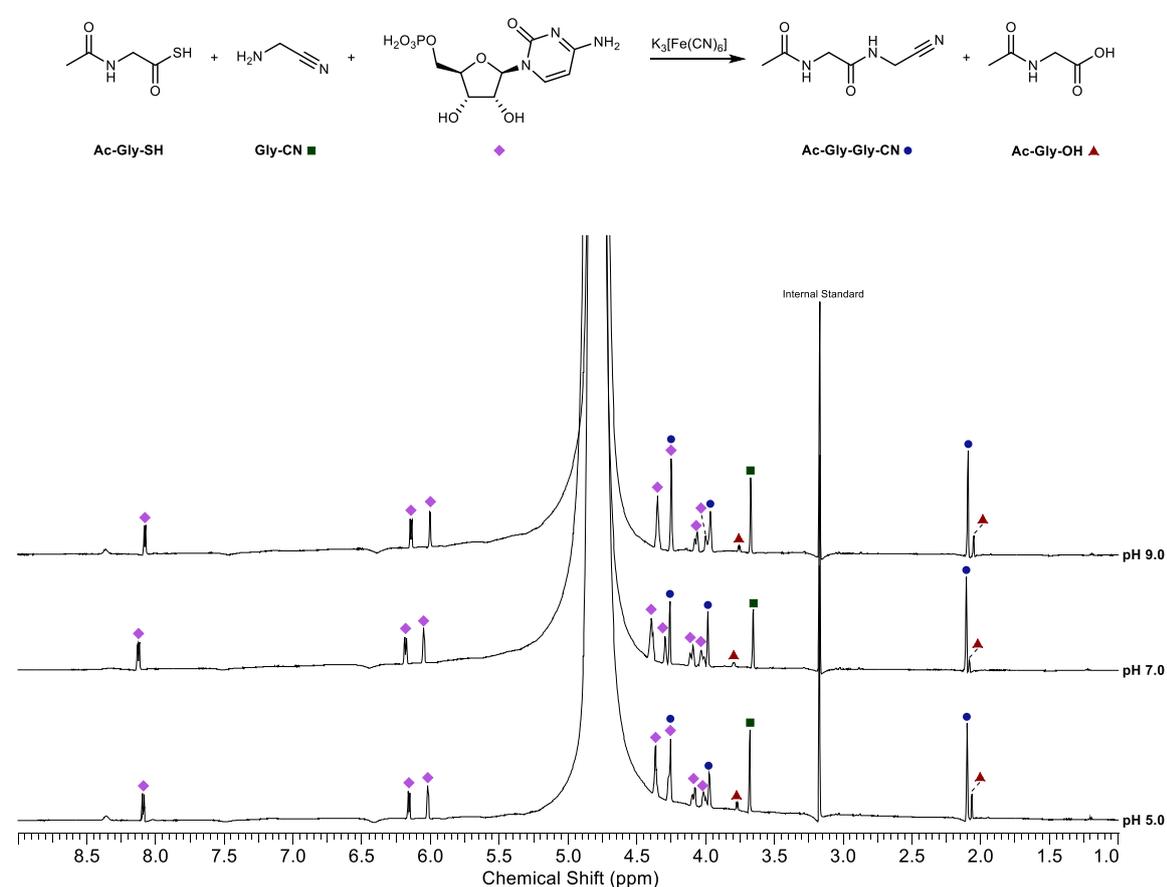
$^1\text{H}$  NMR (600 MHz, 2%  $\text{D}_2\text{O}/\text{H}_2\text{O}$  pH 9.0, 4.40 – 2.00 ppm):

glycine nitrile **XXX** (■)  $\delta$  3.68 (s, 2H, (C2)- $\text{H}_2$ ); *N*-Acetylglycyl glycine nitrile **Ac-Gly-Gly-CN** (●)  $\delta$  4.25 (s, 2H, (GlyCN2)-(C2)- $\text{H}_2$ ), 3.97 (s, 2H, (Gly1)-(C2)- $\text{H}_2$ ), 2.10 (s, 3H, (CO)- $\text{CH}_3$ );

*N*-Acetylglycine **Ac-Gly-OH** (▲)  $\delta$  3.77 (d,  $J = 6.3$  Hz, 2H, (C2)- $\text{H}_2$ ), 2.06 (s, 3H, (CO)- $\text{CH}_3$ );

Glycine nitrile **Gly-CN**  $\delta$

Cytidinemonophosphate (◆):  $\delta$  8.09 (d,  $J = 7.7$  Hz, 1H, (C6)-H), 6.16 (d,  $J = 7.7$  Hz, 1H, (C5)-H), 6.02 (br s, 1H, (C1')-H), 4.38 – 4.34 (obs. m, 2H, (C2')-H, (C3')-H), 4.29 (br s, 1H, (C4')-H), 4.09 (br d,  $J = 13.9$  Hz, 1H, (C5')-H), 4.00 (obs. d, 1H, (C5')-H').



**Experimental figure 67:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 1.00 – 9.00 ppm) spectra to show the reaction of *N*-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), glycine nitrile (**Gly-CN**; 100 mM), cytidine-5'-monophosphate (100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and the specified pH (5.0, 7.0 or 9.0) after 10 min.

### Competition between glycine nitrile and adenosine-5'-monophosphate

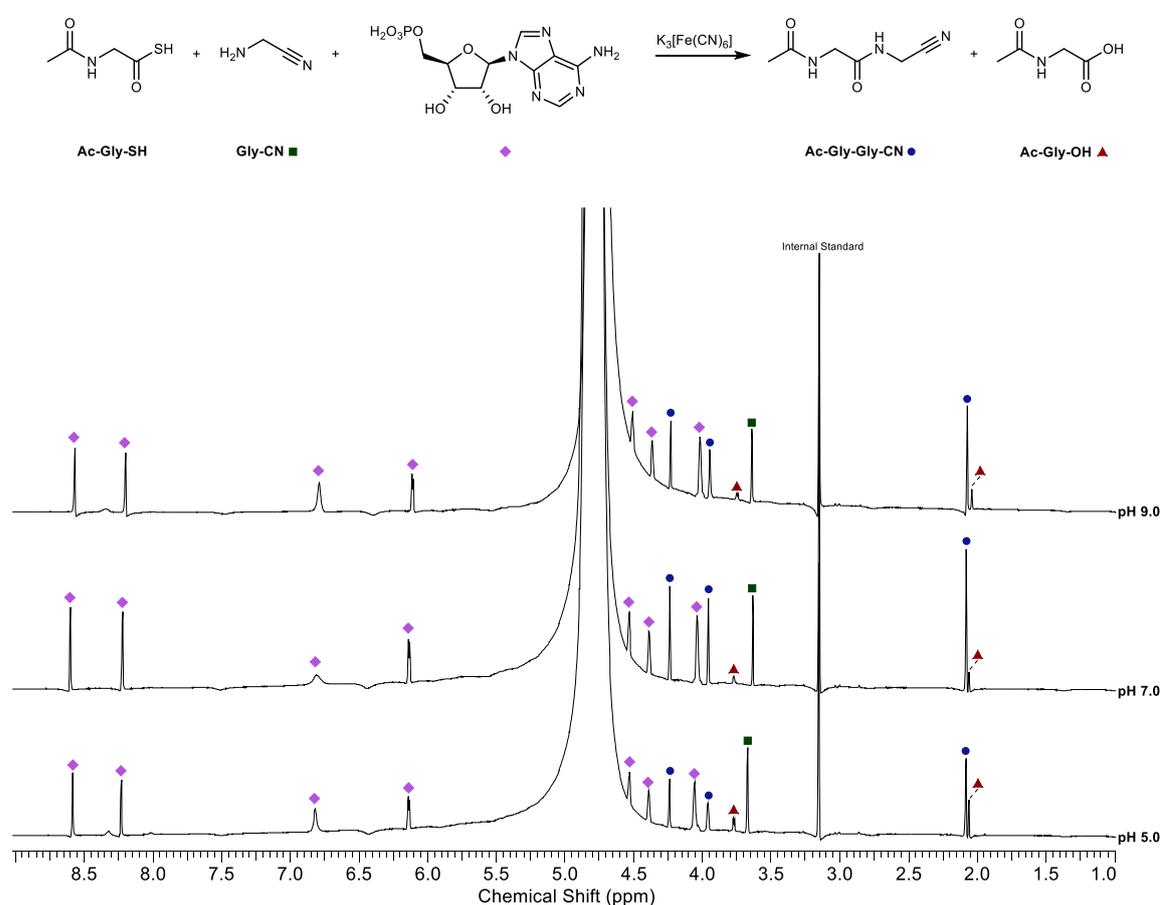
$^1\text{H}$  NMR (600 MHz, 2%  $\text{D}_2\text{O}/\text{H}_2\text{O}$  pH 9.0, 4.40 – 2.00 ppm):

glycine nitrile **XXX** (■)  $\delta$  3.67 (s, 2H, (C2)-H<sub>2</sub>); N-Acetylglycyl glycine nitrile **Ac-Gly-Gly-CN** (●)  $\delta$  4.24 (s, 2H, (GlyCN2)-(C2)-H<sub>2</sub>), 3.96 (s, 2H, (Gly1)-(C2)-H<sub>2</sub>), 2.08 (s, 3H, (CO)-CH<sub>3</sub>);

N-Acetylglycine **Ac-Gly-OH** (▲)  $\delta$  3.77 (d,  $J = 6.3$  Hz, 2H, (C2)-H<sub>2</sub>), 2.06 (s, 3H, (CO)-CH<sub>3</sub>).

Glycine nitrile **Gly-CN**  $\delta$

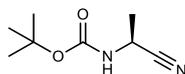
Adenosine monomonophosphate (◆):  $\delta$  8.59 (s, 1H, (C8)-H), 8.23 (s, 1H, (C2)-H), 6.82 (br s, 2H, NH<sub>2</sub>), 6.14 (d, 1H,  $J = 5.8$  Hz, (C1')-H), 4.53 (br s, 1H, (C2')-H), 4.39 (br s, 1H, (C4')-H), 4.06 (br s, 2H, (C5')-H<sub>2</sub>);



**Experimental figure 68:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 9.00 – 1.00 ppm) spectra to show the reaction of N-acetylglycine thioacid (**Ac-Gly-SH**; 50 mM), glycine nitrile (**Gly-CN**, 100 mM), adenosine-5'-monophosphate (100 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (150 mM) at room temperature and the specified pH (5.0, 7.0 or 9.0) after 10 min.

### 9. 2. 7. Enantioretention studies

#### *N*-Boc-L-alanine nitrile **Boc-L-Ala-CN**



To a solution of L-alaninamide hydrochloride **L-Ala-NH<sub>2</sub>·HCl** (5.00 g, 40.14 mmol) and triethylamine (20.31 g, 200.7 mmol, 28.00 mL) in THF (400 mL) at 0 °C was added di-*tert*-butyl dicarbonate (13.14 g, 60.21 mmol) portionwise over 10 minutes. The reaction was stirred at 0 °C for 2 h, after which time trifluoroacetic anhydride (25.29 g, 120.42 mmol, 16.97 mL) was added dropwise over 30 min. The reaction was stirred at 0 °C for 1 h, and then at room temperature for a further 2 h. The reaction was concentrated to dryness *in vacuo*, diluted in water (200 mL) and extracted with diethyl ether (5 × 200 mL). The combined organic layers were washed with 0.1 M HCl (3 × 200 mL), saturated potassium carbonate (3 × 200 mL) and brine (200 mL). The crude residue was recrystallized from diethyl ether to give *N*-Boc-L-alanine nitrile **Boc-L-Ala-CN** as a white crystalline solid (4.87 g, 28.62 mmol, 71%).

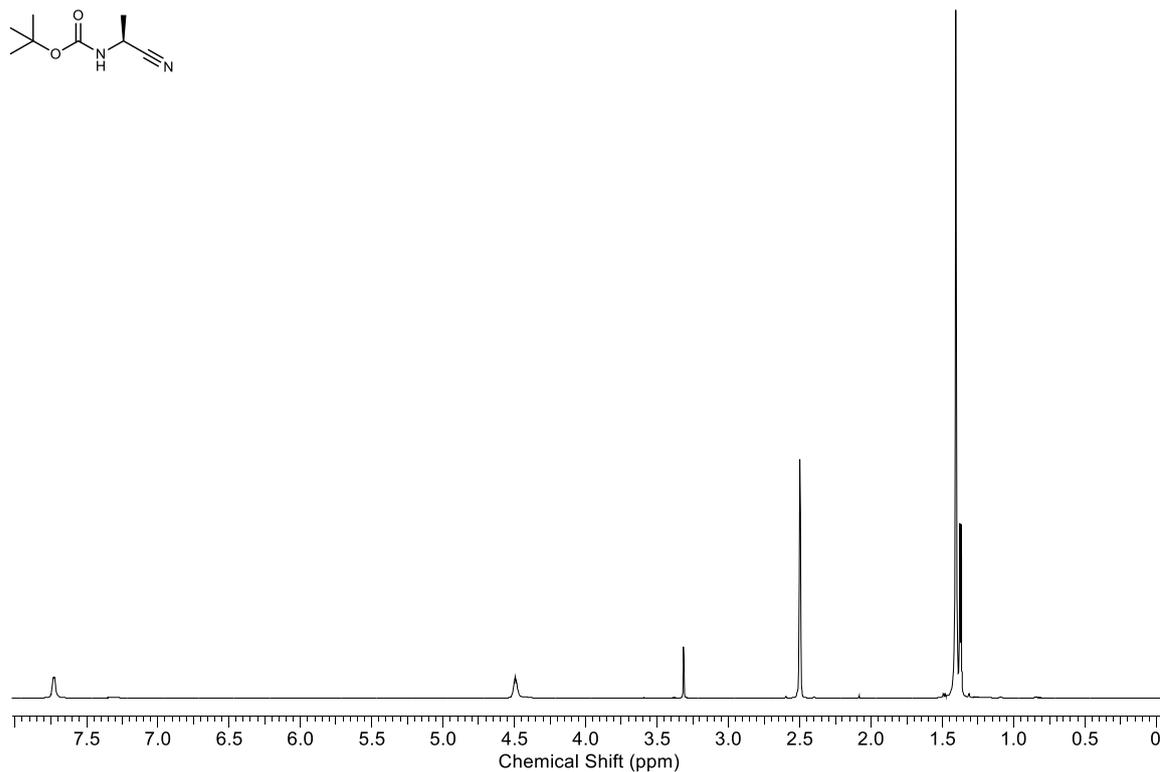
**<sup>1</sup>H NMR** (700 MHz, DMSO-*d*<sub>6</sub>): δ 7.73 (br s, 1H, NH), 4.49 (br m, 1H, (C2)-H), 1.40 (s, 9H, (C)-CH<sub>3</sub>), 1.37 (d, *J* = 7.1 Hz, 3H, (C3)-H<sub>3</sub>).

**<sup>13</sup>C NMR** (176 MHz, DMSO-*d*<sub>6</sub>): δ 154.5 (CO), 120.7 (C1), 79.2 (C(CH<sub>3</sub>)<sub>3</sub>), 37.3 (C2), 28.0 (C(CH<sub>3</sub>)<sub>3</sub>), 18.3 (C3).

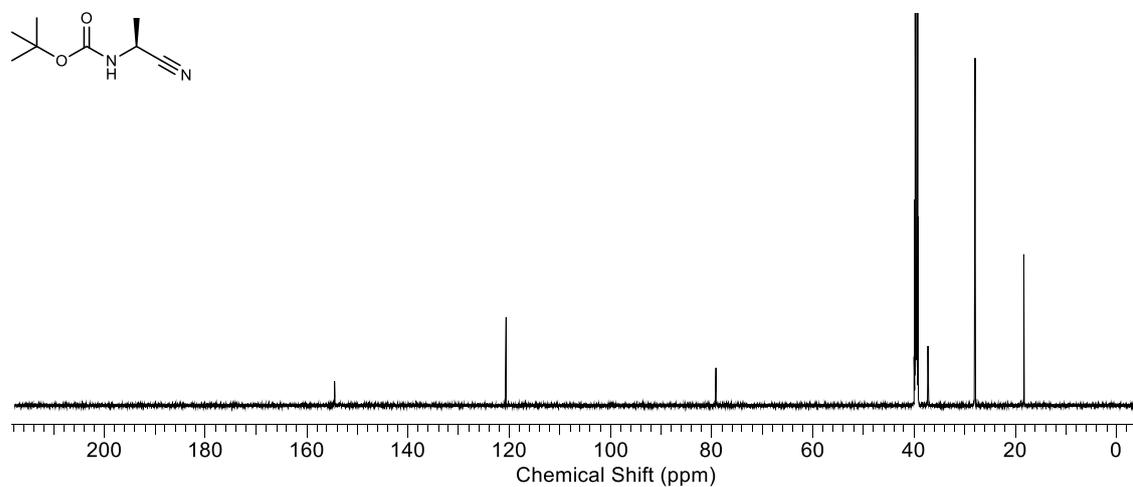
**HRMS** (ES<sup>+</sup>): calcd. for [C<sub>8</sub>H<sub>14</sub>N<sub>2</sub>O<sub>2</sub> + Na]<sup>+</sup>: 193.0947; observed: 193.0945.

**IR** (solid): 3339, 3316, 3005, 2978, 2946, 2192, 1737, 1704, 1677, 1524 cm<sup>-1</sup>.

**M.p.** 106.1 – 106.8 °C.

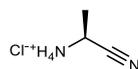


**Experimental figure 69:** <sup>1</sup>H NMR (700 MHz, DMSO-d<sub>6</sub>, 8.00 – 0.00 ppm) spectrum to show *N*-Boc-L-alanine nitrile **Boc-L-Ala-CN**.



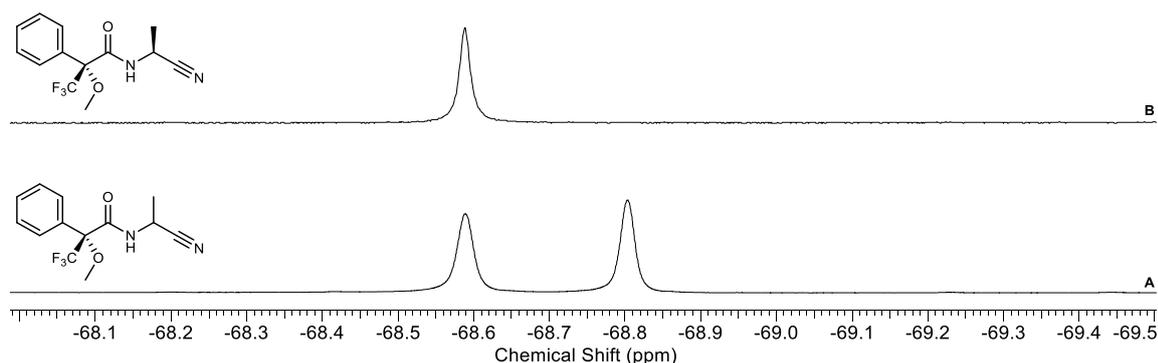
**Experimental figure 70:** <sup>13</sup>C NMR (176 MHz, DMSO-d<sub>6</sub>, 220 – 0 ppm) spectrum to show *N*-Boc-L-alanine nitrile **Boc-L-Ala-CN**.

*L*-alanine nitrile hydrochloride **L-Ala-CN·HCl**



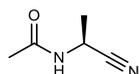
*N*-Boc-*L*-alanine nitrile **Boc-L-Ala-CN** (2.00 g, 11.75 mmol) was stirred in formic acid (50 mL) at room temperature for 1 h. The reaction was concentrated *in vacuo* and purified by flash column chromatography (SiO<sub>2</sub>; EtOAc/petroleum ether 0:100 → 100:0 + 0.1% MeOH). Fractions containing *L*-alanine nitrile **L-Ala-CN** were collected and stirred at -78 °C. 2 M HCl in Et<sub>2</sub>O (6 mL, 12.0 mmol) was added, and the resultant precipitate was filtered, washed with Et<sub>2</sub>O (3 × 100 mL), and dried to afford *L*-alanine nitrile hydrochloride **L-Ala-CN·HCl** as a white powder (712 mg, 6.68 mmol, 51%). All spectral data were found to be identical to those obtained for *rac*-alanine nitrile hydrochloride **Ala-CN·HCl** prepared following *General Procedure E*.

The enantiomeric purity of *L*-alanine nitrile hydrochloride **L-Ala-CN·HCl** was quantified by derivatisation with (*R*)-(+)- $\alpha$ -methoxy- $\alpha$ -(trifluoromethyl)phenylacetyl chloride (Mosher's acyl chloride) according to the following procedure. To a solution of *L*-alanine nitrile hydrochloride **L-Ala-CN·HCl** (10.7 mg, 0.10 mmol) and triethylamine (24.5 mg, 0.24 mmol, 33.5  $\mu$ L) in DMSO (1 mL) was added (*R*)-(+)- $\alpha$ -methoxy- $\alpha$ -(trifluoromethyl)phenylacetyl chloride (126 mg, 0.50 mmol, 93.6  $\mu$ L). The solution was stirred at room temperature for 10 min, and <sup>19</sup>F NMR data were acquired. The procedure was repeated using *rac*-alanine nitrile hydrochloride **Ala-CN·HCl** as starting material for comparison. The enantiomeric ratio of *L*-alanine nitrile hydrochloride **L-Ala-CN·HCl** was found to be > 99:1.



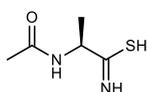
**Experimental figure 71:** <sup>19</sup>F NMR (376 Hz, DMSO-d<sub>6</sub>, -68.0 – -69.5 ppm) to show the reaction of triethylamine (240 mM) and (*R*)-(+)- $\alpha$ -methoxy- $\alpha$ -(trifluoromethyl)phenylacetyl chloride (500 mM) at room temperature with **A**: alanine nitrile hydrochloride **Ala-CN·HCl** (100 mM); **B**: *L*-alanine nitrile hydrochloride **L-Ala-CN·HCl** (100 mM).

*N*-acetyl-L-alanine nitrile **Ac-L-Ala-CN**

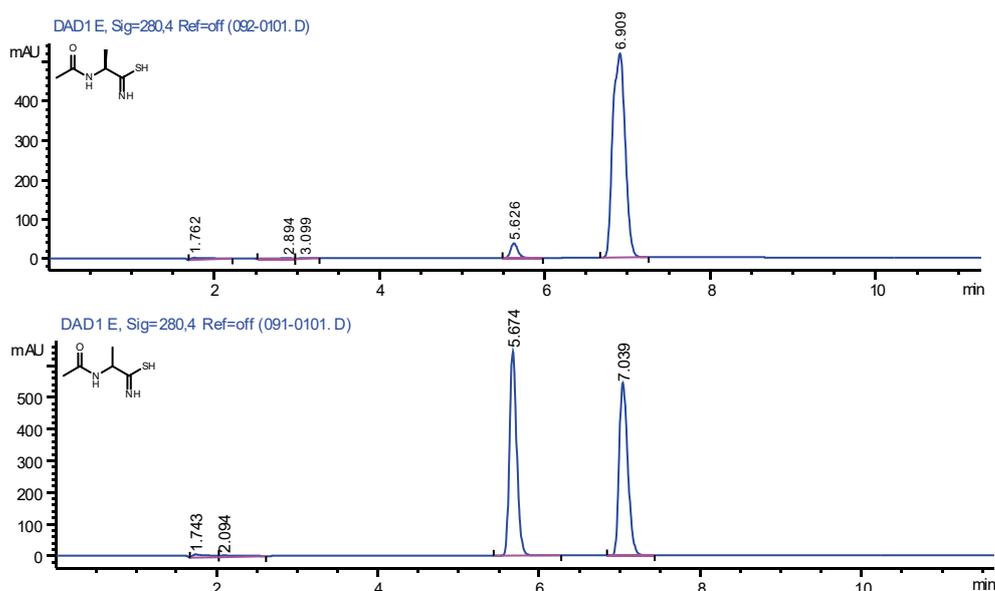


*N*-acetyl-L-alanine nitrile **Ac-L-Ala-CN** was prepared following *General procedure C* using L-alanine nitrile hydrochloride **L-Ala-CN**·HCl (1.07 g, 10 mmol). The reaction mixture was concentrated *in vacuo* and the residue purified by flash column chromatography (SiO<sub>2</sub>; EtOAc/petroleum ether 0:100 → 100:0) to afford *N*-acetyl aminonitrile **Ac-L-Ala-CN** as a white, crystalline solid (473 mg, 4.22 mmol, 42%). All spectral data were found to be identical to those obtained for *rac*-*N*-acetyl-alanine nitrile **Ac-Ala-CN** prepared following *General Procedure B*.

*N*-acetyl-L-alanine thioamide **Ac-L-Ala-SNH<sub>2</sub>**

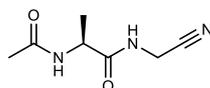


*N*-acetyl-L-alanine thioamide **Ac-L-Ala-SNH<sub>2</sub>** was prepared following *General procedure C* using *N*-acetyl-L-alanine nitrile **Ac-L-Ala-CN** (280.3 mg, 2.50 mmol). The reaction afforded *N*-acetyl alanine thioamide **Ac-L-Ala-SNH<sub>2</sub>** as a white solid (277.1 mg, 1.90 mmol, 76%). All NMR spectral data were found to be identical to those obtained for *rac*-*N*-acetyl-alanine thioamide **Ac-Ala-SNH<sub>2</sub>**. The ratio of enantiomers was determined to be 98:2 by chiral HPLC analysis (YMC chiral amylose – SA 5 μm, isocratic 100% ethanol gradient, 1.0 mL · min<sup>-1</sup>, λ = 280.4 nm).

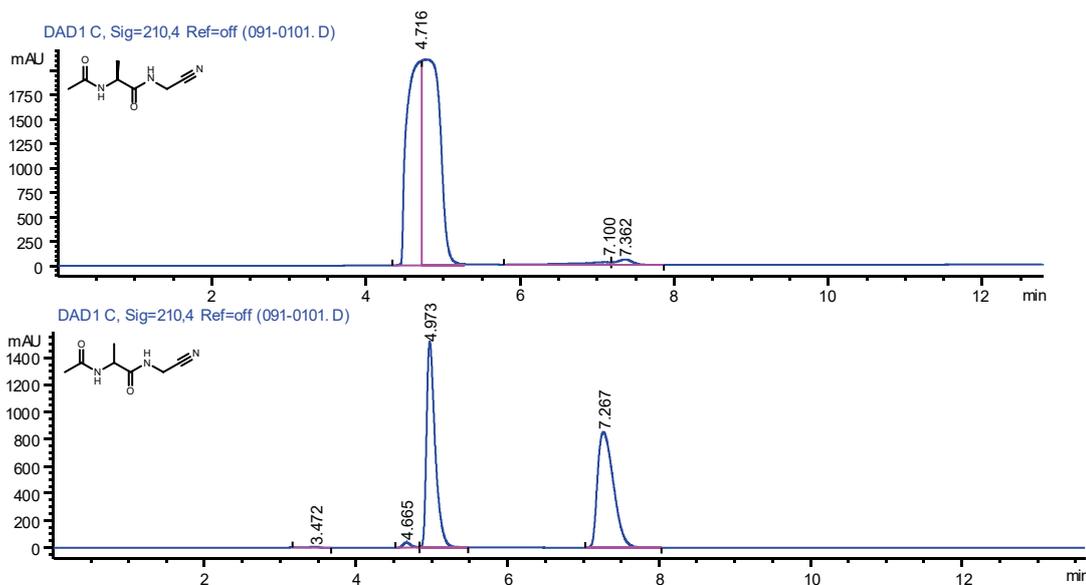


**Experimental figure 72:** Chiral HPLC traces for *rac*-*N*-acetyl-alanine thioamide **Ac-Ala-SNH<sub>2</sub>** (bottom) and *N*-acetyl-L-alanine thioamide **Ac-L-Ala-SNH<sub>2</sub>** (top) obtained from prebiotic thiolysis of the corresponding *N*-acetyl aminonitrile.

*N*-acetyl-L-alanyl-glycine nitrile **Ac-L-Ala-Gly-CN**

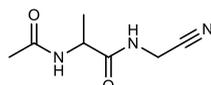


*N*-acetyl-L-alanine thioacid **Ac-L-Ala-SH** was prepared following *General procedure C* using *N*-acetyl-L-alanine thioamide **Ac-L-Ala-SNH<sub>2</sub>** (146 mg, 1.00 mmol). The resulting thioacid was ligated with glycine nitrile **Gly-CN** (5.00 mmol) following *General Procedure D*. The reaction was concentrated to dryness *in vacuo* and the residue purified by flash column chromatography (SiO<sub>2</sub>; MeOH/DCM 0:100 → 5:95). Subsequent recrystallisation from MeOH/EtO<sub>2</sub> afforded *N*-acetyl-L-alanyl-glycine nitrile **Ac-L-Ala-Gly-CN** as a white, crystalline solid (14.6 mg 0.09 mmol, 9%). The supernatant was concentrated to dryness *in vacuo* and analysed by <sup>1</sup>H NMR to ensure no peptide nitrile was lost in the recrystallisation process. All NMR spectral data were found to be identical to those obtained for *rac*-*N*-acetylalanyl-glycine nitrile. The ratio of enantiomers was determined to be 98:2 by chiral HPLC analysis (YMC chiral amylose – SA 5 μm, isocratic 100% ethanol gradient, 0.7 mL · min<sup>-1</sup>, λ = 210.4 nm).



**Experimental figure 73:** Chiral HPLC traces for rac-*N*-acetylalanine thioamide Ac-Ala-SNH<sub>2</sub> (bottom) and *N*-acetyl-L-alanine thioamide Ac-L-Ala-SNH<sub>2</sub> (top) obtained from prebiotic thiolysis of the corresponding *N*-acetyl aminonitrile following *General procedure C*.

#### *N*-acetylalanyl glycine nitrile **Ac-Ala-Gly-CN**



*N*-acetylalanine **Ac-Ala-OH** (1.00 g, 7.63 mmol) and triethylamine (3.86 g, 38.13 mmol, 5.32 mL) were dissolved in dry DCM (50 mL) and stirred at 0 °C under argon. 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (2.96 g, 19.08 mmol) was added portionwise over 5 min, and the reaction was stirred for 10 min at 0 °C. Glycine nitrile hydrochloride **Gly-CN**·HCl (1.77 g, 19.08 mmol) was added, and the reaction was stirred for 2 h while warming up to room temperature. After that time, the reaction was concentrated to dryness *in vacuo* and the residue washed with DCM (5 × 50 mL). The resulting solid was purified by flash column chromatography (SiO<sub>2</sub>; MeOH/DCM 0:100 → 5:95) to afford *N*-acetylalanyl-glycine nitrile **Ac-Ala-Gly-CN** as a white solid (942 mg, 14.6 mg 5.57 mmol, 74%).

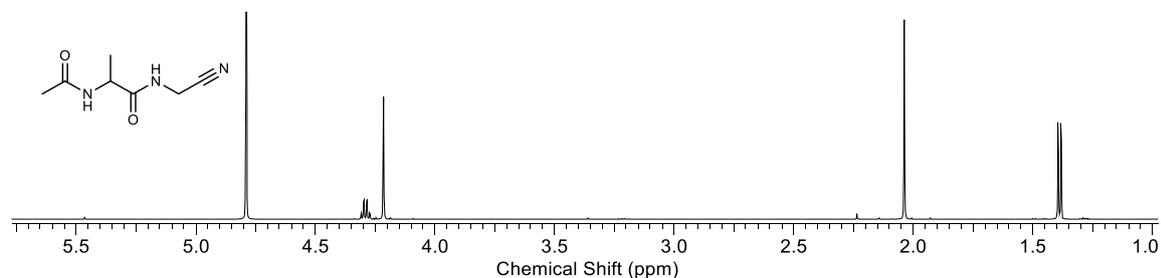
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.29 (q, *J* = 7.3 Hz, 1H, Ala1-(C2)-H), 4.21 (s, 2H, Gly2-(C2)-H<sub>2</sub>), 2.04 (s, 3H, (CO)-CH<sub>3</sub>), 1.39 (d, *J* = 7.3 Hz, 3H, Ala1-(C3)-H<sub>3</sub>).

$^{13}\text{C}$  NMR (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  176.5 ( $\text{COCH}_3$ ), 174.9 (Ala1-(C1)), 117.5 (Gly2-(C1)), 50.4 (Ala1-(C2)), 28.3 (Gly2-(C2)), 22.3 ( $\text{COCH}_3$ ), 17.0 (Ala1-(C3)).

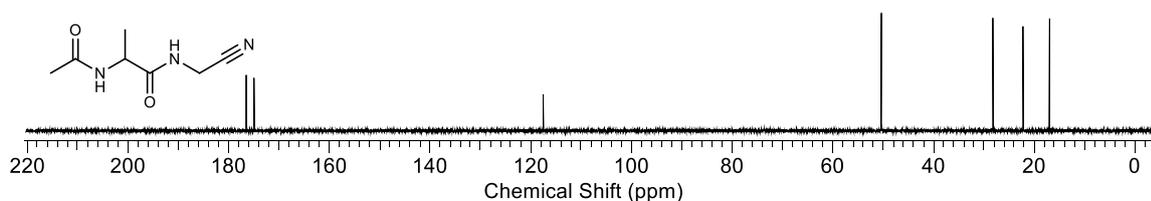
HRMS ( $\text{ES}^+$ ): calcd. for  $[\text{C}_7\text{H}_{11}\text{N}_3\text{O}_2 + \text{H}]^+$ : 170.0924; observed: 170.0927.

IR (solid): 3291, 3212, 3063, 2989, 2945, 2812, 2246, 1746, 1666, 1630, 1566, 1538  $\text{cm}^{-1}$ .

M.p.: 156.0 – 157.1  $^\circ\text{C}$ .



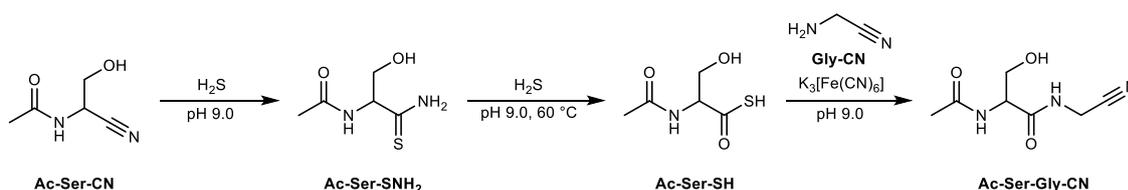
**Experimental figure 74:**  $^1\text{H}$  NMR (600 MHz,  $\text{D}_2\text{O}$ , 5.75 – 1.00 ppm) spectrum to show *N*-acetylalanyl glycine nitrile **Ac-Ala-Gly-CN**.



**Experimental figure 75:**  $^{13}\text{C}$  NMR (151 MHz,  $\text{D}_2\text{O}$ , 220 – 0 ppm) spectrum to show *N*-acetylalanyl glycine nitrile **Ac-Ala-Gly-CN**.

## 9. 2. 8. One-pot experiments

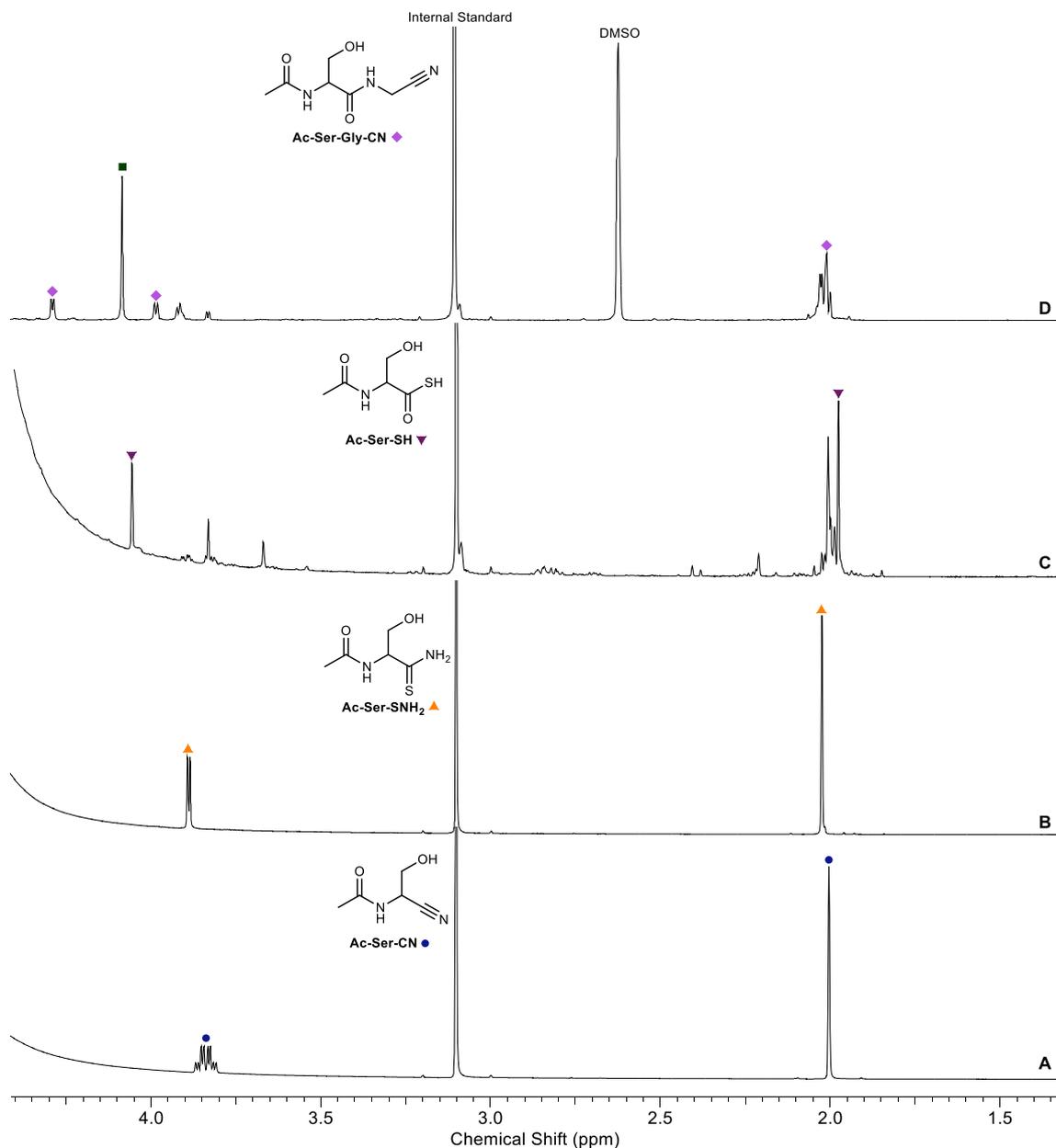
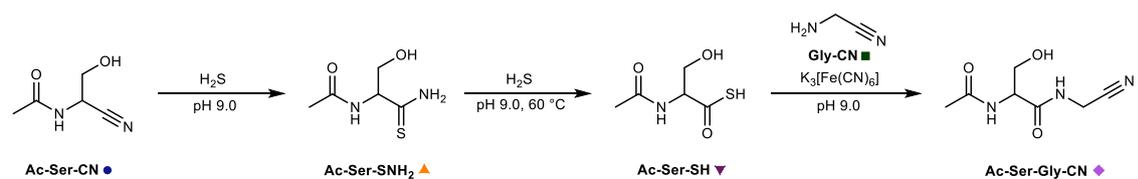
### Sequential elongation of *N*-acetylserine nitrile



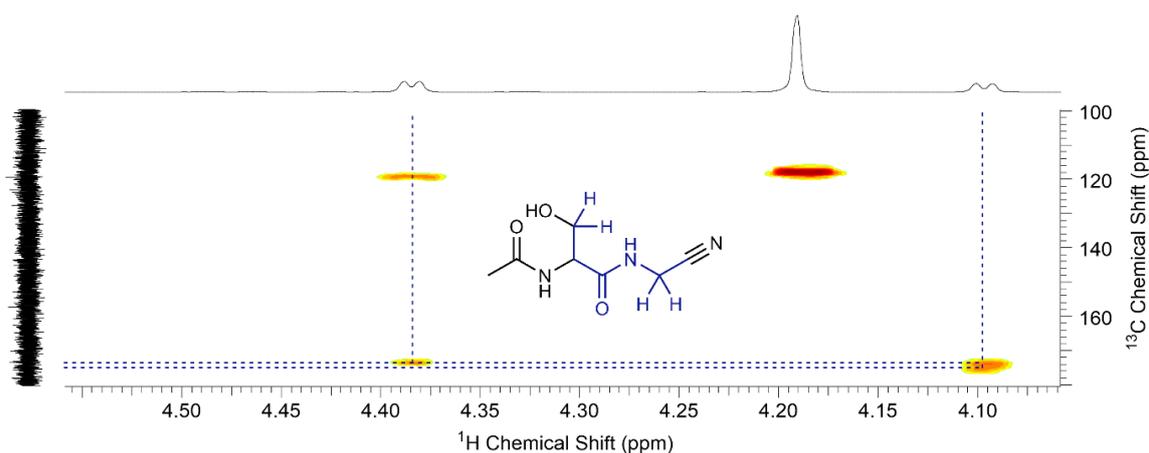
*N*-Acetylserine nitrile **Ac-Ser-CN** (6.4 mg, 0.05 mmol),  $\text{NaSH}\cdot x\text{H}_2\text{O}$  (56 mg, 0.50 mmol) and methylsulfonylmethane (0.05 mmol) were dissolved in degassed  $\text{H}_2\text{O}/\text{D}_2\text{O}$  (98:2, 1 mL), and the solution was adjusted to pH 9.0 with  $\text{NaOH}/\text{HCl}$ . The solution was incubated at room temperature whilst maintaining the solution at pH 9.0 with  $\text{NaOH}/\text{HCl}$ . After 24 h, NMR analysis showed quantitative conversion of **Ac-Ser-CN** to *N*-acetylserine thioamide **Ac-Ser-SNH<sub>2</sub>**, which was confirmed by  $^1\text{H}$  –  $^{13}\text{C}$  HMBC NMR analysis. The yield was quantified using the methylsulfonylmethane signal as an internal standard (>99%). The resulting solution was adjusted

to pH 9.0 with NaOH/HCl and incubated at 60 °C for a further 24 h. NMR analysis showed complete consumption of *N*-acetylserine thioamide **Ac-Ser-SNH<sub>2</sub>**. The presence of *N*-acetylserine thioacid **Ac-Ser-SH** was confirmed by <sup>1</sup>H – <sup>13</sup>C HMBC NMR analysis. The yield was quantified using the methylsulfonylmethane signal as an internal standard (61%). The solution was sparged with argon for 30 min whilst maintaining the solution at pH 7.0 with NaOH/HCl. Glycine nitrile **Gly-CN** (0.061 mmol) was added, and the solution was adjusted to pH 9.0 with NaOH/HCl, followed by the addition of potassium hexacyanoferrate(III) (30 mg, 0.09 mmol). The solution was stirred at room temperature for 20 minutes. The resulting suspension was then adjusted to pH 9.0 and cooled to 0 °C before dilution with DMSO-*d*<sub>6</sub> (1 mL). NMR spectra were acquired and the presence of *N*-acetylserylglycine nitrile **Ac-Ser-Gly-CN** was confirmed by <sup>1</sup>H – <sup>13</sup>C HMBC NMR analysis and high-resolution mass spectrometry. The yield of **Ac-Ser-Gly-CN** was determined by integration against the methylsulfonylmethane signal.

<b>Experimental table 9: sequential elongation of <i>N</i>-acetylserine nitrile</b>			
<b>Entry</b>	<b>Step</b>	<b>Product</b>	<b>Yield (%)</b>
1	Thiolysis	<b>Ac-Ser-SNH<sub>2</sub></b>	99
2	Hydrolysis	<b>Ac-Ser-SH</b>	61
3	Ligation	<b>Ac-Ser-Gly-CN</b>	87



**Experimental figure 76:** <sup>1</sup>H NMR (700 MHz, pH 9.0, 1.35 – 4.40 ppm) spectra to show the one-pot thiolysis, hydrolysis and oxidative ligation with glycine nitrile **Gly-CN** of *N*-acetylserine nitrile **Ac-Ser-CN**. **A:** pure **Ac-Ser-CN**, H<sub>2</sub>O/D<sub>2</sub>O 98:2; **B:** Thiolysis of **Ac-Ser-CN** to **Ac-Ser-SNH<sub>2</sub>**, H<sub>2</sub>O/D<sub>2</sub>O 98:2; **C:** Hydrolysis of **Ac-Ser-SNH<sub>2</sub>** to **Ac-Ser-SH**, H<sub>2</sub>O/D<sub>2</sub>O 98:2; **D:** Coupling of **Ac-Ser-SH** with **Gly-CN**, water suppressed, D<sub>2</sub>O/H<sub>2</sub>O/DMSO-*d*<sub>6</sub> 1:49:50.



**Experimental figure 77:**  $^1\text{H}$  –  $^{13}\text{C}$  HMBC ( $^1\text{H}$  – 700 MHz [4.05 – 4.55 ppm],  $^{13}\text{C}$  – 176 MHz [100 – 180 ppm],  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-}d_6$  1:49:50) spectrum to show the reaction of *N*-acetylseryne thioacid **Ac-Ser-SH** (31 mM), glycine nitrile **Gly-CN** (61 mM) and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  (93 mM) in  $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. After 20 min, the suspension was adjusted to pH 9.0, diluted with  $\text{DMSO-}d_6$  (1 mL), and NMR spectra were acquired. The characteristic  $^3J_{\text{CH}}$  coupling between glycine nitrile proton resonance (4.38 ppm, d,  $J = 5.0$  Hz, 2H, Gly<sub>2</sub>-(C2)–H<sub>2</sub>), and serine amide carbon resonance (174.6 ppm, Ser1-(C1)) that demonstrates peptide ligation is highlighted with overlaid blue dotted lines.

$^1\text{H}$  NMR (700 MHz, pH 9.0):

**Experimental figure 76-A** (pure starting material,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, partial assignment):

*N*-Acetylseryne nitrile **Ac-Ser-CN** (•)  $\delta$  3.85 (ABX,  $J = 5.3, 11.2$  Hz, 1H, (C2)–H), 3.83 (ABX,  $J = 5.3, 11.2$  Hz, 1H, (C2)–H'), 2.02 (s, 3H, (CO)CH<sub>3</sub>).

**Experimental figure 76-B:** (thiolysis,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, pH 9.0, partial assignment):

*N*-Acetylseryne thioamide **Ac-Ser-SNH<sub>2</sub>** (▲)  $\delta$  3.86 (d,  $J = 5.4$  Hz, 2H, (C2)–H<sub>2</sub>), 1.97 (s, 3H, (CO)CH<sub>3</sub>).

**Experimental figure 76-C:** (hydrolysis,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, partial assignment):

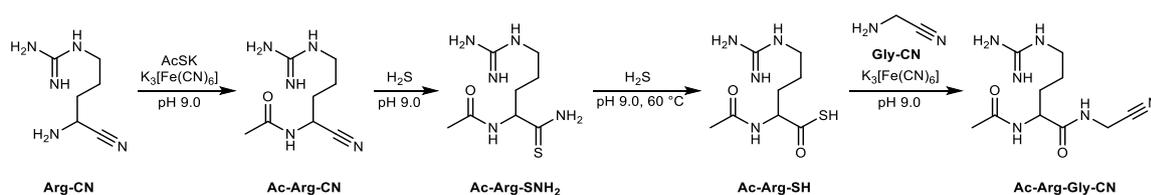
*N*-Acetylseryne thioacid **Ac-Ser-SH** (▼)  $\delta$  4.08 (br s, 2H, (C2)–H<sub>2</sub>), 1.99 (s, 3H, (CO)CH<sub>3</sub>).

**Experimental figure 76-D:** (ligation,  $\text{H}_2\text{O}/\text{D}_2\text{O}/\text{DMSO-}d_6$  49:1:50, partial assignment):

*N*-Acetylserylglycine nitrile **Ac-Ser-Gly-CN** (◆)  $\delta$  4.23 (br d, 5.0 Hz, 2H, Gly-(C2)–H<sub>2</sub>), 3.96 (obs. d, 2H, Ser-(C2)–H<sub>2</sub>), 1.98 (s, 3H, (CO)CH<sub>3</sub>).

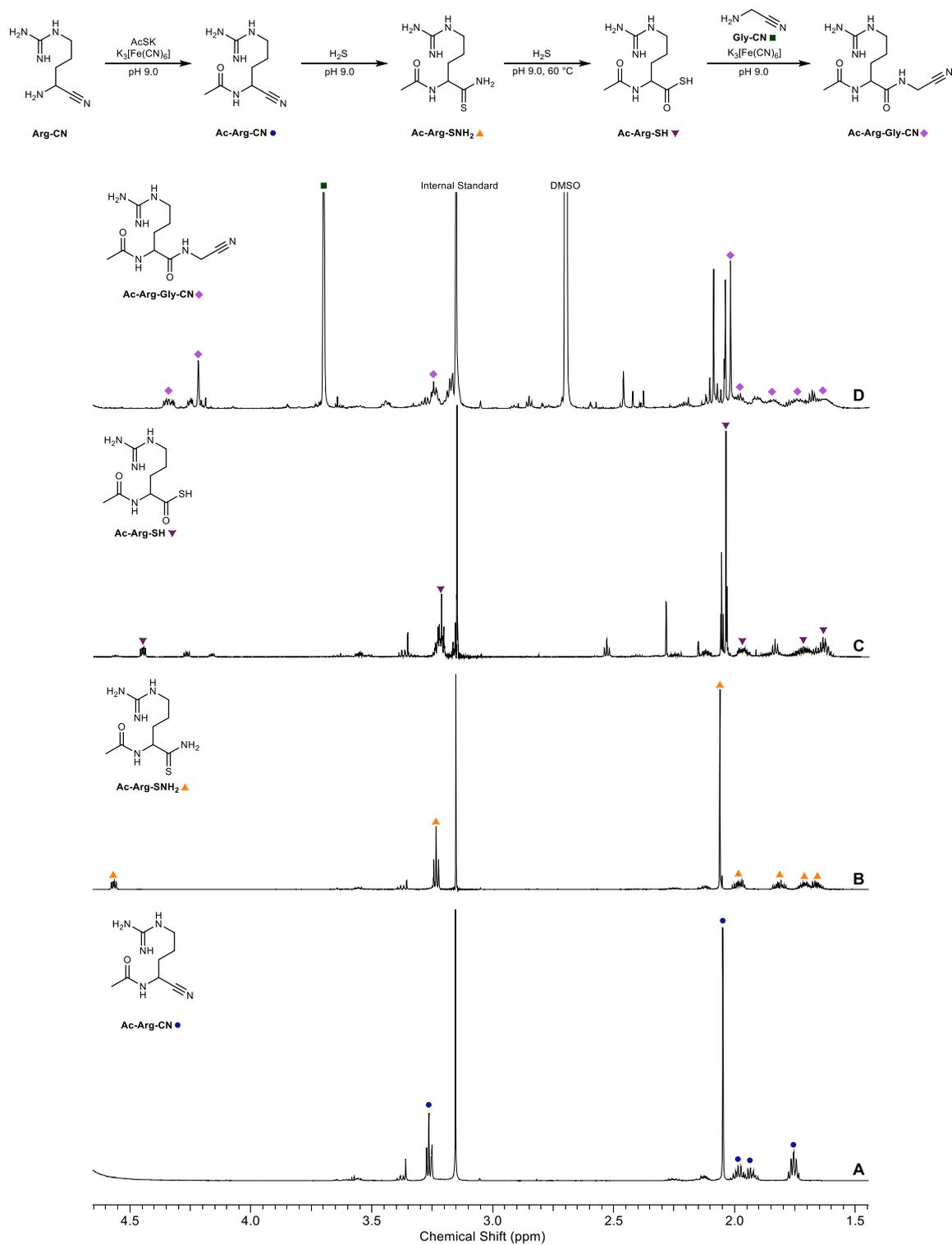
Glycine nitrile **Gly-CN** (■)  $\delta$  3.95 (s, 2H, (C2)–H<sub>2</sub>).

### Sequential elongation of arginine nitrile

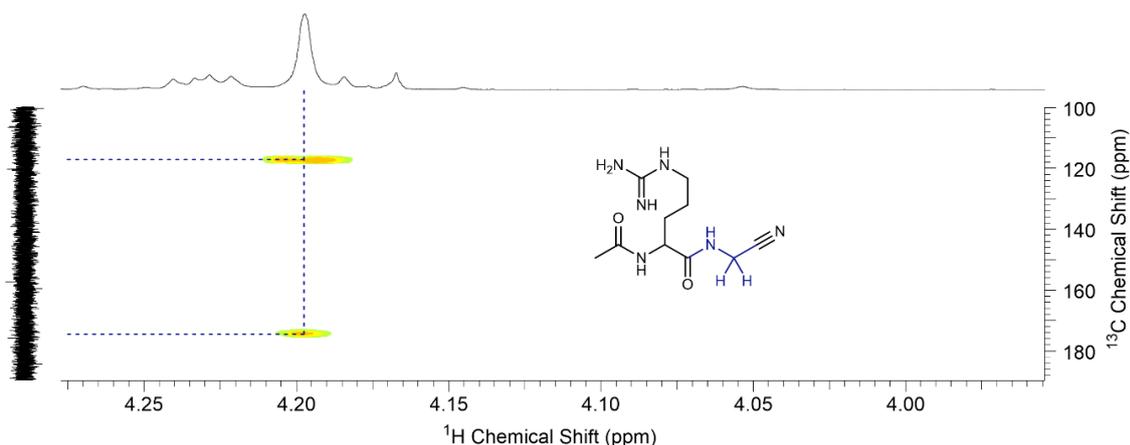


Arginine nitrile (**Arg-CN**; 0.10 mmol), potassium thioacetate (0.30 mmol) and methylsulfonylmethane (0.10 mmol) were dissolved in degassed  $\text{H}_2\text{O}/\text{D}_2\text{O}$  (98:2, 1 mL), and the solution was adjusted to  $\text{pH } 9.0$  with  $\text{NaOH}/\text{HCl}$ . Potassium hexacyanoferrate(III) (0.90 mmol) was added, and the solution was stirred at room temperature for 20 minutes. After that time, the suspension was adjusted to  $\text{pH } 2.0$ , and the reaction was lyophilised. NMR analysis of the residue showed conversion of arginine nitrile **Arg-CN** to *N*-acetylarginine nitrile **Arg-CN** (89%).  $\text{NaSH}\cdot x\text{H}_2\text{O}$  (112 mg, 1.00 mmol) was added, the solution was adjusted to  $\text{pH } 9.0$  with  $\text{NaOH}/\text{HCl}$ , and incubated at room temperature whilst maintaining the solution at  $\text{pH } 9.0$  with  $\text{NaOH}/\text{HCl}$ . After 36 h, NMR analysis showed quantitative conversion of *N*-acetylarginine nitrile **Ac-Arg-CN** to *N*-acetylarginine thioamide **Ac-Arg-SNH<sub>2</sub>**. The resulting solution was adjusted to  $\text{pH } 9.0$  with  $\text{NaOH}/\text{HCl}$  and incubated at  $60\text{ }^\circ\text{C}$  for 96 h, after which time NMR analysis showed complete hydrolysis of *N*-acetylarginine thioamide **Ac-Arg-SNH<sub>2</sub>**. The solution was sparged of hydrogen sulfide with argon for 30 min whilst maintaining the solution at  $\text{pH } 7.0$  with  $\text{NaOH}/\text{HCl}$ . Glycine nitrile **Gly-CN** (0.103 mmol) was added, and the solution was adjusted to  $\text{pH } 9.0$  with  $\text{NaOH}/\text{HCl}$ . Potassium hexacyanoferrate(III) (0.063 mmol) was added, and the solution was stirred at room temperature for 20 minutes.  $\text{DMSO}-d_6$  (1 mL) was added dropwise whilst keep stirring the reaction in an ice bath, and NMR spectra were acquired. The presence of *N*-acetylarginylglycine nitrile **Ac-Arg-Gly-CN** (64%) was confirmed by  $^1\text{H}-^{13}\text{C}$  HMBC NMR analysis and high resolution mass spectrometry. The reaction mixtures were quantified by integration against methylsulfonylmethane as the internal standard.

Entry	Step	Product	Yield (%)
1	Acetylation	<b>Ac-Arg-CN</b>	99
2	Thiolysis	<b>Ac-Arg-SNH<sub>2</sub></b>	99
3	Hydrolysis	<b>Ac-Arg-SH</b>	63
4	Ligation	<b>Ac-Arg-Gly-CN</b>	64



**Experimental figure 78:** <sup>1</sup>H NMR (700 MHz, pH 9.0, 1.45 – 4.65 ppm) spectra to show the sequential acetylation, thiolysis, hydrolysis and oxidative ligation of arginine nitrile **Arg-CN** with glycine nitrile **Gly-CN**. **A:** Acetylation of **Arg-CN**, H<sub>2</sub>O/D<sub>2</sub>O 98:2; **B:** Thiolysis of **Ac-Arg-CN** to **Ac-Arg-SNH<sub>2</sub>**, H<sub>2</sub>O/D<sub>2</sub>O 98:2; **C:** Hydrolysis of **Ac-Arg-SNH<sub>2</sub>**, H<sub>2</sub>O/D<sub>2</sub>O 98:2; **E:** Coupling of **Ac-Arg-SH** with **Gly-CN**, H<sub>2</sub>O/D<sub>2</sub>O/DMSO-*d*<sub>6</sub> 49:1:50.



**Experimental figure 79:**  $^1\text{H}$  –  $^{13}\text{C}$  HMBC ( $^1\text{H}$  – 700 MHz [4.25 – 3.95 ppm],  $^{13}\text{C}$  – 176 MHz [100 – 180 ppm],  $\text{H}_2\text{O}/\text{D}_2\text{O}/\text{DMSO-}d_6$  50:50) spectrum to show the reaction of *N*-acetylgarginine thioacid (51 mM) and glycine nitrile (102 mM).

**Experimental figure 78-A** (acetylation,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, pH 9.0, partial assignment):

*N*-Acetylgarginine nitrile **Ac-Arg-CN** (•)  $\delta$  3.26 (t,  $J = 7.2$  Hz, 2H, (C5)–H<sub>2</sub>), 2.05 (s, 3H, (CO)CH<sub>3</sub>), 2.02 – 1.95 (m, 1H, (C3)–H), 1.95 – 1.88 (m, 1H, (C3)–H'), 1.79 – 1.71 (m, 2H, (C4)–H<sub>2</sub>).

**Experimental figure 78-B** (thiolysis,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, pH 9.0, partial assignment):

*N*-Acetylgarginine thioamide **Ac-Arg-SNH<sub>2</sub>** (▲)  $\delta$  4.57 (dd,  $J = 5.3, 8.8$  Hz, 1H, (C2)–H), 3.24 (t,  $J = 7.2$  Hz, 2H, (C5)–H<sub>2</sub>), 2.07 (s, 3H, (CO)CH<sub>3</sub>), 2.02 – 1.95 (m, 1H, (C3)–H), 1.85 – 1.77 (m, 1H, (C3)–H'), 1.76 – 1.68 (m, 1H, (C4)–H), 1.76 – 1.63 (m, 1H, (C4)–H').

**Experimental figure 78-C** (hydrolysis,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, pH 9.0, partial assignment):

*N*-Acetylgarginine thioacid **Ac-Arg-SH** (▼)  $\delta$  4.47 (dd,  $J = 4.9, 8.6$  Hz, 1H, (C2)–H), 3.24 (obs. t, 2H, (C5)–H<sub>2</sub>), 2.05 (s, 3H, (CO)CH<sub>3</sub>), 2.00 – 1.94 (m, 1H, (C3)–H), 1.77 – 1.69 (m, 1H, (C3)–H'), 1.69 – 1.60 (m, 1H, (C4)–H<sub>2</sub>).

**Experimental figure 78-D** (ligation, 700 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}/\text{DMSO-}d_6$  49:1:50, pH 9.0, partial assignment):

*N*-Acetylgarginylglycine nitrile **Ac-Arg-Gly-CN** (◆)  $\delta$  4.37 – 4.32 (m, 1H, Arg-(C2)–H), 4.22 (s, 2H, Gly<sub>2</sub>-(C2)–H<sub>2</sub>), 3.25 (obs. t, 2H, Arg-(C5)–H<sub>2</sub>), 2.02 (s, 3H, (CO)CH<sub>3</sub>), 2.01– 1.96 (obs. m, 1H, Arg-(C3)–H), 1.88 – 1.80 (obs. m, 1H, (C3)–H'), 1.80 – 1.70 (obs. m, 1H, (C4)–H), 1.65 – 1.58 (obs. m, 1H, (C4)–H').

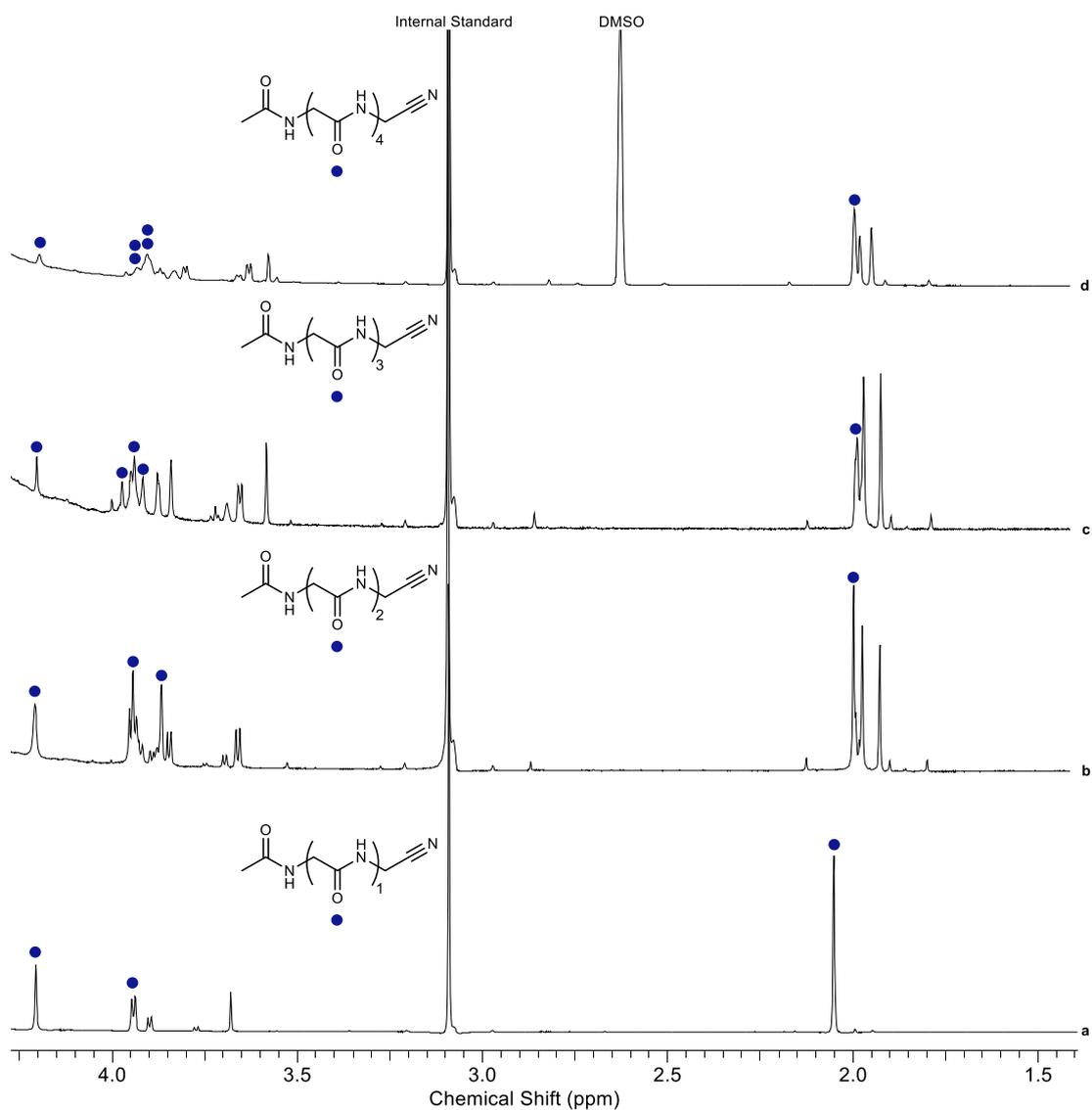
Glycine nitrile **Gly-CN** (■)  $\delta$  3.72 (s, 2H, (C2)–H<sub>2</sub>).

### **Iterative elongation of *N*-acetylglycine nitrile**

The homologation of *N*-acetylglycine nitrile **Ac-Gly-CN** was performed by iteration of the following protocol: **Ac-(Gly)<sub>n</sub>-CN** was dissolved in H<sub>2</sub>O/D<sub>2</sub>O (98:2, 100 mM) and Dowex® (SH<sup>-</sup> form) (0.1 mg/mmol **Ac-(Gly)<sub>n</sub>-CN**) was added. The suspension was adjusted to pH 9.0 and H<sub>2</sub>S was bubbled through the solution for 2 h. The reaction was stirred at room temperature for 6 h, filtered, adjusted to pH 9.0 with KOH/HCl. Quantitative conversion of **Ac-(Gly)<sub>n</sub>-CN** to **Ac-(Gly)<sub>n</sub>-SNH<sub>2</sub>** was observed by NMR. The solution was then heated at 60 °C for 24 h. Quantitative conversion of **Ac-(Gly)<sub>n</sub>-SNH<sub>2</sub>** was observed. Residual H<sub>2</sub>S was removed by sparging with argon for 30 min at pH 5.0. Glycine nitrile hydrochloride **Gly-CN·HCl** (1.2 eq.) was then added and the solution adjusted to pH 9.0 with KOH. Potassium hexacyanoferrate(III) (3.0 eq.) was then added. The solution was stirred at room temperature for 30 min. After this time, the reaction mixture was centrifuged, and the supernatant analysed by NMR.

The first iteration was performed using **Ac-Gly-CN** (981 mg, 10.0 mmol) and methylsulfonylmethane (MSM) (941.3 mg, 10.0 mmol) in H<sub>2</sub>O/D<sub>2</sub>O (98:2, 100 mL, 100 mM). After each iteration, <sup>1</sup>H-<sup>13</sup>C HMBC NMR spectra were acquired. The yield of **Ac-(Gly)<sub>n</sub>-CN** was determined by integration against the internal standard (MSM). The next iteration was carried out immediately thereafter on an aliquot of the reaction mixture (50% of the total volume). After four complete cycles, significant precipitation of **Ac-(Gly)<sub>5</sub>-CN** was observed. The solution was cooled to 0 °C and diluted 2-fold by slow addition of DMSO-*d*<sub>6</sub> (12.5 mL). The resulting solution was filtered sequentially through Dowex® (OH<sup>-</sup>) and (H<sup>+</sup>), and analysed by NMR spectroscopy. The presence of **Ac-(Gly)<sub>5</sub>-CN** was confirmed by signal amplification upon addition of a synthetic standard.

<b>Experimental table 11: iterative elongation of <i>N</i>-acetylglycine nitrile</b>				
<b>Iteration</b>	<b>Starting material</b>	<b>Product</b>	<b>Yield (step)</b>	<b>Yield (from Ac-Gly-CN)</b>
1	<b>Ac-Gly-CN</b> 10.00 mmol	<b>Ac-(Gly)<sub>2</sub>-CN</b>	71	71%
2	<b>Ac-(Gly)<sub>2</sub>-CN</b> 3.55 mmol	<b>Ac-(Gly)<sub>3</sub>-CN</b>	71	51
3	<b>Ac-(Gly)<sub>3</sub>-CN</b> 1.27 mmol	<b>Ac-(Gly)<sub>4</sub>-CN</b>	63	32
4	<b>Ac-(Gly)<sub>4</sub>-CN</b> 0.40 mmol	<b>Ac-(Gly)<sub>5</sub>-CN</b>	41	13



**Experimental figure 80:**  $^1\text{H}$  NMR (600 MHz, pH 9.0, 1.40 – 4.50 ppm) spectra to show the iterative elongation of *N*-acetylglycine nitrile **Ac-Gly-CN** through iterative thiolysis, hydrolysis, and ligation.

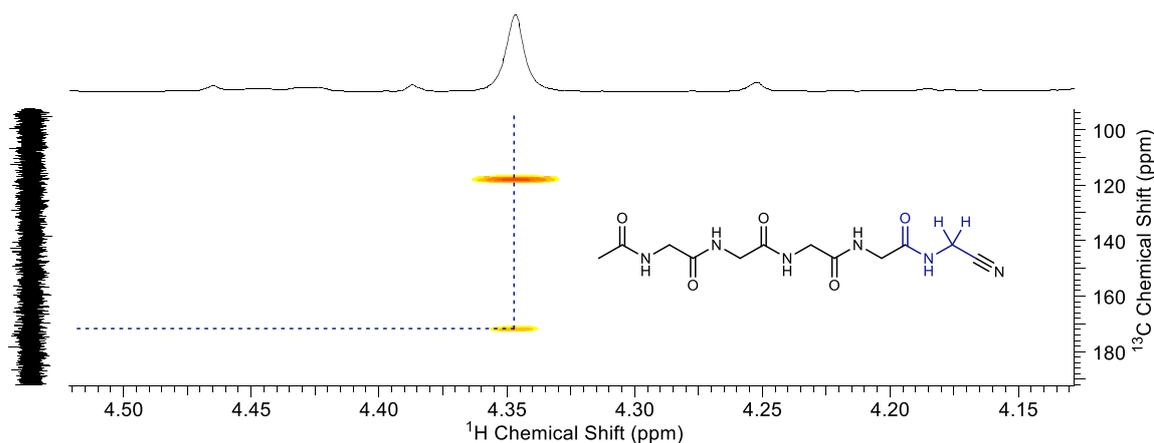
**Data for the ligation product after each iterative cycle**

**Iteration 1 – Experimental figure 80-a:** ( $^1\text{H}$  NMR, 600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, pH 9.0): *N*-Acetylglycylglycine nitrile **Ac-(Gly)<sub>2</sub>-CN** (•)  $\delta$  4.20 (s, 2H, Gly<sub>2</sub>-(C2)-H<sub>2</sub>), 3.95 (app. d,  $J$  = 6.9 Hz, 2H, Gly<sub>1</sub>-(C2)-H<sub>2</sub>), 2.05 (s, 3H, (CO)CH<sub>3</sub>).

**Iteration 2 – Experimental figure 80-b:** ( $^1\text{H}$  NMR, 600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, pH 9.0): *N*-Acetylglycylglycylglycine nitrile **Ac-(Gly)<sub>3</sub>-CN** (•)  $\delta$  4.20 (s, 2H, Gly<sub>3</sub>-(C2)-H<sub>2</sub>), 3.95 (obs. s, 2H, Gly<sub>2</sub>-(C2)-H<sub>2</sub>), 3.95 (s, 2H, Gly<sub>1</sub>-(C2)-H<sub>2</sub>), 1.98 (s, 3H, (CO)CH<sub>3</sub>).

**Iteration 3 – Experimental figure 80-c:** ( $^1\text{H}$  NMR, 600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, pH 9.0): *N*-Acetylglycylglycylglycylglycine nitrile **Ac-(Gly)<sub>4</sub>-CN** (•)  $\delta$  4.20 (s, 2H, Gly<sub>4</sub>-(C2)-H<sub>2</sub>), 3.98 (obs. s, 2H, Gly<sub>3</sub>-(C2)-H<sub>2</sub>), 3.95 (obs. s, 2H, Gly<sub>1</sub>-(C2)-H<sub>2</sub>), 3.94 (s, 2H, Gly<sub>2</sub>-(C2)-H<sub>2</sub>), 1.97 (obs. s, 3H, (CO)CH<sub>3</sub>).

**Iteration 4 – Experimental figure 80-d:** ( $^1\text{H}$  NMR, 600 MHz,  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-}d_6$  1:49:50): *N*-Acetylglycylglycylglycylglycylglycine nitrile **Ac-(Gly)<sub>5</sub>-CN** (•)  $\delta$  4.20 (s, 2H, Gly<sub>5</sub>-(C2)-H<sub>2</sub>), 3.93 (obs. s, 4H, Gly<sub>3</sub>-(C2)-H<sub>2</sub>), Gly<sub>4</sub>-(C2)-H<sub>2</sub>), 3.90 (obs. s, 4H, Gly<sub>2</sub>-(C2)-H<sub>2</sub>), Gly<sub>1</sub>-(C2)-H<sub>2</sub>), 1.99 (obs. s, 3H, (CO)CH<sub>3</sub>).



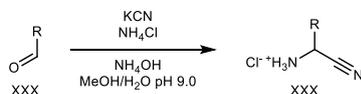
**Experimental figure 81:**  $^1\text{H}$  –  $^{13}\text{C}$  HMBC ( $^1\text{H}$  – 600 MHz [4.50 – 4.15 ppm],  $^{13}\text{C}$  – 151 MHz [100 – 190 ppm],  $\text{D}_2\text{O}/\text{H}_2\text{O}/\text{DMSO-}d_6$  1:49:50) spectrum to show the reaction of **Ac-(Gly)<sub>4</sub>-SH**, glycine nitrile **Gly-CN** and  $\text{K}_3[\text{Fe}(\text{CN})_6]$  ( $\text{D}_2\text{O}/\text{H}_2\text{O}$  (2:98, 1 mL) at pH 9.0 and room temperature. After 30 min, the suspension was adjusted to pH 9.0, diluted with  $\text{DMSO-}d_6$  (1 mL), and NMR spectra were acquired. The characteristic  $^3J_{\text{CH}}$  coupling between Gly<sub>5</sub>-CN nitrile proton resonance (4.35 ppm, s, 2H, Gly<sub>5</sub>-(C2)-H<sub>2</sub>), and Gly<sub>4</sub> amide carbon resonance (171.9 ppm, Gly<sub>4</sub>-(C1)) that demonstrates peptide ligation is highlighted with overlaid blue dotted lines.

**HRMS (ES<sup>+</sup>):** calcd. for  $[\text{C}_{12}\text{H}_{18}\text{N}_6\text{O}_5+\text{H}]^+$  327.1411; observed: 327.1422.

### 9. 3. Preparation of intermediates and synthetic standards

#### 9. 3. 1. Synthesis and characterisation of aminonitriles

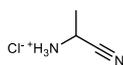
General procedure E: Strecker synthesis of aminonitrile hydrochlorides



Potassium cyanide (4.04 g, 31.25 mmol) and ammonium chloride (6.69 g, 125 mmol) were dissolved in 13 M ammonium hydroxide (25 mL) at room temperature. The specified aldehyde (25 mmol) dissolved in methanol (10 mL) was added dropwise over 5 min, and the reaction was stirred at room temperature for 2 h. Methanol and residual ammonia were evaporated *in vacuo*. The crude mixture was then extracted with ether (5 × 50 mL). The organics were washed with brine (sat., 50 mL), dried over MgSO<sub>4</sub> and filtered. 2 M HCl in diethyl ether (15 mL, 30 mmol) was then added to the organics at 0 °C. The resultant precipitate was filtered, washed with Et<sub>2</sub>O (3 × 100 mL), and dried to afford aminonitrile hydrochloride **AA-CN·HCl** as a white powder.

Entry	Residue	R	Yield AA-CN
1	Ala	CH <sub>3</sub>	31%
2	Val	CH-(CH <sub>3</sub> ) <sub>2</sub>	91%
3	Leu	CH <sub>2</sub> -CH <sub>2</sub> -(CH <sub>3</sub> ) <sub>2</sub>	87%
4	Ile	CH-(CH <sub>3</sub> )-CH <sub>2</sub> -(CH <sub>3</sub> )	63%
5	Met	CH <sub>2</sub> -CH <sub>2</sub> -SCH <sub>3</sub>	58%
6	Phe	CH <sub>2</sub> -Ph	95%
7	β-Ala	CH <sub>2</sub> -CH <sub>3</sub>	43%
8	Ser	CH <sub>2</sub> -OH	44%

Alanine nitrile hydrochloride **Ala-CN·HCl**



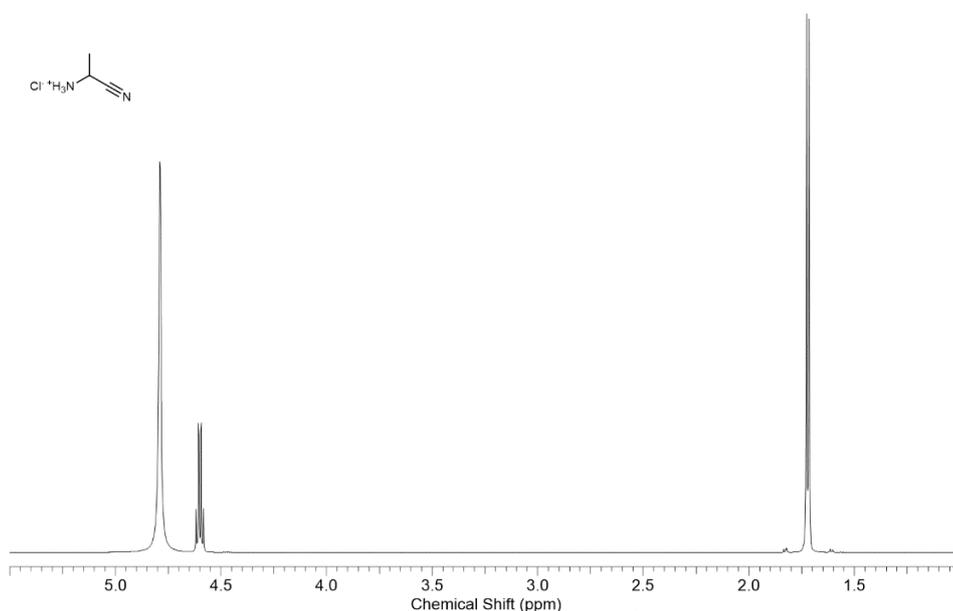
Alanine nitrile hydrochloride **Ala-CN·HCl** was prepared following *General procedure E* using acetaldehyde (1.10 g, 25.0 mmol), affording alanine nitrile hydrochloride **Ala-CN·HCl** as a white solid (815 mg, 7.86 mmol, 31%).

$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  4.60 (q,  $J = 7.1$  Hz, 1H, (C2)–H), 1.72 (d,  $J = 7.1$  Hz, 3H, (C3)– $\text{H}_3$ ).

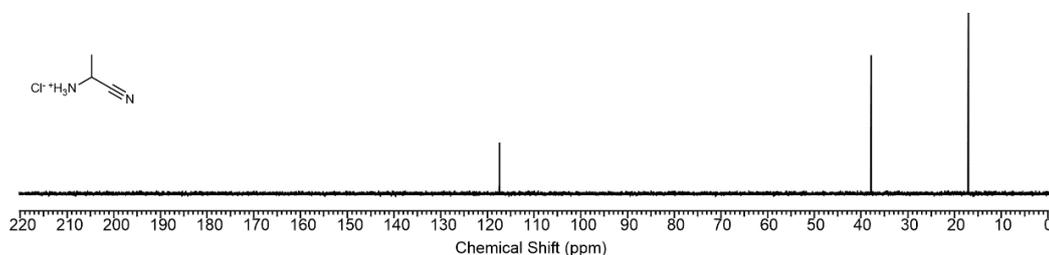
$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  114.4 (C1), 37.8 (C2), 17.1 (C3).

IR (solid,  $\text{cm}^{-1}$ ): 2962, 2871, 2696, 2634, 2495, 2023, 1618, 1585, 1500.

M.p. 135 – 136 °C (decomposes), lit. m.p. 135 – 137 °C (decomposes)<sup>6</sup>.



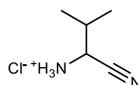
**Experimental figure 82:**  $^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ , 5.50 – 1.00 ppm) spectrum to show alanine nitrile hydrochloride **Ala-CN·HCl**.



**Experimental figure 83:**  $^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ , 220 – 0 ppm) spectrum to show alanine nitrile hydrochloride **Ala-CN·HCl**.

<sup>6</sup> K. Kawashiro *et al.*, Bull. Chem. Soc. Jpn., 1977, vol. 50, p. 2956 – 2960.

Valine nitrile hydrochloride **Val-CN·HCl**



Valine nitrile hydrochloride **Val-CN·HCl** was prepared following *General procedure E* using isobutyraldehyde (1.80 g, 2.28 mL, 25 mmol), affording valine nitrile hydrochloride **Val-CN·HCl** as a white solid (3.38 g, 22.74 mmol, 91%).

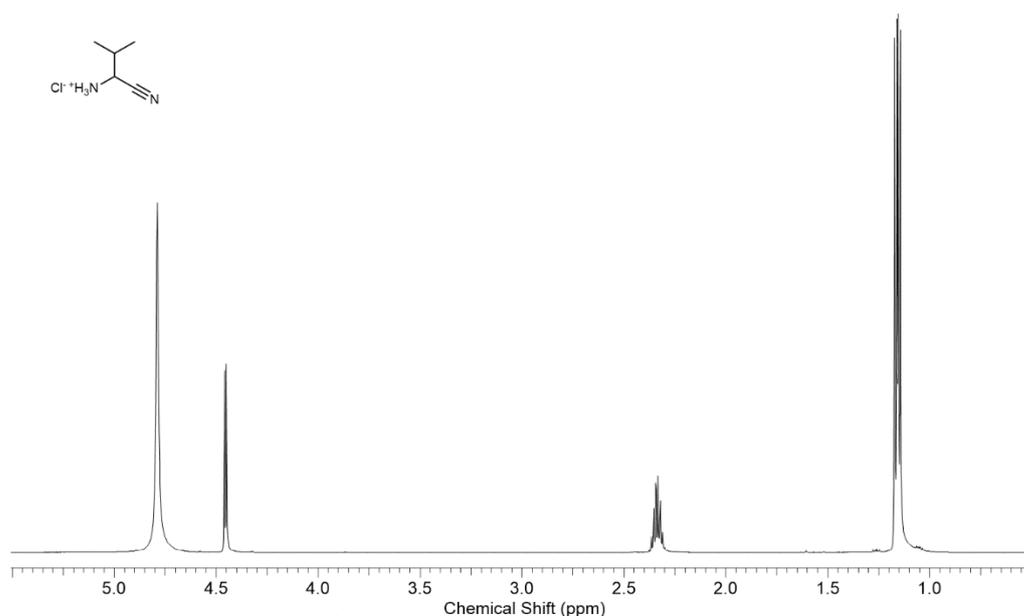
**<sup>1</sup>H NMR** (600 MHz, D<sub>2</sub>O): δ 4.45 (d, *J* = 5.9 Hz, 1H, (C2)-H), 2.38 – 2.29 (m, 1H, (C3)-H), 1.17 (d, *J* = 6.8 Hz, 3H, (C4)-H<sub>3</sub>), 1.15 (d, *J* = 6.8 Hz, 3H, (C4')-H<sub>3</sub>).

**<sup>13</sup>C NMR** (151 MHz, D<sub>2</sub>O): δ 115.7 (C1), 48.1 (C2), 30.3 (C3), 18.6 (C4), 17.1 (C4').

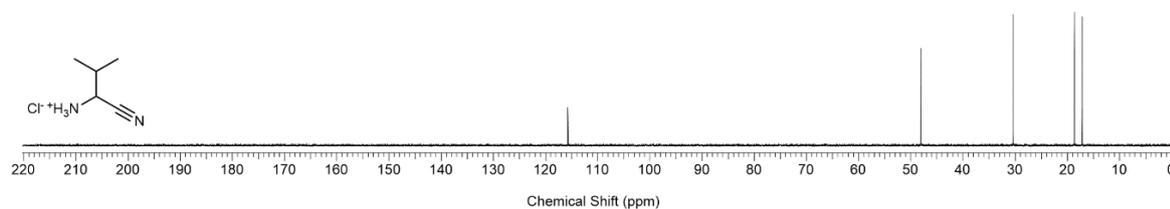
**HRMS** (CI): calcd. for [C<sub>5</sub>H<sub>10</sub>N<sub>2</sub> + NH<sub>4</sub>]<sup>+</sup>: 116.1182; observed: 116.1183.

**IR** (solid): 2848, 2636, 2574, 1583, 1556, 1516 cm<sup>-1</sup>.

**M.p.** 203.3 – 206.5 °C (decomposes), lit. m.p. 200 °C (decomposes)<sup>7</sup>.



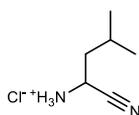
**Experimental figure 84:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 5.50 – 0.50 ppm) spectrum to show valine nitrile hydrochloride **Val-CN·HCl**.



**Experimental figure 85:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 220 – 0 ppm) spectrum to show valine nitrile hydrochloride **Val-CN·HCl**.

<sup>7</sup> J. Lukszo, R. Tyka, Pol. J. Chem., 1978, vol. 52, p. 959 – 963.

Leucine nitrile hydrochloride **Leu-CN·HCl**



Leucine nitrile hydrochloride **Leu-CN·HCl** was prepared following *General procedure E* using isovaleraldehyde (2.15 g, 2.74 mL, 25.0 mmol), affording leucine nitrile hydrochloride **Leu-CN·HCl** as a white solid (3.20 g, 21.5 mmol, 87%).

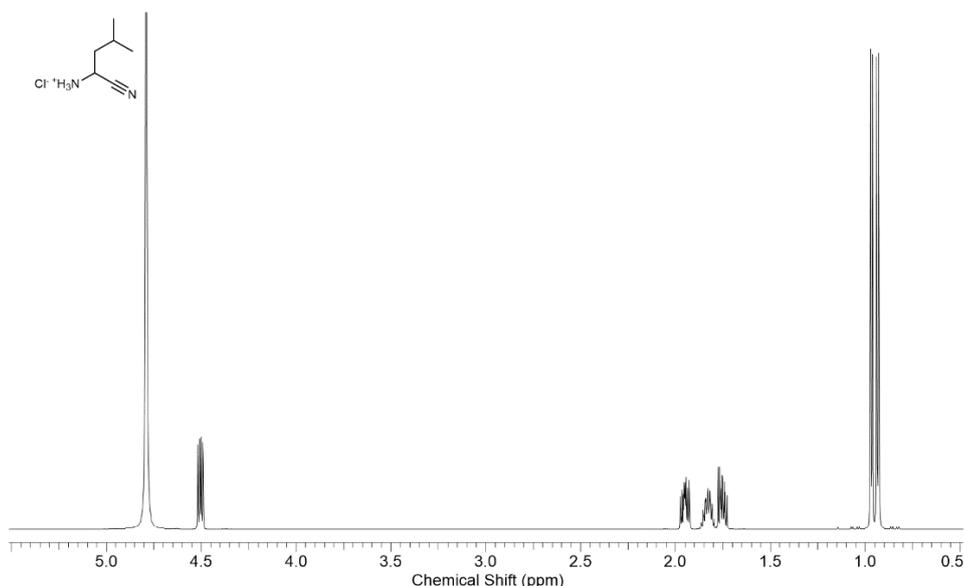
$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  4.50 (dd,  $J = 5.9, 10.2$  Hz, 1H, (C2)-H), 1.89 – 1.79 (m, 1H, (C4)-H), 1.78 – 1.72 (m, 2H, (C3)-H<sub>2</sub>), 0.96 d,  $J = 6.3$  Hz, 3H, (C5)-H<sub>3</sub>), 0.93 (d,  $J = 6.3$  Hz, 3H, (C5')-H<sub>3</sub>).

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  116.7 (C1), 40.1 (C2), 39.4 (C3), 25.2 (C4), 22.2 (C5), 21.1 (C5').

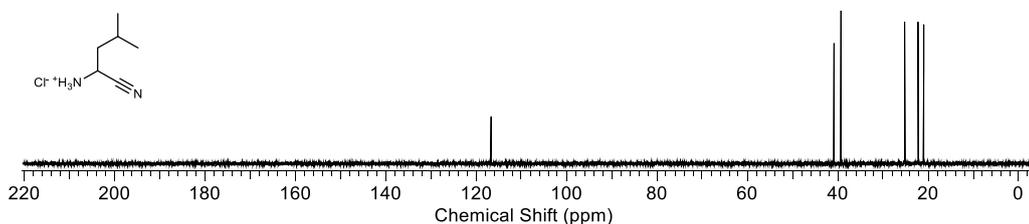
**HRMS** ( $\text{ES}^+$ ): calcd. for  $[\text{C}_6\text{H}_{13}\text{N}_2 + \text{H}]^+$ : 113.0715; observed: 113.0719.

**IR** (solid): 2954, 2870, 2665, 2561, 1622, 1589, 1570  $\text{cm}^{-1}$ .

**M.p.** 207.0 – 208.5 °C (decomposes), lit. m.p. 203 – 208 °C (decomposes)<sup>8</sup>.



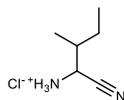
**Experimental figure 86:**  $^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ , 5.50 – 0.50 ppm) spectrum to show leucine nitrile hydrochloride **Leu-CN·HCl**.



**Experimental figure 87:**  $^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ , 220 – 0 ppm) spectrum to show leucine nitrile hydrochloride **Leu-CN·HCl**.

<sup>8</sup> K. Kawashiro *et al.*, *Bull. Chem. Soc. Jpn.*, 1977, vol. 50, p. 2956 – 2960.

Isoleucine nitrile hydrochloride **Ile-CN·HCl**



Isoleucine nitrile hydrochloride **Ile-CN·HCl** was prepared following *General procedure E* using 2-methylbutanal (2.15 g, 2.69 mL, 25 mmol), affording isoleucine nitrile hydrochloride **Ile-CN·HCl** as a white solid (2.29 g, 15.65 mmol, 63%).

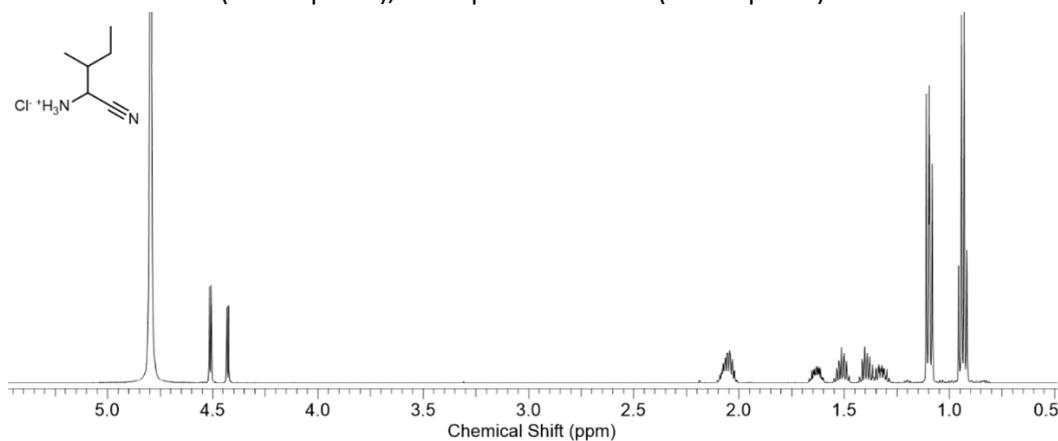
**<sup>1</sup>H NMR** (600 MHz, D<sub>2</sub>O, 56:44 mixture of diastereoisomers a/b): diastereoisomer a: δ 4.50 (d, *J* = 5.1 Hz, 1H, (C2)-H), 2.09 – 2.00 (m, 1H, (C3)-H), 1.66 – 1.60 (m, 1H, (C4)-H), 1.54 – 1.47 (m, 1H, (C4')-H), 1.06 (d, *J* = 6.8 Hz, 3H, (C3)-CH<sub>3</sub>), 0.92 (m, 3H, (C5)-H<sub>3</sub>); diastereoisomer b: δ 4.35 (d, *J* = 6.0 Hz, 1H, (C2)-H), 2.09 – 2.00 (m, 1H, (C3)-H), 1.63 – 1.56 (m, 1H, (C4)-H), 1.28 – 1.17 (m, 1H, (C4)-H), 1.01 (d, *J* = 6.8 Hz, 3H, (C3)-CH<sub>3</sub>), 0.92 (m, 3H, (C5)-H<sub>3</sub>).

**<sup>13</sup>C NMR** (151 MHz, D<sub>2</sub>O, 56:44 mixture of diastereoisomers a/b): diastereoisomer a: δ 115.4 (C1), 46.8 (C2), 36.4 (C3), 24.6 (C4), 14.2 (C3)-CH<sub>3</sub>, 10.7 (C5); diastereoisomer b: δ 115.9 (C1), 47.1 (C2), 36.4 (C3), 26.2 (C4), 14.8 (C3)-CH<sub>3</sub>, 10.9 (C5).

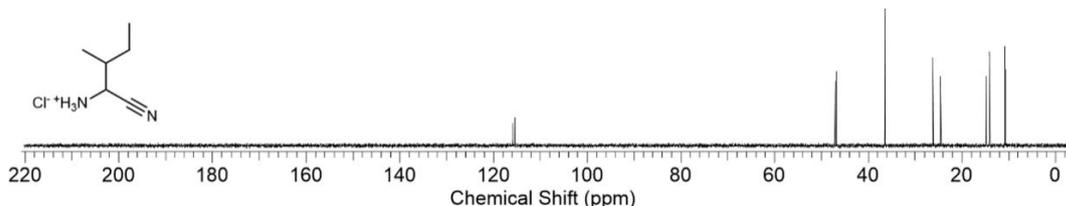
**HRMS** (ES<sup>+</sup>): calcd. for [C<sub>6</sub>H<sub>13</sub>N<sub>2</sub> + H]<sup>+</sup>: 113.0715; observed: 113.0715.

**IR** (solid): 2964, 2933, 2916, 2877, 2858, 1583, 1558 cm<sup>-1</sup>.

**M.p.** 199.3 – 201.0 °C (decomposes), lit. m.p. 190 – 195 °C (decomposes)<sup>9</sup>.



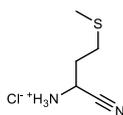
**Experimental figure 88:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 5.50 – 0.50 ppm) spectrum to show isoleucine nitrile hydrochloride **Ile-CN·HCl**.



**Experimental figure 89:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 220 – 0 ppm) spectrum to show isoleucine nitrile hydrochloride **Ile-CN·HCl**.

<sup>9</sup> Gallagher *et al.*, J. Chem. Soc., 1952, p. 4870 – 4873.

*Methionine nitrile hydrochloride* **Met-CN·HCl**



Methionine nitrile hydrochloride **Met-CN·HCl** was prepared following *General procedure E* using methional (2.60 g, 2.50 mL, 25.0 mmol), affording methionine nitrile hydrochloride **Met-CN·HCl** as an off-white solid (2.40 g, 14.38 mmol, 58%).

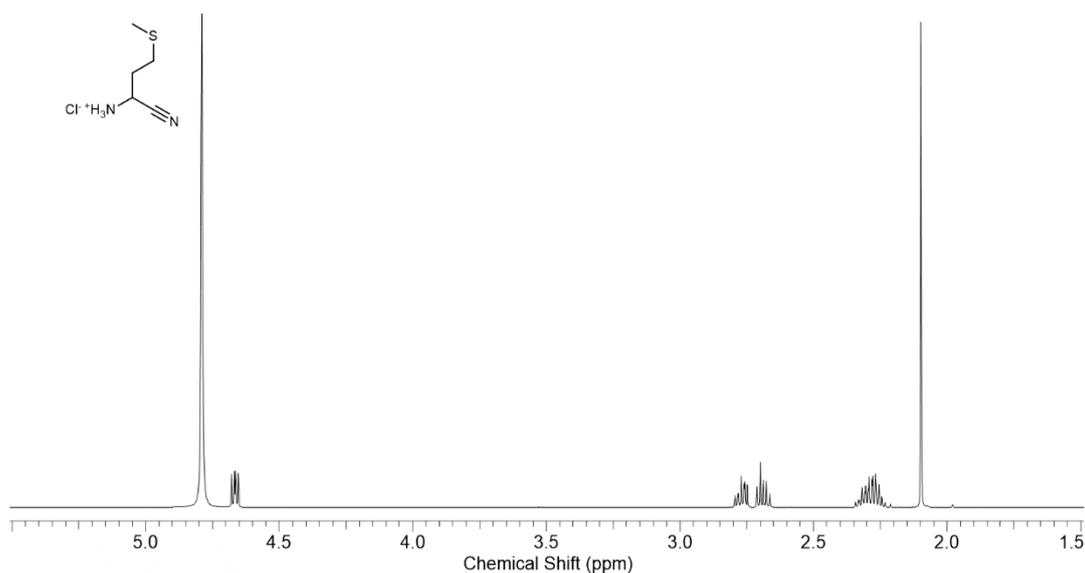
$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  4.67 (dd,  $J = 5.9, 8.9$  Hz, 1H, (C2)-H), 2.84 (ddd,  $J = 5.7, 7.7, 13.6$  Hz, 1H, (C4)-H), 2.76 (m, 1H, (C4')-H), 2.35 – 2.23 (m, 2H, (C3)-H<sub>2</sub>), 2.10 (s, 3H, (SCH<sub>3</sub>)).

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  118.3 (C1), 41.1 (C2), 30.1 (C3), 28.9 (C4), 14.5 (SCH<sub>3</sub>).

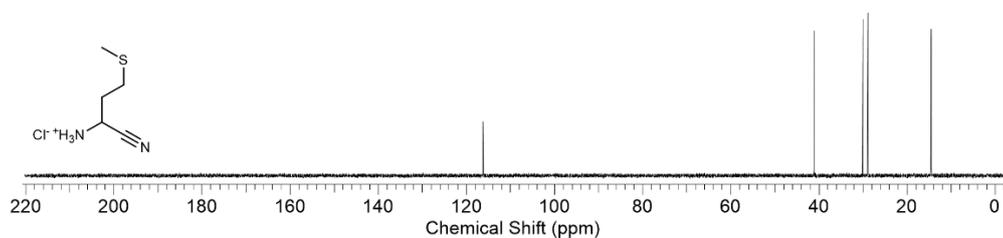
**HRMS** (ESI):  $[\text{M} + \text{H}]^+$  calcd. for  $\text{C}_5\text{H}_{11}\text{N}_2\text{S}$ : 131.0643; observed: 131.0643.

**IR** (solid): 2918, 2280, 1600, 1436  $\text{cm}^{-1}$ .

**M.p.** 120 – 122 °C (decomposes), lit. m.p. 112 – 115 °C (decomposes)<sup>10</sup>.



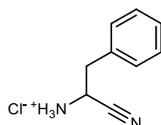
**Experimental figure 90:**  $^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ , 5.80 – 0.00 ppm) spectrum to show methionine nitrile hydrochloride **Met-CN·HCl**.



**Experimental figure 91:**  $^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ , 220 – 0 ppm) spectrum to show methionine nitrile hydrochloride **Met-CN·HCl**.

<sup>10</sup> K. Kawashiro *et al.*, *Bull. Chem. Soc. Jpn.*, 1977, vol. 50, p. 2956 – 2960.

Phenylalanine nitrile hydrochloride **Phe-CN·HCl**



Methionine nitrile hydrochloride **Phe-CN·HCl** was prepared following *General procedure E* using phenylacetaldehyde (3.00 g, 2.78 mL, 25.0 mmol), affording phenylalanine nitrile hydrochloride **Phe-CN·HCl** as a white solid (4.35 g, 23.84 mmol, 95%).

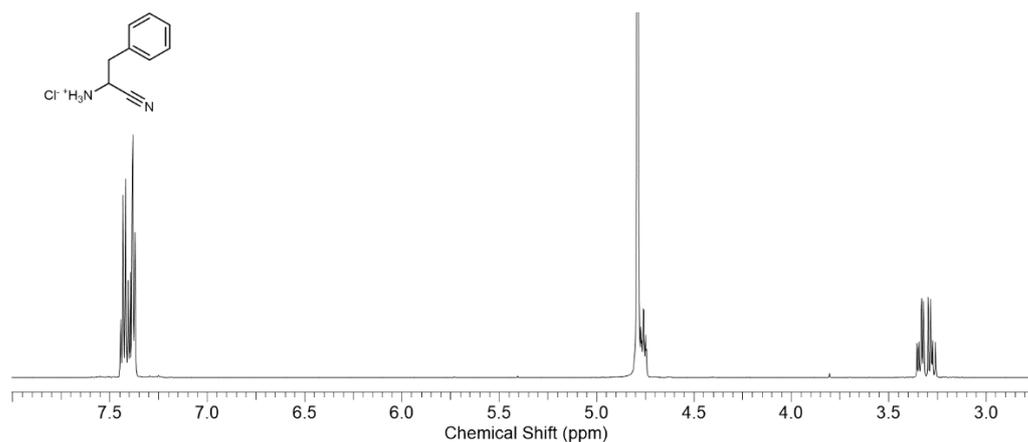
$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  7.45 – 7.36 (m, 5H, Ar), 4.78 – 4.74 (m, 1H, (C2)-H), 3.36 (ABX,  $J = 13.8$ , 6.4, 1H, (C3)-H), 3.28 (ABX,  $J = 13.8$ , 6.4, 1H, (C3)-H').

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  133.1 (Ar), 130.3 (Ar  $\times$  2), 130.0 (Ar  $\times$  2), 129.2 (Ar), 116.2 (C1), 43.3 (C2), 36.6 (C3).

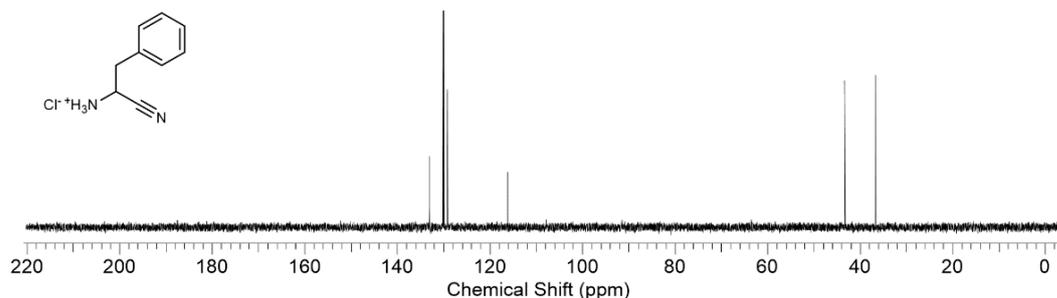
**HRMS** ( $\text{ES}^+$ ): calcd. for  $[\text{C}_9\text{H}_{10}\text{N}_2 + \text{H}]^+$ : 147.0922; observed: 147.0923.

**IR** (solid): 3200, 2927, 2032  $\text{cm}^{-1}$ .

**M.p.** 163.5 – 165.1  $^\circ\text{C}$  (decomposes), lit. m.p. 112 – 115  $^\circ\text{C}$  (decomposes)<sup>11</sup>.



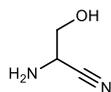
**Experimental figure 92:**  $^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ , 8.00 – 2.75 ppm) spectrum to show phenylalanine nitrile hydrochloride **Phe-CN·HCl**.



**Experimental figure 93:**  $^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ , 220 – 0 ppm) spectrum to show phenylalanine nitrile hydrochloride **Phe-CN·HCl**.

<sup>11</sup> K. Kawashiro *et al.*, Bull. Chem. Soc. Jpn., 1977, vol. 50, p. 2956 – 2960.

**Serine nitrile Ser-CN**



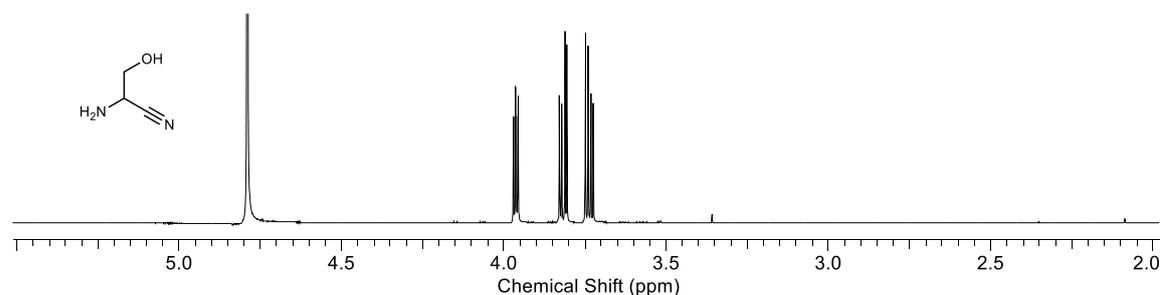
Serine nitrile **Ser-CN** was prepared following *General procedure E* using glycolaldehyde dimer (1.50 g, 25.0 mmol), affording serine nitrile **Ser-CN** as a white solid (939 mg, 10.91 mmol, 44%).

**<sup>1</sup>H NMR** (600 MHz, D<sub>2</sub>O): δ 3.96 (dd, *J* = 5.8, 4.7 Hz, 1H, (C2)-H), 3.81 (ABX, *J* = 15.8, 5.8 Hz, 2H, (C3)-H), 3.74 (ABX, *J* = 15.8, 5.8 Hz, 2H, (C3)-H').

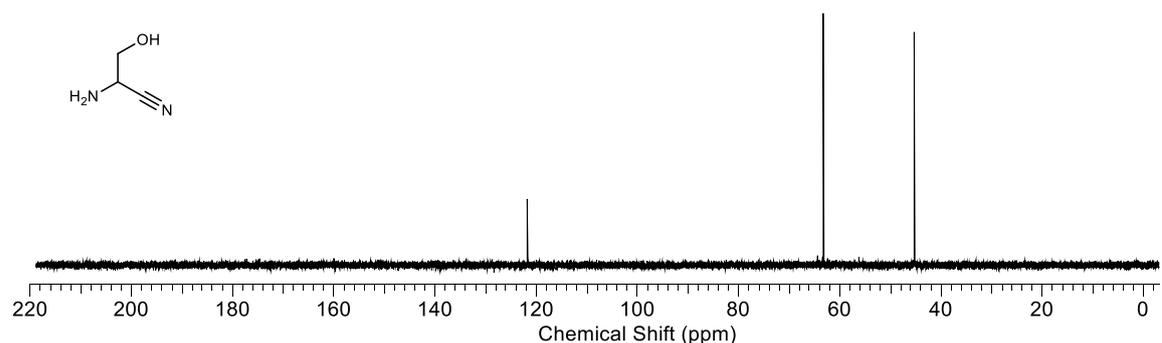
**<sup>13</sup>C NMR** (176 MHz, D<sub>2</sub>O): δ 121.8 (C1), 63.2 (C3), 45.2 (C2).

**IR** (solid): 3330, 3281, 3184, 2935, 2839, 2232, 1739, 1606 cm<sup>-1</sup>.

**M.p.** 53.5 – 54.0 °C, lit. m.p. 58.5 °C<sup>12</sup>.



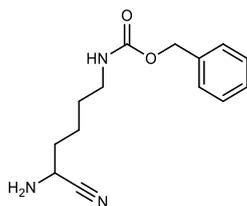
**Experimental figure 94:** <sup>1</sup>H NMR spectrum (700 MHz, D<sub>2</sub>O, 5.50 – 2.00 ppm) to show serine nitrile **Ser-CN**.



**Experimental figure 95:** <sup>13</sup>C NMR (176 MHz, D<sub>2</sub>O, 220 – 0 ppm) spectrum to show serine nitrile **Ser-CN**.

<sup>12</sup> E. Wagner, Y.-B., Xiang, K. Baumann, J. Gueck A. Eschenmoser, *Helv. Chim. Acta*, 1990, vol. 73, p. 1391 – 1409.

*ε*-Cbz-lysine nitrile **ε-Cbz-Lys-CN**



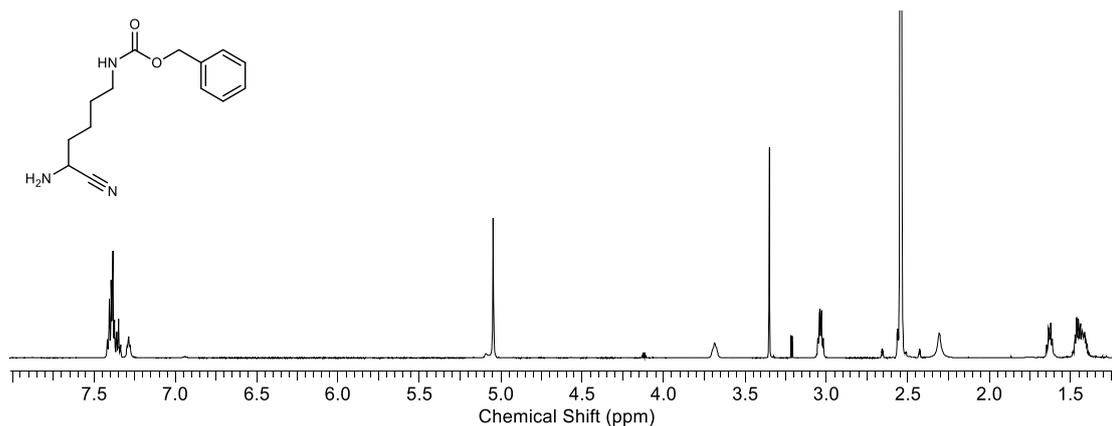
To a solution of oxalyl chloride (1.63 ml, 19.24 mmol) in dry dichloromethane (50 ml) under argon was slowly added a solution of dimethylsulfoxide (2.6 ml, 36.40 mmol) in dry dichloromethane (20 ml), and the solution was stirred for 5 min at -70 °C. After that time, a solution of benzyl (5-hydroxypentyl)carbamate (2.00 g, 8.43 mmol) in dry dichloromethane (50 ml) was added dropwise over 5 min. The mixture was stirred for 25 min at -70°C, and triethylamine (22.40 ml, 16.88 mmol) was added over 5 min. The reaction mixture was stirred for 20 min at -70°C, and for a further 15 min at room temperature. The reaction was then quenched by addition of iced water (200 mL). The aqueous phase was extracted with diethyl ether (5 × 100 ml). The combined organic layers were dried over MgSO<sub>4</sub>, filtered and concentrated *in vacuo* to afford benzyl (5-oxopentyl)carbamate (1.98 g) as a yellow oil, which was used in the next step without further purification. *ε*-Cbz-Lysine nitrile **ε-Cbz-Lys-CN** was prepared following *General procedure A* using benzyl (5-oxopentyl)carbamate (1.98 g, assumed 8.41 mmol), affording *ε*-Cbz-Lysine nitrile **ε-Cbz-Lys-CN** as a clear oil (1.35 g, 5.17 mmol, 61%).

<sup>1</sup>H NMR (600 MHz, DMSO-d<sub>6</sub>): δ 7.42 – 7.33 (m, 5H, Ar), 7.29 (t, *J* = 5.7 Hz, 1H, NH), 5.02 (s, 2H, (Ar)-CH<sub>2</sub>), 3.69 (m, 1H, (C2)-H), 3.03 (q, *J* = 6.2 Hz, 2H, (C6)-H<sub>2</sub>), 2.31 (br s, 2H, NH<sub>2</sub>), 1.63 (m, 2H, (C3)-H<sub>2</sub>), 1.49 – 1.38 (m, 4H, (C4)-H<sub>2</sub> + (C5)-H<sub>2</sub>).

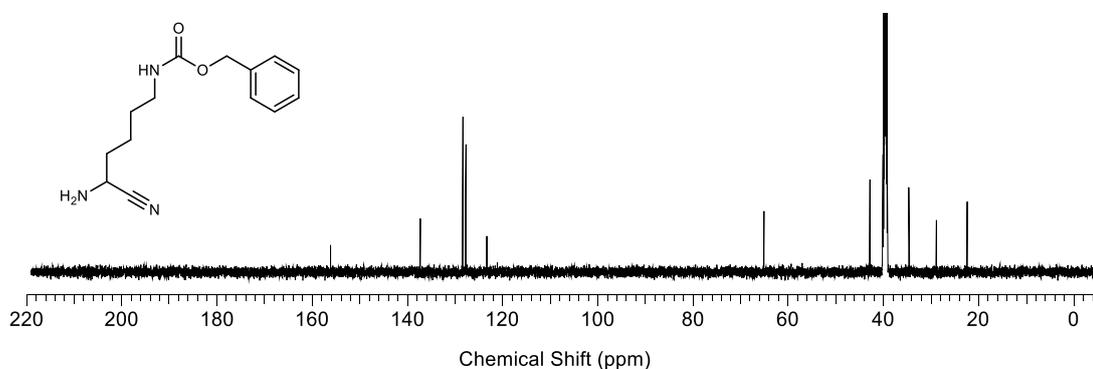
<sup>13</sup>C NMR (151 MHz, DMSO-d<sub>6</sub>): δ 156.1 (CO), 137.3 (Ar), 128.4 (Ar), 127.8 (Ar), 127.7 (Ar), 123.3 (C1), 65.1 ((Ar)-CH<sub>2</sub>), 42.8 (C2), 40.12 (C6), 34.6 (C3), 28.9 (C5), 22.4 (C4).

HRMS (ES<sup>+</sup>): calcd. for [C<sub>14</sub>H<sub>19</sub>N<sub>3</sub>O<sub>2</sub> + H]<sup>+</sup>: 262.1550; observed: 262.1548.

IR (solid): 3018, 2939, 2865, 2234, 1517 cm<sup>-1</sup>.

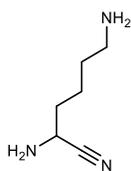


**Experimental figure 96:**  $^1\text{H}$  NMR spectrum (600 MHz,  $\text{DMSO-d}_6$ , 8.00 – 1.25 ppm) to show  $\delta$ -Cbz-lysine nitrile  $\epsilon$ -Cbz-Lys-CN.



**Experimental figure 97:**  $^{13}\text{C}$  NMR (151 MHz,  $\text{DMSO-d}_6$ , 220 – 0 ppm) spectrum to show  $\epsilon$ -Cbz-lysine nitrile  $\epsilon$ -Cbz-Lys-CN.

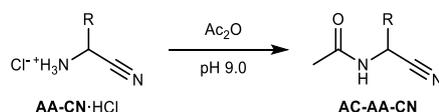
#### Lysine nitrile **Lys-CN**



To a suspension of sodium iodide (2.14 g, 14.33 mmol) in dry acetonitrile (10 ml) under argon was added chlorotrimethylsilane (1.21 ml, 9.55 mmol), and the mixture was stirred for 30 min at room temperature.  $\epsilon$ -Cbz-Lysine nitrile  **$\epsilon$ -Cbz-Lys-CN** (500 mg, 1.91 mmol) was added dropwise over 5 min, and the reaction mixture was stirred for 4 h at room temperature. After that time, water (1.91 ml) and dichloromethane (20 ml) were added to the reaction. The two layers were separated, and the aqueous layer was washed with dichloromethane (50  $\times$  10 ml). The aqueous layer was collected and evaporated under reduced pressure (100 mbar, RT) to remove the remaining traces of organic solvents to afford lysine nitrile **Lys-CN** which was stored as a 1 M stock solution in water (243 mg, 1.91 mmol, > 99%).

### 9. 3. 2. Synthesis and characterisation of *N*-acetyl aminonitriles

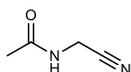
General procedure F: Preparative acetylation of aminonitriles



Aminonitrile hydrochloride **AA-CN·HCl** (10 mmol) was dissolved in H<sub>2</sub>O (50 mL, 0.2 M) and the solution was adjusted to pH 9.0 with NaOH. Acetic anhydride (3.06 g, 2.86 mL, 30 mmol) was added dropwise over 10 min, whilst maintaining the solution at pH 9 with NaOH/HCl. The reaction mixture was concentrated *in vacuo* and the residue purified by flash column chromatography (SiO<sub>2</sub>; eluting with a gradient of EtOAc/petroleum ether (40-60) 0:100 → 100:0) to afford *N*-acetylamino nitrile **Ac-AA-CN**.

Entry	Residue	Isolated yield (%)
1	Gly	> 99
2	Ala	> 99
3	Arg	-
4	Leu	98
5	Ile	95
6	Lys	-
7	Met	98
8	Phe	93
9	Pro	74
10	Ser	94
11	Val	96

*N*-Acetylglycine nitrile **Ac-Gly-CN**



*N*-Acetylglycine nitrile **Ac-Gly-CN** was prepared following *General procedure F* using glycine nitrile hydrochloride **Gly-CN**·HCl (925 mg, 10.0 mmol) afforded *N*-acetylglycine nitrile **Ac-Gly-CN** as a white, crystalline solid (973 mg, 9.92 mmol, > 99%).

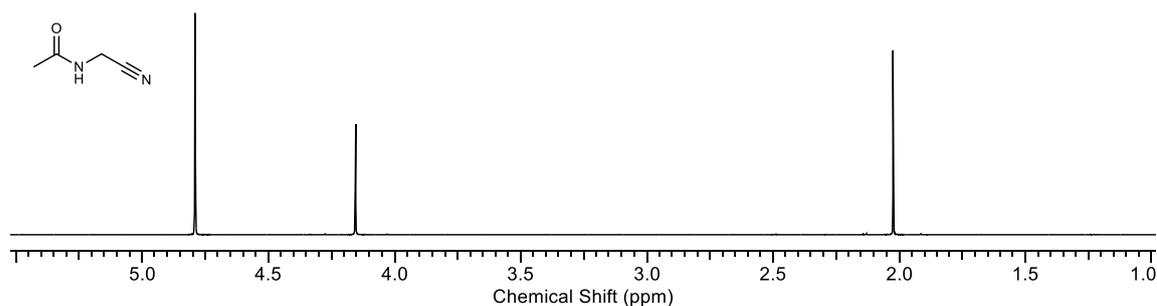
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.15 (s, 2H, (C2)-H<sub>2</sub>), 2.03 (s, 3H, (CO)CH<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 175.4 (CO), 117.8 (C1), 28.3 (C2), 22.2 (CO)CH<sub>3</sub>.

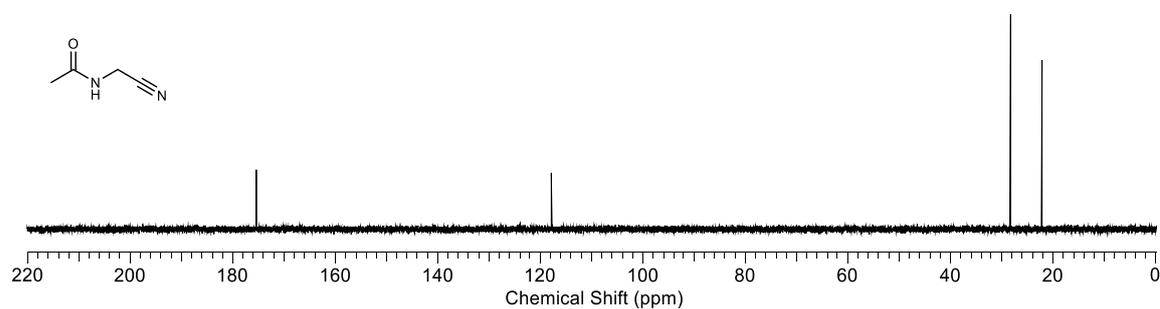
HRMS (CI) calcd. for [C<sub>4</sub>H<sub>6</sub>N<sub>2</sub>O+NH<sub>4</sub>]<sup>+</sup> 116.0818; observed: 116.0819.

IR (solid): 3226, 3056, 2989, 2935, 2843, 2650, 2031, 1635, 1448, 1415 cm<sup>-1</sup>.

M.p. 76.3–77.0 °C ; lit. m.p. 77 – 79 °C.<sup>13</sup>



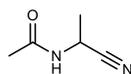
**Experimental figure 98:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 1.00 – 5.50 ppm) spectrum to show *N*-acetylglycine nitrile **Ac-Gly-CN**.



**Experimental figure 99:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show *N*-acetylglycine nitrile **Ac-Gly-CN**.

<sup>13</sup> C. Elias, M. E. Gelpi, R. A.-J. Cadenas, Carbohyd. Chem. 1995 vol. 14, 8 p. 1209 – 1216.

*N*-Acetylalanine nitrile **Ac-Ala-CN**



*N*-acetylalanine nitrile **Ac-Ala-CN** was prepared following *General procedure F* using alanine nitrile hydrochloride **Ala-CN**·HCl (1.07 g, 10 mmol) afforded *N*-acetylalanine nitrile **Ac-Ala-CN** as a white, crystalline solid (1.12 g, 9.99 mmol, > 99%).

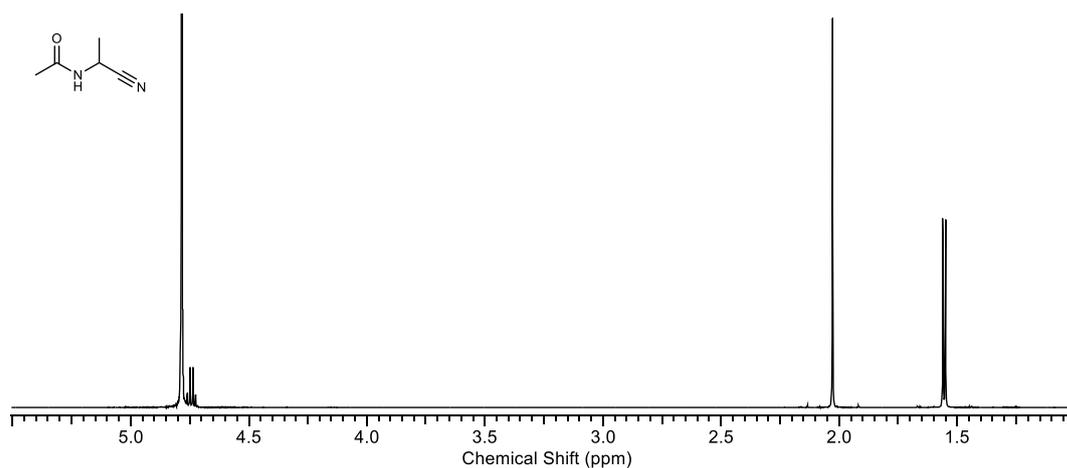
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.75 (q, *J* = 7.2 Hz, 1H, (C2)–H), 2.03 (s, 3H, (CO)CH<sub>3</sub>), 1.56 (d, *J* = 7.2 Hz, 3H, (C3)–H<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 174.5 (CO), 120.6 (C1), 37.0 (C2), 22.2 ((CO)CH<sub>3</sub>), 17.2 (C3).

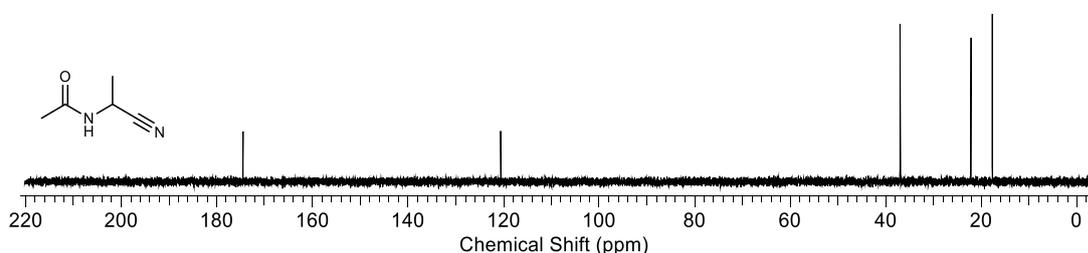
HRMS (EI<sup>+</sup>): calcd. for [C<sub>5</sub>H<sub>8</sub>N<sub>2</sub>O]<sup>+</sup> 112.0631; observed: 112.0631.

IR (solid): 3238, 3048, 3011, 2954, 2838, 2807, 2243, 1734, 1633, 1357 cm<sup>-1</sup>.

M.p. 93.5 – 95.2 °C; lit. m.p. 98–101 °C.<sup>14</sup>



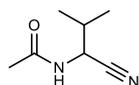
**Experimental figure 100:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 1.00 – 5.50 ppm) spectrum to show *N*-acetylalanine nitrile **Ac-Ala-CN**.



**Experimental figure 101:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show *N*-acetylalanine nitrile **Ac-Ala-CN**.

<sup>14</sup> T. Shirai, S. Kurashige, *J. Synth. Org. Chem. Jpn.* 1972 vol. 30, p. 1019 – 1023.

*N*-Acetylvaline nitrile **Ac-Val-CN**



*N*-Acetylvaline nitrile **Ac-Val-CN** was prepared following *General procedure F* using valine nitrile hydrochloride **Val-CN·HCl** (1.35 g, 10 mmol) afforded *N*-acetylvaline nitrile **Ac-Val-CN** as a white solid (1.40 g, 9.97 mmol, 96%).

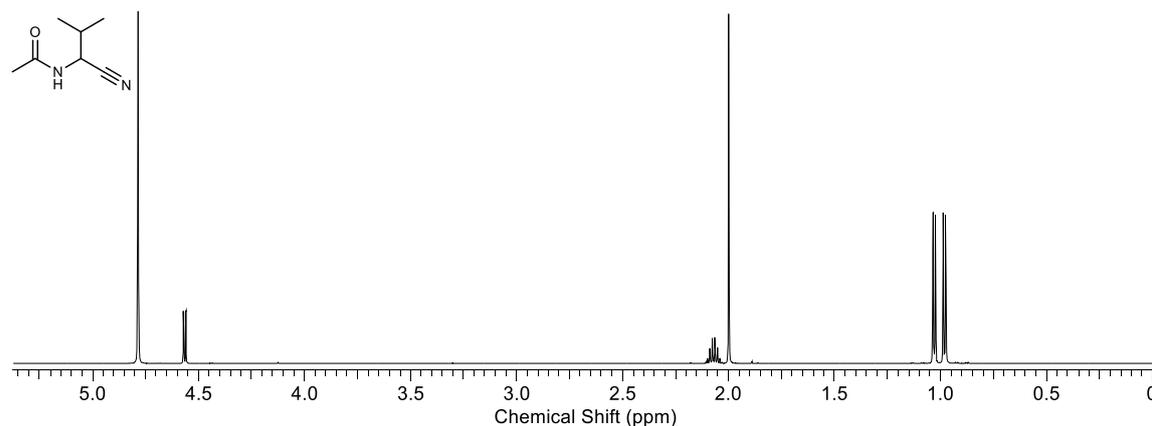
$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  4.57 (d,  $J = 7.0$  Hz, 1H, (C2)–H), 2.12 – 2.03 (m, 1H, (C3)–H), 1.03 (d,  $J = 6.8$  Hz, 3H, (C4)–H<sub>3</sub>), 0.98 (d,  $J = 6.8$  Hz, 3H, (C4')–H<sub>3</sub>).

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  174.6 (CO), 119.2 (C1), 47.8 (C2), 31.2 (C3), 22.2 (CO)CH<sub>3</sub>, 18.4 (C4), 18.1 (C4').

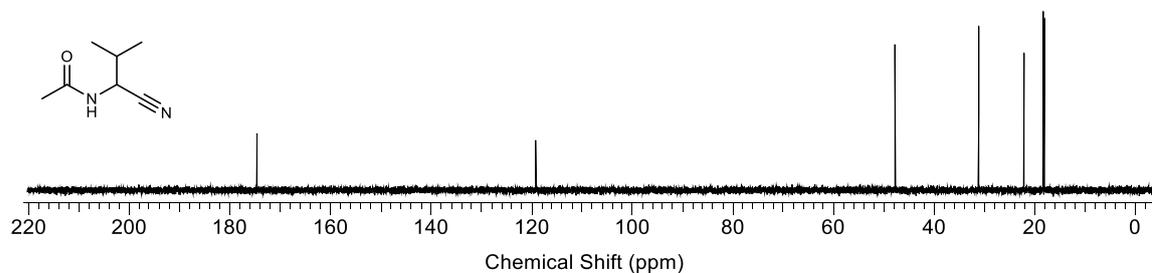
**HRMS** (CI): calcd. for  $[\text{C}_7\text{H}_{12}\text{N}_2\text{O}+\text{H}]^+$ : 141.1022; observed: 141.1022.

**IR** (solid): 3454, 2970, 2849, 2636, 2575, 2133, 1736, 1583, 1555, 1514  $\text{cm}^{-1}$ .

**M.p.** 100.9 – 102.1  $^\circ\text{C}$ .

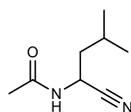


**Experimental figure 102:**  $^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ , 0.00 – 5.85 ppm) spectrum to show *N*-acetylvaline nitrile **Ac-Val-CN**.



**Experimental figure 103:**  $^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ , 0 – 220 ppm) spectrum to show *N*-acetylvaline nitrile **Ac-Val-CN**.

*N*-Acetylleucine nitrile **Ac-Leu-CN**



*N*-Acetylleucine nitrile **Ac-Leu-CN** was prepared following *General procedure F* using leucine nitrile hydrochloride **Leu-CN·HCl** (1.49 g, 10 mmol) afforded *N*-acetylleucine nitrile **Ac-Leu-CN** as a white solid (1.51 g, 9.79 mmol, 98%).

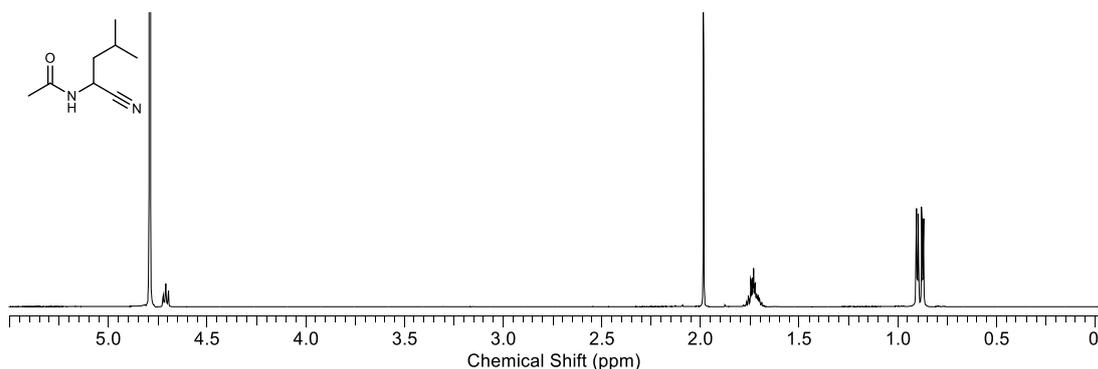
$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  4.71 (t,  $J = 6.1$  Hz, 1H, (C2)–H), 1.98 (s, 3H, (CO)CH<sub>3</sub>), 1.78 – 1.68 (m, 3H, (C3)–H<sub>2</sub>, (C4)–H), 0.90 (d,  $J = 6.1$ , 3H, (C5)–H<sub>3</sub>), 0.87 (d,  $J = 6.1$ , 3H, (C5')–H<sub>3</sub>).

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  174.6 (CO), 120.2 (C1), 40.3 (C2), 39.9 (C3), 24.9 (C4), 24.2 (CO)CH<sub>3</sub>, 21.9 (C5). 21.5 (C5').

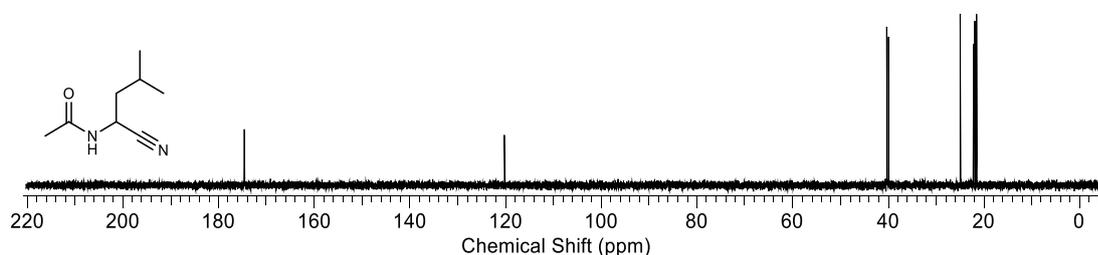
**HRMS** (CI): calcd. for  $[\text{C}_8\text{H}_{14}\text{N}_2\text{O}+\text{H}]^+$  155.1179; observed 155.1176.

**IR** (solid): 3285, 3046, 2965, 2938, 2876, 2156, 1657, 1531  $\text{cm}^{-1}$ .

**M.p.** 50.2 – 53.1 °C; lit. m.p. 48 – 49 °C.<sup>15</sup>



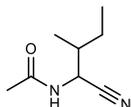
**Experimental figure 104:**  $^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ , 0.00 – 5.50 ppm) spectrum to show *N*-acetylleucine nitrile **Ac-Leu-CN**.



**Experimental figure 105:**  $^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ , 0 – 220 ppm) spectrum to shown *N*-acetylleucine nitrile **Ac-Leu-CN**.

<sup>15</sup> B. H. Lipshutz, K. E. McCarthy, *J. Am. Chem. Soc.*, 1983, 105 (26), pp 7703–7713

*N*-Acetylisoleucine nitrile **Ac-Ile-CN**



*N*-Acetylisoleucine nitrile **Ac-Ile-CN** was prepared following *General procedure F* using isoleucine nitrile hydrochloride **Ile-CN**·HCl (297 mg, 2.0 mmol) afforded *N*-acetylisoleucine nitrile **Ac-Ile-CN** as a white solid (295 mg, 1.91 mmol, 95%).

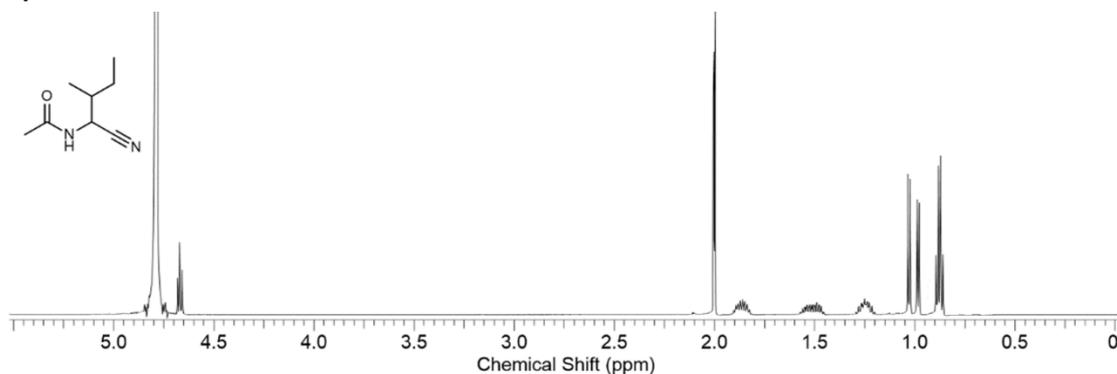
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 55:45 mixture of diastereoisomer a/b): diastereoisomer a: δ 4.67, (obs. t, *J* = 7.2 Hz, 1H, (C2)-H), 1.99 (s, 3H, (CO)-CH<sub>3</sub>), 1.92 – 1.82 (m, 1H, (C3)-H), 1.58 – 1.45 (m, 1H, (C4)-H), 1.29 – 1.21 (m, 1H, (C4')-H), 1.03 (d, *J* = 7.0 Hz, 3H, (C3)-CH<sub>3</sub>), 0.90 – 0.85 (t, *J* = 7.5 Hz, 3H, (C5)-H<sub>3</sub>); diastereoisomer b: δ 4.67, (obs. t, *J* = 7.2 Hz, 1H, (C2)-H), 2.00 (s, 3H, (CO)-CH<sub>3</sub>), 1.92 – 1.82 (m, 1H, (C3)-H), 1.58 – 1.45 (m, 1H, (C4)-H), 1.29 – 1.21 (m, 1H, (C4')-H), 0.98 (d, *J* = 7.0 Hz, 3H, (C3)-CH<sub>3</sub>), 0.90 – 0.85 (t, *J* = 7.5 Hz, 3H, (C5)-H<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 56:44 mixture of diastereoisomers a/b): diastereoisomer a: δ 174.6 (CO), 119.2 (C1), 46.3 (C2), 37.4 (C3), 25.5 (C4), 22.2 ((CO)-CH<sub>3</sub>), 15.1 ((C3)-H<sub>3</sub>), 10.8 (C5); diastereoisomer b: δ 174.5 (CO), 119.5 (C1), 46.6 (C2), 37.2 (C3), 25.7 (C4), 22.2 ((CO)-CH<sub>3</sub>), 15.0 ((C3)-H<sub>3</sub>), 10.9 (C5).

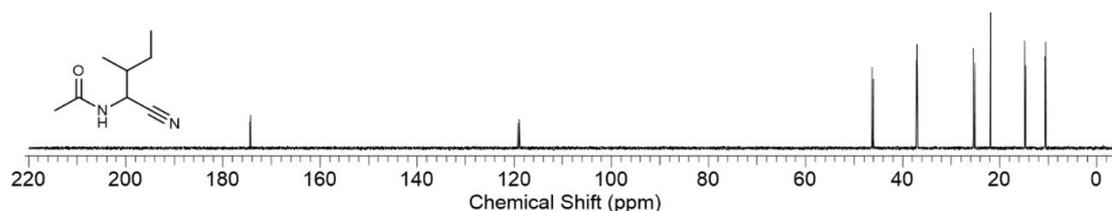
HRMS (CI): calcd. for [C<sub>8</sub>H<sub>14</sub>N<sub>2</sub>O + H]<sup>+</sup>: 155.1179; observed: 155.1178.

IR (solid): 3292, 3041, 2953, 2905, 2841, 2263, 1666, 1531 cm<sup>-1</sup>.

M.p. 54.6 – 55.9 °C.

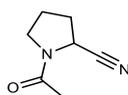


**Experimental figure 106:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 0.00 – 5.50 ppm) spectrum to show *N*-acetylisoleucine nitrile **Ac-Ile-CN**.



**Experimental figure 107:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to shown *N*-acetylisoleucine nitrile **Ac-Ile-CN**.

*N*-Acetylproline nitrile **Ac-Pro-CN**



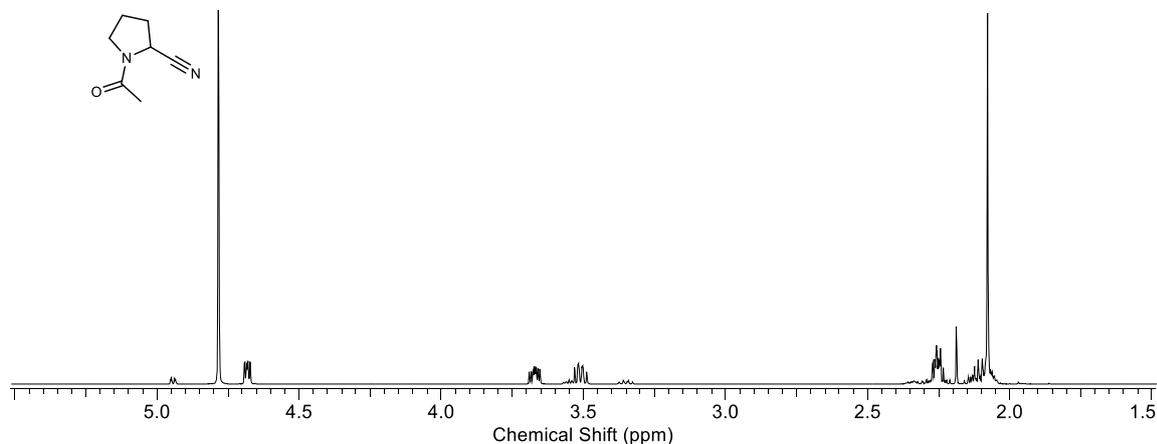
*N*-Acetylproline nitrile **Ac-Pro-CN** was prepared following *General procedure F* using proline nitrile hydrochloride **Pro-CN**·HCl (265 mg, 2 mmol) afforded *N*-acetylproline nitrile **Ac-Pro-CN** as a clear oil (205 mg, 1.48 mmol, 74%).

<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 85:15 mixture of rotamers a/b): rotamer a: δ 4.69 (dd, *J* = 4.3, 7.2 Hz, 1H, (C2)–H), 3.70 – 3.66 (m, 1H, (C3)–H), 3.54 – 3.49 (m, 1H, (C3')–H), 2.28 – 2.23 (m, 2H, (C5)–H<sub>2</sub>), 2.15 – 2.06 (m, 2H, (C4)–H<sub>2</sub>), 2.08 (s, 3H, (CO)CH<sub>3</sub>); rotamer b: δ 4.96 – 4.94 (m, 1H, (C2)–H), 3.58 – 3.54 (m, 1H, (C3)–H), 3.38 – 3.34 (m, 1H, (C3')–H), 2.36 – 2.28 (m, 2H, (C5)–H<sub>2</sub>), 2.19 (s, 3H, (CO)CH<sub>3</sub>), 2.15 – 2.06 (m, 2H, (C4)–H<sub>2</sub>).

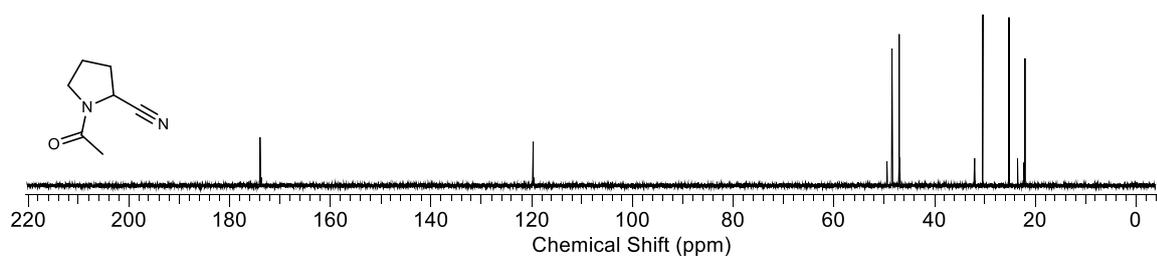
<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 85:15 mixture of rotamers a/b): rotamer a: δ 173.9 (CO), 119.8 (C1), 48.4 (C2), 47.1 (C3), 30.5 (C5), 25.2 (C4), 22.0 (CO)CH<sub>3</sub>; rotamer b: δ 173.8 (CO), 119.6 (C1), 49.4 (C2), 46.7 (C3), 32.1 (C5), 23.6 (C4), 22.2 (CO)CH<sub>3</sub>).

HRMS (ES<sup>+</sup>): calcd. for [C<sub>7</sub>H<sub>10</sub>N<sub>2</sub>O + Na]<sup>+</sup>: 161.0685; observed: 161.0668.

IR (solid): 3291, 3037, 2959, 2951, 2870, 2800, 2251 1671, 1610, 1551, 1501 cm<sup>-1</sup>.

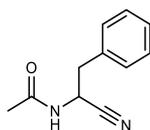


**Experimental figure 108:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 1.50 – 5.50 ppm) spectrum to show *N*-acetylproline nitrile **Ac-Pro-CN**.



**Experimental figure 109:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show *N*-acetylproline nitrile **Ac-Pro-CN**.

*N*-Acetylphenylalanine nitrile **Ac-Phe-CN**



*N*-Acetylphenylalanine nitrile **Ac-Phe-CN** was prepared following *General procedure F* using phenylalanine nitrile hydrochloride **Phe-CN**·HCl (1.83 g, 10 mmol) afforded *N*-acetylphenylalanine nitrile **Ac-Phe-CN** as a white solid (1.84 mg, 9.78 mmol, 93%).

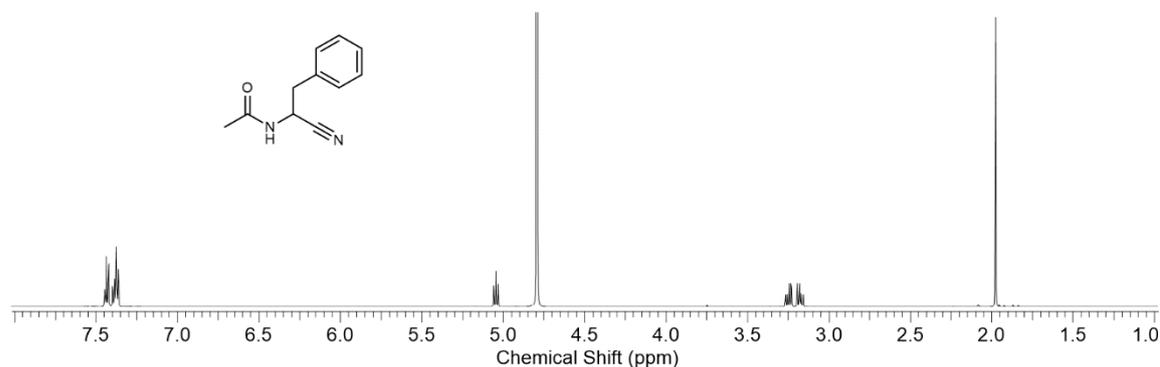
$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  7.45 – 7.35 (m, 5H, Ar), 5.04 (t,  $J = 7.4$  Hz, 1H, (C2)–H), 3.25 (ABX,  $J = 7.4$ , 13.8 Hz, 1H, (C3)–H), 3.17 (ABX,  $J = 7.4$ , 13.8 Hz, 1H, (C3)–H'), 1.97 (s, 3H, (CO)CH<sub>3</sub>).

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  174.4 (CO), 135.7 (Ar), 130.1 (Ar  $\times$  2), 129.6 (Ar  $\times$  2), 128.4 (Ar), 119.5 (C1), 42.8 (C2), 38.0 (C3), 22.2 (CO)CH<sub>3</sub>).

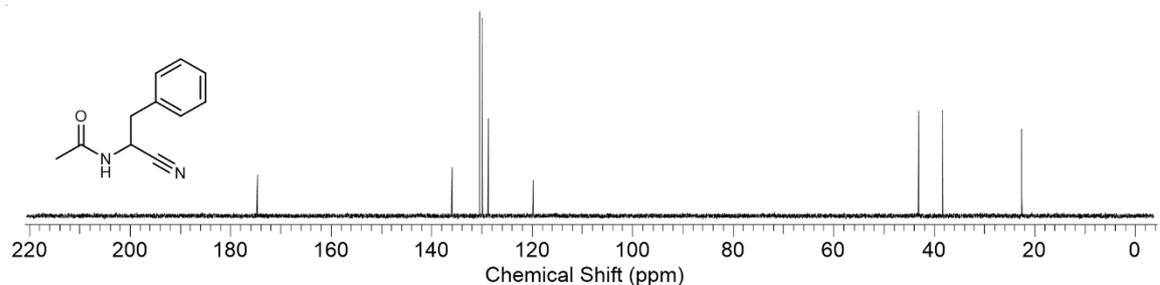
**HRMS:** ( $\text{ES}^+$ ): calcd. for  $[\text{C}_{11}\text{H}_{13}\text{N}_2\text{O}+\text{H}]^+$ : 189.1028; observed: 189.1032.

**IR** (solid): 3282, 3063, 3029, 2966, 2930, 2239, 1654, 1602, 1583, 1537  $\text{cm}^{-1}$ .

**M.p.** 95.8 – 97.1  $^\circ\text{C}$ ; lit. m.p. 94 – 95  $^\circ\text{C}$ .<sup>16</sup>



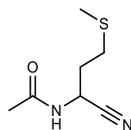
**Experimental figure 110:**  $^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ , 1.00 – 8.00 ppm) spectrum to show *N*-acetylphenylalanine nitrile **Ac-Phe-CN**.



**Experimental figure 111:**  $^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ , 0 – 220 ppm) spectrum to show *N*-acetylphenylalanine nitrile **Ac-Phe-CN**.

<sup>16</sup> Pengju Zeng, Yuefei Hu & Hongwen Hu, *Synthetic Communications* 1997 vol. 27, 5 p. 939 - 944

*N*-Acetylmethionine nitrile **Ac-Met-CN**



*N*-Acetylmethionine nitrile **Ac-Met-CN** was prepared following *General procedure F* using methionine nitrile hydrochloride **Met-CN**·HCl (1.83 g, 10 mmol) afforded *N*-acetylmethionine nitrile **Ac-Met-CN** as a white solid (1.84 g, 9.78 mmol, 98%).

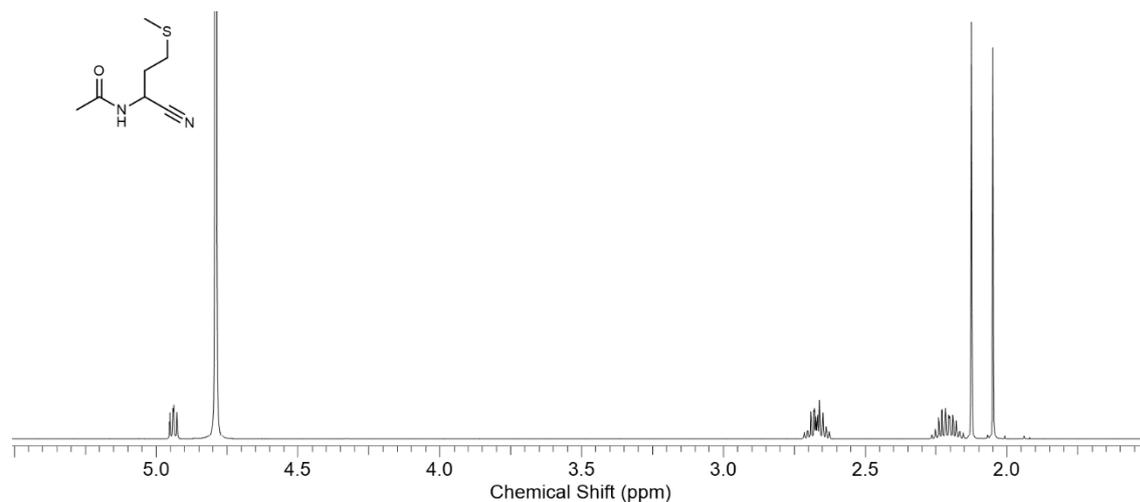
$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  4.94 (dd,  $J = 6.8, 8.3$  Hz, 1H, (C2)–H), 2.72 – 2.62 (m, 2H, (C3)–H<sub>2</sub>), 2.27 – 2.15 (m, 2H, (C4)–H<sub>2</sub>), 2.13 (s, 3H, SCH<sub>3</sub>), 2.05 (s, 3H, (CO)CH<sub>3</sub>).

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  174.7 (CO), 119.6 (C1), 40.2 (C2), 31.2 (C3), 29.4 (C4), 22.3 (CO)CH<sub>3</sub>, 14.7 (SCH<sub>3</sub>).

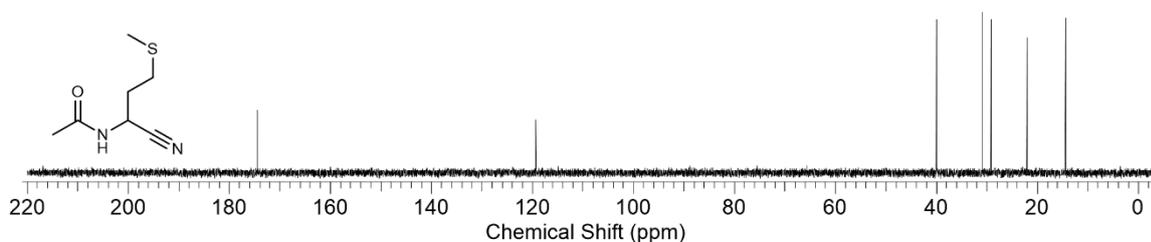
**HRMS** (CI): calcd. for  $[\text{C}_7\text{H}_{12}\text{N}_2\text{OS}+\text{H}]^+$ : 173.0743; observed: 173.0745.

**IR** (solid): 2966, 2769, 2186, 1736, 1590, 1569, 1503  $\text{cm}^{-1}$ .

**M.p.** 43.0 – 43.9 °C.

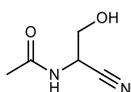


**Experimental figure 112:**  $^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ , 1.50 – 5.50 ppm) spectrum to show *N*-acetylmethionine nitrile **Ac-Met-CN**.



**Experimental figure 113:**  $^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ , 0 – 220 ppm) spectrum to show *N*-acetylmethionine nitrile **Ac-Met-CN**.

*N*-Acetylserine nitrile **Ac-Ser-CN**



*N*-Acetylserine nitrile **Ac-Ser-CN** was prepared following *General procedure F* using serine nitrile **Ser-CN** (258 mg, 3.00 mmol) afforded *N*-acetylserine nitrile **Ac-Ser-CN** as white, crystalline solid (363 mg, 2.83 mmol, 94%).

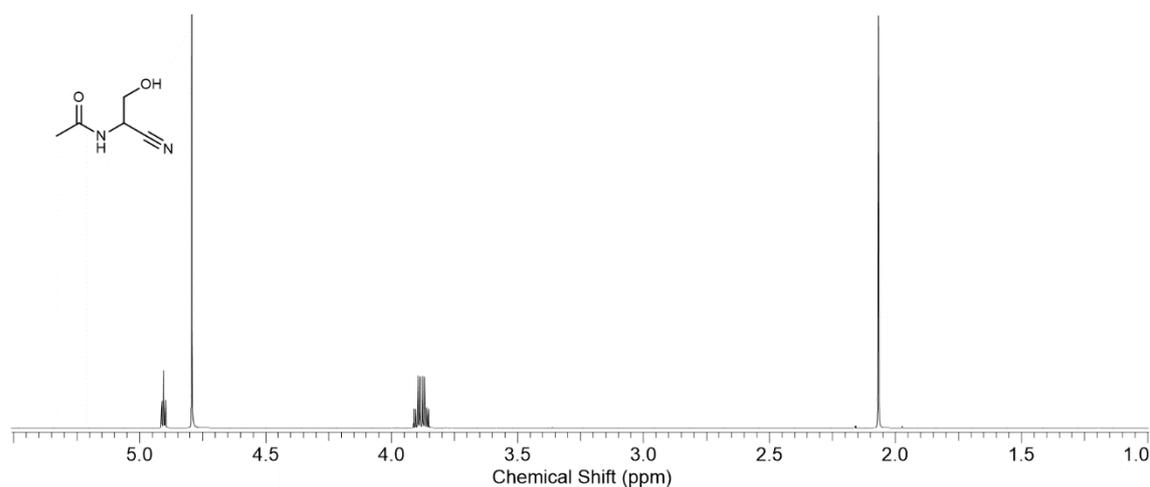
$^1\text{H NMR}$  (700 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  4.90 (app. t,  $J = 5.3$  Hz, 1H, (C2)–H), 3.90 (ABX,  $J = 5.3, 11.2$  Hz, 1H, (C2)–H), 3.86 (ABX,  $J = 5.3, 11.2$  Hz, 1H, (C2)–H'), 2.06 (s, 3H, (CO)CH<sub>3</sub>).

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  174.6 (CO), 118.3 (C1), 61.1 (C3), 43.6 (C2), 22.1 (CO)CH<sub>3</sub>.

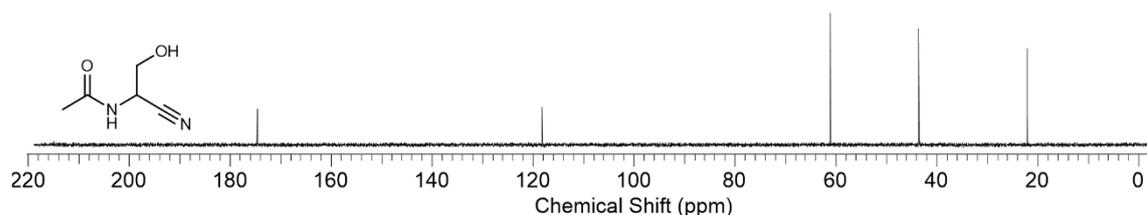
**HRMS** (CI): calcd. for [ $\text{C}_5\text{H}_8\text{N}_2\text{O}_2+\text{H}$ ]<sup>+</sup>: 129.0659; observed: 129.0658.

**IR** (solid): 3456, 3262, 3016, 2970, 2970, 3016, 2970, 2939, 2241, 2125, 1738, 1620, 1527  $\text{cm}^{-1}$ .

**M.p.** 116.5 – 117.9 °C; lit. m.p. 111 – 112 °C.<sup>17</sup>



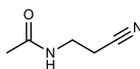
**Experimental figure 114:**  $^1\text{H NMR}$  (700 MHz,  $\text{D}_2\text{O}$ , 1.00 – 5.50 ppm) spectrum to show *N*-acetylserine nitrile **Ac-Ser-CN**.



**Experimental figure 115:**  $^{13}\text{C NMR}$  (176 MHz,  $\text{D}_2\text{O}$ , 0 – 220 ppm) spectrum to show *N*-acetylserine nitrile **Ac-Ser-CN**.

<sup>17</sup> Rolf Paul, Richard P. Williams, and Elliott Cohen, *J. Org. Chem.*, 1975, 40 (11), pp 1653 –1656

*N*-Acetyl- $\beta$ -alanine nitrile **Ac- $\beta$ -Ala-CN**



*N*-Acetyl- $\beta$ -alanine nitrile **Ac- $\beta$ -Ala-CN** was prepared following *general procedure F* using  $\beta$ -alanine nitrile  **$\beta$ -Ala-CN** (701 mg, 731  $\mu$ L, 10.0 mmol). The reaction afforded *N*-acetyl- $\beta$ -alanine nitrile **Ac- $\beta$ -Ala-CN** as a white solid (912 mg, 8.13 mmol, 83%).

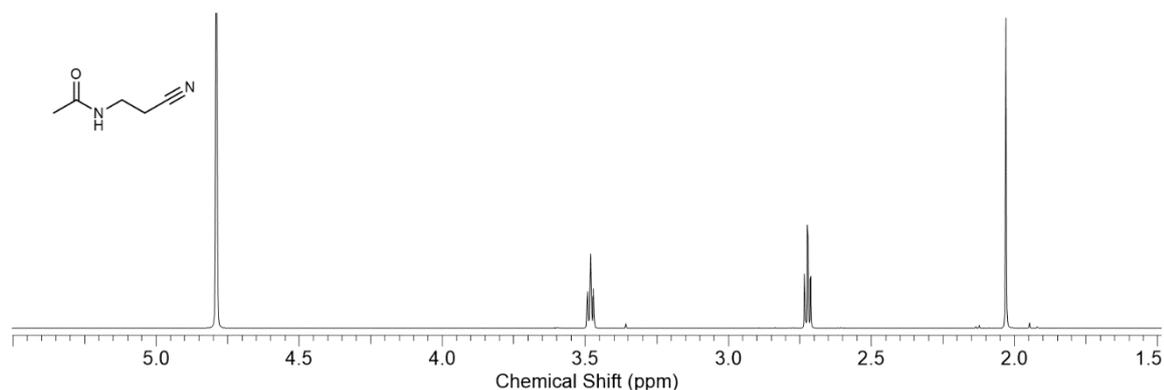
$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  3.48 (t,  $J = 6.8$  Hz, 2H, (C2)- $\text{H}_2$ ), 2.73 (t,  $J = 6.8$  Hz, 2H, (C3)- $\text{H}_2$ ), 2.03 (s, 3H, (CO) $\text{CH}_3$ ).

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  175.3 (CO), 120.5 (C1), 35.7 (C2), 22.5 (C3), 18.3 (CO) $\text{CH}_3$ .

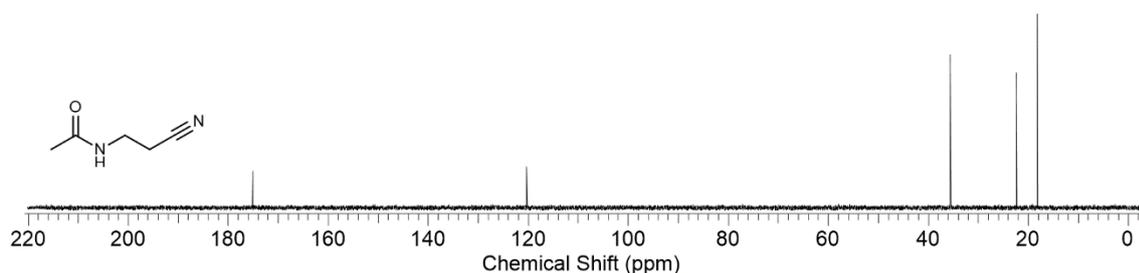
**HRMS** (CI): calcd. for  $[\text{C}_5\text{H}_8\text{N}_2\text{O}+\text{NH}_4]^+$ : 130.0975; observed: 130.0975.

**IR** (solid): 3253, 3071, 2950, 2929, 2856, 2240, 1652, 1556  $\text{cm}^{-1}$ .

**M.p.** 60.7 – 61.5  $^\circ\text{C}$ ; lit. m.p. 62 – 63  $^\circ\text{C}$ .<sup>18</sup>



**Experimental figure 116:**  $^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ , 1.50 – 5.50 ppm) spectrum to show *N*-acetyl- $\beta$ -alanine nitrile **Ac- $\beta$ -Ala-CN**.

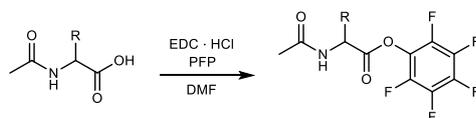


**Experimental figure 117:**  $^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ , 0 – 220 ppm) spectrum to show *N*-acetyl- $\beta$ -alanine nitrile **Ac- $\beta$ -Ala-CN**.

<sup>18</sup> Denis M. Bailey, John E. Bowers, and C. David Gutsche, *J. Org. Chem.*, 1963, 28 (3), pp 610–614

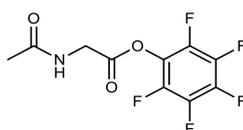
### 9. 3. 3. Synthesis and characterisation of *N*-acetyl perfluorophenyl esters

General procedure G: Preparative acetylation of *N*-acetyl perfluorophenyl esters



*N*-Acetyl amino acid **Ac-AA-OH** (100 mmol) and pentafluorophenol (PFP, 27.61 g, 150 mmol) were dissolved in dry DMF (200 mL). 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC · HCl, 23.00 g, 120 mmol) was added portionwise over 10 min. The reaction was stirred at room temperature for 2 h and then concentrated *in vacuo*. The residue was purified by flash column chromatography (SiO<sub>2</sub>; eluting with a gradient of EtOAc/petroleum ether (60-40) 0:100 → 100:00) to afford *N*-acetyl amino acid pentafluorophenyl ester **Ac-AA-PFP**.

*N*-Acetyl glycine pentafluorophenyl ester **Ac-Gly-PFP**



*N*-Acetylglycine **Ac-Gly-OH** (100 mmol) and pentafluorophenol (PFP, 27.61 g, 150 mmol) were dissolved in dry DMF (200 mL). 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC·HCl, 23.00 g, 120 mmol) was added portionwise over 10 min. The reaction was stirred at room temperature for 2 h and then concentrated *in vacuo*. The residue was purified by flash column chromatography (SiO<sub>2</sub>; eluting with a gradient of EtOAc/petroleum ether (60-40) 0:100 → 100:00) to afford *N*-acetyl glycine pentafluorophenyl ester **Ac-Gly-PFP** (23.9 g, 84.6 mmol, 85%).

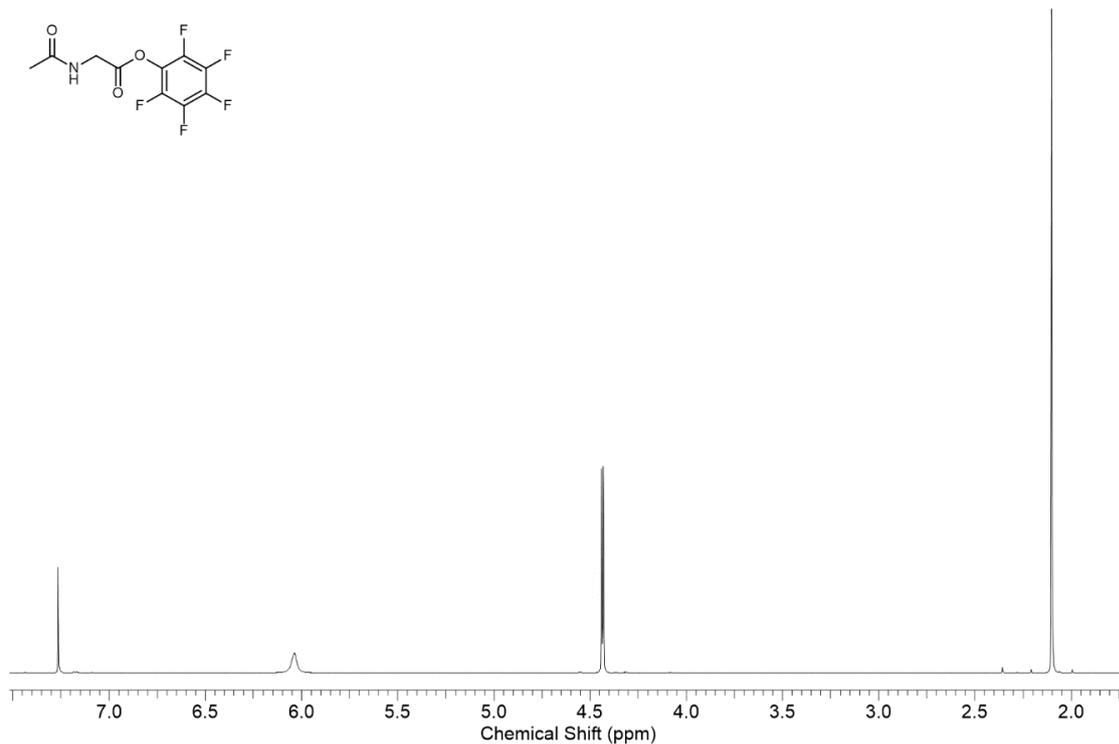
<sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>): δ 6.04 (br. s, 1H, (C2)-NH), 4.44 (d, *J* = 5.6 Hz, 2H, (C2)-H<sub>2</sub>), 2.10 (s, 3H, (CO)-H<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, CDCl<sub>3</sub>): δ 170.4 (CO), 166.6 (CO), 141.9 – 124.6 (Ar), 40.8 (C2), 23.0 ((CO)-CH<sub>3</sub>).

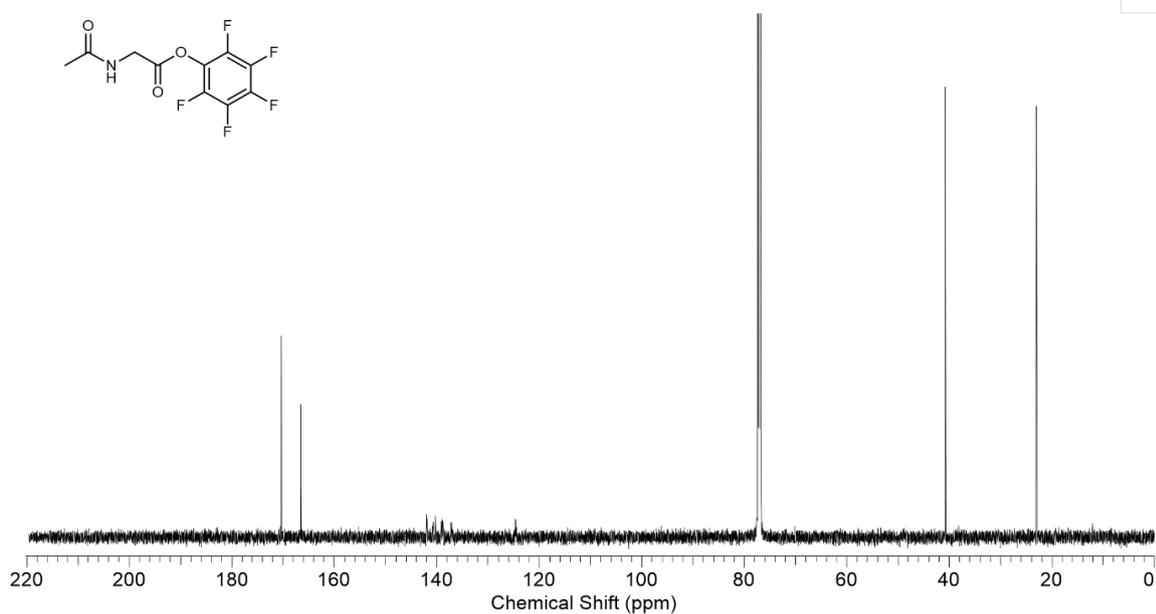
HRMS (CI): calcd. for [C<sub>10</sub>H<sub>6</sub>F<sub>5</sub>NO<sub>3</sub> + NH<sub>4</sub>]<sup>+</sup>: 301.0606; observed: 301.0607.

IR (solid): 3260, 3075, 2962, 1796, 1785, 1716, 1514 cm<sup>-1</sup>.

M. p. (solid): 112.8 – 114.1 °C.



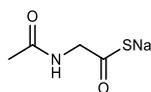
**Experimental figure 118:** <sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>, 7.50 – 1.75 ppm) spectrum to show *N*-acetyl glycine pentafluorophenyl ester **Ac-Gly-PFP**.



**Experimental figure 119:** <sup>13</sup>C NMR (151 MHz, CDCl<sub>3</sub>, 220 – 0 ppm) spectrum to show *N*-acetyl glycine pentafluorophenyl ester **Ac-Gly-PFP**.

### 9. 3. 4. Synthesis and characterisation of *N*-acetyl aminothioacids

#### Sodium 2-acetamidoethanethioate **Ac-Gly-SNa**



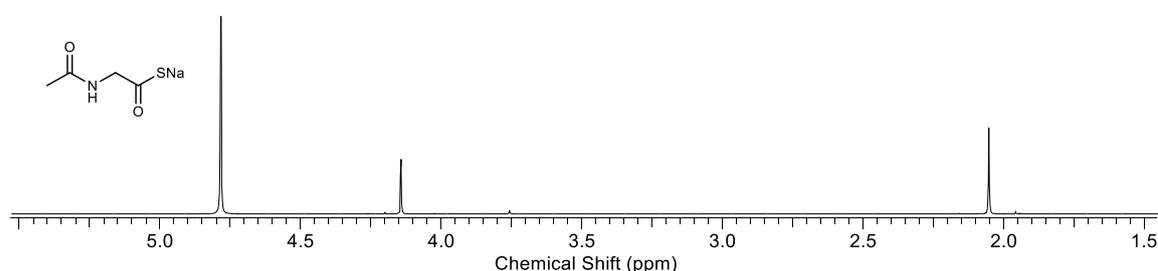
Sodium hydrosulfide hydrate (NaSH·xH<sub>2</sub>O; 807 mg, 6 mmol) was dissolved in degassed H<sub>2</sub>O (50 mL), and the solution was adjusted to pH 9.0. *N*-Acetyl glycine pentafluorophenyl ester (1.42 g, 5 mmol) was added portion-wise whilst the solution was maintained at pH 9.0 through the addition of NaOH. The reaction was vigorously stirred at room temperature for 1 h. The solution was lyophilised and the resultant residue was washed with diethyl ether (5 × 50 mL) and dried *in vacuo* to afford sodium 2-acetamidoethanethioate **Ac-Gly-SNa** as a white solid (1.53 g), which was used without further purification.

<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.15 (s, 2H, (C2)-H<sub>2</sub>), 2.06 (s, 3H, (CO)-CH<sub>3</sub>).

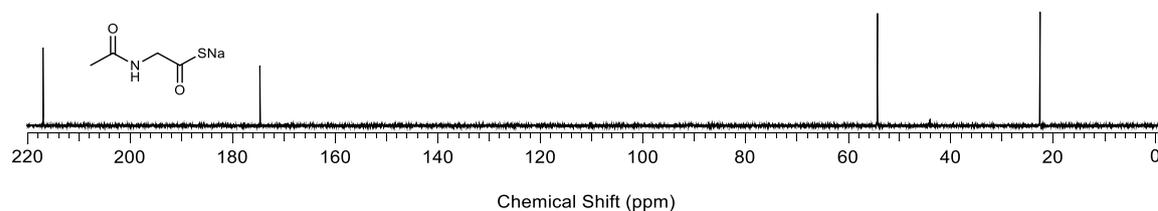
<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 217.0 (C1), 174.7 (CO), 54.3 (C2), 22.6 (CO)CH<sub>3</sub>.

HRMS (ES<sup>-</sup>): calcd. for [C<sub>4</sub>H<sub>7</sub>NO<sub>2</sub>S-H]<sup>-</sup>: 132.0119; observed: 132.0117.

IR (solid): 3403, 3254, 2985, 1629, 1568, 1518 cm<sup>-1</sup>.



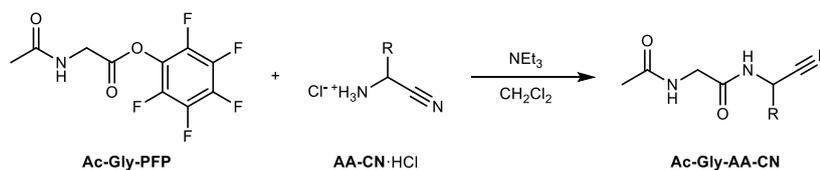
**Experimental figure 120:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 1.50 – 5.50 ppm) spectrum to show 2-acetamidoethanethioate **Ac-Gly-SNa**.



**Experimental figure 121:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show 2-acetamidoethanethioate **Ac-Gly-SNa**.

### 9. 3. 5. Synthesis and characterisation of *N*-acetyl dipeptide nitriles

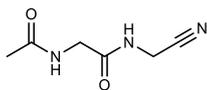
General procedure H: Preparative acetylation of *N*-acetyl dipeptide nitriles



*N*-Acetyl glycine pentafluorophenyl ester **Ac-Gly-PFP** (708 mg, 2.5 mmol) was added to aminonitrile hydrochloride **AA-CN·HCl** (3.0 mmol) and triethylamine (607 mg, 837  $\mu\text{L}$ , 6.0 mmol) in  $\text{CH}_2\text{Cl}_2$  (50 mL). The solution was stirred for 16 h and then concentrated *in vacuo*. The solids were washed with ether (5  $\times$  50 mL) and then purified by flash column chromatography ( $\text{SiO}_2$ ; eluting with a gradient of  $\text{MeOH}/\text{CH}_2\text{Cl}_2$  0:100  $\rightarrow$  10:90) to afford *N*-acetyl peptide nitrile **Ac-Gly-AA-CN** as a white solid.

Experimental table 14: preparation of <i>N</i> -acetyl dipeptide nitriles from <i>N</i> -acetyl pentafluorophenyl esters		
Entry	Residue	Yield Ac-Gly-AA-CN (%)
1	Gly	91
2	Ala	82
3	Val	77
4	Leu	83
5	Ile	80
6	Met	62
7	Phe	84
9	Pro	70
10	$\beta$ -Ala	36

*N*-acetylglycyl glycine nitrile **Ac-Gly-Gly-CN**



*N*-Acetyl glycyl glycine nitrile **Ac-Gly-Gly-CN** was prepared following *General procedure H* using glycine nitrile hydrochloride **Gly-CN**·HCl (277.6 mg, 3.0 mmol). The reaction afforded *N*-acetylglycyl glycine nitrile **Ac-Gly-Gly-CN** as a white solid (351.3 mg, 2.26 mmol, 91%).

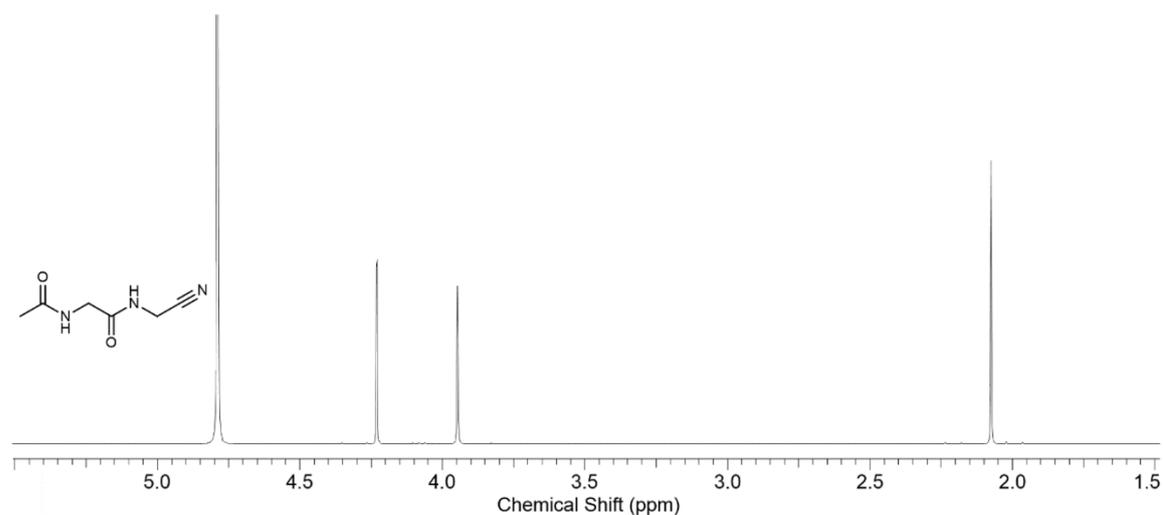
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.23 (s, 2H, Gly2-(C2)-H<sub>2</sub>), 3.95 (s, 2H, Gly1-(C2)-H<sub>2</sub>), 2.07 (s, 3H, (CO)-CH<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 175.7 (Gly1-(C1)), 172.8 (COCH<sub>3</sub>), 117.4 (Gly2-(C1)), 43.0 (Gly2-(C2)), 28.1 (Gly1-(C2)), 22.3 ((CO)-CH<sub>3</sub>).

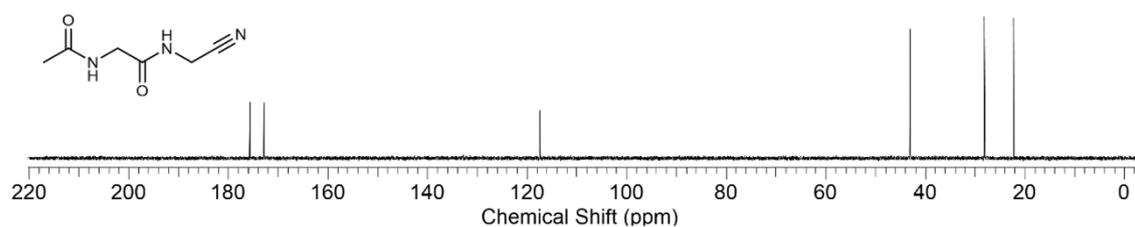
HRMS (ES<sup>-</sup>): calcd. for [C<sub>6</sub>H<sub>9</sub>N<sub>3</sub>O<sub>2</sub>·H]<sup>-</sup>: 154.0616; observed: 154.0613.

IR (solid): 3321, 3201, 3047, 2843, 2803, 2020, 1674, 1631, 1552 cm<sup>-1</sup>.

M.p. 157.1 – 159.3 °C.

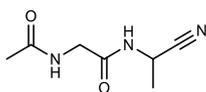


**Experimental figure 122:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 5.50 – 1.50 ppm) spectrum to show *N*-acetylglycyl glycine nitrile **Ac-Gly-Gly-CN**.



**Experimental figure 123:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 220 – 0 ppm) spectrum to show *N*-acetyl glycyl glycine nitrile **Ac-Gly-Gly-CN**.

*N*-acetylglycyl alanine nitrile **Ac-Gly-Ala-CN**



*N*-Acetylglycyl alanine nitrile **Ac-Gly-Ala-CN** was prepared following *General procedure H* using alanine nitrile hydrochloride **Ala-CN**·HCl (319.7 mg, 3.0 mmol). The reaction afforded *N*-acetylglycyl alanine nitrile **Ac-Gly-Ala-CN** as a white solid (388.7 mg, 2.06 mmol, 82%).

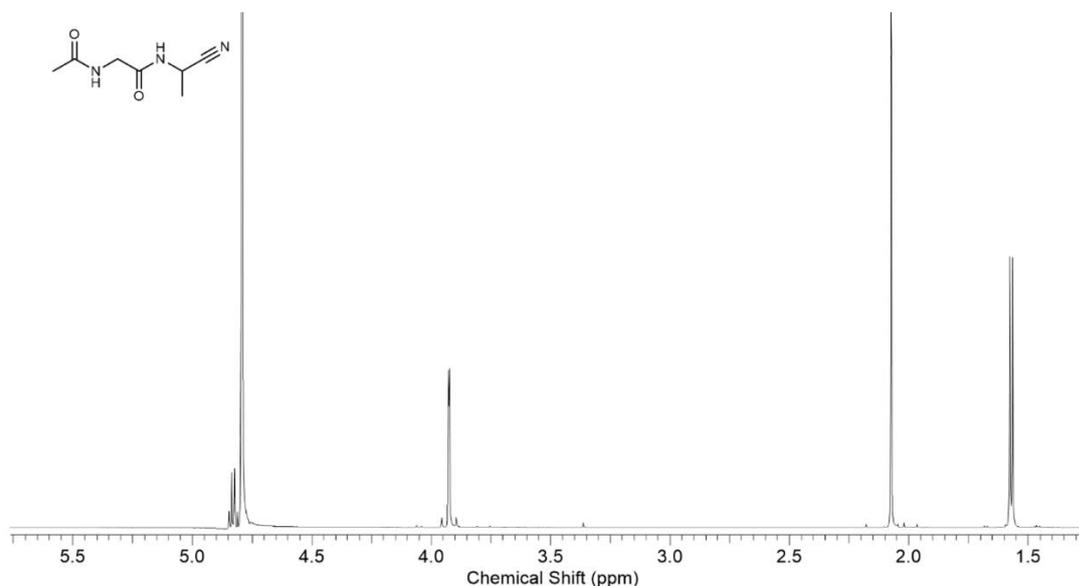
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.82 (q, *J* = 7.2 Hz, 1H, Ala2-(C2)-H), 3.92, 3.91 (ABq, *J* = 16.2 Hz, 2H, Gly1-(C2)-H<sub>2</sub>), 2.07 (s, 3H, (CO)-CH<sub>3</sub>), 1.64 (d, *J* = 7.2 Hz, 3H, Ala2-(C3)-H<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 175.6 (Gly1-(C1)), 171.9 (CO), 120.4 (Ala2-(C1)), 43.1 (Ala2-(C2)), 37.1 (Gly1-(C2)), 22.3 ((CO)-CH<sub>3</sub>), 17.9 (Ala2-(C3)).

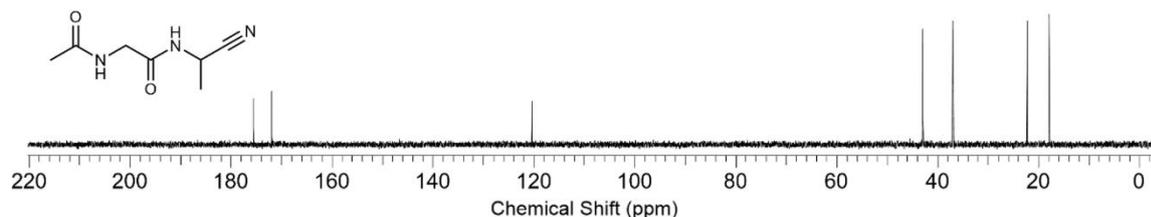
HRMS (ES): calcd. for [C<sub>7</sub>H<sub>11</sub>N<sub>3</sub>O<sub>2</sub>·H]<sup>+</sup>: 168.0773; observed: 168.0778.

IR (solid): 3316, 3198, 3041, 2944, 2032, 1670, 1642 cm<sup>-1</sup>.

M.p. 155.5 – 157.1 °C.

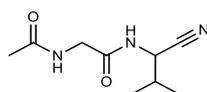


**Experimental figure 124:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 5.75 – 1.25 ppm) spectrum to show *N*-acetylglycyl alanine nitrile **Ac-Gly-Ala-CN**.



**Experimental figure 125:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 220 – 0 ppm) spectrum to show *N*-acetylglycyl alanine nitrile **Ac-Gly-Ala-CN**.

*N*-acetylglycyl valine nitrile **Ac-Gly-Val-CN**



*N*-Acetyl glycyl valine nitrile **Ac-Gly-Val-CN** was prepared following *General procedure H* using valine nitrile hydrochloride **Val-CN**·HCl (403.8 mg, 3.0 mmol). The reaction afforded *N*-acetylglycyl valine nitrile **Ac-Gly-Val-CN** as a white solid (379.6 mg, 1.92 mmol, 77%).

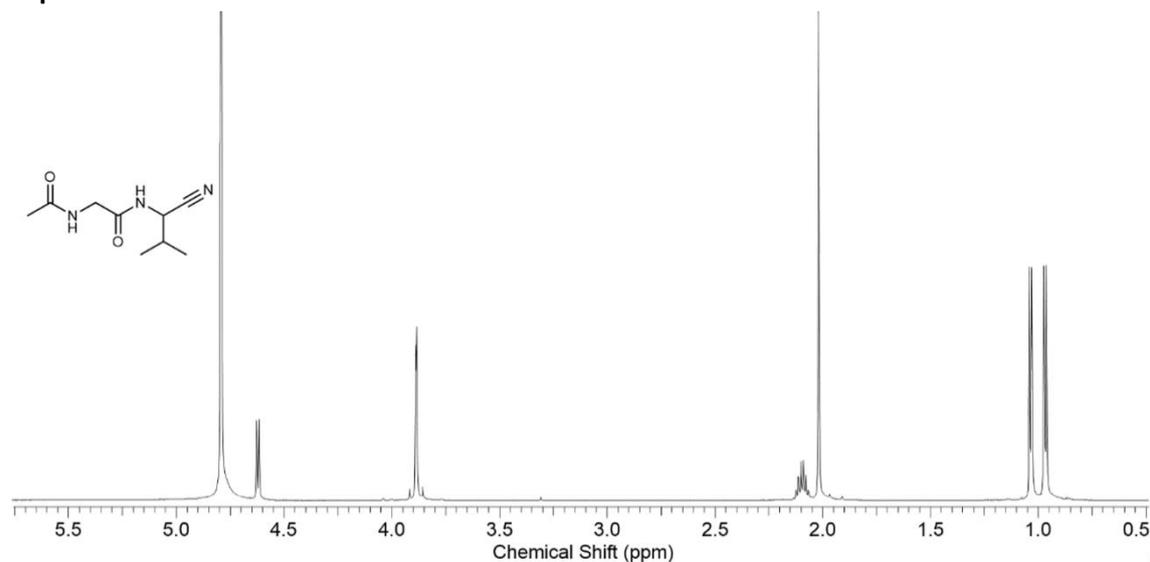
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.62 (d, *J* = 7.4 Hz, 1H, Val2-(C2)-H), 3.88, 3.87 (ABq, *J* = 15.9 Hz, 2H, Gly1-(C2)-H<sub>2</sub>), 2.13 – 2.06 (m, 1H, Val2-(C3)-H), 2.02 (s, 3H, (CO)-CH<sub>3</sub>), 1.03 (d, *J* = 6.7 Hz, 3H, Val2-(C4)-H<sub>3</sub>), 0.97 (d, *J* = 6.7 Hz, 3H, Val2-(C4')-H<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 175.6 (Gly1-C1), 172.0 ((CO)-CH<sub>3</sub>), 119.0 (Val2-C1), 47.7 (Val2-C2), 43.1 (Gly1-C2), 31.4 (Val2-C3), 22.3 (Val2-C4), 18.3 (Val2-C4').

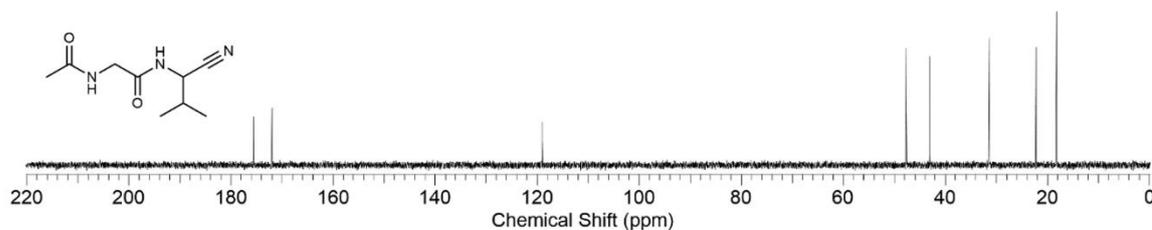
HRMS (ES<sup>-</sup>): calcd. for [C<sub>9</sub>H<sub>15</sub>N<sub>3</sub>O<sub>2</sub>-H]<sup>-</sup>: 196.1086; observed: 196.1084.

IR (solid): 3318, 3193, 3034, 2971, 2877, 2154, 1667, 1639, 1557 cm<sup>-1</sup>.

M.p. 124.0 – 124.8 °C.

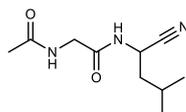


**Experimental figure 126:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 5.75 – 0.50 ppm) spectrum to show *N*-acetylglycyl valine nitrile **Ac-Gly-Val-CN**.



**Experimental figure 127:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 220 – 0 ppm) spectrum to show *N*-acetylglycyl valine nitrile **Ac-Gly-Val-CN**.

*N*-acetylglycyl leucine nitrile **Ac-Gly-Leu-CN**



*N*-acetylglycyl leucine nitrile **Ac-Gly-Leu-CN** was prepared following *General procedure H* using leucine nitrile hydrochloride **Leu-CN**·HCl (445.9 mg, 3.0 mmol). The reaction afforded *N*-acetylglycyl leucine nitrile **Ac-Gly-Leu-CN** as a white solid (441.0 mg, 2.09 mmol, 83%).

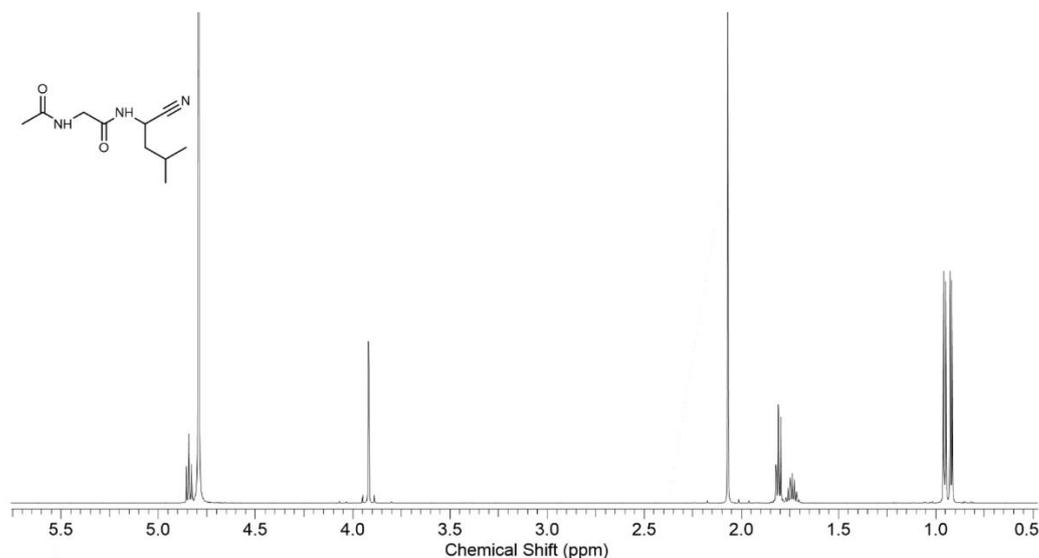
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.84 (t, *J* = 7.9 Hz, 1H, (Leu2)-(C2)-H), 3.92 (s, 2H, (Gly1)-(C2)-H<sub>2</sub>), 2.07 (s, 3H, (CO)-CH<sub>3</sub>), 1.83 – 1.78 (m, 2H, (Leu2)-(C3)-H<sub>2</sub>), 1.77 – 1.70 (m, 1H, (Leu2)-(C4)-H), 0.95 (d, *J* = 6.5 Hz, 1H, (Leu2)-(C5)-H<sub>3</sub>), 0.92 (d, *J* = 6.5 Hz, 1H, (Leu2)-(C5')-H<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 175.6 ((Gly1)-C1), 172.0 (CO), 119.0 ((Leu2)-C1), 43.2 ((Gly1)-C2), 40.5 ((Leu2)-C2), 39.9 ((Leu2)-C3), 24.9 ((Leu2)-C4), 22.3 ((Leu2)-C5), 22.0 (CO), 21.5 ((Leu2)-C5').

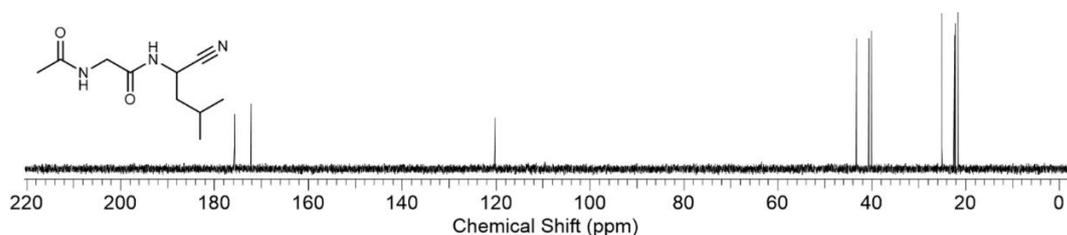
HRMS (ES<sup>+</sup>): calcd. for [C<sub>10</sub>H<sub>17</sub>N<sub>3</sub>O<sub>2</sub> + Na]<sup>+</sup>: 234.1213; observed: 234.1212.

IR (solid): 3357, 3320, 3301, 3294, 3087, 2999, 2920, 2227, 1668, 1599 cm<sup>-1</sup>.

M.p. 109.4 – 111.8 °C.

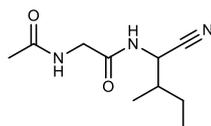


**Experimental figure 128:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 5.75 – 0.50) spectrum to show *N*-acetylglycyl leucine nitrile **Ac-Gly-Leu-CN**.



**Experimental figure 129:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 220 – 0 ppm) spectrum to show *N*-acetylglycyl leucine nitrile **Ac-Gly-Leu-CN**.

*N*-acetylglycyl leucine nitrile **Ac-Gly-Ile-CN**



*N*-acetylglycyl isoleucine nitrile **Ac-Gly-Ile-CN** was prepared following *General procedure H* using isoleucine nitrile hydrochloride **Ile-CN**·HCl (445.9 mg, 3.0 mmol). The reaction afforded *N*-acetylglycyl isoleucine nitrile **Ac-Gly-Ile-CN** as a white solid (424.4 mg, 2.01 mmol, 80%).

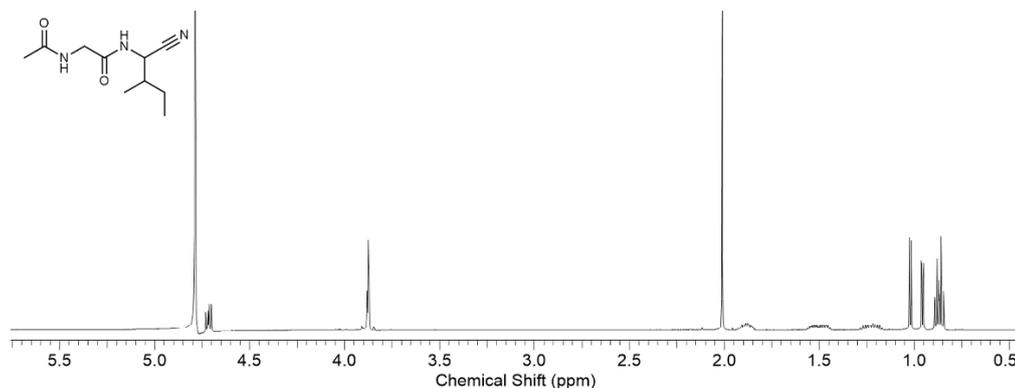
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 57:43 mixture of diastereoisomers a/b): diastereoisomer a: δ 4.71 (obs. d, *J* = 10.9 Hz, 1H, (Ile2)-(C2)-H), 3.88 (s, 2H, (Gly1)-(C2)-H<sub>2</sub>), 2.06 (s, 3H, (CO)-CH<sub>3</sub>), 1.93 – 1.84 (m, 1H, (Ile2)-(C3)-H), 1.57 – 1.45 (m, 1H, (Ile2)-(C4)-H), 1.29 – 1.16 (m, 1H, (Ile2)-(C4')-H), 1.02 (d, *J* = 6.8 Hz, 3H, (Ile2)-(C3)-CH<sub>3</sub>), 0.90 – 0.94 (m, 3H, Ile2-(C5)-H<sub>3</sub>); diastereoisomer b: δ 4.73 (obs. d, *J* = 7.1 Hz, 1H, (Ile2)-(C2)-H), 3.88 (s, 2H, (Gly1)-(C2)-H<sub>2</sub>), 2.06 (s, 3H, (CO)-CH<sub>3</sub>), 1.93 – 1.84 (m, 1H, (Ile2)-(C3)-H), 1.57 – 1.45 (m, 1H, (Ile2)-(C4)-H), 1.29 – 1.16 (m, 1H, (Ile2)-(C4')-H), 0.96 (d, *J* = 6.8 Hz, 3H, (Ile2)-(C3)-CH<sub>3</sub>), 0.90 – 0.94 (m, 3H, (Ile2)-(C5)-H<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 57:43 mixture of diastereoisomers a/b): diastereoisomer a: δ 175.5 ((Gly1)-C2), 172.0 (CO), 119.0 ((Ile2)-C1), 46.5 ((Ile2)-C2), 43.1 ((Gly1)-C2), 37.4 ((Ile2)-C3), 25.5 ((Ile)-C4), 22.2 ((CO)-CH<sub>3</sub>), 15.2, ((Ile2)-(C3)-CH<sub>3</sub>), 10.6 ((Ile2)-(C5)); diastereoisomer b: δ 175.5 ((Gly1)-C2), 172.1 (CO), 119.3 ((Ile2)-C1), 46.2 ((Ile2)-C2), 43.1 ((Gly1)-C2), 37.6 ((Ile2)-C3), 25.6 ((Ile)-C4), 22.2 ((CO)-CH<sub>3</sub>), 14.9, ((Ile2)-(C3)-CH<sub>3</sub>), 10.9 ((Ile2)-(C5)).

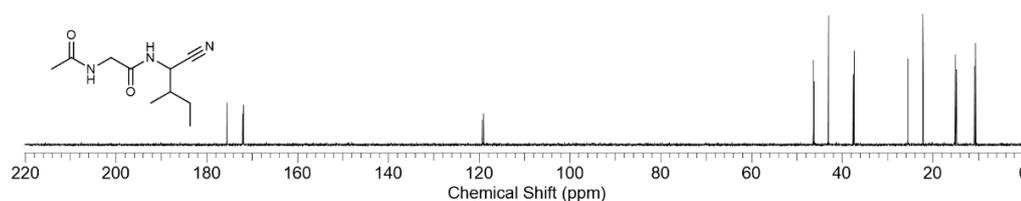
HRMS (ES<sup>+</sup>): calcd. for [C<sub>10</sub>H<sub>17</sub>N<sub>3</sub>O<sub>2</sub> + H]<sup>+</sup>: 212.1399; observed: 212.1388.

IR (solid): 3340, 3294, 3240, 3216, 3048, 2959, 2945, 2241, 1965, 1680, 1644, 1543 cm<sup>-1</sup>.

M.p. 112.1 – 113.0 °C.

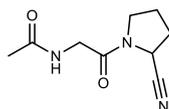


**Experimental figure 130:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 5.75 – 0.50 ppm) spectrum to show *N*-acetyl glycyl isoleucine nitrile **Ac-Gly-Ile-CN**.



**Experimental figure 131:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 220 – 0 ppm) spectrum to show *N*-acetyl glycyl isoleucine nitrile **Ac-Gly-Ile-CN**.

*N*-acetylglycyl proline nitrile **Ac-Gly-Pro-CN**



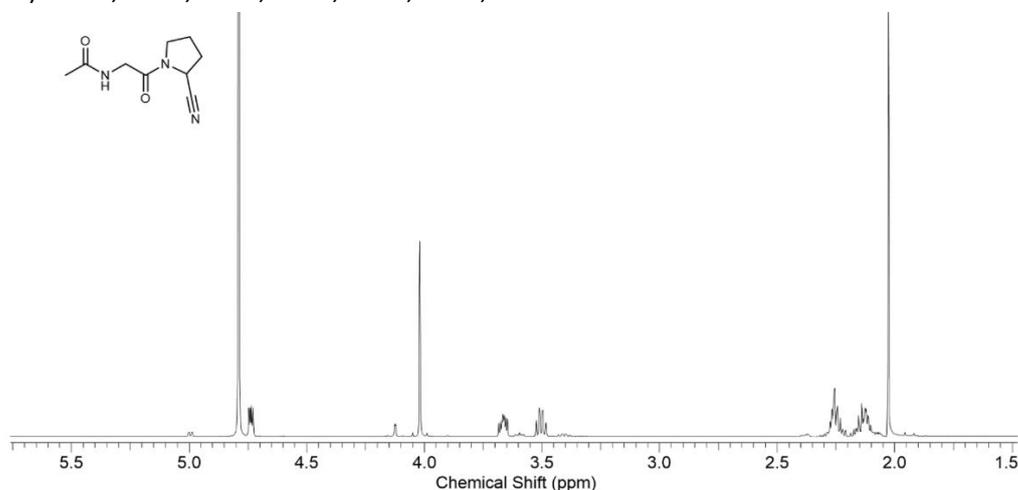
*N*-acetylglycyl proline nitrile **Ac-Gly-Pro-CN** was prepared following *General procedure H* using proline nitrile hydrochloride **Pro-CN**·HCl (159.1, 1.2 mmol). The reaction afforded *N*-acetylglycyl proline nitrile **Ac-Gly-Pro-CN** as a clear oil (136.2 mg, 0.70 mmol, 70%).

<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 90:10 mixture of rotamers a/b): rotamer a: δ 4.74 (dd, *J* = 4.0, 7.5 Hz, 1H, Pro2-(C2)-H), 4.02 (s, 2H, Gly1-(C2)-H), 3.66 (ddd, *J* = 4.2, 7.5, 9.9 Hz, 1H, Pro2-(C5)-H), 3.55 (ddd, *J* = 7.0, 8.3, 9.9 Hz, 1H, Pro2-(C5')-H), 2.30 – 2.20 (m, 2H, Pro2-(C3)-H<sub>2</sub>), 2.19 – 2.08 (m, 2H, Pro2-(C4)-H<sub>2</sub>), 2.02 (s, 3H, (CO)-CH<sub>3</sub>); rotamer b: δ 4.98 (dd, *J* = 1.9, 7.9 Hz, 1H, Pro2-(C2)-H), 4.12 (s, 2H, Gly1-(C2)-H), 3.62 – 3.57 (m, 1H, Pro2-(C5)-H), 3.44 – 3.38 (m, 1H, Pro2-(C5')-H), 2.30 – 2.20 (m, 2H, Pro2-(C3)-H<sub>2</sub>), 2.19 – 2.08 (m, 2H, Pro2-(C4)-H<sub>2</sub>), 2.13 (obs. s, 3H, (CO)-CH<sub>3</sub>).

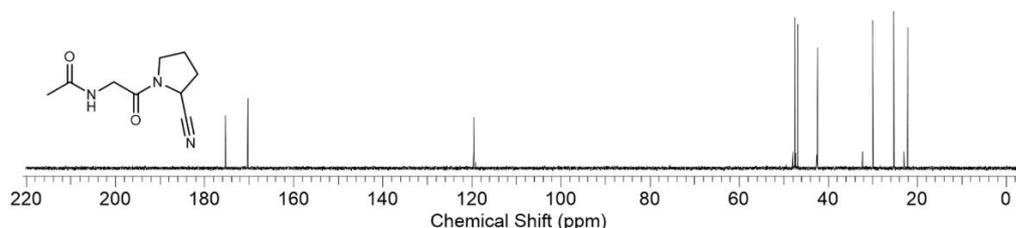
<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 90:10 mixture of rotamers a/b): rotamer a: δ 175.2 (CO), 170.3 (Gly1-C1), 119.6 (Pro-C1), 47.6 (Pro-C2), 46.8 (Pro-C5), 42.4 (Gly1-C2), 30.0 (Pro-C3), 25.3 (Pro-C4), 22.2 ((CO)-CH<sub>3</sub>); rotamer b: δ 175.3 (CO), 170.3 (Gly1-C1), 119.2 (Pro-C1), 47.9 (Pro-C2), 46.8 (Pro-C5), 42.6 (Gly1-C2), 32.7 (Pro-C3), 25.4 (Pro-C4), 23.0 ((CO)-CH<sub>3</sub>).

HRMS (ES<sup>+</sup>): calcd. for [C<sub>9</sub>H<sub>13</sub>N<sub>3</sub>O<sub>2</sub> + H]<sup>+</sup>: 196.1081; observed: 196.1081.

IR (solid): 3307, 3093, 2957, 2883, 2243, 1636, 1537 cm<sup>-1</sup>.

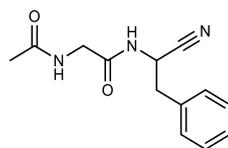


**Experimental figure 132:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 5.75 – 1.50 ppm) spectrum to show *N*-acetyl glycyl proline nitrile **Ac-Gly-Pro-CN**.



**Experimental figure 133:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 220 – 0 ppm) spectrum to show *N*-acetyl glycyl proline nitrile **Ac-Gly-Pro-CN**.

*N*-acetylglycyl phenylalanine nitrile **Ac-Gly-Phe-CN**



*N*-acetylglycyl phenylalanine nitrile **Ac-Gly-Phe-CN** was prepared following *General procedure H* using phenylalanine nitrile hydrochloride **Phe-CN**·HCl (547.9 mg, 3.0 mmol). The reaction afforded *N*-acetylglycyl phenylalanine nitrile **Ac-Gly-Phe-CN** as a white solid (218.3 mg, 0.89 mmol, 84%).

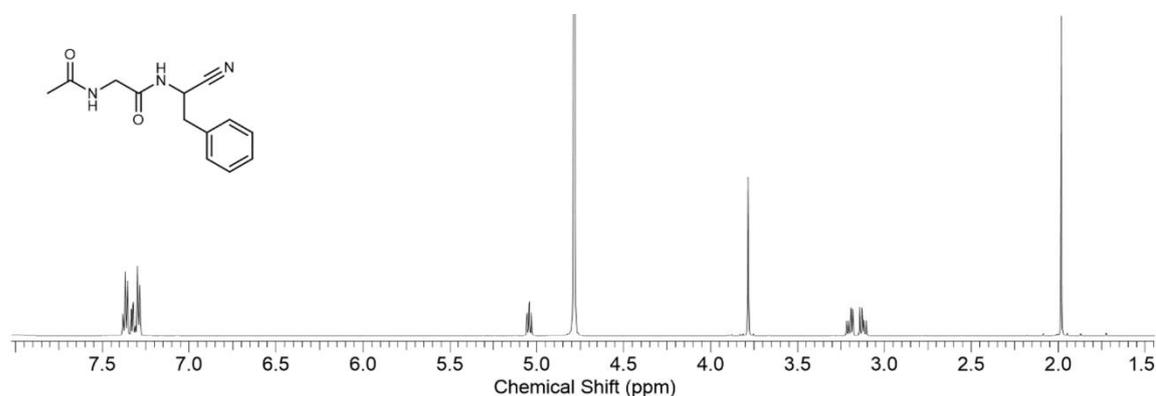
$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  7.39 – 7.28 (m, 5H, Ar), 5.05 (dd,  $J = 6.9, 8.2$  Hz, 1H, Phe2-(C1)-H), 3.79 (s, 2H, Gly1-(C2)-H<sub>2</sub>), 3.21 (ABX,  $J = 6.9, 13.8$  Hz, 1H, (Phe2)-(C3)-H), 3.13 (ABX,  $J = 8.2, 13.8$  Hz, 1H, (Phe2)-(C3')-H), 1.98 (s, 3H, (CO)-CH<sub>3</sub>).

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  175.5 (CO), 171.9 (Gly1-C2), 135.4 (Ar), 130.1 (Ar  $\times$  2), 129.5 (Ar  $\times$  2), 128.4 (Ar), 119.2 (Phe2-C1), 43.0 (Phe2-C2), 42.6 (Phe2-C3), 38.0 (Gly1-C2), 22.2 ((CO)-CH<sub>3</sub>).

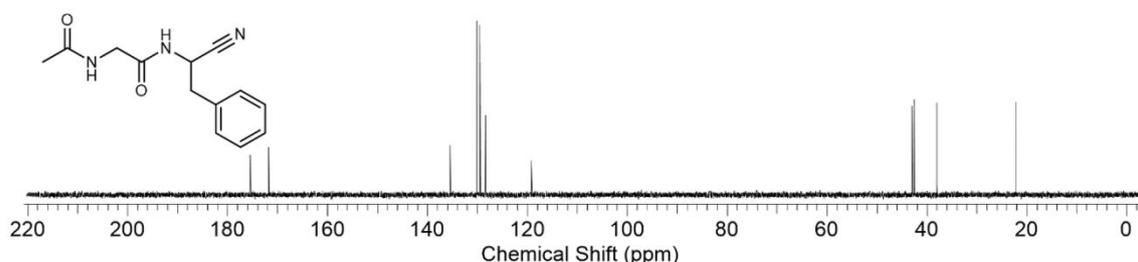
**HRMS** (ES<sup>+</sup>): calcd. for [C<sub>13</sub>H<sub>15</sub>N<sub>3</sub>O<sub>2</sub> + H]<sup>+</sup>: 246.1237; observed: 246.1244.

**IR** (solid): 3285, 3227, 3098, 3057, 2982, 2930, 2862, 2231, 1679, 1631, 1554, 1501 cm<sup>-1</sup>.

**M.p.** 144.2 – 145.8 °C.

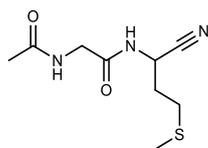


**Experimental figure 134:**  $^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ , 7.75 – 1.50 ppm) spectrum to show *N*-acetylglycyl phenylalanine nitrile **Ac-Gly-Phe-CN**.



**Experimental figure 135:**  $^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ , 220 – 0 ppm) spectrum to show *N*-acetylglycyl phenylalanine nitrile **Ac-Gly-Phe-CN**.

*N*-acetylglycyl methionine nitrile **Ac-Gly-Met-CN**



*N*-acetyl glycyl methionine nitrile **Ac-Gly-Met-CN** was prepared following *General procedure H* using methionine nitrile hydrochloride **Met-CN**·HCl (500.0 mg, 3.0 mmol). The reaction afforded *N*-acetylglycyl methionine nitrile **Ac-Gly-Met-CN** as a white solid (306 mg, 1.34 mmol, 53%).

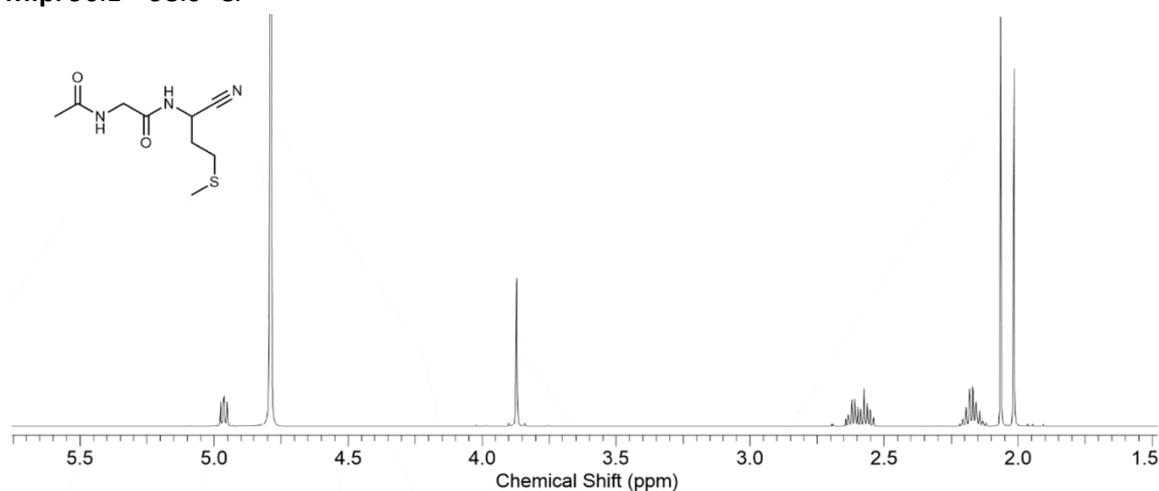
$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  4.96 (dd,  $J = 6.5, 8.5$  1H, Met2-(C2)-H), 3.81 (s, 2H, Gly1-(C2)-H<sub>2</sub>), 2.65 – 2.53 (m, 2H, Met2-(C3)-H<sub>2</sub>), 2.22 – 2.11 (m, 2H, Met2-(C4)-H<sub>2</sub>), 2.06 (s, 3H, Met2-SCH<sub>3</sub>), 2.02 (s, 3H, (CO)-CH<sub>3</sub>).

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  176.5 (CO-CH<sub>3</sub>), 172.2 (Gly1-C1), 119.2 (Met2-C1), 43.1 (Met2-C2), 40.1 (Gly1-C2), 31.1 (Met2-C4), 29.3 (Met2-C3), 22.3 ((CO)-CH<sub>3</sub>), 14.6 (Met2-SCH<sub>3</sub>).

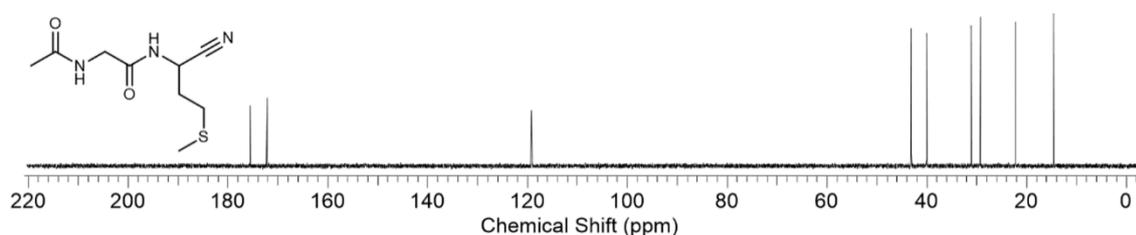
**HRMS** ( $\text{ES}^+$ ): calcd. for  $[\text{C}_9\text{H}_{15}\text{N}_3\text{O}_2\text{S} + \text{H}]^+$ : 230.0958 observed: 230.0962.

**IR** (solid): 3242, 3077, 3043, 2952, 2924, 2856, 2240, 1679, 1651, 1636, 1563, 1534  $\text{cm}^{-1}$ .

**M.p.** 96.1 – 98.0 °C.



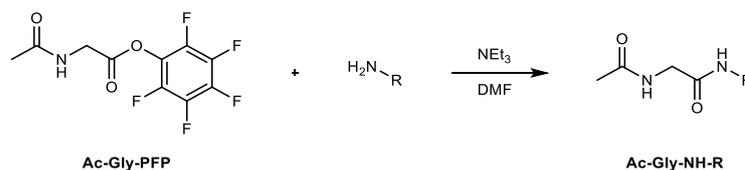
**Experimental figure 136:**  $^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ , 5.75 – 1.50 ppm) spectrum to show *N*-acetylglycyl methionine nitrile **Ac-Gly-Met-CN**.



**Experimental figure 137:**  $^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ , 220 – 0 ppm) spectrum to show *N*-acetylglycyl methionine nitrile **Ac-Gly-Met-CN**.

### 9. 3. 6. Synthesis and characterisation of competitor ligation byproducts

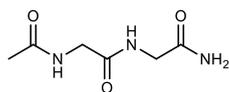
General procedure I: Preparative acetylation of competitor ligation byproducts



Triethylamine (607 mg, 837  $\mu\text{L}$ , 6.0 mmol) and the appropriate amine (3.0 mmol) were dissolved in DMF (30 mL). *N*-Acetyl glycine pentafluorophenyl ester **Ac-Gly-PFP** (708 mg, 2.5 mmol) was added, and the solution was stirred for 16 h. The solution was then concentrated *in vacuo* and solids washed with ether (3  $\times$  50 mL) and  $\text{CH}_2\text{Cl}_2$  (3  $\times$  50 mL). The solid residue was then purification by flash column chromatography ( $\text{SiO}_2$ ; eluting with a gradient of  $\text{MeOH}/\text{CH}_2\text{Cl}_2$  0:100  $\rightarrow$  10:90) to afford amide **Ac-Gly-NH-R** as a white solid.

Experimental table 15: preparation of <i>competitor ligation byproducts</i>			
Entry	Residue	R	Yield dipeptide (%)
1	Gly-NH <sub>2</sub>	CH <sub>2</sub> -CONH <sub>2</sub>	91%
2	Gly-OH	CH <sub>2</sub> -CO <sub>2</sub> H	82%
3	-	H	41%
4	$\beta$ -Ala	CH <sub>2</sub> -CH <sub>2</sub> -CO <sub>2</sub> H	83%
5	-	CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>3</sub>	80%

*N*-acetylglycyl glycinamide **Ac-Gly-Gly-NH<sub>2</sub>**



*N*-Acetylglycyl glycinamide **Ac-Gly-Gly-NH<sub>2</sub>** was prepared following *General procedure 1* using glycine nitrile hydrochloride **Gly-NH<sub>2</sub>·HCl** (331.6 mg, 3.0 mmol). The reaction afforded *N*-acetylglycyl glycinamide **Ac-Gly-Gly-NH<sub>2</sub>** as a white solid (392.6 mg, 2.27 mmol, 91%).

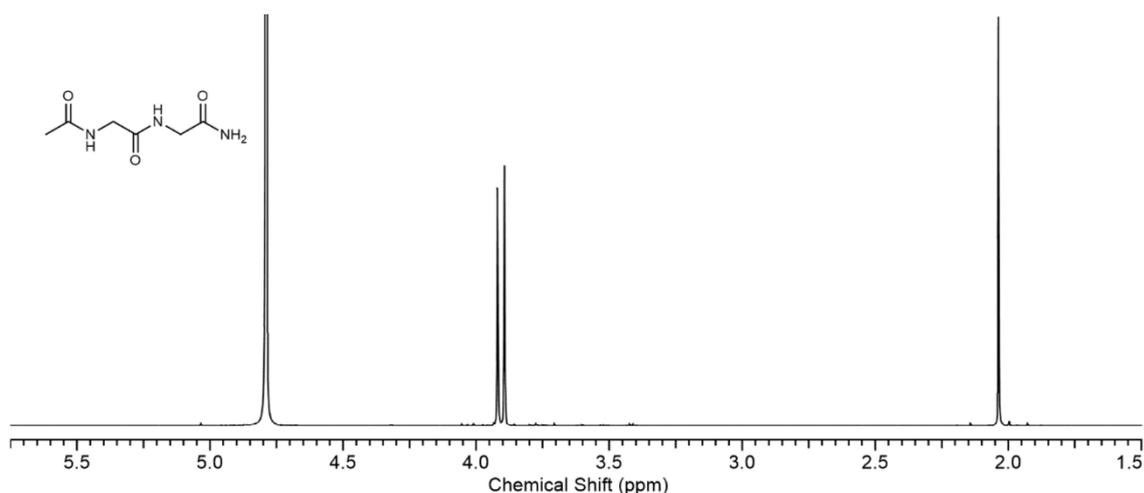
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 3.92 (s, 2H, Gly2-(C2)-H<sub>2</sub>), 3.89 (s, 2H, Gly1-(C2)-H<sub>2</sub>), 2.04 (s, 3H, (CO)-CH<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 175.8 (Gly2-(C1)), 174.9 (Gly1-(C1)), 173.1 (COCH<sub>3</sub>), 43.3 (Gly2-(C2)), 42.6 (Gly1-(C2)), 22.4 (COCH<sub>3</sub>).

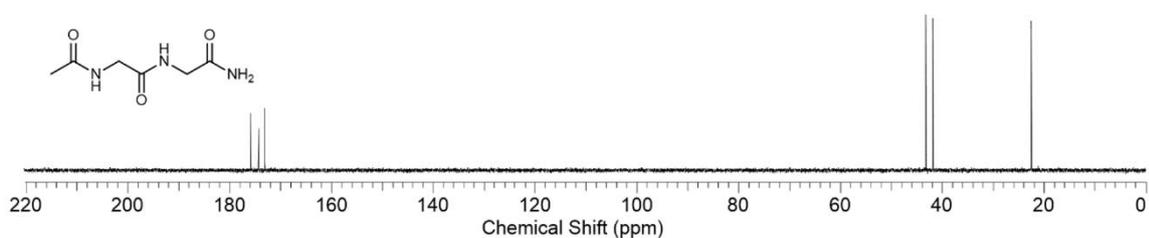
HRMS (ES<sup>-</sup>): calcd. for [C<sub>6</sub>H<sub>11</sub>N<sub>3</sub>O<sub>3</sub>-H]<sup>-</sup>: 172.0722; observed: 172.0725.

IR (solid): 3382, 3294, 3194, 3079, 2877, 2939, 2603, 2493, 1651, 2634, 1548 cm<sup>-1</sup>.

M.p. 197.1 – 199.9 °C.

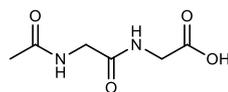


**Experimental figure 138:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 5.75 – 1.50 ppm) spectrum to show *N*-acetylglycyl glycinamide **Ac-Gly-Gly-NH<sub>2</sub>**.



**Experimental figure 139:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 220 – 0 ppm) spectrum to show *N*-acetylglycyl glycinamide **Ac-Gly-Gly-NH<sub>2</sub>**.

Synthesis of *N*-acetylglycyl glycine **Ac-Gly-Gly-OH**



*N*-Acetylglycyl glycine **Ac-Gly-Gly-OH** was prepared following *General procedure I* using glycine **Gly-OH** (225.2 mg, 3.0 mmol). The reaction afforded *N*-acetylglycyl glycine **Ac-Gly-Gly-OH** as a white solid (358.6 mg, 2.06 mmol, 91%).

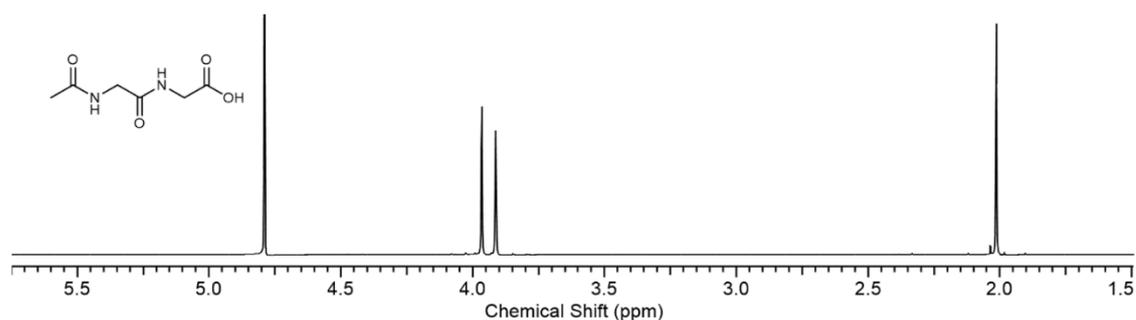
$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  3.97 (s, 2H, Gly1-(C2)-H<sub>2</sub>), 3.91 (s, 2H, Gly2-(C2)-H<sub>2</sub>), 2.02 (s, 3H, (CO)-CH<sub>3</sub>).

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  175.6 (Gly2-(C1)), 173.9 (Gly1-(C1)), 172.8 (COCH<sub>3</sub>), 43.0 (Gly1-(C2)), 41.6 (Gly2-(C2)), 22.3 (COCH<sub>3</sub>).

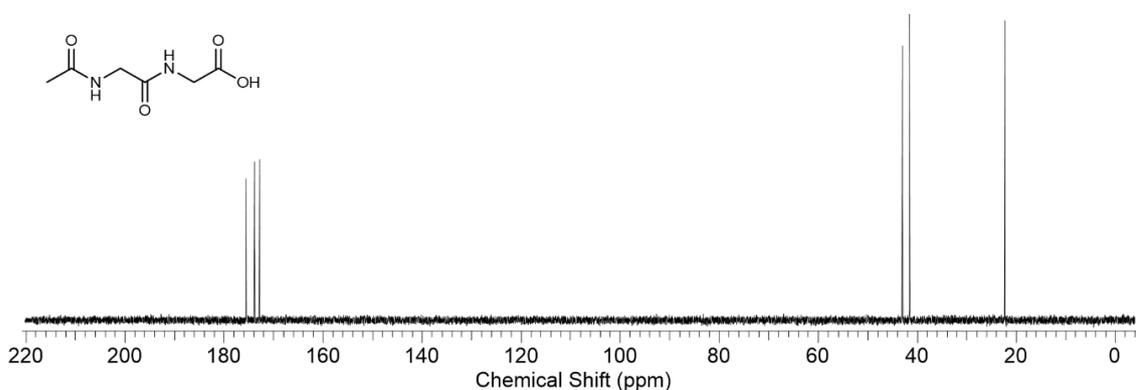
**HRMS** ( $\text{ES}^+$ ): calcd. for [ $\text{C}_6\text{H}_{10}\text{N}_2\text{O}_4 + \text{H}$ ]<sup>+</sup>: 175.0713; observed: 175.0714.

**IR** (solid): 3291, 3105, 1701, 1690, 1669, 1603, 1585, 1550  $\text{cm}^{-1}$ .

**M.p.** 183.5 – 186.0 °C.

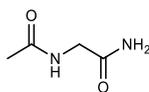


**Experimental figure 140:**  $^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ , 5.75 – 1.50 ppm) spectrum to show *N*-acetylglycyl glycine **Ac-Gly-Gly-OH**.



**Experimental figure 141:**  $^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ , 220 – 0 ppm) spectrum to show *N*-acetylglycyl glycine **Ac-Gly-Gly-OH**.

Synthesis of *N*-acetylglycinamide **Ac-Gly-NH<sub>2</sub>**



*N*-Acetylglycinamide **Ac-Gly-NH<sub>2</sub>** was prepared following *General procedure I* using ammonium chloride (160.467 mg, 3.0 mmol). The reaction afforded *N*-acetylglycinamide **Ac-Gly-NH<sub>2</sub>** as a white solid (118.0 mg, 1.02 mmol, 41%).

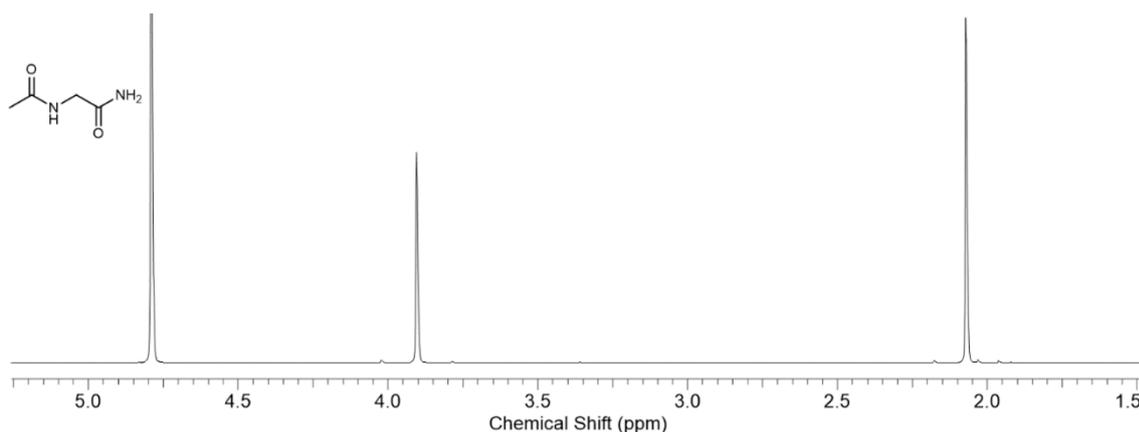
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 3.90 (s, 2H, (C2)-H<sub>2</sub>), 2.07 (s, 3H, (CO)-CH<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 175.6 (C1), 175.2 (COCH<sub>3</sub>), 42.8 (C2), 22.4 (COCH<sub>3</sub>).

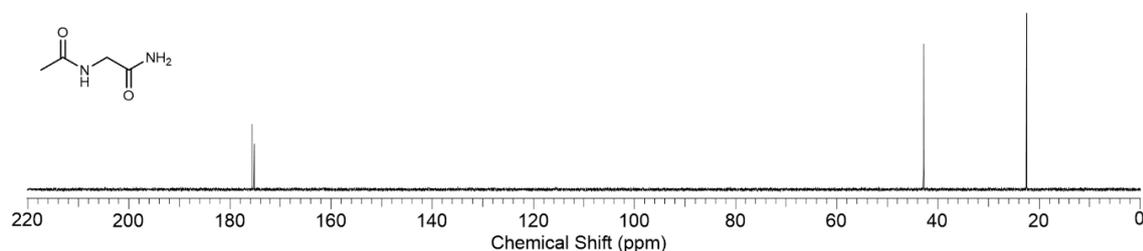
HRMS (ES<sup>+</sup>): calcd. for [C<sub>4</sub>H<sub>8</sub>N<sub>2</sub>O<sub>2</sub> + H]<sup>+</sup>: 117.0664; observed: 117.0665.

IR (solid): 3350, 3179, 3064, 2973, 2930, 2851, 2782, 1683, 1638, 1551 cm<sup>-1</sup>.

M.p. 138.2 – 139.4 °C.

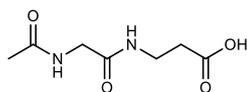


**Experimental figure 142:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 5.25 – 1.50 ppm) spectrum to show *N*-acetylglycinamide **Ac-Gly-NH<sub>2</sub>**.



**Experimental figure 143:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 220 – 0 ppm) spectrum to show *N*-acetylglycinamide **Ac-Gly-NH<sub>2</sub>**.

Synthesis of *N*-acetylglycyl  $\beta$ -alanine **Ac-Gly- $\beta$ -Ala**



*N*-Acetylglycyl  $\beta$ -alanine **Ac-Gly- $\beta$ -Ala** was prepared following *General procedure I* using  $\beta$ -alanine  **$\beta$ -Ala** (222.7 mg, 3.0 mmol). The reaction afforded *N*-acetylglycyl  $\beta$ -alanine **Ac-Gly- $\beta$ -Ala** as a white solid (388.6 mg, 2.07 mmol, 83%).

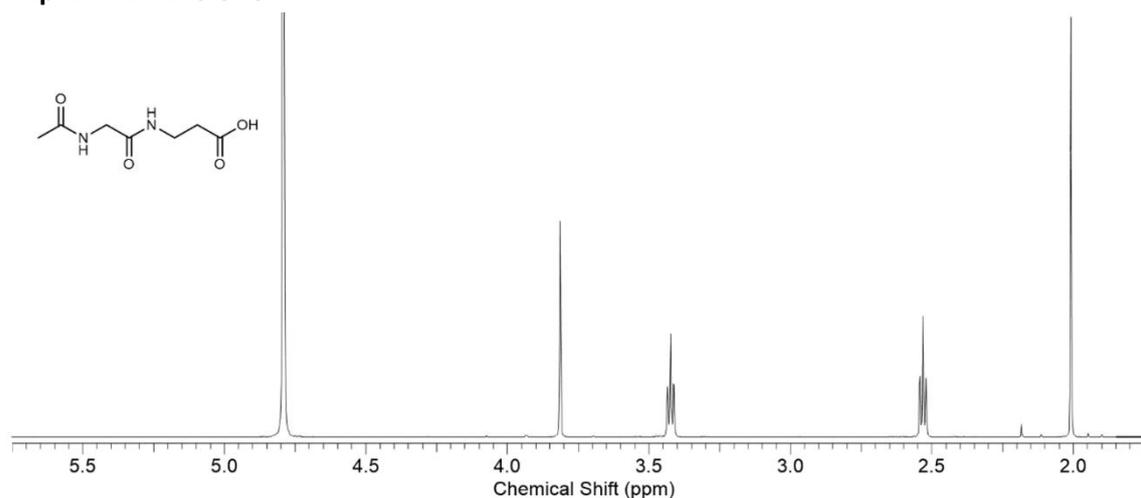
$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  3.81 (s, 2H, Gly1-(C2)-H<sub>2</sub>), 3.42 (t,  $J = 7.0$  Hz, 2H,  $\beta$ -Ala2-(C2)-H<sub>2</sub>), 2.53 (t,  $J = 7.0$  Hz, 2H,  $\beta$ -Ala2-(C3)-H<sub>2</sub>), 2.00 (s, 3H, (CO)-CH<sub>3</sub>).

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  176.9 ( $\beta$ -Ala2-(C1)), 175.5 (COCH<sub>3</sub>), 172.2 (Gly1-(C1)), 43.2 (Gly1-(C2)), 35.7 ( $\beta$ -Ala2-(C3)), 34.1 ( $\beta$ -Ala2-(C2)), 22.3 (COCH<sub>3</sub>).

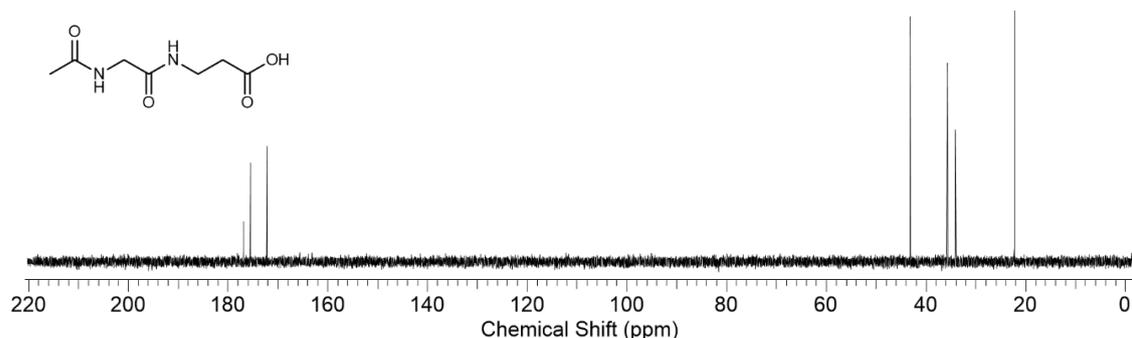
HRMS (ES<sup>-</sup>): calcd. for [C<sub>7</sub>H<sub>12</sub>N<sub>2</sub>O<sub>4</sub>-H]<sup>-</sup>: 187.0719; observed: 187.0718.

IR (solid): 3336, 3302, 3094, 2943, 2495. 1958, 1699, 1654, 1606, 1549 cm<sup>-1</sup>.

M.p. 144.1 – 145.8 °C.

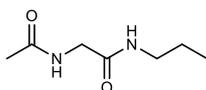


**Experimental figure 144:**  $^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ , 5.75 – 1.75 ppm) spectrum to show *N*-acetylglycyl  $\beta$ -alanine **Ac-Gly- $\beta$ -Ala**.



**Experimental figure 145:**  $^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ , 220 – 0 ppm) spectrum to show *N*-acetylglycyl  $\beta$ -alanine **Ac-Gly- $\beta$ -Ala**.

Synthesis of *N*-acetylglycyl propylamine **Ac-Gly-Prop**



*N*-Acetylglycyl propylamine **Ac-Gly-Prop** was prepared following *General procedure 1* using *n*-propylamine (177.3 mg, 247  $\mu$ L, 3.0 mmol) as starting material. The reaction afforded *N*-acetylglycyl propylamine **Ac-Gly-Prop** as a white solid (316.0 mg, 2.0 mmol, 80%).

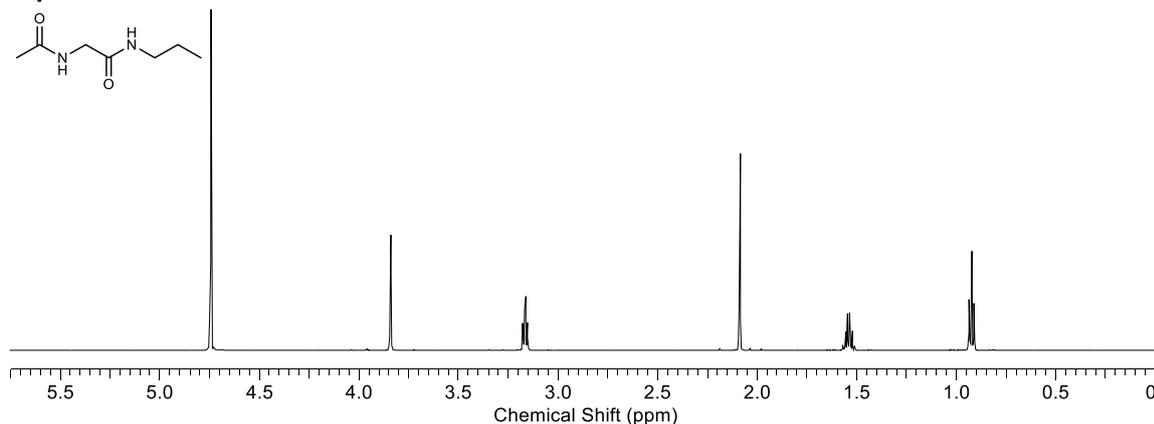
$^1\text{H NMR}$  (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  3.83 (s, 2H, (Gly1)-(C2)-H<sub>2</sub>), 3.17 (t,  $J = 6.9$  Hz, 2H, (Prop2)-(C1)-H<sub>2</sub>), 2.07 (s, 3H, (CO)-CH<sub>3</sub>), 1.56 – 1.50 (m, 2H, (Prop2)-(C2)-H<sub>2</sub>), 0.92 (t,  $J = 7.4$  Hz, 3H, (Prop2)-(C3)-H<sub>3</sub>).

$^{13}\text{C NMR}$  (151 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  175.6 ((Gly1)-C1), 172.1 (CO), 43.4 ((Gly1)-C2), 41.8 (Prop2)-C1), 22.5 (Prop2)-C2), 22.4 ((CO)-CH<sub>3</sub>), 11.2 (Prop2)-C3).

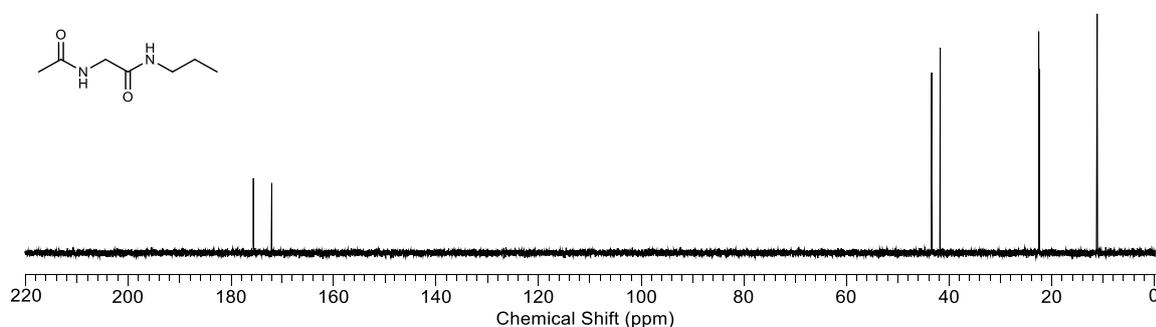
**HRMS** ( $\text{ES}^+$ ): calcd. for  $[\text{C}_7\text{H}_{14}\text{N}_2\text{O}_2 + \text{H}]^+$ : 159.1134; observed: 159.1130.

**IR** (solid): 3269, 3088, 2967, 2933, 2874, 1636, 1554  $\text{cm}^{-1}$ .

**M.p.** 139.1 – 141.7  $^\circ\text{C}$ .



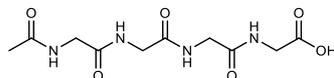
**Experimental figure 146:**  $^1\text{H NMR}$  spectrum (600 MHz,  $\text{D}_2\text{O}$ , 5.75 – 0.00 ppm) to show *N*-acetylglycyl propylamine **Ac-Gly-Prop**.



**Experimental figure 147:**  $^{13}\text{C NMR}$  spectrum (151 MHz,  $\text{D}_2\text{O}$ , 220 – 0 ppm) to show *N*-acetylglycyl propylamine **Ac-Gly-Prop**.

### 9. 3. 7. Synthesis and characterisation of glycine oligomers

#### *N*-Acetylglycylglycylglycylglycine **Ac-Gly-Gly-Gly-Gly-OH**



Glycylglycylglycine (1.0 g, 5.27 mmol) and triethylamine (642 mg, 0.88 mL, 6.34 mmol) were dissolved in DMF (50 mL). *N*-Acetylglycine pentafluorophenyl ester (1.49 g, 5.27 mmol) was added, and the reaction was stirred at room temperature for 2 h. After that time, the reaction was concentrated *in vacuo*. The crude residue was dissolved in H<sub>2</sub>O at pH 1.0 (10 mL) and extracted with DCM (3 × 50 mL). The aqueous later was kept at 4 °C for 24 h. After that time, a white precipitate formed, which was isolated by filtration and washed with H<sub>2</sub>O (1 × 20 mL), methanol (1 × 20 mL) and diethyl ether (1 × 50 mL) to afford *N*-Acetylglycylglycylglycylglycine **Ac-Gly-Gly-Gly-Gly-OH** as a white powder (1.09, 3.78 mmol, 71%).

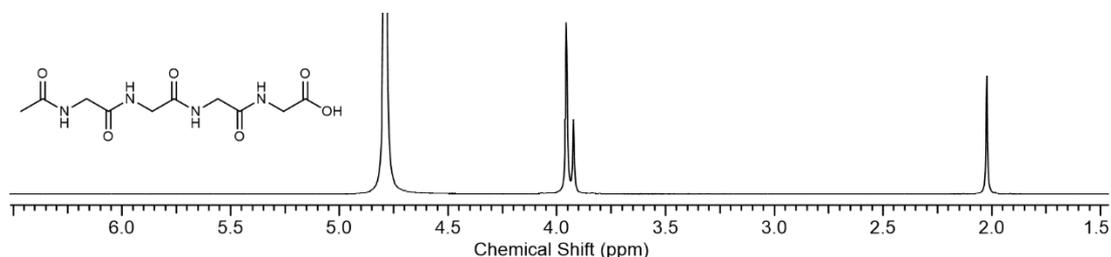
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 3.96 (br s, 6H, Gly<sub>2</sub>-(C2)-H<sub>2</sub>), (Gly<sub>3</sub>-(C2)-H<sub>2</sub>), (Gly<sub>4</sub>-(C2)-H<sub>2</sub>), 3.92 (s, 2H, Gly<sub>1</sub>-(C2)-H<sub>2</sub>), 2.03 (s, 3H, (CO)CH<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 175.6 (CO)CH<sub>3</sub>, 174.1 (Gly<sub>1</sub>-(C1)), 173.1 (Gly<sub>2</sub>-(C1)), 172.7 (Gly<sub>3</sub>-(C1)), 172.4 (Gly<sub>1</sub>-(C2)), 43.1 (Gly<sub>1</sub>-(C2)), 43.0 (Gly<sub>3</sub>-(C2)), (Gly<sub>2</sub>-(C2)), 41.7 (Gly<sub>4</sub>-(C2)), 22.6 (CO)CH<sub>3</sub>.

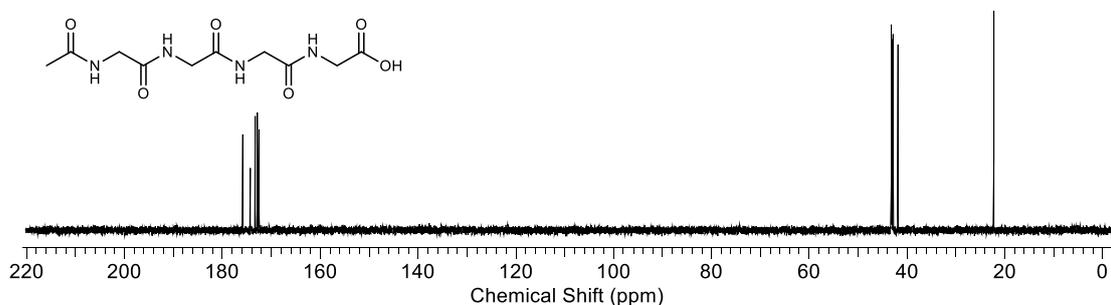
HRMS (ES<sup>+</sup>): calcd. for [C<sub>10</sub>H<sub>16</sub>N<sub>4</sub>O<sub>6</sub>Na]<sup>+</sup>: 311.0962; observed: 311.0967.

IR (solid): 3330, 3260, 3083, 2978, 2937, 1707, 1637, 1551 cm<sup>-1</sup>.

M.p.: 245.0 – 246.6 °C (decomp.).

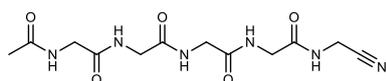


**Experimental figure 148:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 1.50 – 6.50 ppm) spectrum to show *N*-acetylglycylglycylglycylglycine **Ac-Gly-Gly-Gly-Gly-OH**.



**Experimental figure 149:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show *N*-acetylglycylglycylglycylglycine **Ac-Gly-Gly-Gly-Gly-OH**.

*N*-Acetylglycylglycylglycylglycylglycine nitrile **Ac-Gly-Gly-Gly-Gly-Gly-CN**



*N*-Acetylglycylglycylglycylglycylglycine **Ac-Gly-Gly-Gly-Gly-OH** (500 mg, 1.73 mmol), glycine nitrile hydrochloride **Gly-CN·HCl** (480 mg, 5.19 mmol) and triethylamine (875 mg, 1.21 mL, 8.65 mmol) were dissolved in DMF (25 mL). EDC hydrochloride (497 mg, 2.60 mmol) was added, and the reaction was stirred at room temperature for 4 h. After that time, the reaction was concentrated *in vacuo*. The crude residue was dissolved in H<sub>2</sub>O (20 mL) and refrigerated at 4 °C for 24 h. After that time, a white precipitate formed, which was isolated by filtration and washed with H<sub>2</sub>O (5 × 25 mL), methanol (5 × 25 mL) and ether (5 × 25 mL) to afford *N*-acetylglycylglycylglycylglycylglycine nitrile **Ac-Gly-Gly-Gly-Gly-Gly-CN** as a white powder (203 mg, 0.62 mmol, 36%).

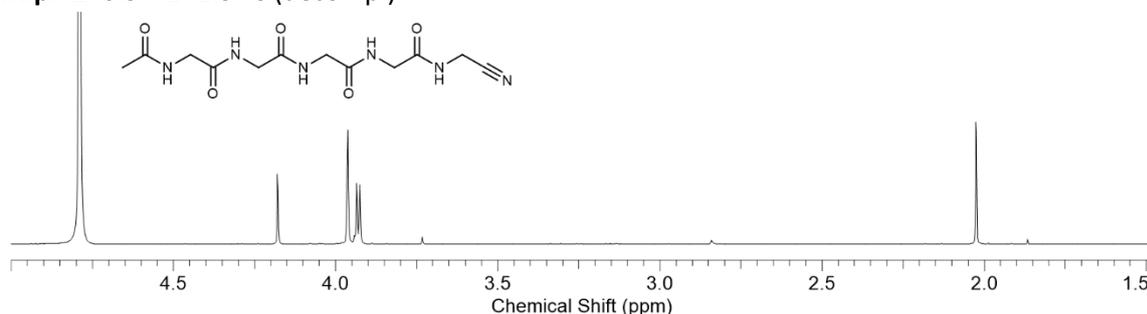
<sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O): δ 4.18 (s, 2H, Gly<sub>5</sub>-(C2)-H<sub>2</sub>), 3.96 (s, 4H, Gly<sub>3</sub>-(C2)-H<sub>2</sub>), Gly<sub>4</sub>-(C2)-H<sub>2</sub>), 3.94 (s, 2H, Gly<sub>2</sub>-(C2)-H<sub>2</sub>), 3.96 (s, 2H, Gly<sub>1</sub>-(C2)-H<sub>2</sub>), 2.02 (s, 3H, (CO)CH<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 175.8 (COCH<sub>3</sub>), 173.3 (Gly<sub>1</sub>-(C1)), 173.0 (Gly<sub>3</sub>-(C1)), 172.8 (Gly<sub>2</sub>-(C1)), 172.5 (Gly<sub>4</sub>-(C1)), 117.4 (Gly<sub>5</sub>-(C1)), 42.2 (Gly<sub>1</sub>-(C2)), 43.1 (Gly<sub>3</sub>-(C2)), 42.9 (Gly<sub>2</sub>-(C2)), 42.8 (Gly<sub>4</sub>-(C2)), 29.1 (Gly<sub>5</sub>-(C2)), 22.3 (COCH<sub>3</sub>).

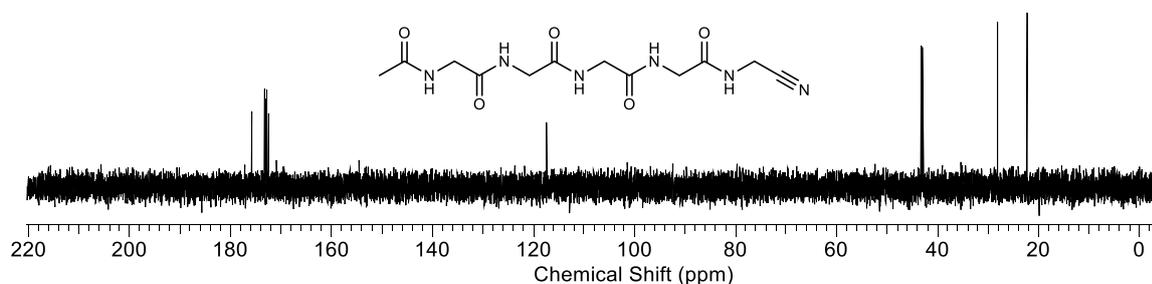
HRMS (ES<sup>+</sup>): calcd. for [C<sub>12</sub>H<sub>18</sub>N<sub>6</sub>O<sub>5</sub>+H]<sup>+</sup> 327.1411; observed: 327.1419.

IR (solid): 3368, 3329, 3312, 3229, 3134, 2198, 1742, 1687, 1565, 1501 cm<sup>-1</sup>.

M.p.: 270.5 – 271.9 °C (decomp.).



**Experimental figure 150:** <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O, 1.50 – 5.00 ppm) spectrum to show *N*-acetylglycylglycylglycylglycylglycine nitrile **Ac-Gly-Gly-Gly-Gly-Gly-CN**.



**Experimental figure 151:** <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O, 0 – 220 ppm) spectrum to show *N*-acetylglycylglycylglycylglycylglycine nitrile **Ac-Gly-Gly-Gly-Gly-Gly-CN**.

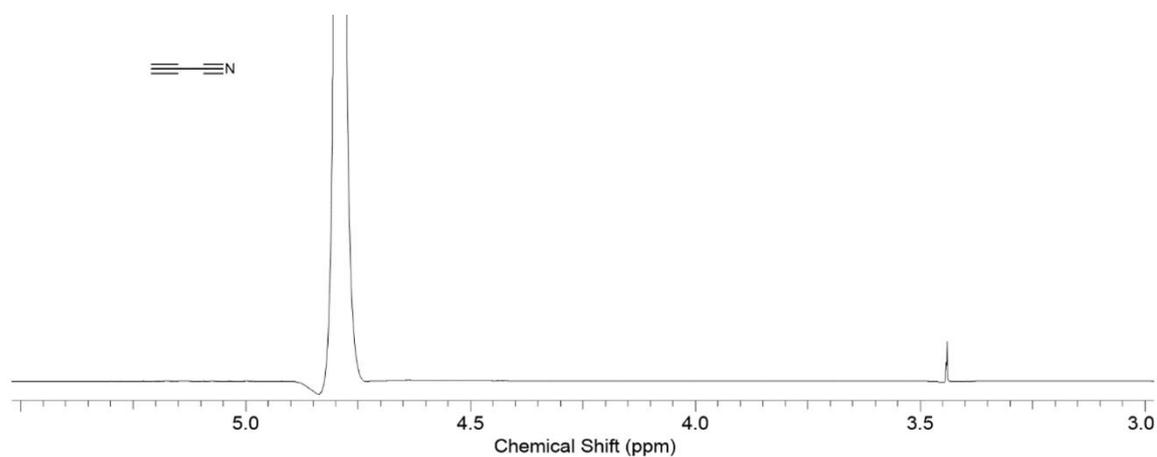
### 9. 3. 8. Preparation of cyanoacetylene



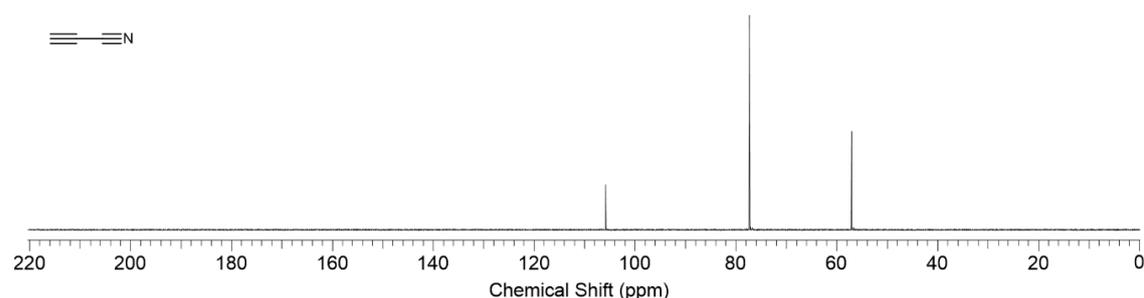
Methyl propiolate (20.0 g, 237.9 mmol) was dissolved in ammonia (200 mL) and the reaction was stirred at -78 °C for 4 h. The solution was then concentrated *in vacuo* to afford propiolamide as a white solid (16.41 g, 237.6 mmol). The resulting propiolamide, phosphorous pentoxide (74.2 g, 261.4 mmol) and sand (260 g) were ground into a homogeneous solid mixture and cyanoacetylene was distilled from the mixture at 130 °C over 2 h, condensed at -78 °C, and stored as an aqueous solution (1 M, 187 mL, 186.9 mmol, 79%).

$^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2):  $\delta$  3.51 (s, 1H, (C3)-H).

$^{13}\text{C}$  NMR (151 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2):  $\delta$  105.6 (C1), 77.3 (C2), 57.0 (C3).



**Experimental figure 152:**  $^1\text{H}$  NMR (600 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 5.50 – 3.00 ppm) spectrum to show cyanoacetylene.



**Experimental figure 153:**  $^{13}\text{C}$  NMR (151 MHz,  $\text{H}_2\text{O}/\text{D}_2\text{O}$  98:2, 220 – 0 ppm) spectrum to show cyanoacetylene.

## 10. Bibliography.

- [1]. Whitesides, G. M., *Angew. Chem.*, **2015**, *54*, 3196-3209.
- [2]. Oldroyd, D. R., *Biol. Philos.*, **1986**, *1*, 133-168.
- [3]. Pross, A., Pascal, R., *Open Biol.*, **2013**, *3*.
- [4]. Alberts B, J. A., Lewis J, *et al.*, *Molecular Biology of the Cell*, 4th edition ed., **2002**.
- [5]. Powner, M. W., Sutherland, J. D., *Philos. Trans. Royal Soc. B*, **2011**, *366*, 2870-2877.
- [6]. Noffke, N. D. C., Hazen, R. M., *Astrobiology*, **2013**, *13*, 1103-1124.
- [7]. Williams, J. P., *Contemp. Phys.*, **2010**, *51*, 381-396.
- [8]. Bonanno, A., Schlattl, H., Paternò, L., *Astron. Astrophys.*, **2002**, *390*, 1115-1118.
- [9]. Amelin, Y., Krot, A. N., Ulyanov, A. A., *Science*, **2002**, *297*, 1678-1683.
- [10]. Alan, P. B., Sandra, A. K., *Astrophys. J. Lett.*, **2010**, *717*.
- [11]. Goldreich, P., Ward, W. R., *Astrophys. J.*, **1973**, *118*, 1051-1062.
- [12]. Orgel, L. E., *Trends. Biochem. Sci.*, **1998**, *23*, 491-495.
- [13]. Jacobs, J. A., *Nature*, **1953**, *172*, 297-298.
- [14]. Canup, R. M., Asphaug, E., *Nature*, **2001**, *412*, 708-712.
- [15]. Canup, R. M., *Annu. Rev. Astron. Astrophys.*, **2004**, *42*, 441-475.
- [16]. Mojzsis, S. J., Harrison, T. M., Pidgeon, R. T., *Nature*, **2001**, *409*, 178-181.
- [17]. Cui, S., *Phys. Chem. Chem. Phys.*, **2010**, *12*, 10147-10153.
- [18]. Gomes, R., Levison, H. F., Tsiganis, K., Morbidelli, A., *Nature*, **2005**, *435*, 466-469.
- [19]. Ryder, G., Koeberl, C., Mojzsis, S. J., *Origin of the Earth and Moon*, **2000**, *1*, 475-492.
- [20]. Abramov, O., Mojzsis, S. J., *Nature*, **2009**, *459*, 419-422.
- [21]. Gifford, D. P., *Advances in archaeological method and theory*, **1981**, *4*, 365-438.
- [22]. Valley, J. W. *et al.*, *Nat. Geosci.*, **2014**, *7*, 219-223.
- [23]. Wacey, D. *et al.*, *Nat. Geosci.*, **2011**, *4*, 698-702.
- [24]. A. M. Satkoskia, N. J. B., W. Lid *et al.*, *Earth Planet. Sci. Lett.*, **2015**, *430*, 43-53.
- [25]. Grosberg, R. K., Strathmann, R. R., *Rev. Ecol. Evol. Syst.*, **2007**, 621-654.
- [26]. Holland, H. D., *Philos. Trans. Royal Soc. B*, **2006**, *361*, 903-915.
- [27]. El Albani, A. *et al.*, *Nature*, **2010**, *466*, 100-104.
- [28]. Nesbitt, S. J. *et al.*, *Biol. Lett.*, **2013**, *9*, 20120949.
- [29]. Feldhamer, G. *et al.*, *C., Ecology*, **2007**, 643.
- [30]. Villmoare, B. *et al.*, *Science*, **2015**, *347*, 1352-1355.
- [31]. Oró, J., Miller, S. L., Lazcano, A., *Annu. Rev. Earth Planet. Sci.*, **1990**, *18*, 317.
- [32]. Frei, R., Gaucher, C., Poulton, S. W., Canfield, D. E., *Nature*, **2009**, *461*, 250-253.
- [33]. Sleep, N. H., *CSH Perspect. Biol.*, **2010**, *2*, a002527.

- [34]. Abelson, P. H., *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *55*, 1365-1372.
- [35]. Shock, E. L., Schulte, M. D., *Nature*, **1990**, *343*, 728-731
- [36]. J. L. Bada, S. L. M., *Science*, **1968**, *159*, 423-425.
- [37]. Cleaves, H. J., *Precambrian Res.*, **2008**, *164*, 111-118.
- [38]. Bada, J. L., Chalmers, J. H., Cleaves, H. J., *Phys. Chem. Chem. Phys.*, **2016**, *18*, 20085-20090.
- [39]. Pasek, M. A. *et al.*, *Angew. Chem.*, **2008**, *47*, 7918-7920.
- [40]. Foustoukos, D. I. *et al.*, *Geochim. Cosmochim. Acta*, **2011**, *75*, 1594-1607.
- [41]. Walde, P., Boiteau, L., *Prebiotic chemistry*, Springer Science & Business Media, **2005**.
- [42]. Olson, K. R., Straub, K. D., *Physiology*, **2016**, *31*, 60-72.
- [43]. Rasmussen, R. *et al.*, *Science*, **1982**, *215*, 665-667.
- [44]. Sanchez, R., Ferris, J., Orgel, L., *Science*, **1966**, *154*, 784-785.
- [45]. Schoonen, M., Smirnov, A., Cohn, C., *AMBIO*, **2004**, *33*, 539-551.
- [46]. Dmowska, R., Holton, J. R., Teisseyre, R., Majewski, E., *Earthquake Thermodynamics and Phase Transformation in the Earth's Interior*, Academic Press, **2000**, Vol. 76.
- [47]. Koshland, D. E., *Science*, **2002**, *295*, 2215-2216.
- [48]. Miller, S. L., *Science*, **1953**, *117*, 528-529.
- [49]. Johnson, A. P. *et al.*, *Science*, **2008**, *322*, 404.
- [50]. Crick, F., *Nature*, **1970**, *227*, 561-563.
- [51]. Mullard, A., *Nat. Rev. Mol. Cell Biol.*, **2008**, *9*, 501-501.
- [52]. Ahlquist, P., *Science*, **2002**, *296*, 1270-1273.
- [53]. McCarthy, B., Holland, J., *Proc. Natl. Acad. Sci. U.S.A.*, **1965**, *54*, 880-886.
- [54]. Crick, F. H., *J. Mol. Biol.*, **1968**, *38*, 367-379.
- [55]. Orgel, L. E., *J. Mol. Biol.*, **1968**, *38*, 381-393.
- [56]. Cech, T. R., *CSH Perspect. Biol.*, **2012**, *4*.
- [57]. Robertson, M. P., Joyce, G. F., *CSH Perspect. Biol.*, **2012**, *4*.
- [58]. Nudler, E., Mironov, A. S., *Trends. Biochem. Sci.*, **2004**, *29*, 11-17.
- [59]. DeRose, V. J., *Curr. Opin. Struct. Biol.*, **2003**, *13*, 317-324.
- [60]. Leslie E. O., *Crit. Rev. Biochem. Mol. Biol.*, **2004**, *39*, 99-123.
- [61]. Cafferty, B. J., Hud, N. V., *Curr. Opin. Chem. Biol.*, **2014**, *22*, 146-157.
- [62]. Powner, M. W., Gerland, B., Sutherland, J. D., *Nature*, **2009**, *459*, 239-242.
- [63]. Gull, M., *Challenges* **2014**, *5*, 193-212.
- [64]. Bishop, M., Lohrmann, R., Orgel, L., *Nature*, **1972**, *237*, 162-164.
- [65]. Ferris, J. P., Ertem, G., *J. Am. Chem. Soc.*, **1993**, *115*, 12270-12275.
- [66]. Attwater, J. *et al.*, *Nat. Commun.*, **2010**, *1*, 76.

- [67]. Engelhart, A. E., Powner, M. W., Szostak, J. W., *Nat. Chem.*, **2013**, *5*, 390-394.
- [68]. Szostak, J. W., *J. Syst. Chem.*, **2012**, *3*.
- [69]. Heuberger, B. D. *et al.*, *J. Am. Chem. Soc.*, **2015**, *137*, 2769-2775.
- [70]. Gargaud, M. *et al.*, *From suns to life: a chronological approach to the history of life on Earth*, Springer, **2006**.
- [71]. van der Gulik, P. T., Speijer, D., *Life* **2015**, *5*, 230-246.
- [72]. Ramakrishnan, V., *Cell* **2002**, *108*, 557-572.
- [73]. Li, L., Francklyn, C., Carter, C. W., *J. Biol. Chem.*, **2013**, *288*, 26856-26863.
- [74]. Fosgerau, K., Hoffmann, T., *Drug Discov. Today*, **2015**, *20*, 122-128.
- [75]. Uhlig, T. *et al.*, *EuPA Open Proteom.*, **2014**, *4*, 58-69.
- [76]. Craik, D. J. *et al.*, *Chem. Biol. Drug Des.*, **2013**, *81*, 136-147.
- [77]. Pattabiraman, V. R., Bode, J. W., *Nature*, **2011**, *480*, 471-479.
- [78]. Lundberg, H. *et al.*, *Chem. Soc. Rev.*, **2014**, *43*, 2714-2742.
- [79]. Valeur, E., Bradley, M., *Chem. Soc. Rev.*, **2009**, *38*, 606-631.
- [80]. Ishihara, K., Ohara, S., Yamamoto, H., *J. Org. Chem.*, **1996**, *61*, 4196-4197.
- [81]. Charville, H. *et al.*, *Chem. Commun.*, **2010**, *46*, 1813-1823.
- [82]. De Sarkar, S., Studer, A., *Org. Lett.*, **2010**, *12*, 1992-1995.
- [83]. Yoo, W.-J., Li, C.-J., *J. Am. Chem. Soc.*, **2006**, *128*, 13064-13065.
- [84]. Foot, J. S., Kanno, H., Giblin, G. M., Taylor, R. J., *Synthesis*, **2003**, *2003*, 1055-1064.
- [85]. Gunanathan, C., Ben-David, Y., Milstein, D., *Science*, **2007**, *317*, 790-792.
- [86]. Shangguan, N. *et al.*, *J. Am. Chem. Soc.*, **2003**, *125*, 7754-7755.
- [87]. Yuan, Y. *et al.*, *J. Pept. Sci.*, **2010**, *94*, 373-384.
- [88]. Talan, R. S., Sanki, A. K., Sucheck, S. J., *Carbohydr. Res.*, **2009**, *344*, 2048-2050.
- [89]. Crich, D., Sasaki, K., *Org. Lett.*, **2009**, *11*, 3514-3517.
- [90]. Dawson, P. E. *et al.*, *Science*, **1994**, 776-776.
- [91]. Berg, J. M., Tymoczko, J. L., Stryer, L., *Biochemistry: International Version*, **2002**.
- [92]. Li, R. *et al.*, *Int. J. Mol. Sci.*, **2015**, *16*, 15872-15902.
- [93]. Funabashi, M. *et al.*, *Nat. Chem. Biol.*, **2010**, *6*, 581-586.
- [94]. Keefe, A. D. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **1995**, *92*, 11904-11906.
- [95]. Ring, D. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 765-768.
- [96]. Jiang, L. *et al.*, *Sci. Rep.*, **2014**, *4*, 6769.
- [97]. Patel, B. H. *et al.*, *Nat. Chem.*, **2015**, *7*, 301-307.
- [98]. Altwegg, K. *et al.*, *Sci. Adv.*, **2016**, *2*.
- [99]. Cronin, J. R., Pizzarello, S., *Adv. Space Res.*, **1983**, *3*, 5-18.

- [100]. Kuan, Y.-J. *et al.*, *Astrophys. J.*, **2003**, 593, 848.
- [101]. Danger, G. *et al.*, *J. Am. Chem. Soc.*, **2006**, 128, 7412-7413.
- [102]. Commeyras, A., *et al.*, *Polym. Int.*, **2002**, 51, 661-665.
- [103]. Leman, L., Orgel, L., Ghadiri, M. R., *Science*, **2004**, 306, 283-286.
- [104]. Leman, L. J., Huang, Z.-Z., Ghadiri, M. R., *Astrobiology*, **2015**, 15, 709-716.
- [105]. Lohrmann, R., Orgel, L., *Nature*, **1973**, 244, 418-420.
- [106]. Tiessen, A., Pérez-Rodríguez, P., Delaye-Arredondo, L. J., *BMC Res. Notes*, **2012**, 5, 1.
- [107]. Forsythe, J. G. *et al.*, *Angew. Chem.*, **2015**, 54, 9871-9875.
- [108]. Maurel, M.-C., Orgel, L. E., *Orig. Life Evol. Biospheres*, **2000**, 30, 423-430.
- [109]. Jöns, N., Bach, W. *Encyclopedia of Marine Geosciences*, Springer Netherlands, **2014**.
- [110]. Leman, L. J., Ghadiri, M. R., *Synlett*, **2017**, 28, 68-72.
- [111]. Chadha, M. S. *et al.*, *Orig. Life Evol. Biospheres*, **1975**, 6, 127-136.
- [112]. Strecker, A., *Liebigs Ann.*, **1850**, 75, 27-45.
- [113]. Chitale, S., *et al.*, *Chem. Commun.*, **2016**, 52, 13147-13150.
- [114]. Wang, J., Liu, X., Feng, X., *Chem. Rev.*, **2011**, 111, 6947-6983.
- [115]. Miller, S. L., *Chem. Scr. B*, **1986**, 26, 5-11.
- [116]. Stevenson, G. W., Williamson, D., *J. Am. Chem. Soc.*, **1958**, 80, 5943-5947.
- [117]. Leow, G. H., Chadha, M. S., Chang, S., *J. Theor. Biol.*, **1972**, 35, 359-373.
- [118]. Kriebel, V. K., Noll, C. I., *J. Am. Chem. Soc.*, **1939**, 61, 560-563.
- [119]. Ritson, D., Sutherland, J. D., *Nat. Chem.* **2012**, 4, 895-899.
- [120]. Rousselet, G., Capdevielle, P., Maumy, M., *Tetrahedron Lett.*, **1993**, 34, 6395-6398.
- [121]. Rosenthal, D., Taylor, T. I., *J. Am. Chem. Soc.*, **1957**, 79, 2684-2690.
- [122]. Butler, E. A., Peters, D. G., Swift, E. H., *Anal. Chem.*, **1958**, 30, 1379-1383.
- [123]. Paventi, M., Edward, J. T., *Can. J. Chem.*, **1987**, 65, 282-289.
- [124]. Danger, G., Plasson, R., Pascal, R., *Chem. Soc. Rev.*, **2012**, 41, 5416-5429.
- [125]. Steinberg, S. M., Bada, J. L., *J. Org. Chem.* **1983**, 48, 2295-2298.
- [126]. Liu, R., Orgel, L. E., *Nature*, **1997**, 389, 52-54.
- [127]. Ree, R., Varland, S., Arnesen, T., *Exp. Mol. Med.*, **2018**, 50.
- [128]. Huber, C., Wächtershäuser, G., *Science*, **1997**, 276, 245-247.
- [129]. Chandru, K. *et al.*, *Sci. Rep.*, **2016**, 6, 29883.
- [130]. Bal, B. S., Childers, W. E. Jr., Pinnick, H. W., *Tetrahedron*, **1981**, 37, 2091-2096.
- [131]. Coggins, A. J., Powner, M. W., *Nat. Chem.*, **2017**, 9, 310-317.
- [132]. Bordier, F. *et al.*, *J. Mol. Catal. B Enzym.*, **2014**, 107, 79-88.

- [133]. Shriver, D. F. *et al.*, *Inorganic Chemistry*, Freeman, W. H., New York, **2006**.
- [134]. Beesley, R. M., Ingold, C. K., Thorpe, J. F., *J. Chem. Soc. Trans.*, **1915**, 107, 1080-1106.
- [135]. Rao, Y. *et al.*, *Tetrahedron Lett.*, **2009**, 50, 6684-6686.
- [136]. Mali, S. M., Gopi, H. N., *J. Org. Chem.* **2014**, 79, 2377-2383.
- [137]. Danger, G., Plasson, R., Pascal, R., *Chem. Soc. Rev.*, **2012**, 41, 5416-5429.
- [138]. Bowler, R. F. *et al.*, *Nat. Chem.*, **2013**, 5, 383-389.
- [139]. Danger, G. *et al.*, *Angew. Chem.*, **2013**, 52, 611-614.
- [140]. Xiang, Y. B. *et al.*, *Helv. Chim. Acta*, **1994**, 77, 2209-2250.
- [141]. Griesser, H., *et al.*, *Angew. Chem.*, **2016**, 56, 1224-1228.
- [142]. Griesser, H., *et al.*, *Angew. Chem.*, **2016**, 56, 1219-1223.
- [143]. Deamer, D. *et al.*, *Philos. Trans. R. Soc. B. Biol. Sci.*, **2006**, 361, 1809-1818.
- [144]. Johnson, A. P. *et al.*, *Science*, **2008**, 322, 404.
- [145]. Breslow, R., Cheng, Z.-L., *Proc. Natl. Acad. Sci. U.S.A.*, **2009**, 106, 9144-9146.
- [146]. Shirom, M., Stein, G., *J. Chem. Phys.*, **1971**, 55, 3379-3382.
- [147]. Ranjan, S. *et al.*, *Astrobiology*, **2018**, 18.
- [148]. Xu, J. *et al.*, *Chem. Commun.*, **2018**, 54, 5566-5569.
- [149]. Spicer, C. D., Davis, B. G., *Nat. Commun.*, **2014**, 5.
- [150]. Buller, A. B., Townsend, C. A., *Proc. Natl. Acad. Sci. U.S.A.*, **2013**, 110, E653-E661.
- [151]. Bonfio, C. *et al.*, *Nat. Chem.*, **2017**, 9, 1229-1234.
- [152]. Kitadai, N., Maruyama, S., *Geosci. Front.*, **2018**, 9, 1117-1153.
- [153]. Nielsen, P. E., Egholm, M., Berg, R. H., Buchardt, O., *Science*, **1991**, 254, 1497-1500.
- [154]. Zimmer, C., *Science*, **2009**, 323, 198-199.