Studies on the Prebiotic Origin of 2-Deoxy-D-ribose

Andrew Mark Steer

Doctor of Philosophy

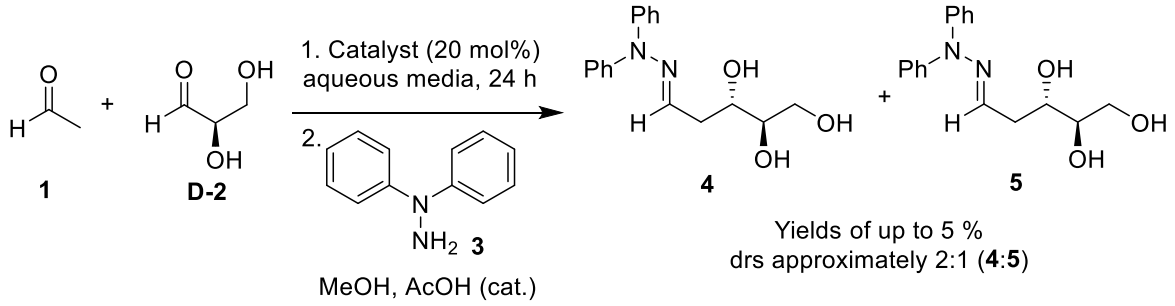
University of York

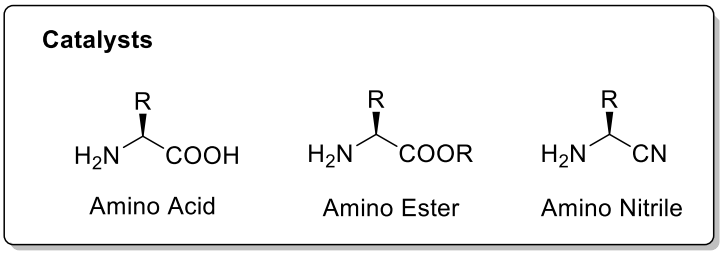
Chemistry

August 2017

Abstract

DNA is an important biological structure necessary for cell proliferation. The origins of cell-like structures and the building blocks of DNA are therefore also of great concern. As of yet the prebiotic origin of 2-deoxy-D-ribose, the sugar of DNA, has no satisfactory explanation. This research attempts to provide a possible explanation to the chemical origin of 2-deoxy-D-ribose *via* an aldol reaction between acetaldehyde **1** and D-glyceraldehyde **D-2** (Error! Reference source not found.). The sugar mixture is trapped with *N,N*-diphenylhydrazine **3** for ease of purification and characterisation. The reaction is promoted by amino acids, amino esters and amino nitriles consistently giving selectivities in favour of 2-deoxy-D-ribose. This is the first example of an amino nitrile promoted reaction.





Potential prebiotic synthesis of 2-deoxy-D-ribose and subsequent trapping with *N*,*N*-diphenyl hydrazine **3**.

The research is developed further by exploring the formation of 2-deoxy-D-ribose in a "protocell" environment – a primitive cell. Here we suggest that primitive cells may have been simple hydrogel systems. A discussion of the characterisation and catalytic ability of small peptide-based supramolecular structures is included.

Contents

List of Figures

List of Schemes

List of Tables

Acknowledgements

First of all I would like to thank both of my supervisors Dr Paul Clarke and Prof David Smith for giving the opportunity to work on such a challenging and exciting project. Over the course of the last three and a half years they have provided a lot of valuable advice and discussions about the direction of the project. It has been a fun project to be part of, if challenging at times, and has given me the chance to develop lab skills in multiple areas of chemistry. I would also like to thank the members of the Clarke group who have made the lab such a great place to work. I would especially like to thank Dr Ian George for all of his invaluable help during my research, in both stimulating conversations regarding my research, looking over my thesis drafts and providing lots of coffee. Special thanks to Sam Griggs, who has really kept me going through my PhD with his positive attitude, amazing banter, great choice of music in the lab and the many squash and departmental football matches we've participated in together. He's been a great friend in and out of the lab.

Particular thanks to Martin Fascione and Richard Spears for teaching me solid phase peptide synthesis, and to Jorge Ruiz and Phil Chivers for their expertise with gel chemistry. Thanks also to Dr William Unsworth for giving me the opportunity to work part-time in his lab during my write-up on other types of chemistry other than prebiotic, and Tom Stephens for helping me fit into the WPU lab straight away. I would also like to mention Heather Fish and Karl Hale for all of their assistance with acquiring NMR and mass spectrometry data and Meg Stark for assistance with acquiring TEM images. Other people to mention are the Wentworth Squash team who have been a great group to play and socialize with and I'll definitely miss our Tuesday afternoon sessions. Finally I would like to thank my amazing girlfriend Katie and my family for all the support and encouragement they have given me during my PhD.

Declaration

I hereby declare that the substance of this thesis has not been submitted, nor is currently being submitted, in candidature for any other degree.

I also declare that the work embodied in this thesis is the result of my own investigations and in the event the work of others has been used this has been fully acknowledged in the text.

Some of the research outlined in this thesis has been published in the following paper:

Prebiotic synthesis of 2-deoxy-D-ribose from interstellar building blocks promoted by amino esters or amino nitriles, A. M. Steer, N. Bia, D. K. Smith and P. A. Clarke, *Chem. Comm.,* **2017**, *53*, 1036

1. Introduction

1.1. The Origin of Life and the RNA World hypothesis

One of the biggest unanswered questions of our time is the origin of Life. There is little evidence surrounding the composition of the prebiotic atmosphere, the temperature of the Earth or the initial building blocks at hand.[1] Despite the lack of evidence, scientists have developed some fascinating theories backed-up with experimental research on this topic but nevertheless we will never know unequivocally how Life began.[1, 2]

The central dogma of molecular biology, first proposed by Crick, explains the flow of genetic information in a biological system (**Figure 1.1**).[3] In this model, DNA, the genetic storage system of the cell, can undergo replication to copy genetic information.

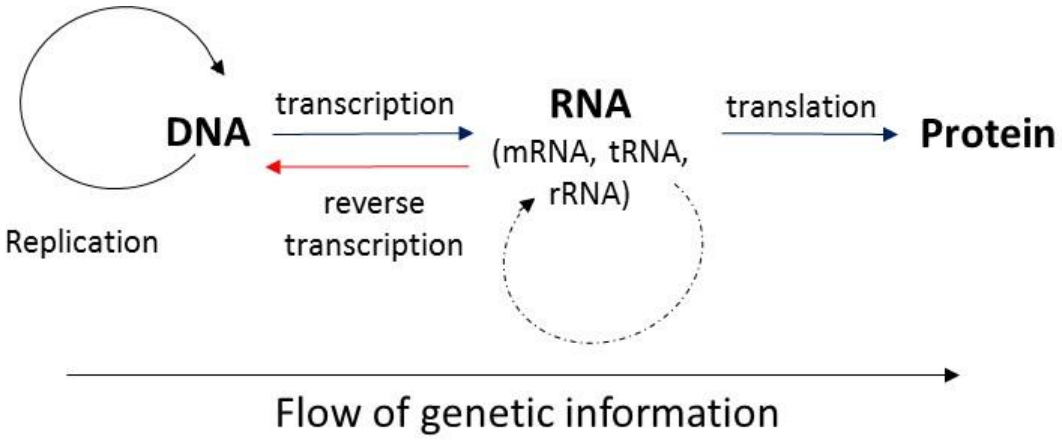


Figure 1.1. Central dogma of molecular biology.

DNA is composed of a double helix structure held together by hydrogen bonding between nucleic bases on opposite strands, known as complementary base pairing. Each strand is made of a phosphate polymer backbone made up of a string of nucleotides. Each nucleotide consists of a sugar (2-deoxy-D-ribose), a phosphate group and a base. The four bases of DNA are; guanine, cytosine, adenine and thymine and it is these recognition elements that allow genetic information to be transferred. DNA can undergo replication to copy the genetic information but can also undergo transcription and translation which leads to protein synthesis. Transcription starts in the nucleus where a section of the DNA is unravelled exposing the bases of the DNA strand. mRNA nucleotides then bind to the complementary exposed base and the enzyme RNA polymerase catalyses the polymerisation of the phosphate backbone effectively making a single strand mirror image (**Figure 1.2**). This strand then leaves the nucleus and moves to the ribozyme where the information is translated into an amino acid sequence which forms the primary structure of a peptide. Astoundingly the sequence of four base pairs selects for 20 different amino acids in the human body.

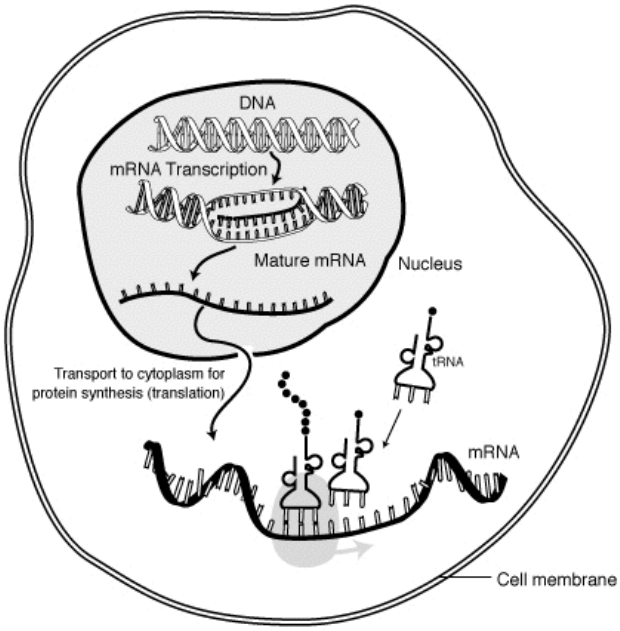
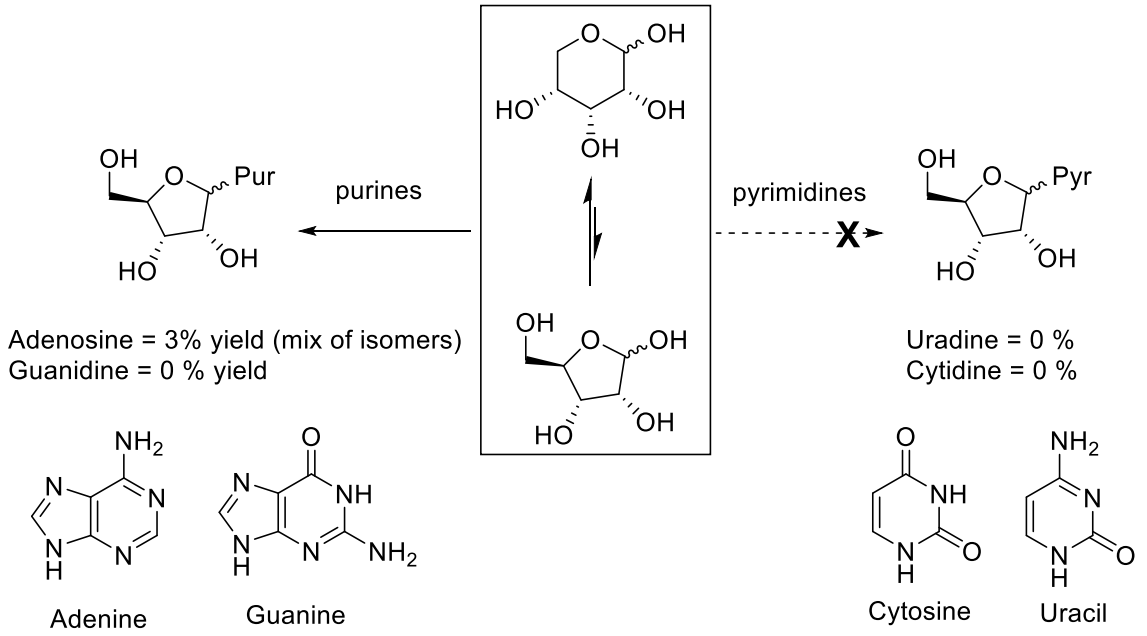


Figure 1.2. The processes of transcription and translation that occur within a mammalian cell.[4]

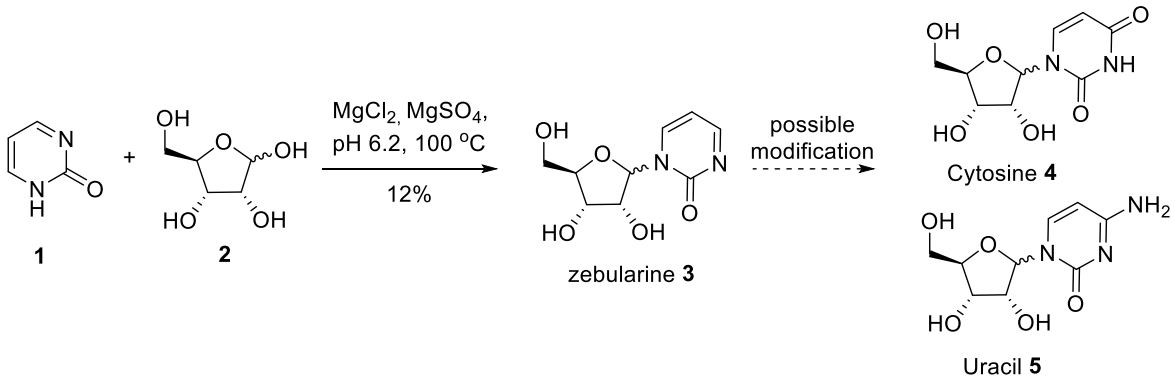
It is a daunting task for the prebiotic community to take the complex biological process of genetic information transfer and strip this back to a primitive form that may have occurred on the early Earth. The general idea that RNA replication proceeded the appearance of proteins and that DNA was a later evolutionary adaption of RNA is a commonly accepted theory amongst the origins of Life community, the so-called "RNA world hypothesis".[5,6,7,8] Although the RNA world theory means different things to different researchers it includes three basic assumptions: (i) Genetic information was passed on by the replication of RNA. (ii) Specific base pairing between uracil / adenine and cytosine / guanine was essential for replication. (iii) Genetically encoded proteins did not play a catalytic role.[9]

There has been a great deal of work into the synthesis of RNA nucleosides under potentially prebiotic conditions. There are two approaches to synthesising RNA nucleotides. The first is a modular approach; building the sugar, phosphate and base as three separate elements and then attempting to construct the nucleoside. The approach may seem optimistic as the DNA and RNA base cytosine can be synthesised by condensation of urea and cyanoacetaldehyde, however, this approach has met with little success.[10] Orgel attempted to form the glycosidic link between ribose and the purine/pyrimidine bases. A 3% mixture of adenosine isomers were obtained but no trace of guanidine, uridine or cytidine were detected when the synthesis was attempted with the relevant bases (**Scheme 1.1.**).[11] The failure of this reaction can be explained in terms of the equilibrium between the 5-membered furanose sugar ribose and the more favoured 6-membered pyranose sugar, thermodynamically the reverse hydrolysis reaction is more favoured than the desired condensation reaction in water. Secondly, the relevant nitrogen lone pair on the nucleosides are not available due to delocalisation.[12]



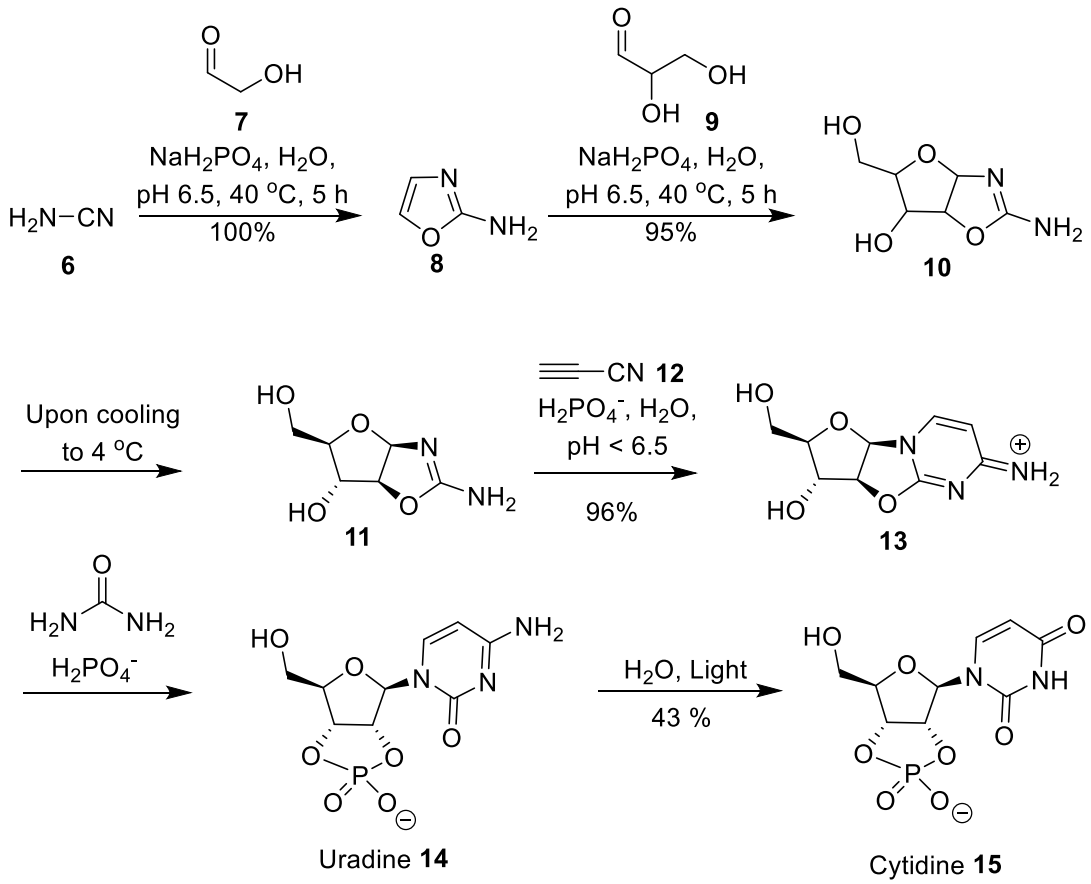
Scheme 1.1. Orgel's attempt at forming the glycosidic linkage between ribose and nucleic bases.[11]

The condensation of nucleic base and ribose was not plausible, however, Hud *et al* showed how the condensation of ribose **2** with a pyrimidinone base **1** was possible to give the glycoside zebularine **3** in a 12 % yield by a drying and heating process (**Scheme 1.1.**). The group postulated that the nucleosides cytosine **4** and uracil **5** could be a further modification of zebularine.[13]



Scheme 1.2. Synthesis of zebularine **3** and postulated post-modification to uracil **5** and cytosine **4**.

An alternative approach to the construction of nucleotides is to the build the molecule as one entity avoiding the need to force building blocks together in unfavourable conditions. Powner and Sutherland have successfully demonstrated this approach.[12,14] Starting from the simple prebiotic molecule cyanoamide **6** they built up a sugar base molecule using aqueous conditions and small building blocks such as glycolaldehyde **7**, glyceraldehyde **9** and cyanoacetylene **11**. The intermediate oxazoline **10** was found to crystallise out of solution as a single diastereomer. Oxazoline **11** and cyanoacetylene **12** were then reacted in phosphate buffer to give **13** in high yield. The reaction of **13** with pyrophosphate and urea in the dry state followed by an intramolecular rearrangement gave uradine **14** in a 46 % yield for the final step. Access to the cytidine nucleotide **15** was obtained from uracil by irradiating in phosphate buffer for three days.



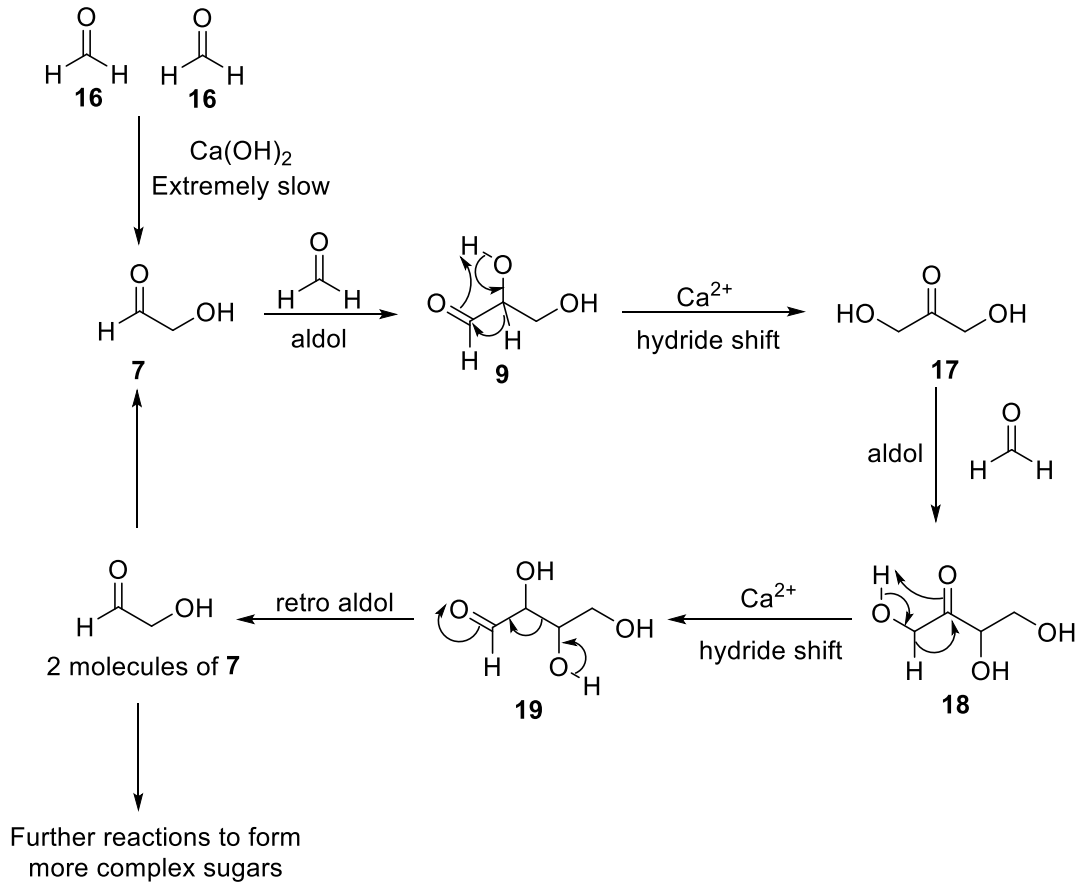
Scheme 1.3. Sutherland's synthesis of cytidine and uridine from small molecules under potential prebiotic conditions.[12]

The research into a prebiotic origin of RNA is impressive and great progress has been made in the last few decades. However, the origin of DNA as a separate entity to RNA has had far less exposure. Other routes to the synthesis of DNA should be explored as it is possible that DNA may have originated in a number of pathways. DNA, after all, contains our genetic information and therefore to truly understand the origin of Life and cells all synthetic routes to DNA should be explored. In this thesis the work towards a synthetic route for the synthesis of 2-deoxy-D-ribose, the sugar of DNA, will be discussed. As this work is at its very beginning the synthesis of DNA will not be discussed, however, forming glycosidic linkages with nucleic bases is a further milestone to consider. The synthesis of RNA has shown this to be challenging and indeed the same challenges would apply to the prebiotic synthesis of DNA nucleotides.

1.2. The Origin of Sugars

It is widely believed that the first sugars arose from the formose reaction; the autocatalytic reaction of small carbonyl building blocks to form larger sugar polymers.[15,16,20] This reaction was first reported in 1861 involving the calcium hydroxide promoted polymerisation of formaldehyde leading to a complex mixture of "sweet tasting" products of which ribose was found in < 1 %.[15,16] The mechanism of the reaction was elucidated by Breslow and is shown in **Scheme 1.4** below.[20,21] The first step of the reaction mechanism is not fully understood. To form glycolaldehyde **7** two molecules of formaldehyde **16** must undergo an unfavourable, "umpolung", self-condensation initiation step. Basic dicationic salts such as calcium hydroxide have been found to be essential to stabalise the ene diol intermediates in the reaction pathway.[17] However, Weiss *et al.* found that the formation of glycolaldehyde **7** from a pure aqueous alkaline solution of formaldehyde was not possible which has led to the

suggestion impure formaldehyde **16**, contaminated with glycolaldehyde **7** is needed for the formose reaction.[18] However, Lambert has argued that formaldehyde obtained in a number of ways (through cracking of paraformaldehyde and commercial formaldehyde) provided the same kinetic profile, also, in the absence of calcium hydroxide the reaction did not occur.[23] Further to this Kopetzki and Antonietti have shown under hydrothermal conditions (up to 200 °C and 100 bar) the intitation step of the formose reaction occurs at a greater rate in the presence of glycolaldehyde **7** but still occurs in the absence of glycoladehyde but in the presence of a dicationic species, calcium acetate, and at an even slower rate with mono cationic sodium acetate.[19]



Scheme 1.4. Reaction pathway for the formose reaction conceived by Breslow.[21]

Upon the addition of a third molecule of formaldehyde **16**, glyceraldehyde **9** is formed. A series of hydride shifts and formaldehyde additions can lead to the formation of tetrose products. These products can undergo retro aldol reactions producing two molecules of glycolaldehyde **7** from one molecule of **19** and the cycle can occur again. An alternative pathway is further addition of formaldehyde to the sugar chain leading to higher linear and branched products.[20,21]. Efforts to stabalise the products of this reaction using mineral additives such as borates[22,23] and silicates[24] has met with some success, particularly as stabilsers for the formation of ribose.[25]

1.3. Prebiotic Relevance of Amino Acids

The origin of amino acids has evoked a lot of interest in recent years dating back to the pioneering work of Miller and Urey.[26] In an attempt to mimic reducing conditions on the primitive Earth, boiling water containing methane, ammonia and hydrogen was continuously circulated through an electrical discharge for one week.[26] The researchers were able to identify amino acids such as glycine **20**, α-alanine **21** and β-alanine **22** and inferred other amino acids are probably formed in smaller quantities. Further investigation revealed many other small molecules such as formic, acetic and lactic acids to be present.[27] The significant findings of this experiment in a prebiotic context has led to further studies into the formation of molecules *via* electrical discharge. For example, a very recent study compared hydrogen-containing starting materials with deuterium-containing starting materials in the Urey-Miller experiments and found notable differences between the range and types of products produced between the two sets of conditions.[28] However, the validity of a reducing atmosphere on the early Earth has since been questioned and instead a more neutral atmosphere, made up of carbon dioxide, water and nitrogen has been suggested.[29] Under these more oxidizing conditions the electrical discharge experiments were much less efficient,[30,31] however, a reassessment of this work, buffering the reaction to pH 7 and using

oxidation inhibitors, gave increased yields of amino acids by 2-15 times compared to the original Miller-Urey experiment.[32,]

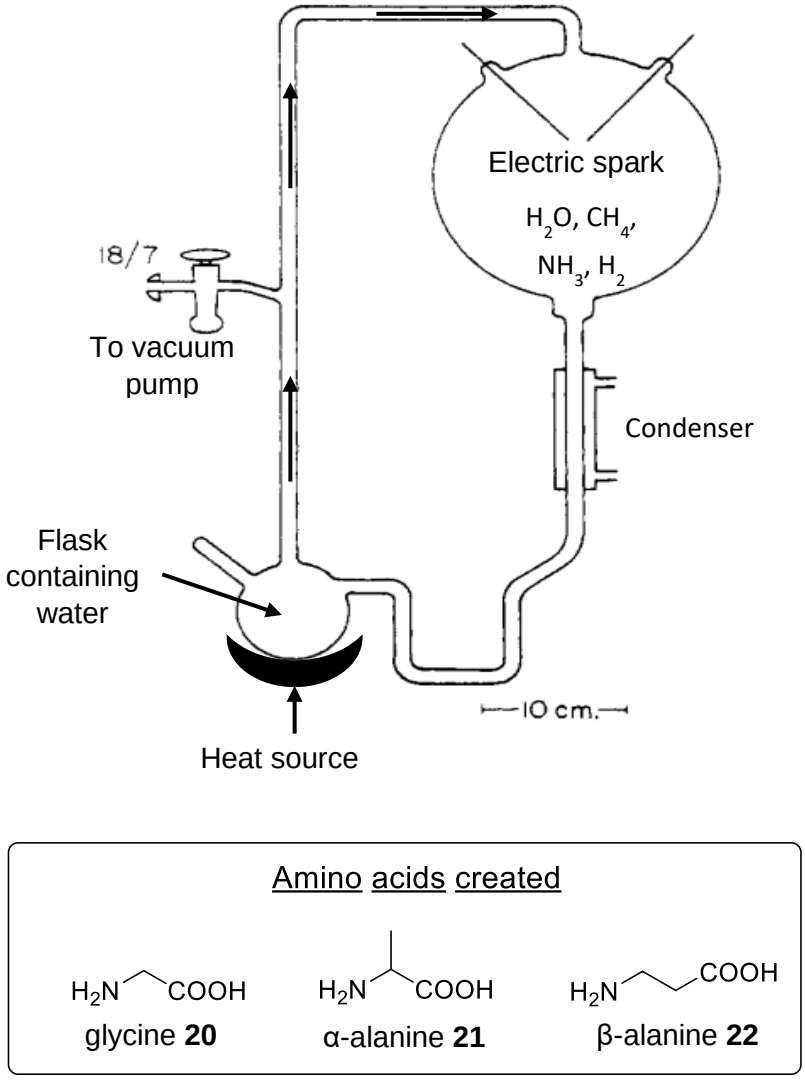
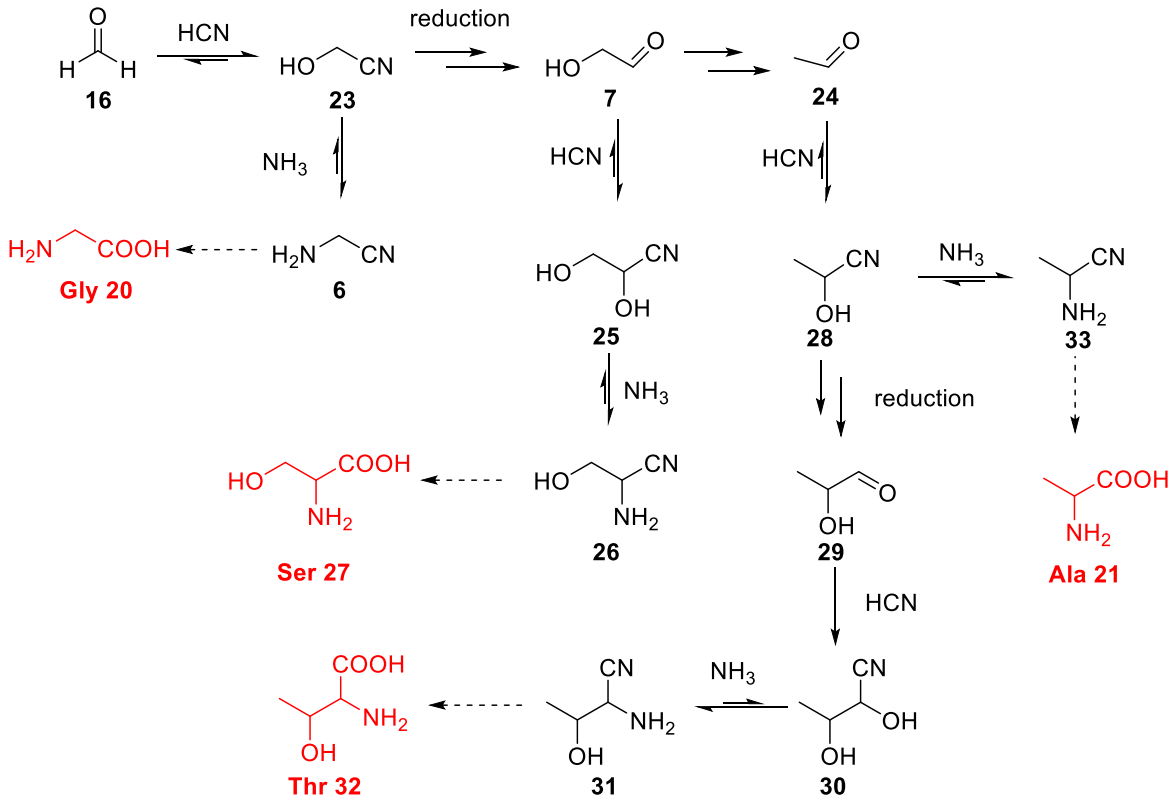


Figure 1.3. Miller-Urey experimental set-up and amino acids synthesised. Reprinted with permission from J. Am .Chem. Soc., 1995, *77*, 2351. Copyright 2017 American Chemical Society.

A second scenario, is that amino acids may have arisen from hydrothermal vents in the seas and oceans. Processes such as the reduction of carbon dioxide to methane at high temperature (150 °C) when in contact with minerals such as pyrite or magnetite to achieve

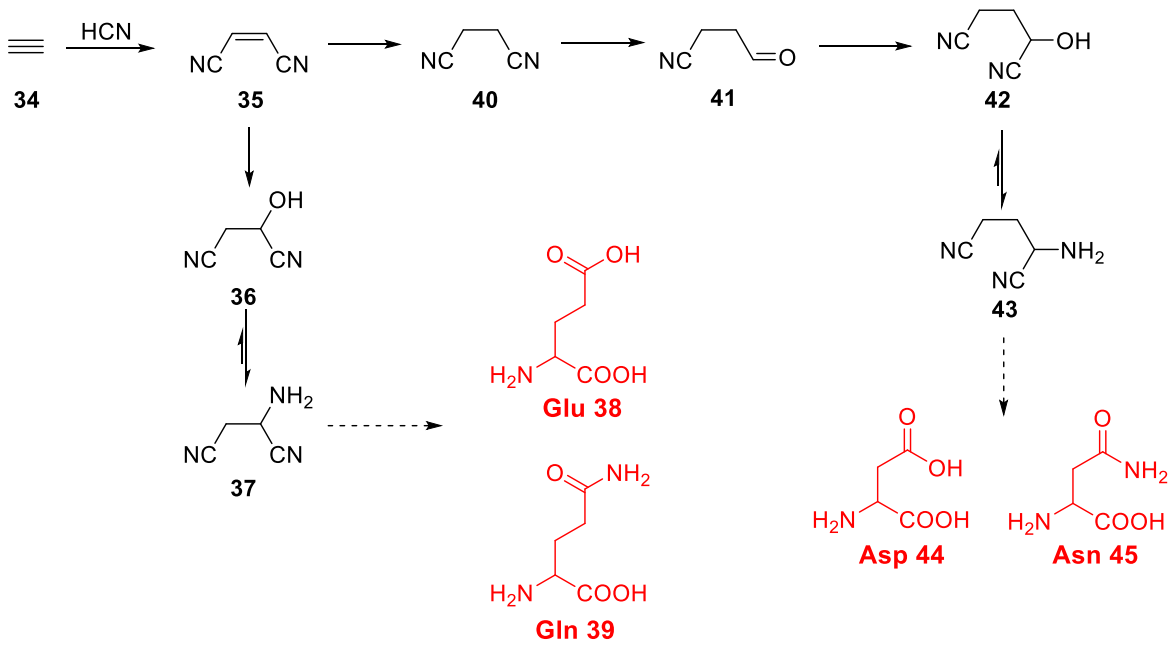
equilibrium, are known to take place in such systems.[33] One study tried to replicate these conditions by conducting experiments in one litre autoclaves at 150 °C and 10 atm with an aqueous phase consisting of potassium cyanide, formaldehyde **16**, ammonium chloride, and HCl. There was also a mineral phase containing pyrite/magnetite or illite and a gaseous atmosphere of hydrogen and carbon dioxide (1:3). After 54 hours samples were acidified with HCl and analysed by GC and HPLC. A number of amino acids (asparagine, serine, glutamic acid, glycine, alanine, cysteine, methionine, isoleucine) were found to be present in milimolar concentrations.[34] The group noted that the concentrations of amino acids were higher than those from the electrical discharge experiments. However, the lack of D-enantiomers of amino acids has led to scrutiny and suggestions that the experiment was contaminated.[35] Other studies based on samples taken from hydrothermal vents in the Pacific Ocean and Toyoha mine, Japan, showed nanomolar and micromolar levels of common amino acids again with some selectivity for L over D.[36,37] Additionally, Yoshino *et al.* have tried to mimic the conditions of hydrothermal vents starting from CO, hydrogen and ammonia at temperaures of 200-700 °C and were able to demonstrate the presence of several amino acids *via* Fischer Tropsch chemistry, albeit in very small 0-0.1 % yields.[38]

Sutherland and coworkers developed a further alternative to the emergence of amino acids as part of their remarkable origins of life systems chemistry. The chemistry stemmed from simple building blocks such as ammonia and hydrogen cyanide and focused around the photoredox reactions of simple copper salts. The copper salts could facilitate the oxidation of hydrogen cyanide (HCN) to cyanogen ($(NC)_2$) to generate the reduction power required for the conversion of HCN to formaldehyde imine, which could then be hydrolysed to formaldehyde **16**.[39] This sparked a series of high yielding iterative conversions to possible amino acid precursors “amino nitriles” shown in **Scheme 1.5**.



Scheme 1.5. Reaction pathway from formaldehyde to amino acids.[39]

Sutherland also showed simple copper-catalysed cross-coupling of acetylene **34** and HCN followed by addition of water and ammonia can lead to more amino nitrile precursors (**Scheme 1.6**).[40] This review only shows a couple of the amino acids that Sutherland alludes to and pathways towards many more amino acids are described in his paper published in 2015.[40]



Scheme 1.6. Reactions pathway from acetylene to amino acids.[40]

Three theories of the origin of amino acids have been discussed so far. Although intriguing, none of these theories discuss one major issue, the origin of chirality of the amino acids, i.e. none of the theories show a preference for the natural L-amino acid over the D-enantiomer.

Another possible origin of amino acids is an extraterrestrial one. Amino acids may have been brought to the Earth on chondritic meteorites during the meteoritic bombardment period of the Earth around 3.5 billion years ago. Evidence for this theory came from a number of meteorites that fell to the Earth in recent years and have been analysed for amino acids amongst other potential small chemical building blocks. An example of this was the Paris meteorite which was found to contain a range of natural and unnatural amino acids albeit at part per billion (ppb) levels.[41] However, the most extensively studied meteorite was the Murchison meteorite, which was found to contain many natural amino acids such as

proline, valine, glycine and glutamic acid in significantly higher concentrations (μg / g of meteorite).[42] In meteoritic studies it has been argued that some degree of biological contamination of the sample could occur.[43] A counter argument to this is that the high abundance of ^{13}C and ^{15}N present in the samples and the very small ee in favor of L-amino acids supports the hypothesis of an extraterrestrial origin.[44]

Within the Murchison meteorite, some α-methyl amino acids were also found.[45,46] The α-methyl amino acids contain a methyl group in place of the α-H found in natural amino acids. **Figure 1.4** shows five α-methyl amino acids (**46-50**) found in the Murchison meteorite all with slight ee.s in favour of the L enantiomer. The 'unnatural' structure of these amino acids suggests an extraterrestrial origin.

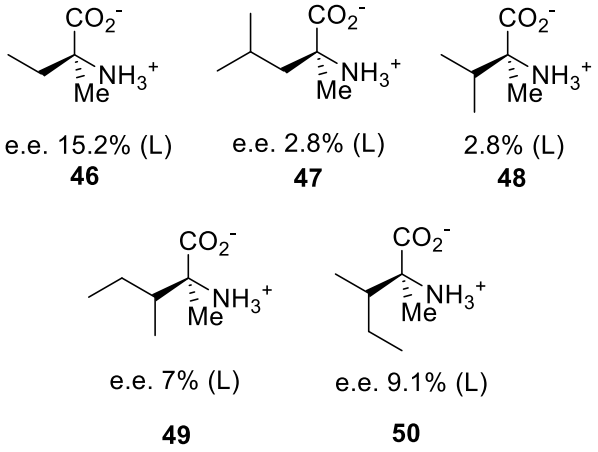
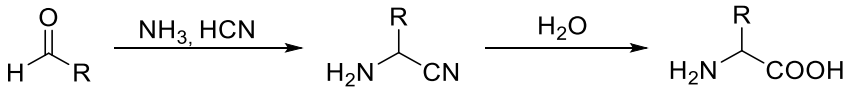


Figure 1.4. Five methyl-α-amino acids found in fragments of the Murchison meteorite.

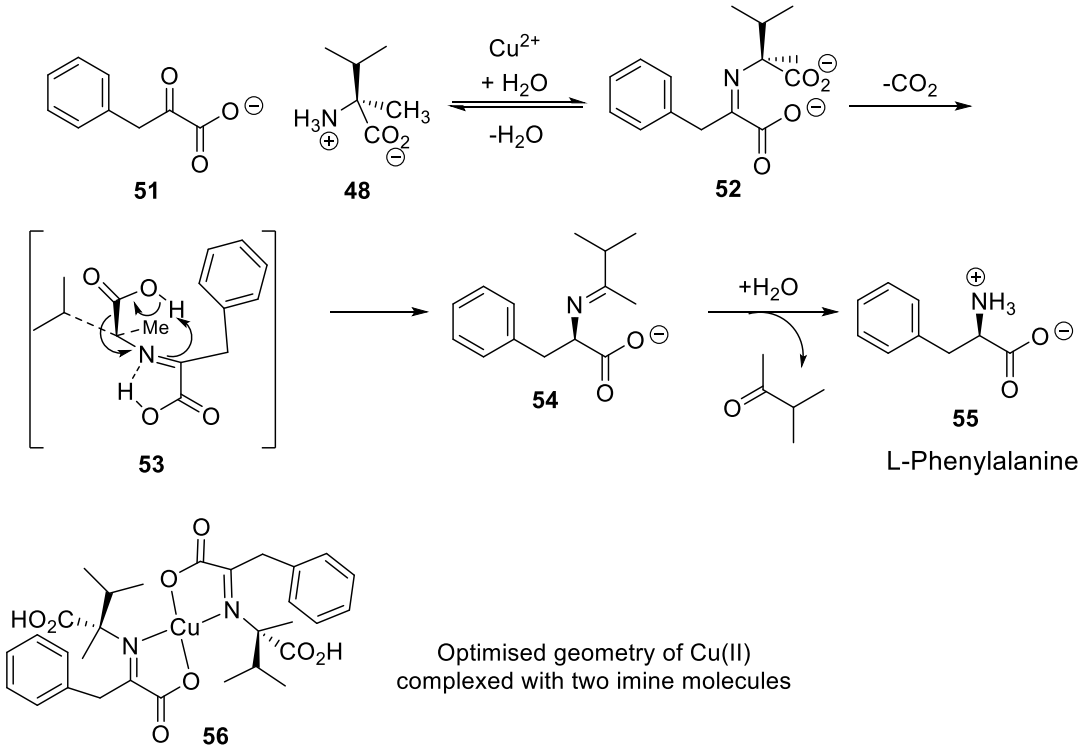
The extraterrestrial origins of amino acids has received much speculation. A plausible theory for the formation of these non-terrestrial amino acids is through Strecker reactions in space on the surface of interstellar dust particles. The reaction, named after Adolph Strecker, who first published the reaction in 1850 demonstrated how the condensation of

an ammonia and an aldehyde in the presence of HCN could form an amino nitrile. Subsequent hydrolysis of the nitrile formed amino acids (**Scheme 1.7**).[47, 48]



Scheme 1.7. Amino acid formation *via* the Strecker reaction.

A popular theory is that the feedstocks came together to form amino acids in the Kuiper belt from where asteroids or meteorites originate.[49] The racemic amino acids then passed through a sector of right-handed polarized light destroying a small amount of the D-enantiomer.[50] There is of yet, no evidence of right-handed polarized light in our section of the universe and many astronomers debate over its origin.[51,52] Nevertheless, this theory has been demonstrated experimentally through the irradiation of a racemic mixture of amino acids with right-handed circularly polarized light, from a laser source, which resulted in a small excess of the L-enantiomer.[53] The problem with this theory is that, in space the amino acids may have undergone proton extraction and addition leading to racemization *via* epimerization. Epimerization, however, cannot occur with α-methyl amino acids as the methyl group cannot be spontaneously lost and hence the small ee created by polarized light would remain.[54] Breslow *et al* have shown how α-methyl amino acids can be used to form α-H amino acids using L-α-methylvaline **48** as an example. They heated four equivalents of L-α-methylvaline **48** with one equivalent of sodium pyruvate **51** in the presence of one equivalent of cupric sulfate gave L-phenylalanine **55** in 37 % ee hence, in part, transferring chirality (**Scheme 1.8**).[55] The stereochemical outcome was explained through DFT calculations of the copper complex intermediate **56** which showed only one face to accessible for protonation. The stereochemical outcome could also be explained by a concerted mechanism shown in **Scheme 1.8**.



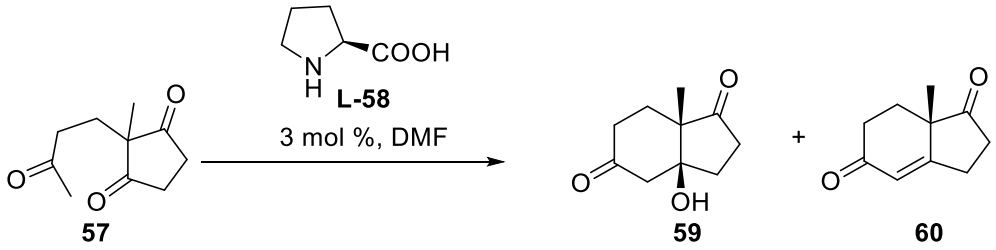
Scheme 1.8. Conversion of L-α-methylvaline to L-phenylalanine by Breslow *et al*.[55]

This section has briefly looked at four different perspectives to the origin of amino acids. The theories of hydrothermal vents, electrical discharge and copper/sulfur photoredox chemistry are all feasible under a reducing environment. If the early atmosphere was of a neutral environment then an extraterrestrial origin of amino acids may be more plausible.

1.4. Amino acids as catalysts in aldol reactions

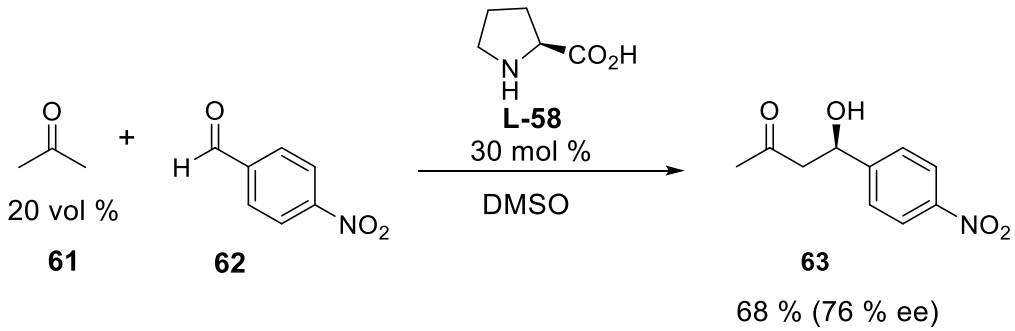
Catalysis is a very important aspect of synthetic chemistry and plays a role in thousands of chemical reactions. The ability of a molecule to reduce the activation energy barrier of a reaction without actually getting consumed itself is something chemists have used for 100s of years replicating the role of enzymes in biological reactions. Recently organocatalysis,

the use of small organic molecules as catalysts to promote asymmetric organic transformations *via* the transfer of chiral information, has become increasingly popular in synthesis.[56] The first use of tertiary amines as catalysts was carried out in the late 1800's but it wasn't until the early 1970's that amino acid organocatalysis was reported.[57] Independently, Eder, Sauer and Weichert and Hajos and Parish reported the asymmetric cyclisation of a cyclic triketone **57** catalysed by the amino acid L-proline **L-58** (**Scheme 1.9**).[58,59] At the time, the mechanism of the reaction was not fully understood and so the scope of the reaction was not explored.



Scheme 1.9. L-Proline **L-55** catalysed synthesis of a bicyclic system.

Amino acid inspired organocatalysis was later revived in 2000 by List *et al.* which involved the natural amino acid L-proline **L-58** and derivatives as catalysts for the aldol reaction between acetone **61** and *para*-nitrobenzaldehyde **62** in a 4 : 1 mixture of DMSO : acetone.[60] The reaction gave the aldol product in a moderate yield of 68 % and 76 % ee after only four hours (**Scheme 1.10**).



Scheme 1.10. L-Proline catalysed aldol reaction carried out by List *et al.*[60]

The same reaction was attempted using a range of amino acids but none gave as high yields as L-proline **L-58** (**Table 1.1**). However, they did find that adding a functional group to the 4-position of the pyrrolidine ring gave slightly improved yields but with little difference to the enantioselectivity.[60]

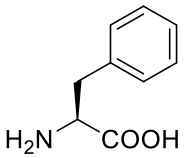
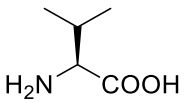
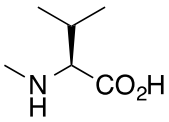
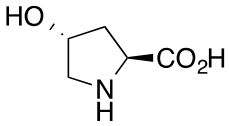
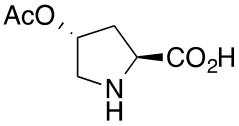
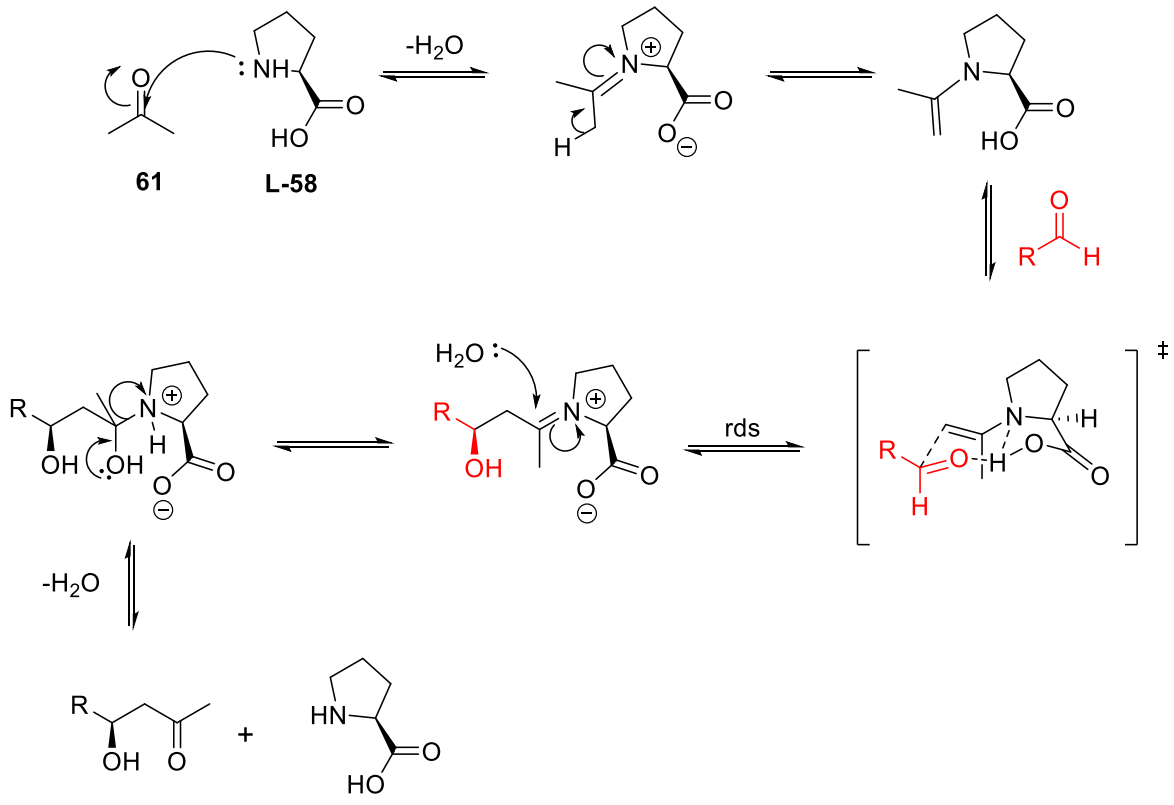
Entry	Compound	Yield (%)	ee (%)
1		< 10	n. d.
2		< 10	n. d.
3		< 10	n. d.
4		85	78
5		70	74

Table 1.1. Amino acids and derivatives used to catalyse the aldol reaction of acetone **61** with para-nitrobenzaldehyde **62** between 4 and 24 hours. N.d. – ee not determined.[60]

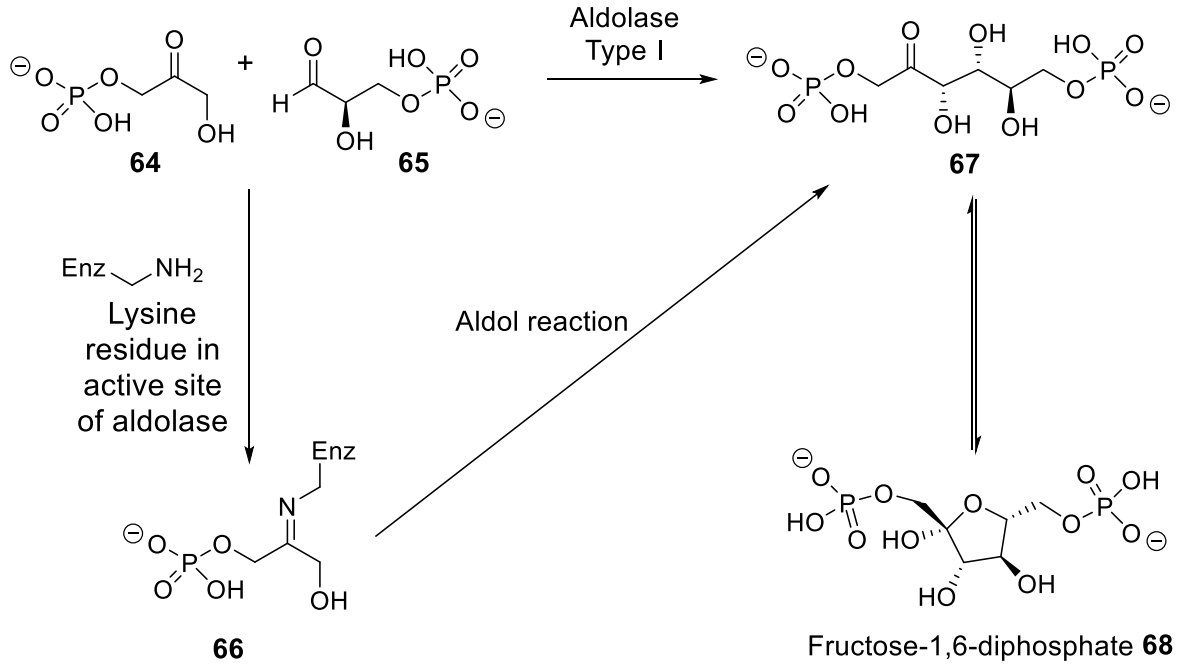
List *et al* postulated that this reaction was occurring *via* a condensation reaction of acetone 61 and L-proline **L-58** to form an enamine intermediate, followed by addition to benzaldehyde **62** to form the aldol product, after hydrolysis of the enamine.[60] Blackmond and co-workers conducted a detailed investigation into the L-proline **L-58** catalysed aldol reaction arriving at the same conclusion that the amino acid promoted the reaction *via* an enamine intermediate.[61] The key step in the mechanism was postulated to involve a chair-like transition state facilitated by hydrogen bonding between the carboxylic acid, the nitrogen of the enamine and the carbonyl of the electrophile (**Scheme 1.11**).[61] This step determines the stereochemistry of the product by holding the electrophile in a specific conformation so the larger R-group on the electrophile is in the pseudo-equatorial position of the chair conformation. Blackmond also found that attack of the electrophile by the enamine in the chair-like transition step to be the rate-determining step of the mechanism.



Scheme 1.11. Proposed mechanism for the proline-catalysed aldol reaction. The rate-determining step (rds) has been found to be the addition of the electrophile to the enamine.[61]

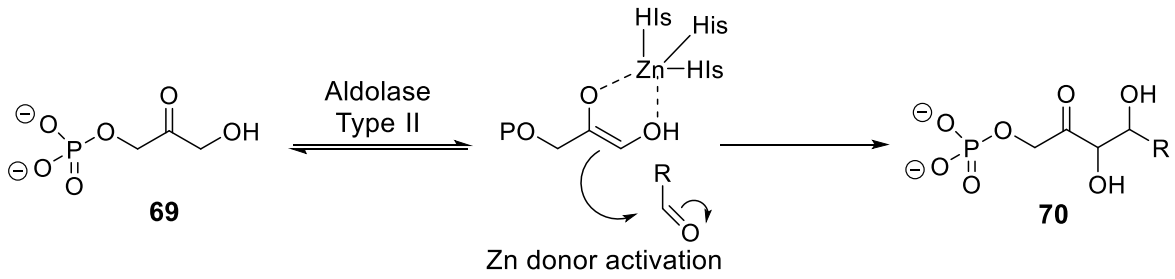
There are many benefits to organocatalytic reactions. The catalysts are less expensive than precious metals such as palladium, gold or rhodium. Organocatalysts are also typically less toxic, do not require inert conditions and many are water soluble (e.g. amino acids) meaning they can readily be removed *via* aqueous extraction upon work up.

Nature has perfected this enamine catalysed aldol reaction. Class I aldolase enzymes contain a lysine residue in the active site of the enzyme that acts as a catalyst for the aldol reaction of hydroxyacetone phosphate **64** and glyceraldehyde phosphate **65** to form fructose,1,4-diphosphate **67** with two new stereocentres in a *syn* configuration (**Scheme 1.12**).[62] This molecule exists in an equilibrium between the straight chain **67** and furanose form **68** in solution.



Scheme 1.12. Aldol reaction of the aldolase type I enzyme.

There is a second type of aldolase (Type II) that uses a zinc-cofactor bound to an enzyme active site. The zinc acts like a Lewis acid activating the substrate making the α-proton more acidic allowing the zinc enolate to form. An example of this mechanism is shown in **Scheme 1.13** using dihydroxyacetone phosphate (DAHP) **69** as the substrate.[62]



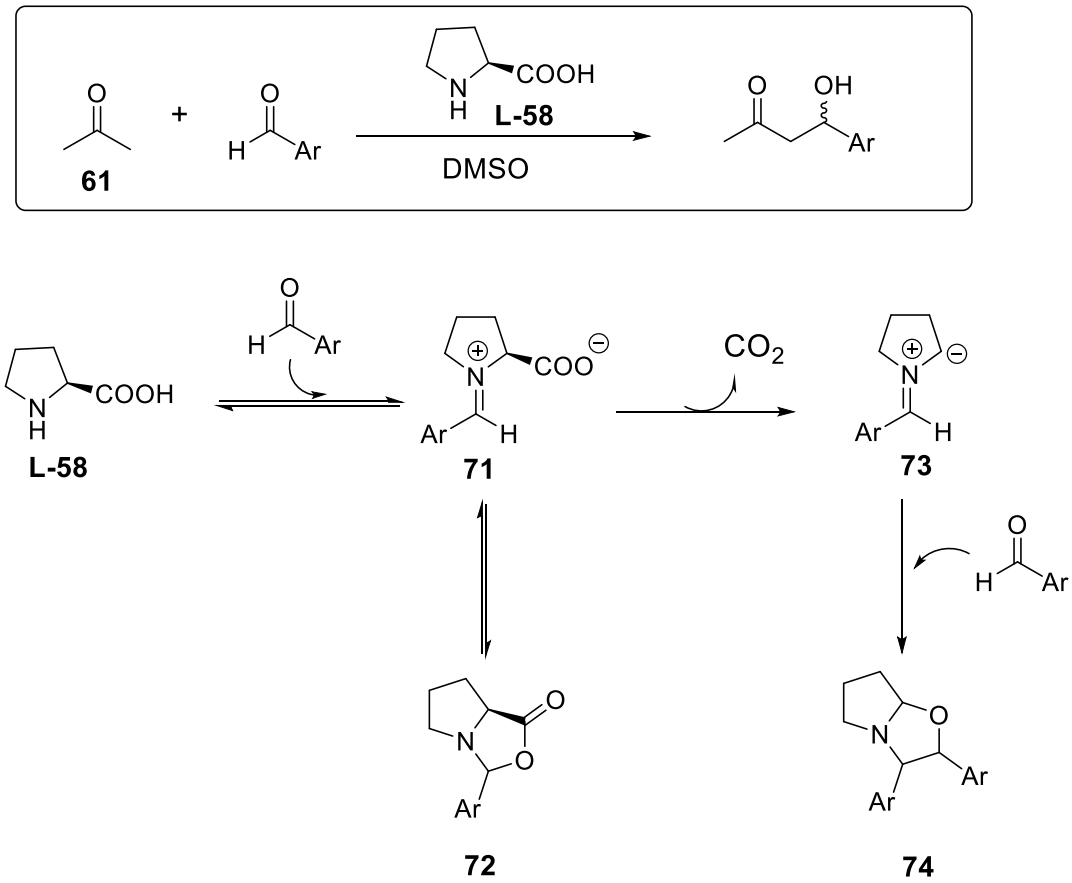
Scheme 1.13. Aldol type II reaction mechanism using a zinc-cofactor.

One of the main differences between the synthetic chemistry shown so far and that of Nature is the solvent these reactions occur in. All biological reactions occur in water, as this is the solvent within the human body. Like-wise, from an origins of life perspective, the solvent on the early Earth is likely to have been water.

1.5. Organocatalysis in water

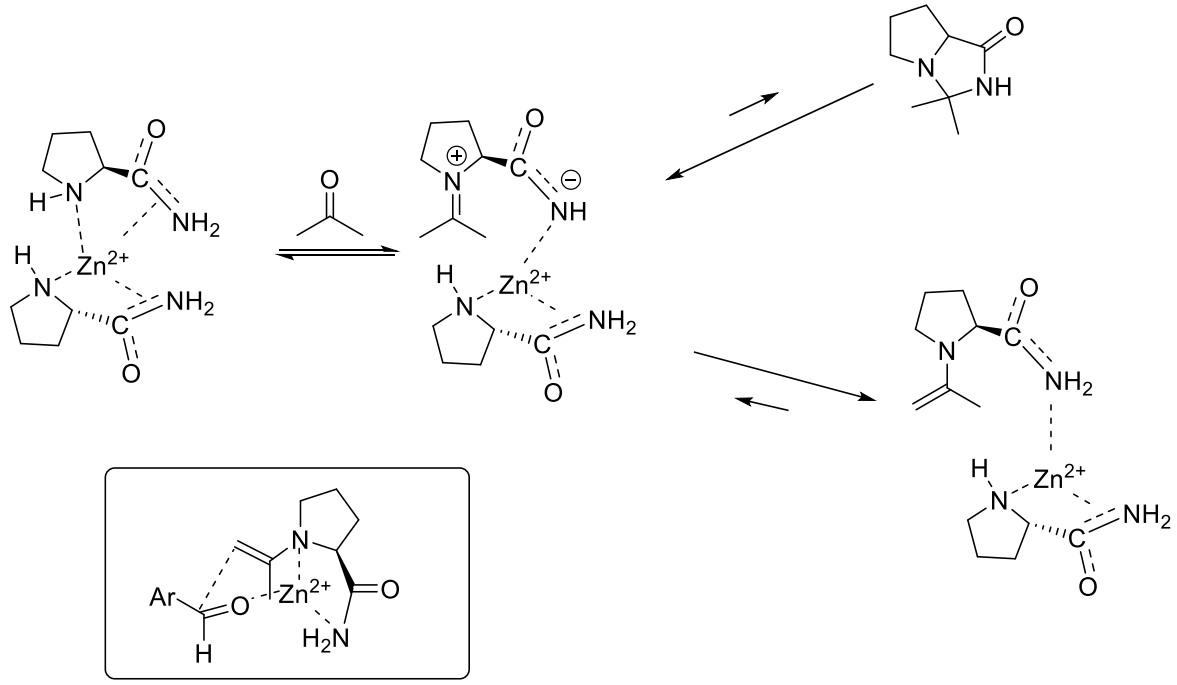
In the first step of an organocatalytic aldol reaction the catalyst, e.g. proline **58**, attacks an aldehyde to undergo a condensation reaction expelling water. The resulting imine is in an equilibrium with the aldehyde starting material. Le Chatelier's principle dictates that the addition of water will shift the reaction back towards starting materials. However, there is also a benefit to having water present. Blackmond *et al.* investigated the kinetics of a proline catalysed reaction of acetone and an aromatic aldehyde (**Scheme 1.14**).[61] Analysis of the

reaction mixture at various points in the reaction, through enthalpy measurements of the reaction calorimeter, found spectator zwitterion **71** formed through the condensation of proline **58** with the aromatic aldehyde. Zwitterion **71** was in equilibrium with its oxazolidinone form **72**. Species **73** could also be formed through irreversible loss of carbon dioxide, which was stabilised through conjugation with the aromatic ring. Blackmond concluded that water was beneficial through reduction of the amount of spectator ion **71**, favouring hydrolysis to L-proline **L-58**, but also hindered the reaction by reduction of the amount of active enamine catalyst.[61]



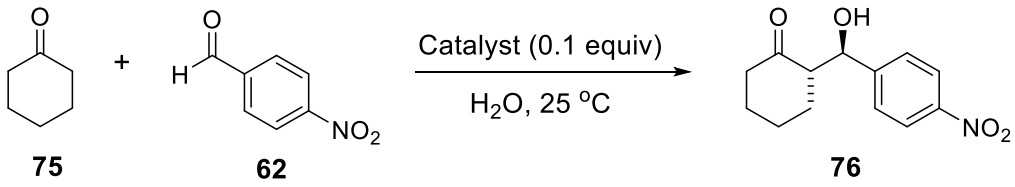
Scheme 1.14. Formation of oxazolidinone byproducts that can take place during organocatalytic aldol reactions.[61]

Based on the mechanism of the aldolase II enzyme, Dabre *et al.* used $Zn(Pro)_2$ to catalyse the aldol reaction between aldehydes and ketones in aqueous medium and found it was higher yielding and gave greater ees than proline alone.[63] Inspired by this work, Asensio and coworkers showed it was possible to use zinc salts as co-catalysts for the aldol reaction of acetone using prolinamide.[64] The zinc salt coordinated to the enamine in a chair-like transition state accounting for the attack at the R_e face (**Scheme 1.15**). This experiment also tackled the problem of the intermediate imidazolidine that Blackmond identified. The zinc ion formed a dimer intermediate (**Scheme 1.15**) which stabalised the amide part of the iminium ion and prevented the cyclization of the "parasitic" imidazolidinone. Although the work of Asensio *et al.* solved the problem of the imidazolidinone byproduct a metal ion was used where as the research decribed in this thesis focuses on the other pathway, Aldolase 1 type approaches. The rest of this review will focus on published chemistry that mimics metal-free type 1 aldolase enzymes.

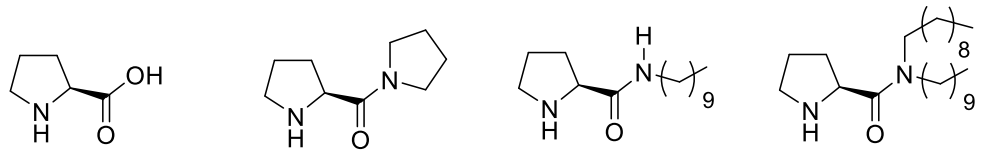


Scheme 1.15. Zinc cofactor forms a more stable intermediate to prevent the formation of the cyclic byproduct.[64]

The presence of water interferes in the six-membered transition state hydrogen bonding between enamine intermediate and electrophile, and a consequence of this is loss of enantioselectivity. Nature has managed to overcome this challenge by using enzymes with active sites located in a hydrophobic binding pocket reducing the contact between water and the reagents/amino acids needed to generate the transition state.[65] Barbas III and coworkers believed the synthetic alternative to this binding pocket was to modify a catalyst by the addition of long alkyl hydrophobic groups to repel water from the site of the reaction. Some of the catalysts are shown in **Table 1.2**.[66]



75 **62** **76**



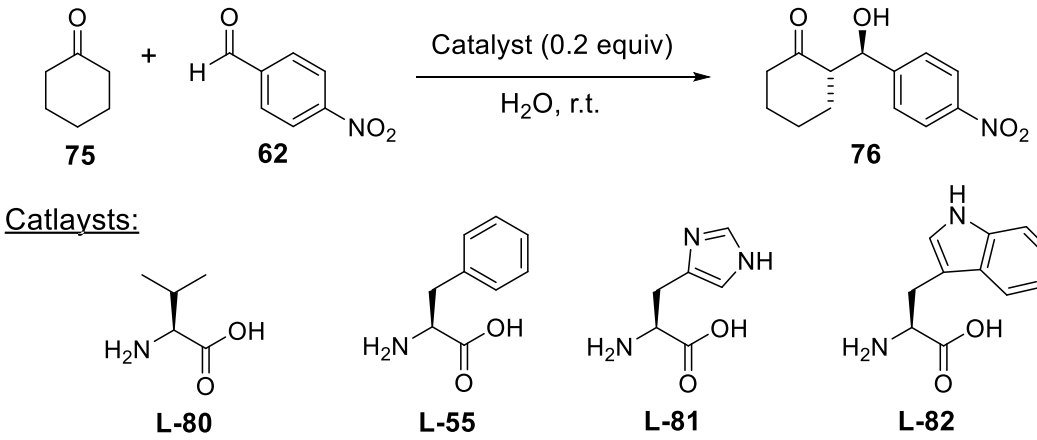
L-58 **L-77** **L-78** **L-79**

Entry	Catalyst	Time (h)	Yield (%)	*anti*:*syn*	ee (%)
1	**L-58**	96	0	-	-
2	**L-77**	5	68	84:16	3
3	**L-78**	5	78	84:16	22
4	**L-79**	5	99	94: 6	1
5	**L-79**[a]	25	99	89:11	94

Table 1.2. Aldol reaction using 75 (2 equiv) and 62 (1 equiv) catalysed by L-proline derivatives. (0.1 equiv). [a] TFA (0.1 equiv) also added.

Barbas III and coworkers studied the reaction of two equivalents of cyclohexanone **75** and one equivalent of *para*-nitrobenzaldehyde **62**. It was found that after four days the reaction, catalysed by L-proline **L-58**, did not yield any product. Modified L-proline derivatives **78** and **79**, however, gave high yields of aldol product, 78 and 99 % respectively (Entries 3 and 4). Barbas reasoned that these long alkyl chains create a hydrophobic environment to exclude water so the reaction could occur much like it would in an organic solvent. The problem of the low enantioselectivites was rectified by addition of a Brønsted-Lowry acid co-catalyst (Entry 5). This dramatically increased the selectivity to 94% as the acid additive helped coordinate the proline catalyst and reagent into the specific chair like transition state similar to the zinc cofactor in **Scheme 1.15**.

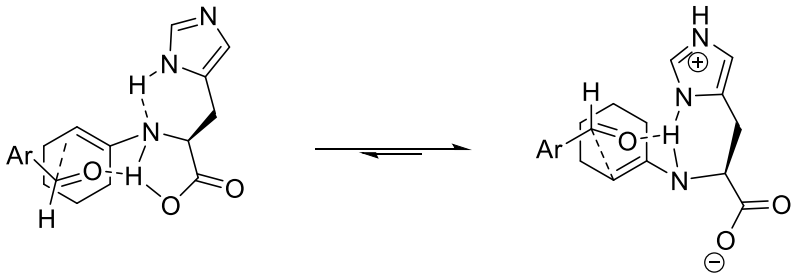
Furthermore, Amedjkouh investigated the use of aromatic amino acids on the same aqueous aldol reaction as Barbas. He compared the catalysis of "hydrophobic" amino acid L-valine **L-80** to those of the aromatic amino acids L-phenylalanine **L-55**, L-histidine **L-81** and L-tryptophan **L-82** (**Table 1.3**).[67] L-Valine **L-80** was unsuccessful at catalysing the aldol reaction after three days, however, phenylalanine **L-55** and tryptophan **L-82** catalysed the reaction in moderate yields and with high selectivity (Entries 2 and 5). Amedjkouh explained this dramatic increase in yield as an implication of the "polar-π" interaction. This is the electrostatically favorable attraction between a π-system (i.e. aromatic ring) and a cation (i.e. protonated iminium ion) more commonly known as a cation-II interaction. This particular interaction has also been shown to occur in proteins.[68] However, this could also be due to π-π stacking interactions between the aromatic ring of **62** and the aromatic ring of the catalyst bringing the active imine catalyst and electrophile into close proximity in order for the reaction to occur. It would be interesting to see if a tangible yield is possible with an aliphatic aldehyde.



Entry	Catalyst	Time (h)	Yield (%)	anti:syn	ee *anti*:*syn* (%)
1	**L-80**	72	0	-	-
2	**L-55**	72	52	19:1	76:-
3[a]	**L-55**	72	79	4:1	54:-
4	**L-81**	72	60	1:1	62:84
5[a]	**L-81**	72	30	1:1	60:80
6	**L-82**	25	62	9:1	88:-

Table 1.3. Aldol reaction using two equiv. of **75** and one equiv. of **62**. [a]Phosphate buffer at pH 7.1 used as solvent.

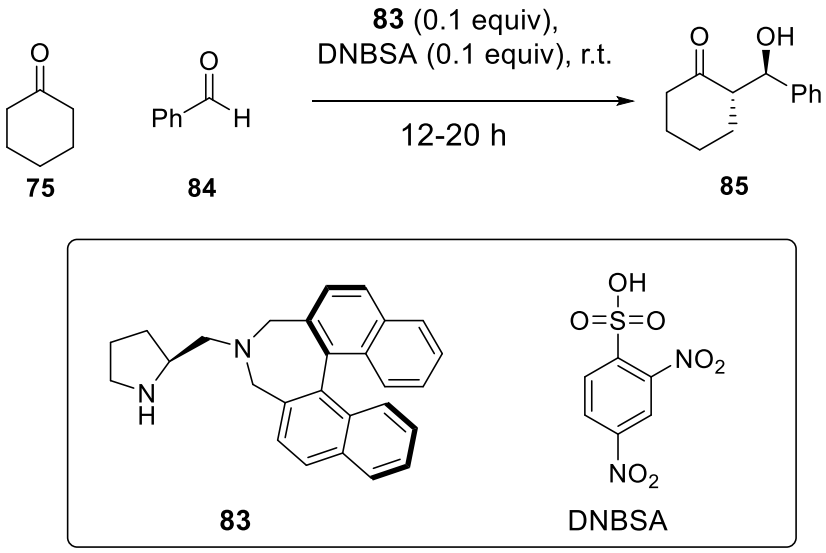
Switching the solvent from water to pH 7 buffer (Entry 3) increased the yield but decreased the selectivity of the products. Furthermore, changing the catalyst to **L-81** changed the enantioselectivity with equal formation of *anti* and *syn* products. The author tried to explain this in terms of the facial attack of the carbonyl in the transition state, arguing that attack from the *Si*-face may be more likely due to stabilization from the imidazole ring (**Figure 1.5**) rather than stabilization from the carboxylic acid group (that would promote R_e-facial attack).[67]



R_e-facial attack of carbonyl S_i-facial attack of carbonyl

Figure 1.5. Proposed transition states leading to S_i-facial attack in the L-histidine **L-81** catalysed aldol reaction.[67]

In 2010, Bisai and Singh tried to incorporate both the large hydrophobic group of a proline derivative with the added benefit of the π- π stacking interactions of aromatic rings to catalyse an aldol reaction (**Scheme 1.16**).[69] A binol–type derivative of prolinamide **83** was used to catalyse the aldol reaction of benzaldehyde **84** and cyclohexanone **75** with an acid cofactor. The reaction was tested over a range of solvents from organic solvents (DMSO, chloroform, THF) to aqueous conditions (water, brine) which all gave high selectivity for the *anti*-product (94:6 and above) and high ees (74 - 84%). The reactions were also conducted in brine and high yields selectivities were observed. It was believed the binaphthyl groups and a salting-out effect created a hydrophobic environment for the active imine catalyst and aromatic aldehyde to come together in a chair configuration through hydrogen bonding.[69]



Scheme 1.16. Compound **83** catalyses the aldol reaction in a range of solvents leading to high yields and selectivities.[69]

Similarly to Bisai and Singh, Pendatella *et al.* also constructed an organocatalyst based on prolinamide, **86**, for the aldol reaction of cycloxhexanone **75** and benzaldehyde derivatives. The catalyst was based on glucosanamine, which itself is a poor catalyst in water. An effective catalyst was designed through protection of the hydroxyl group in the 2-position with a bulky silicon group and addition of L-proline **L-58** to the amine (**Figure 1.6**).[70] The authors conducted the experiments in brine and believed during the reaction the catalyst underwent molecular aggregation, like a surfactant, due to the bulky silyl groups. This created a hydrophobic environment that accumulated the hydrophobic starting materials. [71] The aldol reaction then occured and the slightly more polar products were squeezed out of the binding pockets. After 48 hours and just 2 % catalyst loading, yields of 96 % and ees ranging from 82-99% were achieved.

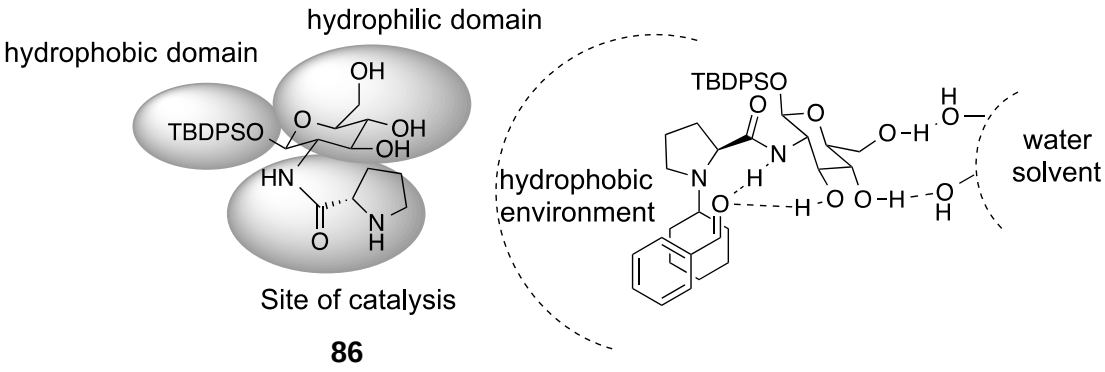
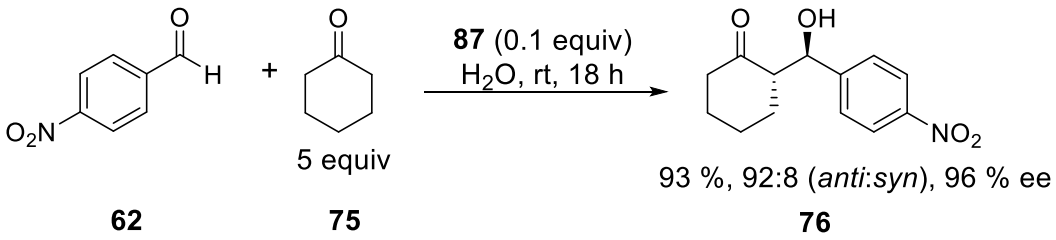


Figure 1.6. Catalyst 86 and relative transition state proposed by Pedatella *et al.*[71]

Lipshutz and Ghoria had a similar idea for the modification of L-proline **L-58**. They made a complex L-proline analogue **87** incorporating the same components of hydrophilic and hydrophobic regions in order to form a binding pocket for the aldol reaction to occur within (**Figure 1.7**).[72] Using *para*-nitrobenzaldehyde **62** and five equivalents of cycloxhexanone **75** as their starting materials the *anti*-aldol product was formed in an excellent yield (93 %) and enantioselectivity (96 %). An added benefit of the reaction was the recyclability. When the reaction was deemed complete all of the products were extracted with ethyl acetate and only the solvent and catalyst remained in the aqueous layer. After three runs of the reaction, extracting the products between each run, the authors demonstrated there was no reduction in yield or selectivity.[72]



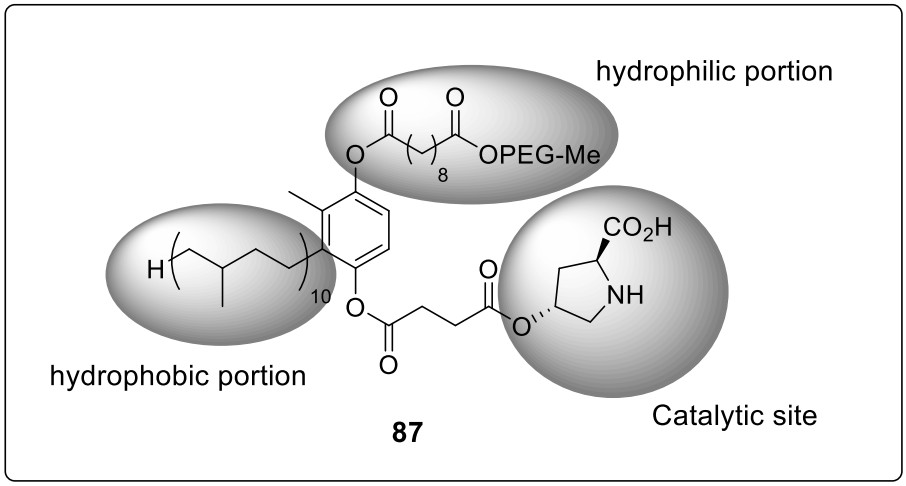
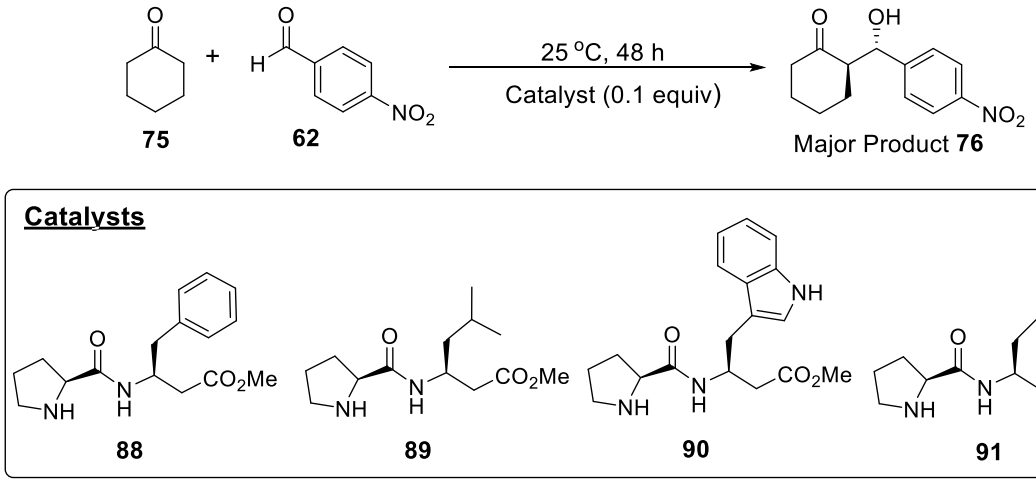


Figure 1.7. Proline derived catalyst **87** synthesised by Lipshutz and Ghorai.[72]

1.6. Peptides in organocatalysis

So far this review has shown that amino acids are poor catalysts for the aldol reaction in aqueous solvents. However, modified amino acids can successfully catalyse aldol reactions of aromatic aldehydes and cyclohexanone in high yields and high selectivities. The next section focuses on the use of dipeptides as catalysts in organocatalysis.

A study by Caputo and coworkers investigated the use of modified dipeptides for the standard aldol reaction of *para*-nitrobenzaldehyde **62** and cyclohexanone **75**. The dipeptides consisted of L-proline **L-58** and a β^3-L-amino acid (**Table 1.4**).[73]



Entry	Catalyst	Solvent	Yield (%)	dr (anti:syn)	ee (%)
1	**88**	Water	72	93:7	83
2	**88**	Brine	98	94:6	88
3	**89**	Water	98	91:9	84
4	**89**	Brine	>99	89:11	76
5	**90**	Water	78	92:8	84
6	**90**	Brine	96	93:7	90
7	**91**	Water	>99	94:6	90
8	**91**	Brine	87	93:7	89

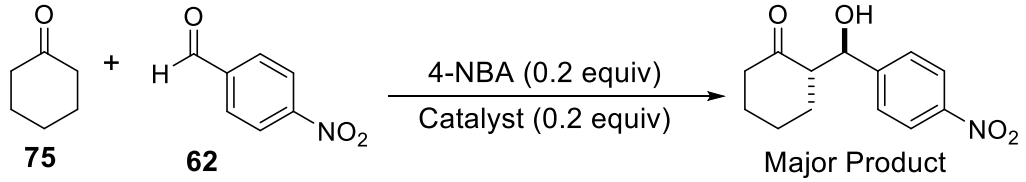
Table 1.4. Aldol reactions of *para*-nitrobenzaldehyde **62** and cyclohexanone **75** catalysed by **88**-**91**.[73]

The group increased the hydrophobicity of L-proline **L-58** by addition of a second hydrophobic amino acid which led to high yields of products all with high selectivity for the *anti* diastereomer. The results were attributed to the increased hydrophobicity of the catalyst and the π-stacking interactions between substrate and catalyst which led to higher selectivities with catalysts **90** and **91**.[73]

In a similar vein, Triandafillidi *et al.* studied dipeptide-like catalysts for the aldol reaction of *para*-nitrobenzaldehyde **62** and 10 equivalents of cyclohexanone **75**. The organocatalysts were based on prolinamide with an attached amino acid residue protected at the C-terminus. The reactions used 0.2 equivalents of catalyst with reaction time lengths of 24

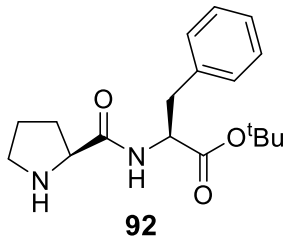
and 72 hours. An acid co-catalyst of 4-nitrobenzoic acid (4-NBA) (0.2 equivalents) was used to increase selectivity and yield. A selection of the results is shown in

Table 1.5.[74]

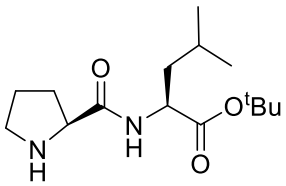


75 62

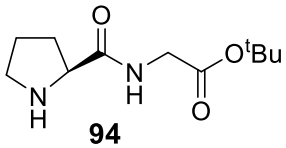
Catalysts



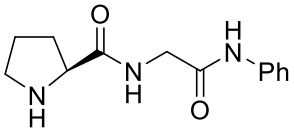
92



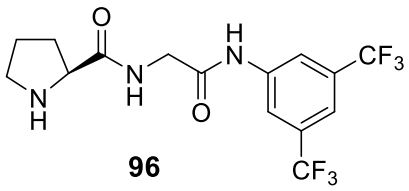
93



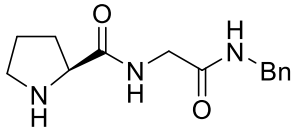
94



95



96



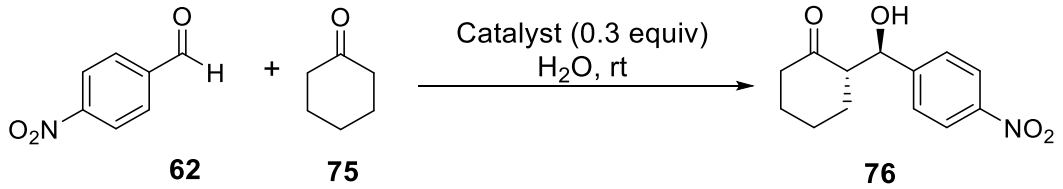
97

Entry	Catalyst	Solvent	Time (h)	Yield (%)	dr (*anti*:*syn*)	ee (%)
1	**92**	Toluene	72	95	75:25	74
2	**92**	Water	24	100	80:20	80
3	**93**	Toluene	48	87	67:33	54
4	**93**	Water	24	100	60:40	57
5	**94**	Toluene	48	87	76:24	87
6	**94**	Water	24	100	90:10	87
7	**95**	Water	48	21	93:7	92
8	**96**	Water	72	0	-	-
9	**97**	Water	48	100	93:7	92

Table 1.5. Catalysts and conditions used to catalsye the aldol reaction of cyclohexanone **75** (10 equiv.) and *para*-nitrobenzaldehyde **62** (1 equiv.).[74]

Catalyst **92**, with a phenylalanine residue, provided excellent yields and good selectivity in both aqueous and organic medium. Changing from an aromatic side chain to an aliphatic side chain, **93** gave similar yields but lower selectivity. The loss of the π-π stacking interaction between substrate and catalyst is a possible explanation for the low selectivity. Removal of the R-group (i.e. using a glycine derivative) seemed to increase the selectivity even further (Entry 5 and 6) and gave higher selectivity in aqueous solvent compared to organic solvent. The authors hypothesised that addition of an achiral amide bond would provide an additional hydrogen bond donor site of the catalyst. However, in these cases low or no yields were obtained (Entries 7 and 8). Surprisingly when the amide-protecting group was changed to a benzyl group, a quantitative yield and high selectivity was obtained. This suggests protecting groups of **95** and **96** act to block the proline catalytic site whereas the additional CH_2 linker is enough to stop the protecting group from blocking the active site. The group concluded that if the catalyst has more hydrogen bonding sites there can be more interactions between electrophile and catalyst holding the complex in a tighter chair-like transition state achieving higher enantioselectivity.[74]

Cordova and coworkers used the same model aldol reaction to test the catalytic ability of unprotected dipeptides in water but achieved only low diastereoselectivities (**Table 1.6**) after several days.[75] After the previous examples of catalyst modifications it is not surprising that Entry 3, with a phenylalanine residue, gave the best enantioselectivity and yield. The best catalyst, Val-Phe, was used in pH 7 buffer (Entries 5 and 6). Buffering the solvent was found to hinder the reaction and after 120 hours the yield was only 27 %, however, the selectivity had increased. The authors do not give a reason for the drop in yield but this could be attributed the lack of competing general acid and base catalysed aldol reactions which would not take place at pH 7. This would also account for the increase in selectivity as only enamine-catalysed aldol reactions would occur.



62 **75** **76**

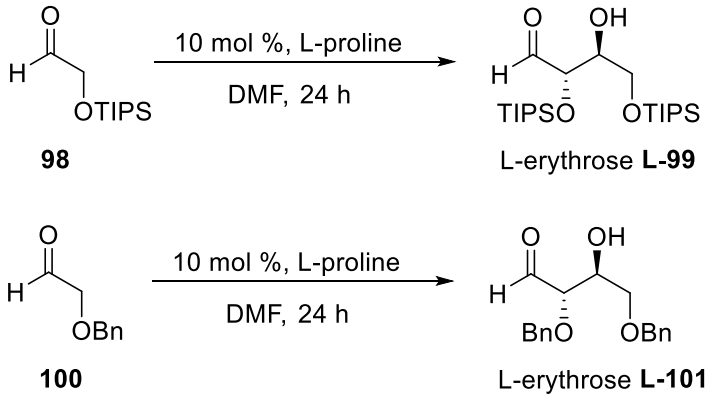
Entry	Catalyst	Time (h)	Yield (%)	dr (*anti:syn*)	ee (%)
1	L-Ala-L-Ala	68	22	2:1	70
2	L-Val-L-Val	76	40	2:1	67
3	L-Val-L-Phe	52	47	3:1	83
4	L-Val-L-Ala	96	30	2:1	70
5	L-Val-L-Phe[a]	120	27	2:1	85
6	L-Val-L-Phe[a]	120	27	2:1	85

Table 1.6. Aldol reaction using **75** (3 equiv.), 62 (1 equiv.), dipeptide catalyst (0.3 equiv.) and SDS additive. [a]Phosphate buffer used a solvent, 40 mM, pH 7.2.

1.7. Amino acid-catalysed sugar chemistry in a prebiotic scenario

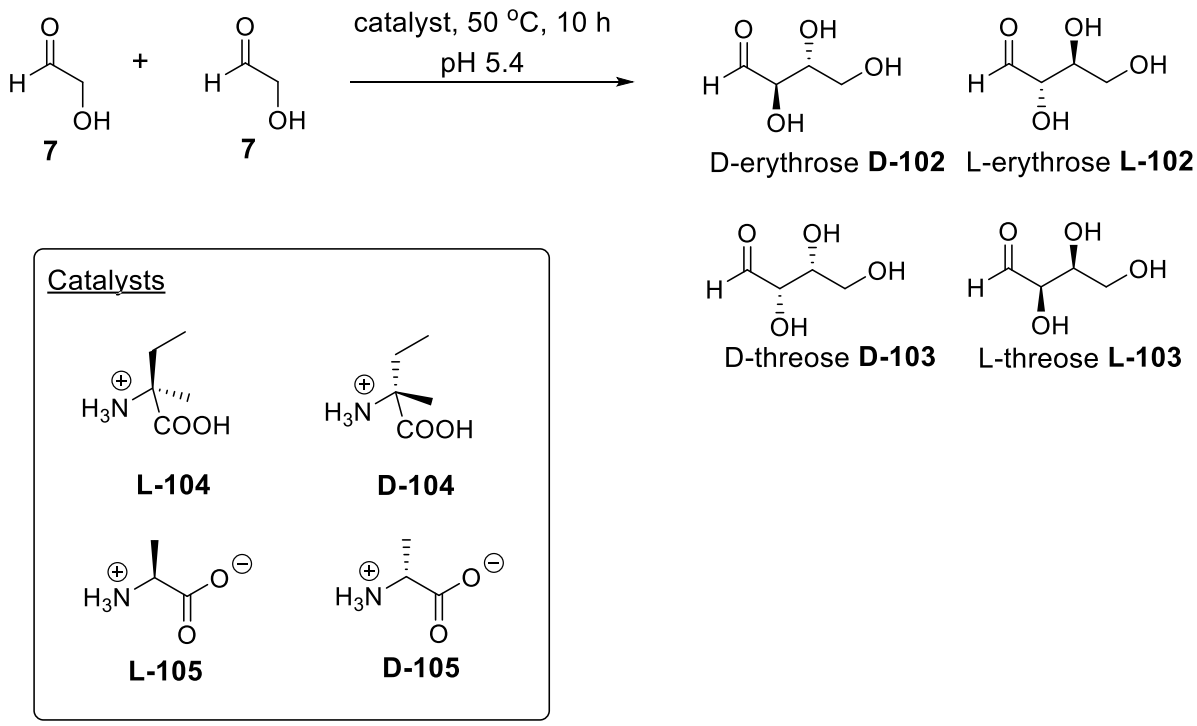
So far the origin of amino acids and the use of amino acids as organocatalysts has been discussed. This next section combines these two topics, exploring the recent research into the use of amino acids and their derivatives as organocatalysts for sugar synthesis in a prebiotic Earth scenario.

In 2004, Northrup and MacMillan showed how proline can be used to catalyse the dimerization of protected glycolaldehyde derivatives **98** and **100** in DMF to give L-erythrose products **L-99** and **L-101** in excellent yields and ees (up to 98 %) (**Scheme 1.17**).[76]



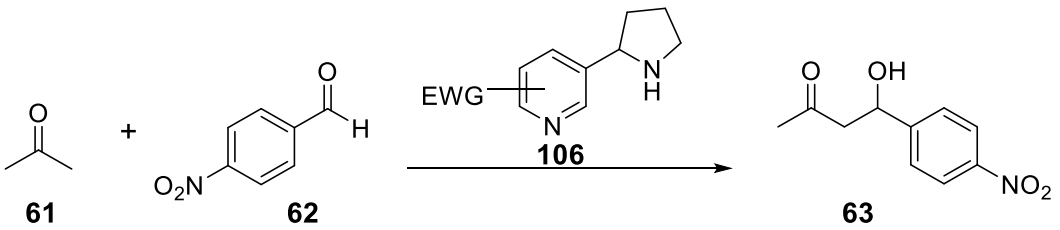
Scheme 1.17. Reactions conducted by Northrup and MacMillan. Using protected glycolaldehyde and L-proline as an asymmetric organocatalyst in DMF gave tetrose products in excellent yields and ees.[76]

Similarly, Pizarello and Weber attempted the dimerisation of unprotected glycolaldehyde in aqueous triethylammonium acetate buffer using enantioenriched amino acids **104** and **105** (**Scheme 1.18**). The desired dimer products were obtained but with low ees (5-12 %) and yields were not stated.[77] It seemed that changing from an organic solvent to water dramatically reduced both the yield and stereochemical control of the reaction. The group's later work showed a vast improvement on these results, using dipeptides as catalysts to obtain ees as high as 78 % for D-erythrose **D-102** but no selectivity for erythrose over threose. The yields were consistently at 20-25 % of tetrose, however, low diastereoselectivities were observed (approximately 1.5:1 erythrose:threose).[78]



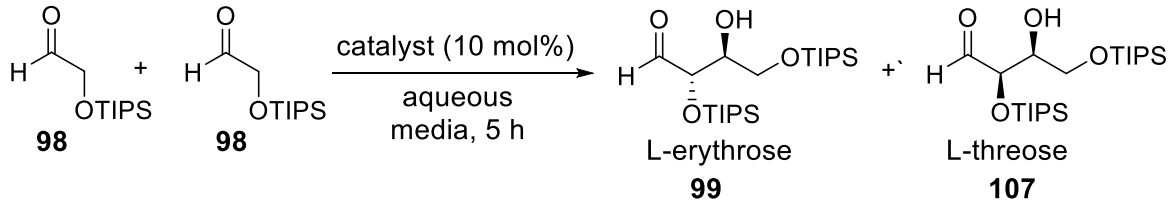
Scheme 1.18. Early work of Pizarello and Weber on the organocatalytic dimerization of glycolaldehyde **7**.[77]

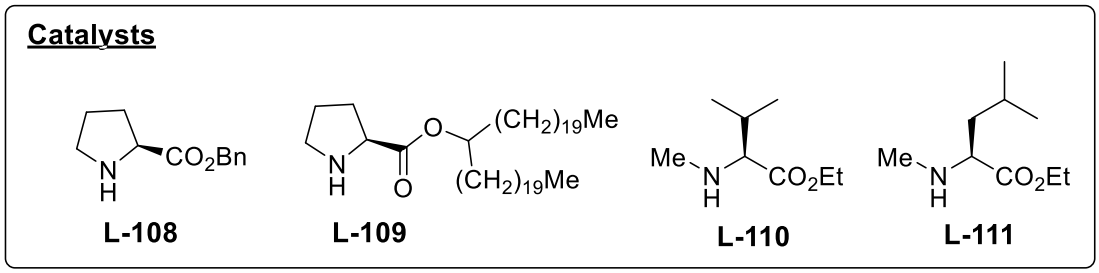
An interesting study by Dickerson and Janda showed that modification of the nicotine metabolite, nornicotine **106**, with electron withdrawing groups increased the rate of the aldol reaction between *para*-nitrobenzaldehyde **62** and acetone **61** in water (**Scheme 1.19**).[79,80] Conversely when electron donating groups were added the rate of reaction decreased. Janda's explanation was that the electron withdrawing groups on the ring lowered the pK_a of the pyrrolidine nitrogen which facilitated the formation of the reactive enamine intermediate. This effectively increased the concentration of the catalyst present in the reaction.



Scheme 1.19. The nornicotine-catalysed aldol reaction investigated by Janda *et al*. Electron withdrawing groups (EWG) increased the rate of the reaction.[79]

Clarke *et al*. were inspired by Janda's work and attempted the dimerization of tri*iso*propylsilyl (TIPS) protected glycolaldehyde in an aqueous solvent with a range of amino esters, where the esters acted as electron withdrawing group.[81] Some of the key results are shown in **Table 5.7**.



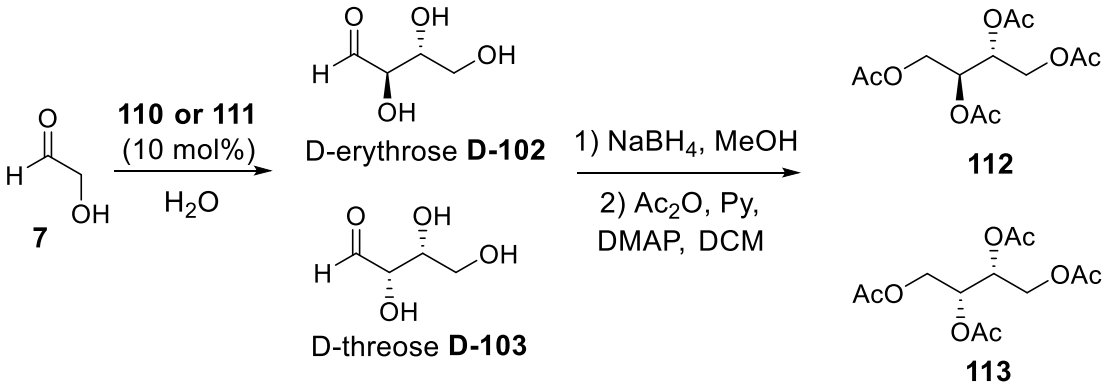


Entry	Catalyst	Solvent	Major Product	Combined Yield (%)	dr	ee % (*anti*)
1	**L-108**	Water	**L-99**	77	1.5:1	18
2	**L-108**	pH 7 buffer	**L-99**	70	1.5:1	47
3	**L-109**	Water	**L-99**	49	2:1	10
4	**L-109**	pH 7 buffer	**L-99**	52	5.5:1	46
5	**L-110**	Water	**D-99**	80	1.5:1	17
6	**L-110**	pH 7 buffer	**D-99**	79	1.5:1	57
7	**L-111**	Water	**D-99**	33	1:1	31
8	**L-111**	pH 7 buffer	**D-99**	40	1.5:1	79

Table 1.7. TIPS protected glycolaldehyde dimerization using amino ester catalysts.[81]

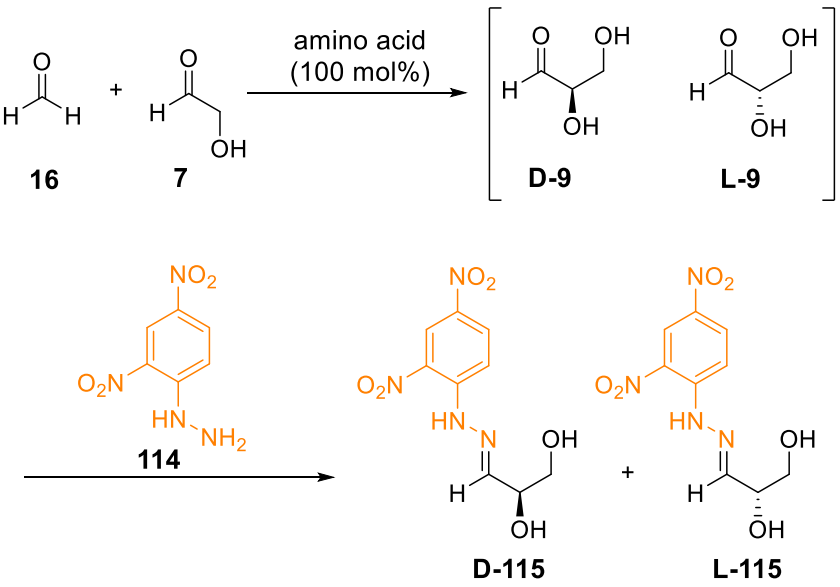
The results from Clarke *et al*. showed that L-proline-based catalysts gave L-erythrose **L-99** as the major product but aliphatic amino esters gave the D-isomer **D-99** as the major product. The authors were surprised by this result and do not give an explanation for this observation. Changing from water to pH 7 phosphate buffer gave a marked increase in enantioselectivity.[81] The authors justified this result by arguing that in water there was also a degree of general base catalysis which had no selectivity for either enantiomer, however, pH 7 buffer should reduce the amount of general acid or general base catalysis and hence catalysis by the amino ester should increase which increases the enantioselectivity.

Clarke *et al*. developed the reaction further through the investigation of protecting group-free glycolaldehyde dimerisation. To aid analysis, the resulting sugar mixture was reduced to tetrols and trapped with acetic anhydride to give protected tetrols **112** and **113** (**Scheme 1.20**). The yield of the reaction was reduced considerably to 0-12 % and the d.r. of products varied considerably from 1:1 at pH 6 with catalyst **111** to 7:1 in pH 7 buffer with catalyst **111**.



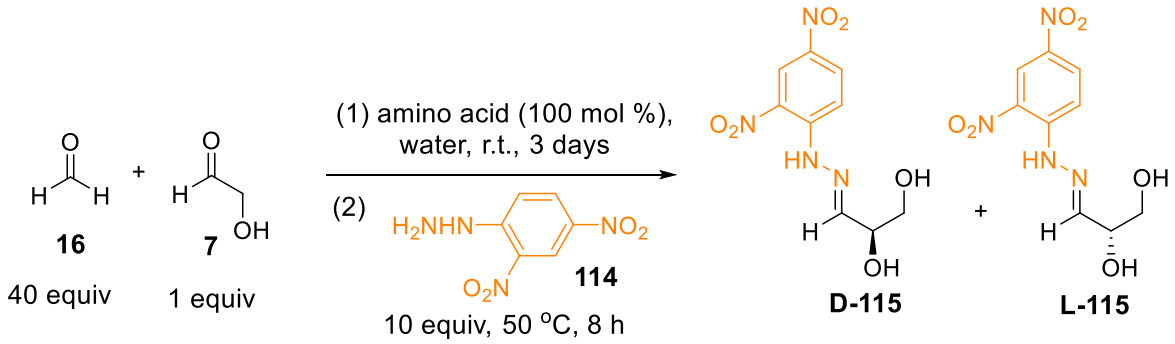
Scheme 1.20. Synthesis of unprotected tetrose products and subsequent trapping by Clarke *et al*.[81]

The prebiotic synthesis of another simple sugar, glyceraldehyde **9**, has been achieved in a similar way *via* an amino acid-catalysed aldol reaction. This reaction, first demonstrated by Breslow and Cheng, used 40 equivalents of formaldehyde **16** and 1 equivalent of glycolaldehyde **7**, both considered interstellar building blocks.[82,83,84] The reaction was stirred for 3-5 days in deionized water and then trapped with 2,4-dinitrophenylhydrazine **114** (**Scheme 1.21**).[85] Out of the seven amino acids tested, five gave a very slight excess of the D-enantiomer product. L-glutamic acid **L-38** gave an excess 21 % in favour of the D-enantiomer of glyceraldehyde **D-9**. Interestingly the authors found when they used L-proline **L-58** as the catalyst an excess of L-glyceraldehyde **L-9** was found (42 %).



Scheme 1.21. Amino acid-catalysed aldol reaction of formaldehyde **16** and glycolaldehyde **7** and *insitu* trapping with 2,-4 dinitrophenyl hydrazine **114**.[85]

Breslow later developed this study further at three different pH ranges, an extract of these results is shown in **Table.5.8**.[86] Noticeably at lower pH values (2.9 – 4.2) there was a large increase in enantioselectivity. For example, L-glutamic acid **L-38** gave an ee of 35 % in favour of **D-9** (an increase of 11 % on the previous work with L-glutamic acid **L-38**). However, this study could not repeat the 42 % ee in favour of L-glyceraldehyde **L-9** when L-proline **L-58** was acting as the catalyst (**Table 1.8**). Interestingly, the ee of the glyceraldehyde **9** product was greatly reduced at higher pH, possibly due to competing general base catalysis, in agreement with the suggestions of Clarke and co-workers.[81]

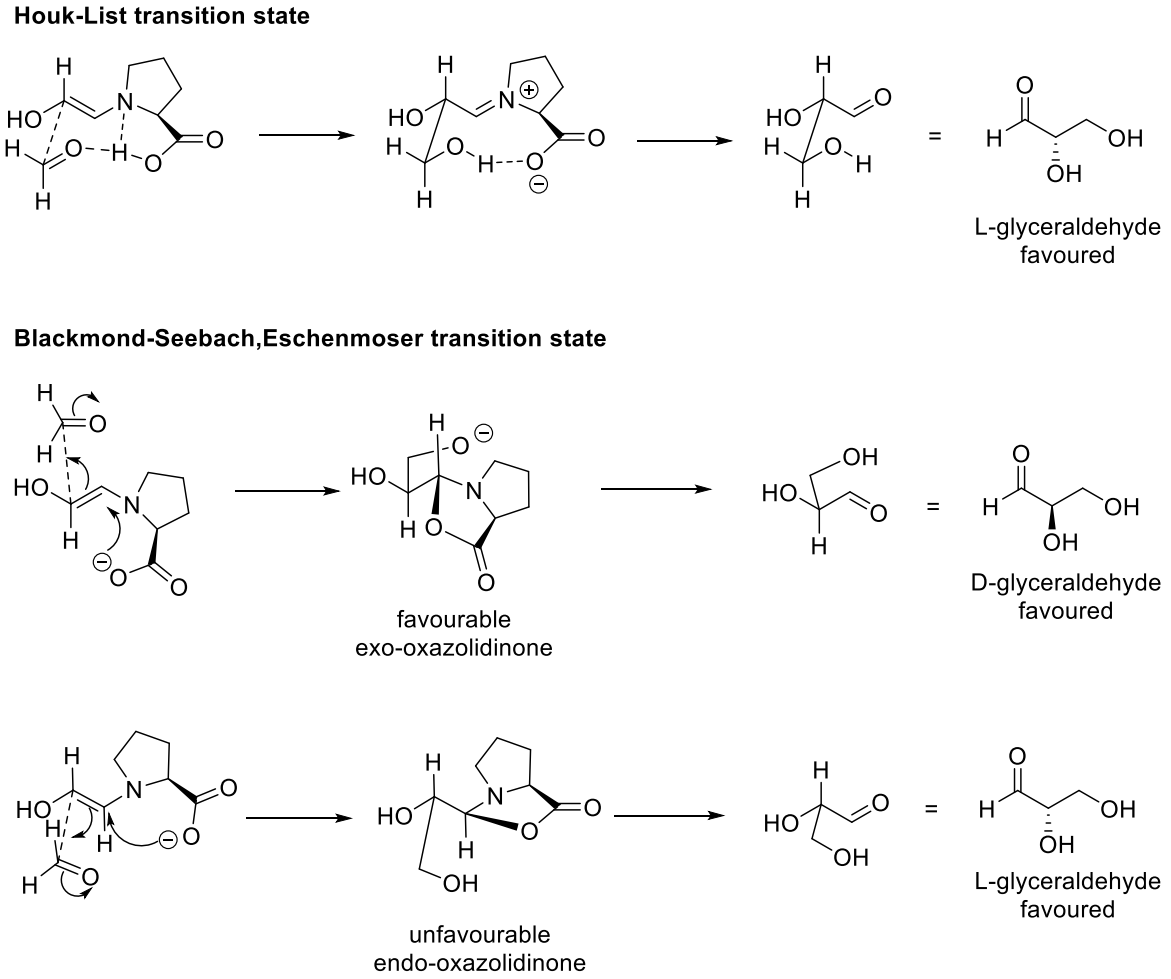


Amino acid	D-glyceraldehyde hydrazone (ee %)		
	pH 2.9-4.2	pH 6.6.-7.6[a]	pH 8.7-9.8[b]
L-glutamic acid	35 ± 2	-3 ± 1	-0.4 ± 0.4
L-aspartic acid	44 ± 1	-1.5 ± 0.2	-1 ± 1
L-leucine	16 ± 1	-2.9 ± 0.6	-3 ± 2
L-valine	19 ± 3	-4 ± 2	-3.0 ± 0.7
L-proline	-20.4 ± 0.3	2.0 ± 0.5	1.5 ± 0.4

Table 1.8. Glyceraldehyde **9** forming reaction carried out by Breslow *et al*.[86] [a] Adjusted by addition of $NaHCO_3$. [b] Adjusted by the addition of Na_2CO_3.

Blackmond and co-workers confirmed that using the same procedure as Breslow, L-proline **L-58** gave L-glyceraldehyde **L-9** with 8 % ee. In the presence of $Bu_4N^+OAc^-$, however, the reaction gave D-glyceraldehyde **D-9** as the major enantiomer in 13 % ee.[87] The apparent link between pH and enantioselectivity with a proline catalyst was explained by Blackmond *et al.*[88] using enamine transition states. In acidic media the carboxylic acid group of L-proline **L-58** is protonated. The aldol reaction of enamine and electrophile proceeds through a Houk-List transition state due to hydrogen bonding between carboxylic acid, amine and electrophile (**Scheme 1.22**).[88] In the presence of base the carboxylic acid is deprotonated and the Houk-List transition state cannot occur. Instead the carboxyanion can attack the enamine to form an oxazolidinone intermediate (**Scheme 1.22**).[89] The more stable exo-oxazolidinone leads to the D-enantiomer of glyceraldehyde **D-9**. Therefore acidic media favours L-glyceraldehyde **L-9** formation and basic media favours D-glyceraldehyde **D-9**

formation. The two transitions states explain the ees observed by both Blackmond and Breslow for the proline-catalysed aldol reaction of glyceraldehyde at varying pH.[86,87]



Scheme 1.22. Transition state rationale of Blackmond *et al*. for the organocatalysed formation of glyceraldehyde.[88]

1.8. Amplification of a Small Enantiomeric Excess

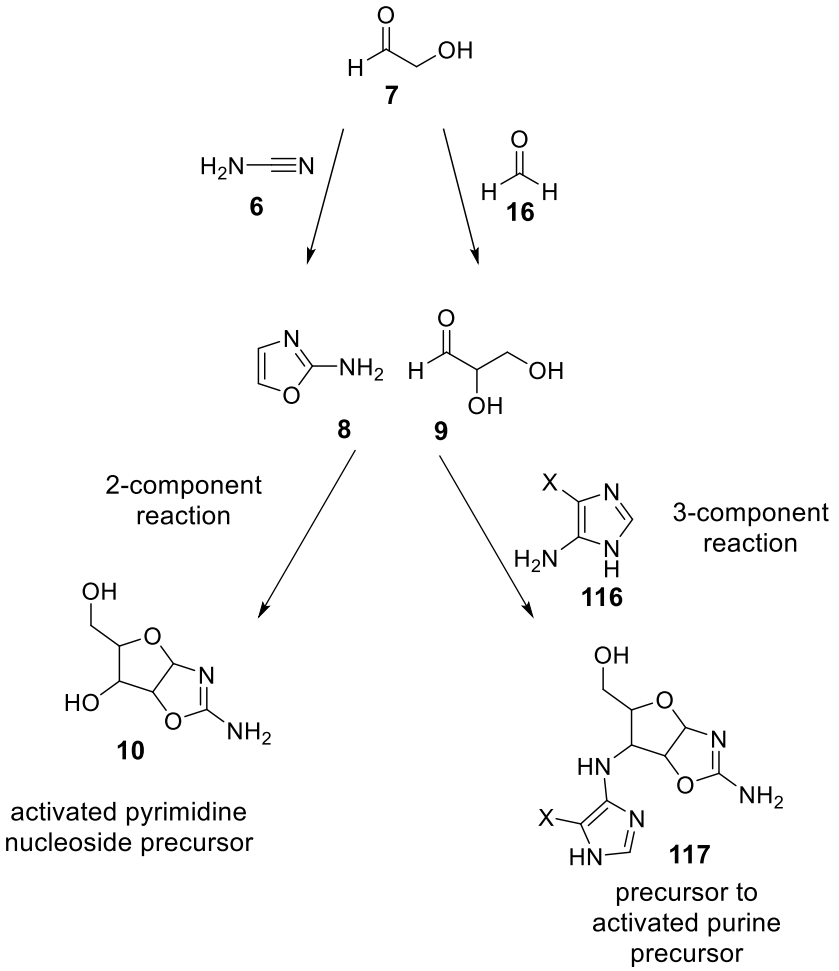
In order to truly achieve a homochiral world there needs to be a means of amplifying the small initial selectivities discussed thus far. Breslow and co-workers found that when glyceraldehyde **9** with a small ee was dissolved in water, the ee of the solution phase was amplified significantly to 94 % ee.[85] It was believed that the racemic dimer of glyceraldehyde **9** is quite insoluble in water leading to its precipitation from solution, hence amplifying the ee of the glyceraldehyde **9** remaining in solution. Similarly, Blackmond demonstrated that amino acids, with just a slight excess of one enantiomer underwent chiral amplification at the so called "eutectic concentration" – the equilibrium point between solid phase and solution phase. If the racemate crystallises out of solution more easily, the molecules left in solution became enantiomerically enriched. Blackmond found this occurred to varying extents with many amino acids (**Table 1.9**).[90]

Amino acid	ee of solution at eutectic concentration (%)	Amino acid	ee of solution at eutectic concentration (%)
Threonine	0	**Methionine**	85
Valine	46	**Leucine**	87
Alanine	60	**Histidine**	93
Phenylalanine	83	**Serine**	>99

Table 1.9. Solution enantiomeric excess of amino acids at eutectic concentration at 25 °C.[90]

Blackmond and co-workers have also achieved amplification of enantiomeric excess by adopting Sutherland's ribonucleotide chemistry. Sutherland's research demonstrated a prebiotically relevant route to nucleotide precursors from glycolaldehyde **7**.[91,92] Reacting glycolaldehyde **7** and cyanamide **6** together formed amino oxazole **8**. Compound **8** was hown to react with glyceraldehyde **9** (either formed *via* the formose reaction or formed

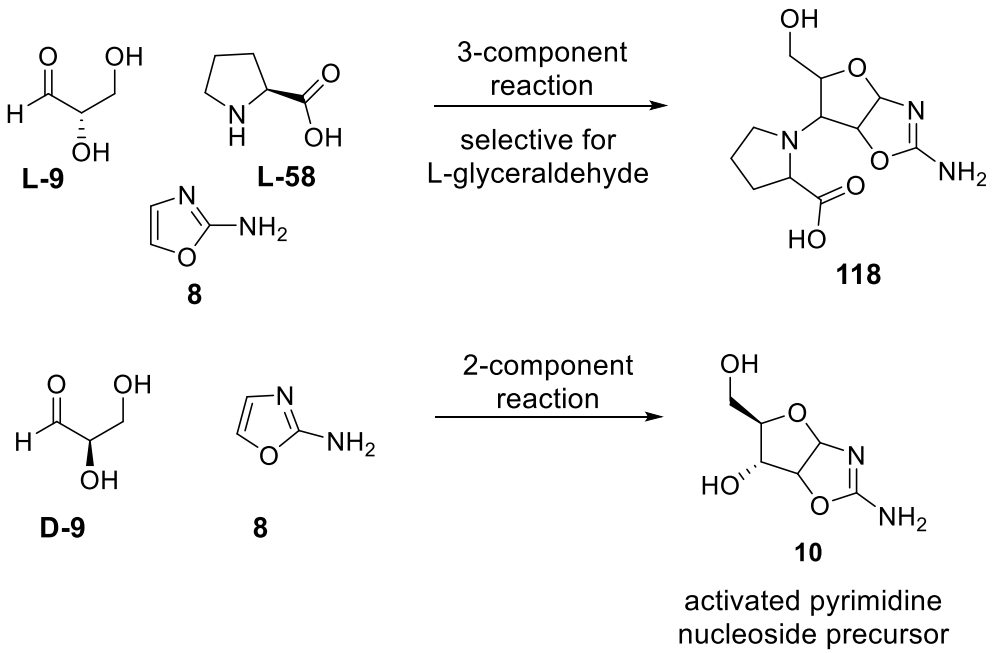
racemically through Sutherland's photoredox chemistry)[39] to form activated pyrimidine nucleoside precursor **10** (**Scheme 1.23**).



Scheme 1.23. Synthesis of nucleotide precursors by Sutherland *et al*.[91,92]

Blackmond has shown the addition of L-proline **L-58** to Sutherland's two component reaction formed adduct **118** (**Scheme 1.24**). L-Proline **L-58** specifically sequestered L-glyceraldehyde **L-9**. Kinetic studies showed that after 40 minutes a solution of racemic glyceraldehyde **9** (1 M), 2-amino-oxazole **8** (1.2 M) and L-proline **L-58** (1 M) gave an enantiomeric excess of D-glyceraldehyde **D-9** of 90 %.[87] This kinetic resolution left a close-to enantiomerically pure solution of D-glyceraldehyde **D-9** which reacted with 2-amino

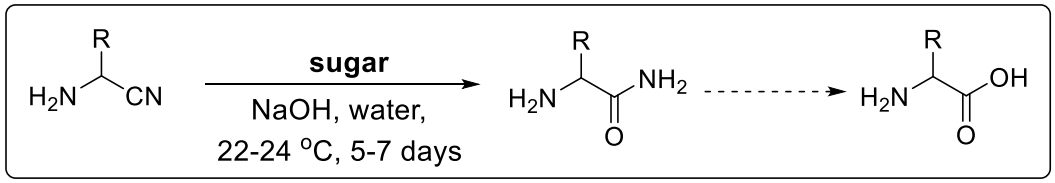
oxazole **8** and transferred the chirality to the pyrimidine nucleoside precursor **10** (**Scheme 1.24**).

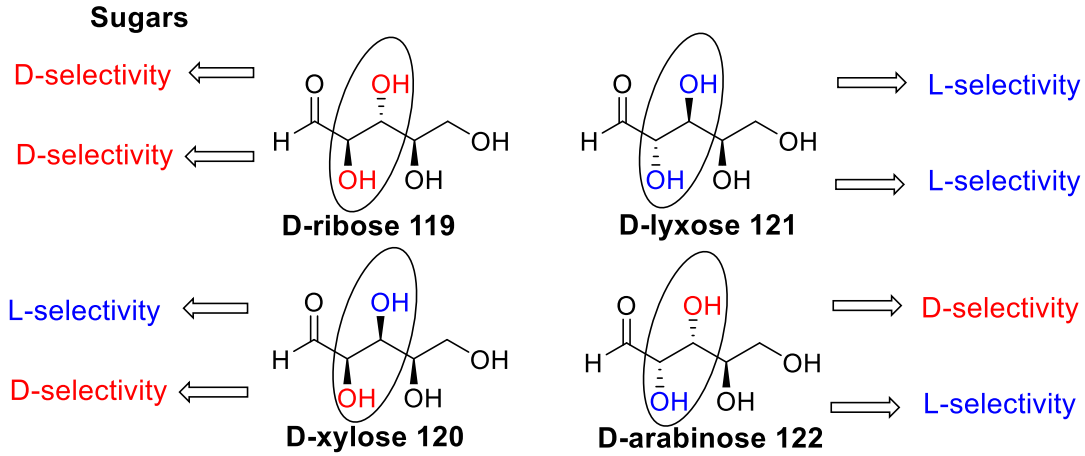


Scheme 1.24. Kinetically resolved synthesis of activated pyrimidine nucleosides.[92]

When dealing with the origins of sugars and amino acids there is a “chicken or egg” argument. Did amino acids help catalyse the formation of enantioselelctive sugars or did sugars help catalyse the formation of amino acids with a degree of enantioselectivity? In an attempt to argue the latter case, Blackmond investigated the influence of sugar chirality on the synthesis of amino acids. Many simple sugar-acids have been found with small ees in meteorites and interstellar ices.[93,94] Blackmond used four sugars, D-ribose **119**, D-xylose **120**, D-lyxose **121** and D-arabinose **122**, for the conversion of prebiotically relevant amino nitrile molecules[40] to amino amides.[95] The reactions were ran for five to seven days and yields of 9 - 29 % were recorded. Through kinetic studies and computer simulations the α and β-hydroxyl groups were found to influence the stereochemistry of the products, with the

α-hydroxyl group having the greatest influence as shown in **Figure 1.8** where the hydroxyl groups are highlighted with a ring and their influence of L and D selectivity stated.





Amino amide (% ee)	Sugar			
	119	**120**	**121**	**122**
alanine	65 (D)	45 (D)	83 (L)	58 (L)
phenylalanine	70 (D)	35 (D)	83 (L)	48 (L)
tryptophan	33 (D)	11 (D)	59 (L)	38 (L)

Figure 1.8. Conversion of amino nitriles to amino amides using pentose sugars and the resultant ee of the products.[95]

Alternatively, chirality may have arisen due to chance. An initial excess of one molecule, whether amino acid or sugar, could have occurred spontaneously, which then through chiral amplification led to exclusively one enantiomer. This can be compared to tossing a coin. The outcome is 50 % chance of getting heads or tails. Nevertheless if you toss a coin ten times one possible outcome is six heads and four tails i.e. six L-amino acids and four D-amino acids and hence an initial slight excess of L-amino acids.

Developing this idea further Frank proposed a scheme to show how a combination of autocatalysis and inhibition in a system of self-replicating chiral molecules allowed stochastic random fluctuations in a racemic mixture to exponentially lead to one enantiomer.[96] Here Frank defined the concentration of one enantiomer, n_1, and a second enantiomer, n_2, in terms of two differential equations (Equations 1.1 / 1.2). Subtracting one from the other gave the exponential Equation 1.3. This showed as long as the initial concentration of one enantiomer, n_{01}, is greater than that of the other, n_{02}, then n_1 increased exponentially in an autocatalytic way.

$$\frac{dn_1}{dt} = (k_1 - k_2 n_2) n_1 \qquad \text{Equation 1.1}$$

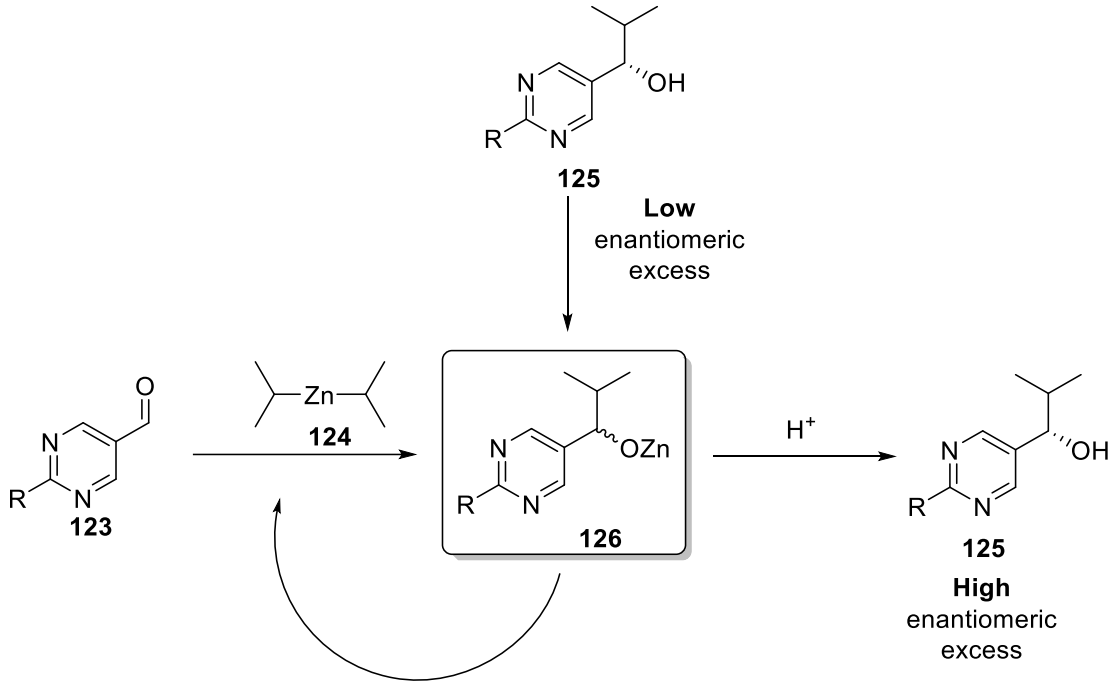
$$\frac{dn_2}{dt} = (k_1 - k_2 n_1) n_2 \qquad \text{Equation 1.2}$$

k_1 / k_2 are positive constants. Subtracting Equation 1.2 from 1.1 gives Equation 1.3.

$$(n_1 - n_2) = (n_{o1} - n_{02}) e^{k_1 t} \qquad \text{Equation 1.3}$$

Frank's principles were demonstrated experimentally forty years later by Soai *et al*. In this experiment the treatment of chiral alcohol **125** (which had a stochastic imbalance of 2 % ee of the *S*-enantiomer) with di*iso*propyl zinc **124** and pyrimidine 5-carboxylate **123**, was amplified to 88 % ee of product without the need to dope the reaction (**Scheme 1.25**).[97] Blackmond later rationalised that it may be the statistical formation of a dimer catalytic species *vs* a heterochiral dimer secies of lower catalytic activity wich accounts for the high enantiomeric excess of the resultant alcohol.[98] Although the Soai reaction is evidence to

support Frank's theory this experiment cannot be thought of as a reaction that could have occurred on the prebiotic Earth.

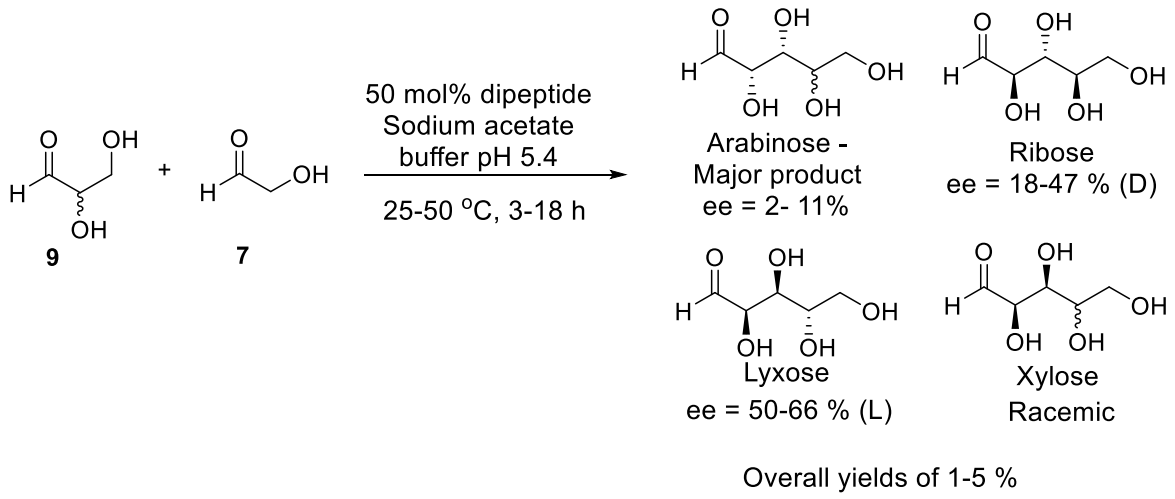


Scheme 1.25. Reaction scheme proposed by Soai *et al.* showing the autocatalysis of 2-methyl-1-(5-pyrimidyl)propan-1-ol leading to amplification of the enantiomeric excess starting from just 2 % ee.[97]

1.9. Approaches to 2-Deoxyribose

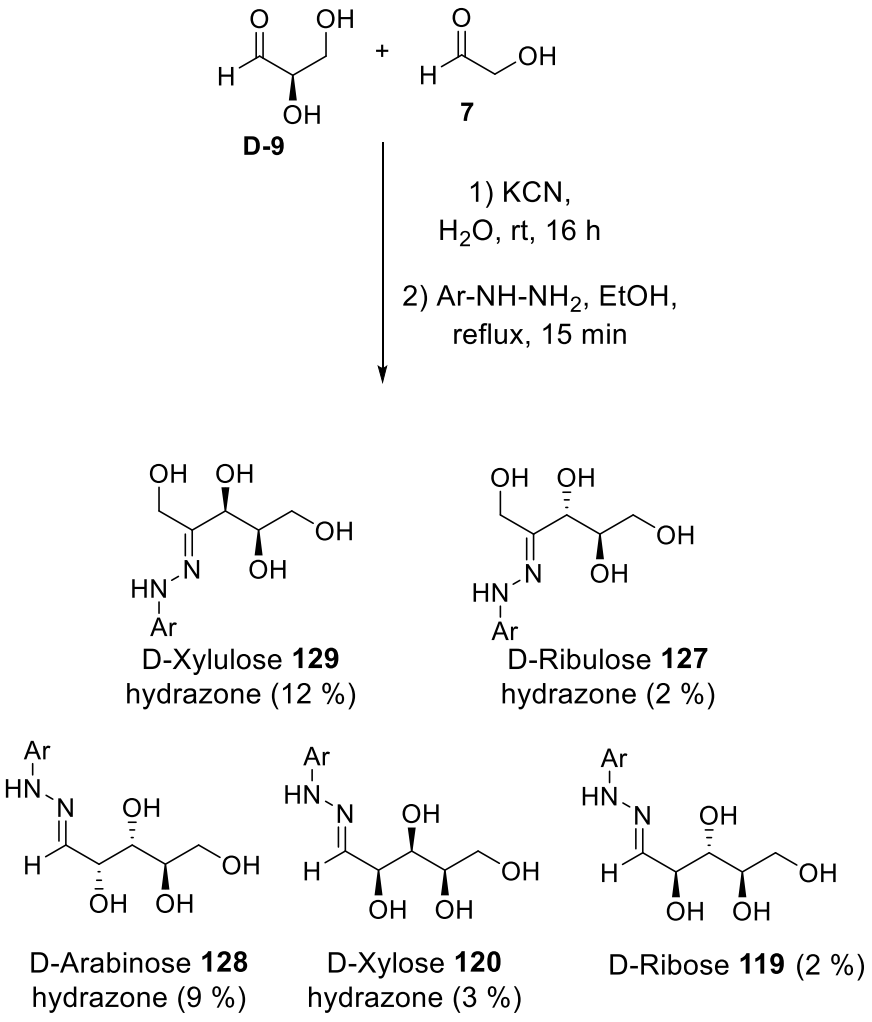
When discussing the prebiotic synthesis of pentose sugars, unlike 3 and 4 membered sugars, there have been far fewer studies into their origins. Benner showed how boron additives could stabalise the pentose products of the formose reaction leading to mixtures of ribulose **127**, arabinose **128** and xylulose **129** pentose sugars as well as tetrose sugars.[23] Pizarello and Weber extended their research of tetrose formation to include pentose formation. The aldol reaction of racemic glyceraldehyde **9** and glycolaldehyde **7**, catalysed by various dipeptides, was found to form pentose sugars (**Scheme 1.26**).[99] The major product formed was arabinose **128** but with a very small enantiomeric excess. D-ribose **D-**

119 and L-lyxose **L-121** were formed in moderate ee.s (35-60 %) when the dipeptide catalysts contained two L-amino acids (L-Val-L-Val and L-Ile-L-Val). When the dipeptide consisted of two D-amino acids enantioselectivity was reversed and L-ribose **L-119** and D-lyxose **D-121** were prevalent. Only a trace amount of racemic xylose **120** was formed.



Scheme 1.26. Dipeptide catalysed synthesis of pentose sugars.[99]

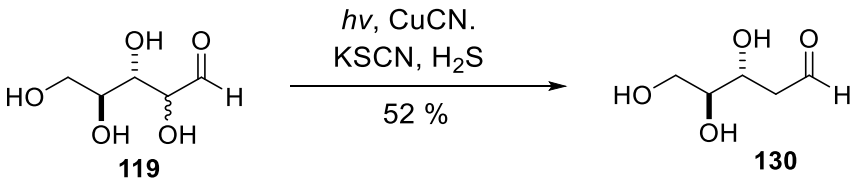
In a similar study Breslow and Appayee investigated the aldol reaction between D-glyceraldehyde **D-9** and glycolaldehyde **7** using potassium cyanide as the catalyst.[100] The resultant sugar mixture was heated in the presence of 2-nitrophenyl hydrazine and the products isolated in their hydrazone forms (**Scheme 1.27**). The major product was found to be xylulose **129** which was not present in Pizarello and Weber's study. Ribulose **127**, arabinose **128**, xylose **120** and ribose **119** were also identified.



Scheme 1.27. Synthesis of pentose products by Breslow and Appayee.[100]

The research described so far has shown possible origins of the pentose sugar ribose, but have not shown selectivity for ribose over the other possible pentoses. The origin of the biological relevant pentose sugar, 2-deoxy-D-ribose, is not described. 2-Deoxy-D-ribose is the sugar present in DNA. Oro and Cox first attempted the prebiotic synthesis of 2-deoxyribose from glyceraldehyde and acetaldehyde. When calcium oxide was used as a base 2-deoxyribose was identified.[101] Although this was a pioneering study at its time, 2-deoxyribose was not isolated nor was it synthesised with any selectivity for the natural D-isomer. Ritson and Sutherland have also demonstrated a possible route to 2-deoxyribose using their photoredox chemistry. [102] The α-deoxygenation of ribose **119** was achieved

using ultraviolet light in the presence of copper cyanide and potassium thiocyanate (**Scheme 1.28**). They proposed that both copper cyanide and thiocyanate are essential to form a more efficient catalyst for the photochemical production of hydrated electrons. The reaction yielded 2-deoxyribose **130** in a very impressive 52 % yield, a large improvement on the 3 % yield quoted by Oro and Cox.[101] Unfortunately the Sutherland chemistry did not account for the stereochemistry of 2-deoxy-D-ribose, as the reaction is racemic. Therefore, as of yet, an enantioselective, prebiotically plausible, route to 2-deoxy-D-ribose **D-130** has not been reported in the literature.



Scheme 1.28. Synthesis of 2-deoxyribose from ribose by Ritson and Sutherland.[102]

1.10. Conclusions and Research Objectives

This literature review has highlighted some of the research that has emerged over the last twenty years in organocatalysis. Many variations and adaptions of amino acids have been used to overcome the aqueous solvent challenge of enamine catalysis in an attempt to replicate the efficiency of aldolase enzymes, and some of this work has met with great success in terms of high yields and good selectivities.

The second half of the review was concerned with discussing recent research into the origins of amino acids and sugars much of which builds on organocatalytic methods. There have been multiple theories put forward to explain the origins of amino acids from

hydrothermal vents and electrical discharge to photoredox chemistry of simple building blocks and extraterrestrial origins. There has also been a great deal of work in understanding the prebiotic origins of sugars. Most chemists believe sugars originated from the formose reaction - the polymerisation of formaldehyde, and a lot of work has centered around this reaction. Several groups have shown the synthesis of the 3-carbon sugar glyceraldehyde and 4-carbon tetrose sugars using aldol reactions catalysed by amino acids and their derivatives. Less research has been published in the area of pentose sugars with a handful of studies showing the isolation of ribose but with no great selectivity. Arguably the most iconic sugar in biology, 2-deoxy-D-ribose, was synthesised by Oro & Cox and Ritson & Sutherland, however, neither of these studies provide any selectivity for the biologically relevant D-enantiomer.[101,102]

The main aims of this research project were to:

i) Provide an alternative explanation for the prebiotic origin of 2-deoxy-D-ribose.

ii) Provide a selective synthesis for 2-deoxy-D-ribose over other diastereomers.

iii) Synthesise 2-deoxy-D-ribose in solution and develop this further into a potential prebiotic supramolecular environment or “protocell”.

2. A Prebiotic Route to 2-Deoxy-D-Ribose

2..1. Initial experimental design

2-Deoxy-D-ribose is a monosaccharide with a characteristic absence of a hydroxyl group at the 2-position of the furanose ring. The sugar, together with phosphate and the bases guanidine, adenine, thymine and cytosine, make up the primary structure of DNA (**Figure 2.1**). Although commonly drawn in its 5-membered furanose form, 2-deoxy-D-ribose **D-130** exists in solution as an equilibrium between its 5-membered furanose, straight chain, and 6-membered pyranose form (**Scheme 2.1**).

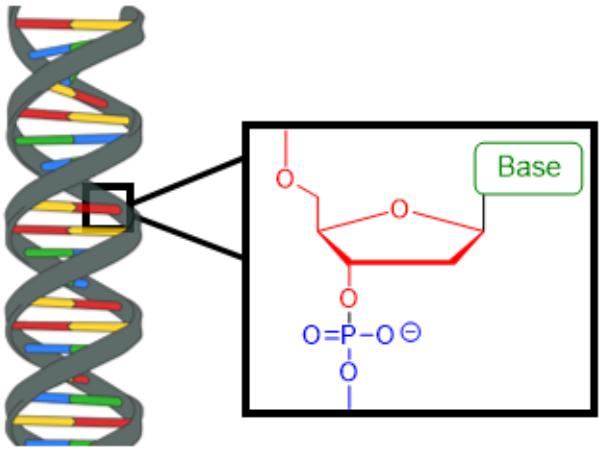
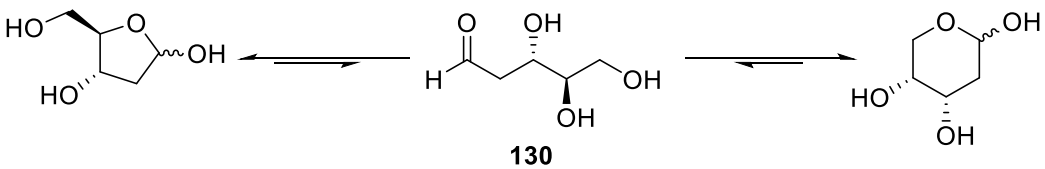


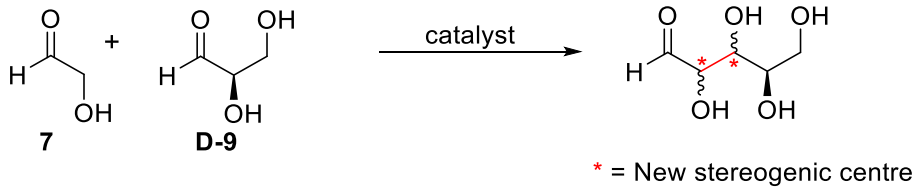
Figure 2.1. DNA anti-parallelhelix and a magnified region of the primary structure of DNA showing deoxyribose, phosphate and a nucleic base.[103]



130

Scheme 2.1. Equilibrium of 2-deoxy-D-ribose **D-130** in solution existing as the furanose, straight chain and pyranose forms.

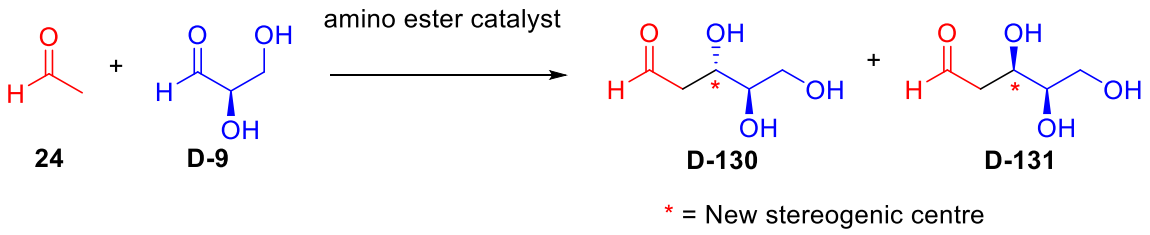
2-Deoxy-D-ribose **D-130** is an essential sugar but its prebiotic origin remains a mystery. The research in this account focuses around a prebiotically plausible synthesis of 2-deoxy-D-ribose **D-130** *via* an aldol reaction. Previously in the literature triose, tetrose and pentose sugars have been formed in aqueous environments *via* aldol reactions catalysed by amino acids and their derivatives.[85,99] A typical example of this is the reaction of glycolaldehyde **7** and D-glyceraldehyde **D-9** which forms a new carbon-carbon bond between the two molecules and generates two new stereogenic centres (**Scheme 2.2**). In this example there could be up to four different diastereomers formed arising from the two chiral centres created.



7 D-9

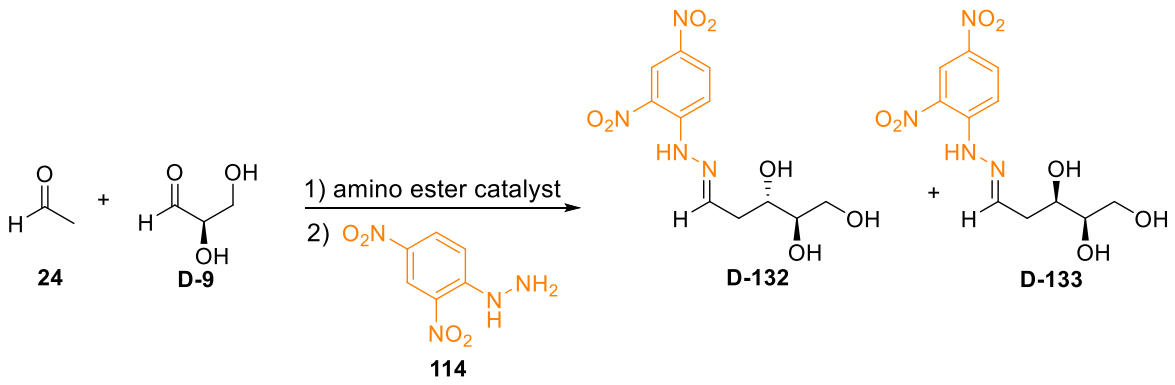
Scheme 2.2. Typical aldol reaction to from larger sugar building blocks.

The same logic was applied to the synthesis of 2-deoxy-D-ribose **D-130**. In this case the two sugar building blocks for the aldol reaction were acetaldehyde **24** and D-glyceraldehyde **D-9**. Two possible pentose products can arise from this reaction, the *anti* diastereomer 2-deoxy-D-ribose **D-130** and the *syn* diastereomer 2-deoxy-D-threopentose **D-131** as only one new stereocentre is generated (**Scheme 2.3**). The initial selection of the catalyst was based on the successful work of Clarke *et al.* who previously used amino acid derivatives, amino esters, for the catalysis of aldol reactions to form tetrose products in aqueous media.[81] These esters should reversibly react with acetaldehyde **24** to form enamine intermediates which can then undergo an aldol reaction with D-glyceraldehyde **D-9** to form 2-deoxypentose sugars **D-130** and **D-131**.



Scheme 2.3. Initial plan for the synthesis of 2-deoxy-D-ribose **D-130** *via* an aldol reaction.

Prebiotic sugar-forming reactions are notorious for producing a broth of molecules including many products in small amounts as well as unreacted starting materials due to the inefficiency of the reactions in water.[15,16,23,26,45] Due to the hydrophilic nature of sugars, simple solvent extraction techniques are ineffective making the resulting broth of sugars difficult to purify and analyse. To aid in the isolation and analysis a simple trapping procedure was to be applied to the sugar mixture which would allow purification and isolation by standard chromatography techniques. Inspired by the work of Breslow *et al.*[85,86] (**Chapter 5.7, Scheme 5.21**) which used 2,4-dinitrophenyl hydrazine **114** as a trap for analysis of the formation of glyceraldehyde **9** from formaldehyde **16** and glycolaldehyde **7**. Therefore 2,4-dintirophenyl hydrazine **114** was proposed as a suitable trap. This modifies the proposed reaction to **Scheme 2.4** which should produce trapped hydrazone products **D-132** and **D-133**.



Scheme 2.4. Modified reaction scheme for the synthesis of 2-deoxy-D-ribose **D-130** with additional trapping step.

1. Synthesis of Standards and Catalysts

To determine the success of the 2-deoxy-D-ribose **D-130** forming reaction, trapped standards of starting materials and products were prepared to aid analysis of the reaction by thin layer chromatography (TLC) and comparison of isolated products by ^{1}H NMR analysis and HPLC. An overview of the hydrazone standards is shown in **Figure 2.2**.

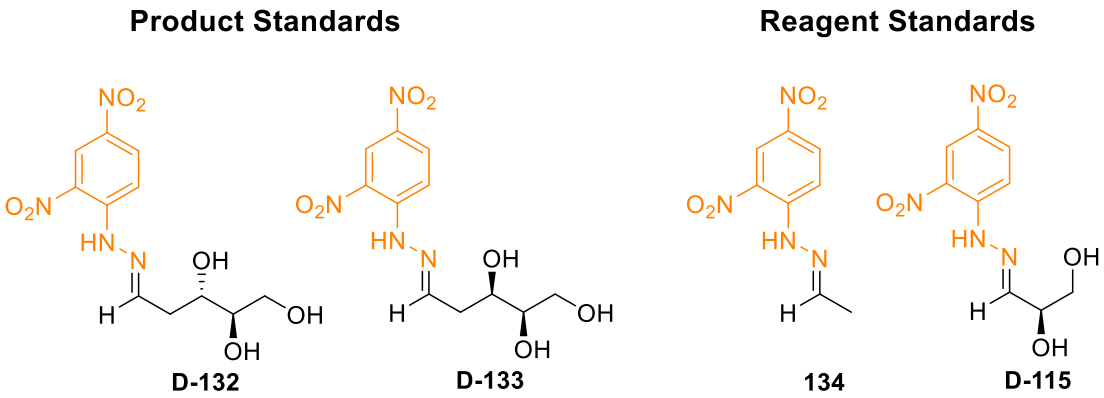
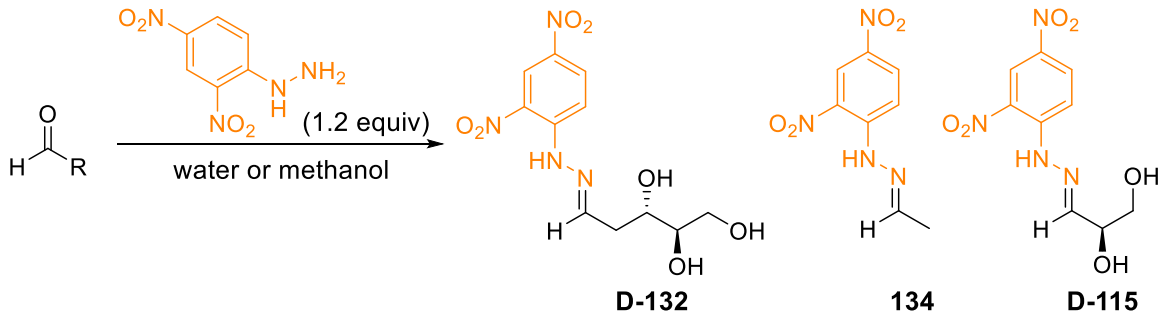


Figure 2.2. Product and reagent hydrazone standards.

The aldehyde precursors of **132**, **134** and **D-115** were commercially available. To form the hydrazone standards the aldehydes were stirred in either methanol or water with 2,4-dinitrophenyl hydrazine **114** for 24 hours (

Table 2.1). As can be seen from

Table 2.1, **134** was obtained in a very good yield and trapped D-glyceraldehyde **D-115** was obtained in a moderate 46 % yield. The trapping of 2-deoxy-D-ribose **D-130** was fist attempted in water and gave **D-132** in a 6 % yield after 24 hours. When the reaction time was increased to 72 hours and the solvent changed to methanol the yield was marginally increased to 8 %.

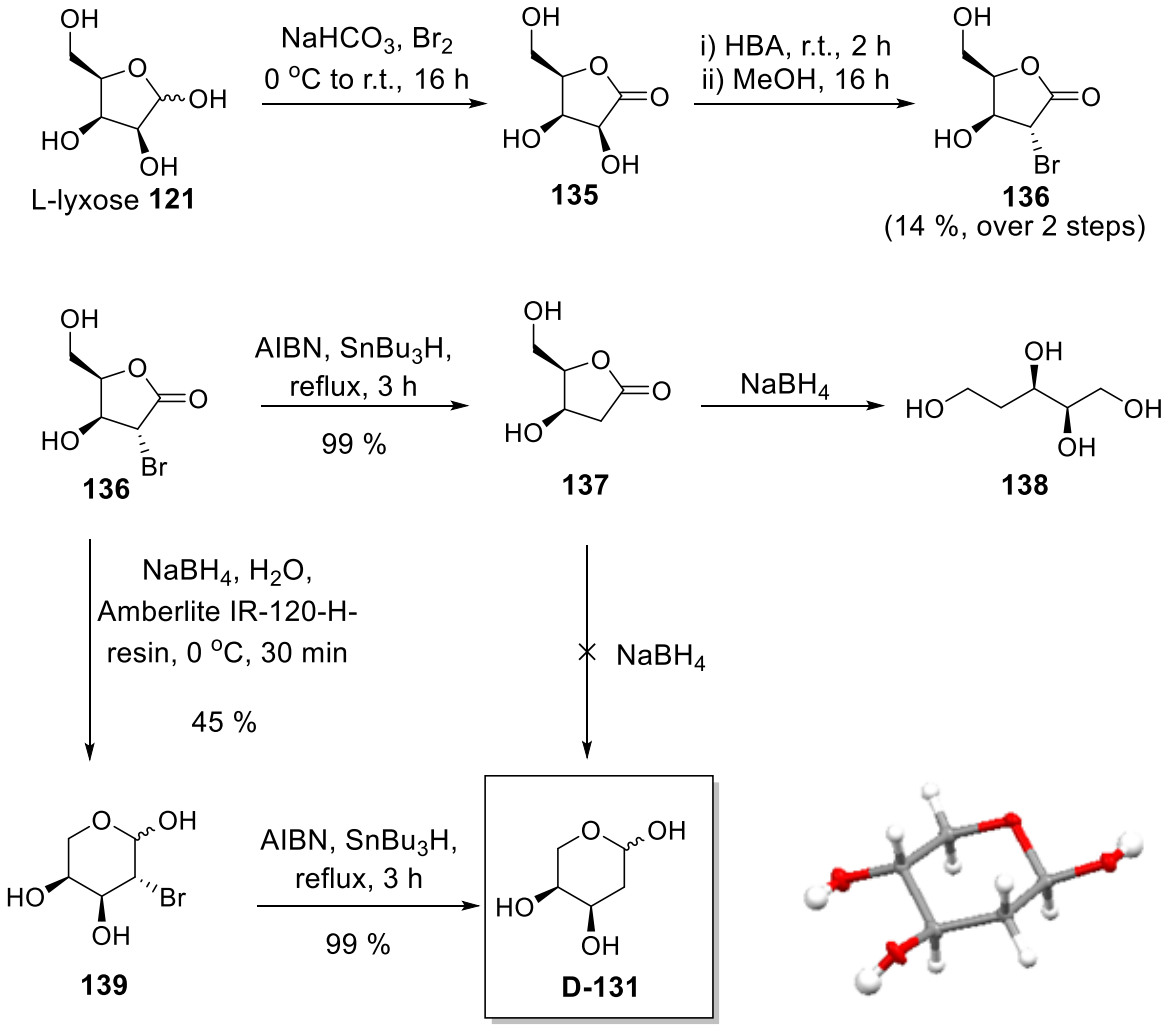


Starting Material	Time	Solvent	Yield (%)
Acetaldehyde **24**	24	water	80
D-Glyceraldehyde **D-9**	24	water	45
2-Deoxy-D-ribose **D-130**	72	methanol	8

Table 2.1. Synthesis of hydrazone standards from commercially available aldehydes.

With hydrazone standards **D-132**, **134** and **D-115** to hand the synthesis of hydrazone **D-133** was attempted. Unfortunately, this sugar was not commercially available and had to be synthesised from commercially available L-lyxose **L-121** through adaptation of previous

literature procedures which ultimately removed the hydroxyl group at the C2 position of the molecule (**Scheme 2.5**).[104,105]

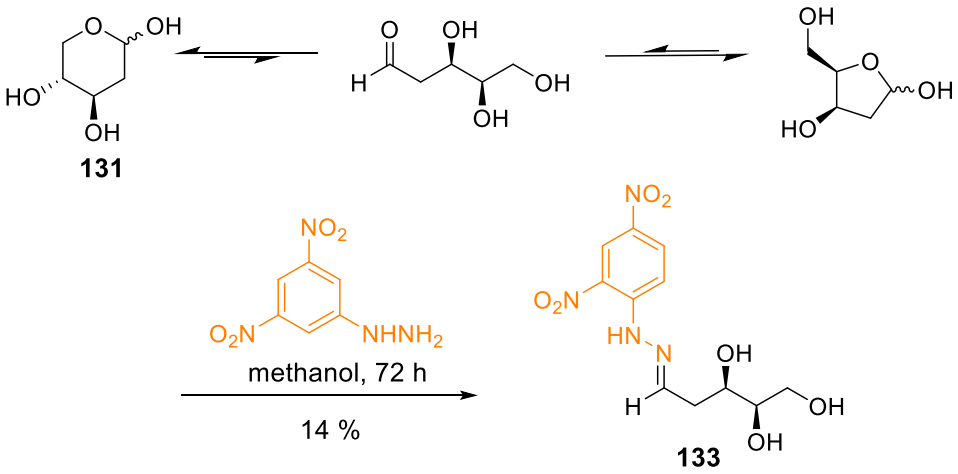


Scheme 2.5. Synthesis of 2-deoxy-D-threopentose **D-131** from commercially available L-lyxose **121** and the crystal structure of **D-131** as 50% eliipsoid.

The first step of the synthesis used sodium bicarbonate as a base with bromine water for the oxidation of L-lyxose **121** to lactone **135**. After work-up, the literature procedure reported that extraction of the product with boiling acetone gave lactone **135** as a white solid in excellent yield.[104] In our hands extraction with boiling acetone gave an extremely poor yield of lactone, < 2 %. Extraction with boiling methanol, however, gave **135** as a beige solid in excess of 100 % yield. As the ^{1}H NMR and ^{13}C NMR spectra of the compound were

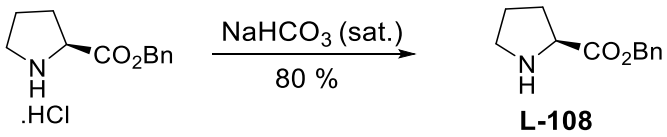
correct and clean it was apparent that inorganic by-products/reagents contaminated the product. Compound **135** was used in the next step of the reaction without further purification, where HBA (hydrobromic acid - a commercially available solution of HBr in acetic acid) induced an S_N2-type reaction which replaced the hydroxyl group at the C2 position with bromine. Compound **136** was obtained in a low 14 % yield over 2 steps. Reduction of **136** to **137** was achieved in excellent yield by radical mediated de-bromination conditions.

With the C2 position now free of heteroatoms the final step of the synthesis was reduction of the lactone down to the lactol to give 2-deoxy-D-threopentose **D-131**. Initially one equivalent of sodium borohydride in water at 0 $^{\circ}$C was used with Amberlite®-120-H resin to keep the pH below 6. After 24 hours only starting material **137** was detected by TLC. The reaction was repeated with 2 equivalents of sodium borohydride and gave the same result after 24 hours. In order to test if the reduction could actually occur the number of equivalents of sodium borohydride was drastically increased to 20. After one hour no starting material was detected by TLC, however the compound had over reduced to tetrol **138**. There was clearly a fine balance between single reduction and over reduction of lactone **137**. A greater degree of control was found by reduction of lactone **136** with sodium borohydride instead. Reduction with 0.5 equivalents of sodium borohydride gave **139** in a 45 % yield with the rest of the mass being a mixture of starting material and over reduced tetrol. Removal of bromine from **139** was then achieved using the same radical conditions as before to give 2-deoxy-D-threopentose **D-131** in a 99 % yield. A crystal structure of **D-131** was obtained which verified the stereochemistry, with the molecule crystallising in the pyranose form. This is shown in **Scheme 2.5**. The final step of the reaction, trapping with 2,4-dinitrophenyl hydrazine **114**, was carried out in methanol and gave the final standard, **D-133**, in a low 14 % yield (**Scheme 2.6**).



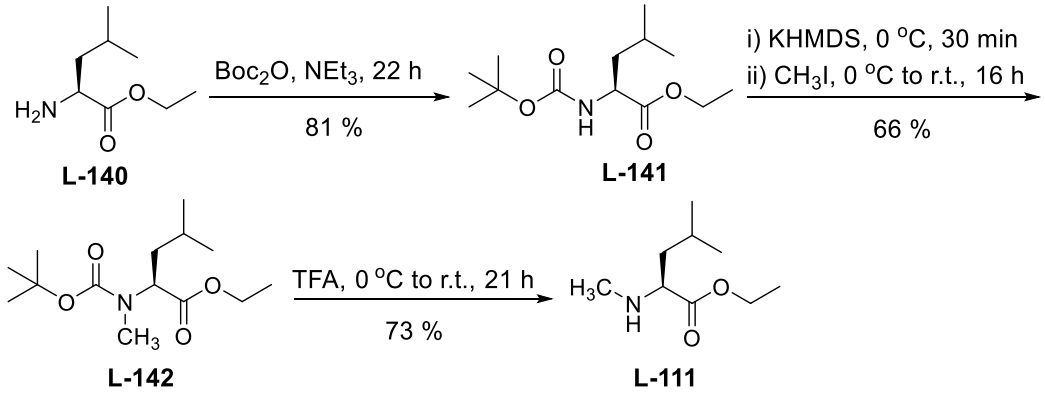
Scheme 2.6. Trapping of 2-deoxy-D-threopentose **D-131** to give hydrazone standard **D-133**.

With all the standards synthesised the synthesis of the amino ester catalysts was attempted. Two successful amino ester catalysts, previously used by Clarke *et al*. in their prebiotically plausible synthesis of tetrose sugars research, were selected as appropriate candidates; L-proline benzyl ester **L-108** and *N*-methyl-L-leucine ethyl ester **L-111**.[81] L-Proline benzyl ester **L-108** was commercially obtained as the hydrochloride salt. The salt was washed with sodium bicarbonate and extracted with dichloromethane (DCM) and upon concentration *in vacuo,* the free amine was obtained as a colourless oil in an 80 % yield (**Scheme 2.7**).



Scheme 2.7. L-Proline benzyl ester **108** was obtained by the neutralisation of the hydrochloride salt.

N-Methyl-L-leucine ethyl ester **L-111** was synthesised from commercially available L-leucine ethyl ester **L-140** by Boc-protection of the free primary amine. Deprotonation of the resultant secondary amine **L-141**, methylation with methyl iodide and subsequent deprotection of the Boc group gave the desired amino ester catalyst **L-111** in an overall 42 % yield (**Scheme 2.8**). The same strategy was used for the synthesis of *N*-methyl-D-leucine ethyl ester **D-111** which allowed access to both enantiomers.



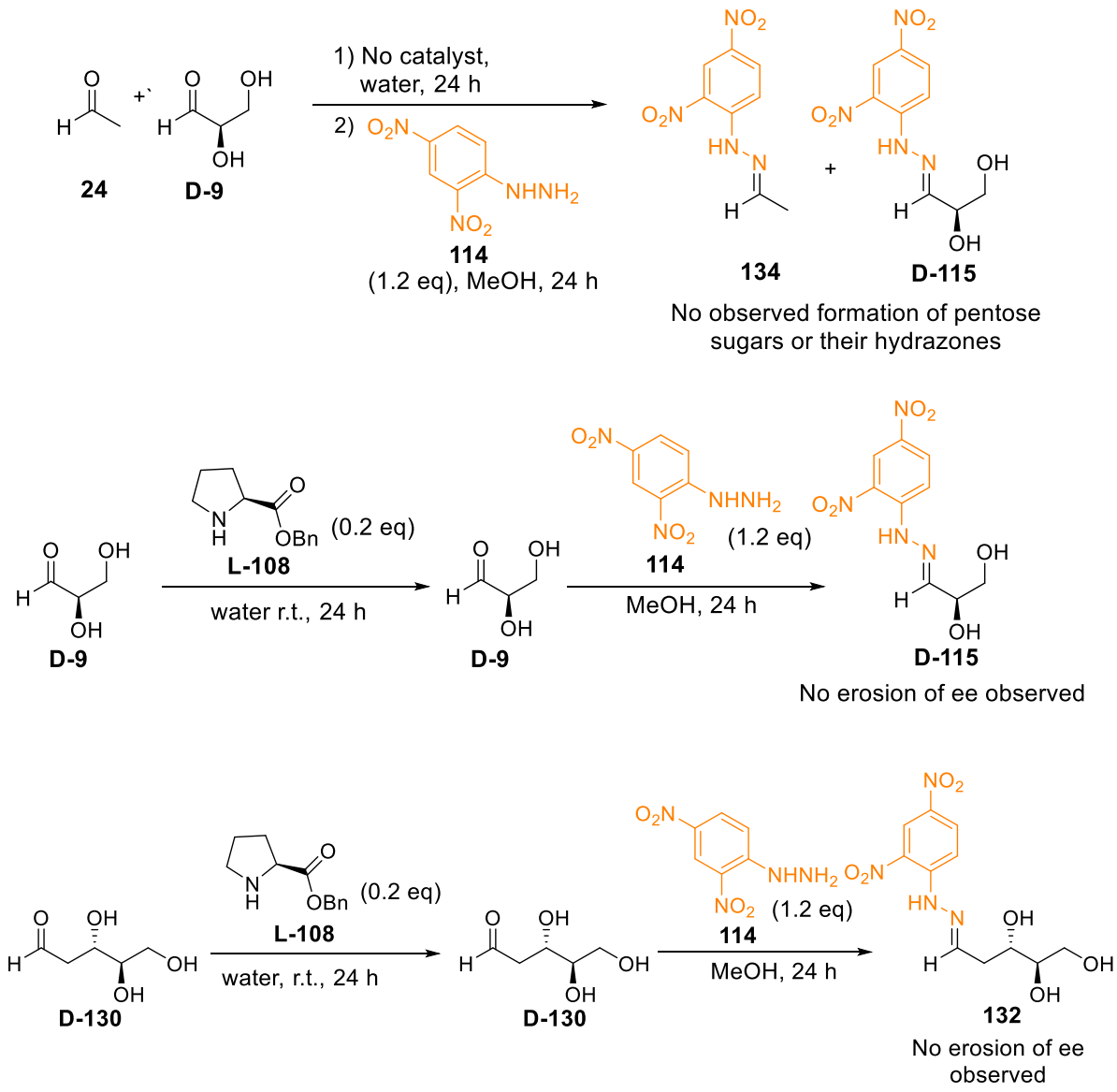
Scheme 2.8. Synthesis of N-methyl-L-leucine ethyl ester **L-111**.

2. Results of the Initial reaction trials

Due to the nature of prebiotic chemistry no one truly knows the conditions or molecules around on the early Earth, however, the building blocks available would have been in low concentrations in an aqueous (water) medium. To make the study as robust as possible challenging criteria were imposed to attempt to reflect the conditions on the early Earth. Firstly, stoichiometries of acetaldehyde **24** (donor) and glyceraldehyde **9** (acceptor) were to be as close to 1:1 as possible to remove any dependence on a single starting material. Secondly, the amino ester compounds were to be sub-stoichiometric. To show the reaction is not dependent on an abundance of amino ester a maximum of 20 mol % promoter was decided. Water or aqueous buffer was to be used as the solvent as this would likely have

been the solvent on the early Earth. Ideally a reaction time period of 24 hours was set, as if products are detectable after just 1 day then one would expect a sufficiently efficient process, over a much longer time period giving validity to the reaction in a prebiotic context. Finally, as previously described, the target compounds were to be isolated and reported with yields, rather than just being detected analytically to show a tangible amount of product had been formed.

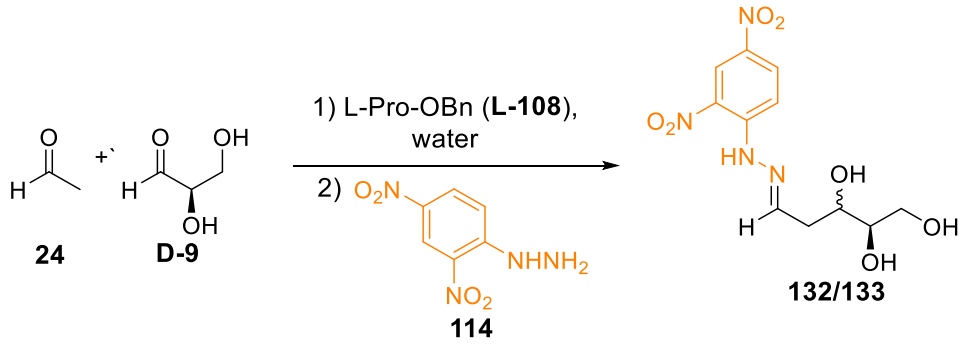
Before embarking on the study a series of control reactions were carried out to establish that any sugar synthesis was indeed a result of amino ester promotion. Firstly acetaldehyde **24** and D-glyceraldehyde **D-9** were stirred together for 24 hours without a promoter to determine whether there was any initial background reaction. The reaction was stirred for 24 hours in water then 2,4-dinitrophenyl hydrazine **114** was added and stirred for a further 24 hours. Only the hydrazones of the starting materials, **134** and **D-115**, were detected by TLC and ^{1}H NMR spectroscopy. There was no detectable product formation and therefore it was concluded that the reaction does not occur in the absence of a promoter (**Scheme 2.9**). In order to determine whether the reaction conditions were compromising the stereochemical integrity, both the glyceraldehyde **D-9** starting material and the sugar product, 2-deoxy-D-ribose **D-130**, were stirred under the reaction conditions for 24 hours before the solvent was evaporated and the trapping conditions applied. HPLC analysis showed no erosion of stereochemistry to either starting material or product.



Scheme 2.9. Control reactions carried out on starting material and prospective products.

Having completed the control reactions and determined the reaction criteria the investigation into the formation of 2-deoxy-D-ribose **D-130** was conducted. Initially 1 equivalent (1 mmol) of reactants **24** and **D-9** were dissolved in water (3 mL) and L-proline benzyl ester **L-108** (0.2 equivalents) was added. The reaction was stirred for 24 hours before the solvent was removed *in vacuo*. The crude sugar mixture was redissolved in methanol (3 mL) and 2,4-dinitrophenyl hydrazine **114** (3 equivalents) was added. After a further 24 hours the solvent was removed *in vacuo* and the detection of hydrazone products

D-132 and **D-133** was attempted through ESI mass spectrometry and TLC analysis by comparison of the crude mixture with the pre-prepared standards. ESI mass spectrometry was preferred to NMR spectroscopy due to its greater sensitivity. The initial results are shown in **Table 2.2**.



Entry	24 (equiv.)	D-9 (equiv.)	L-108 (equiv.)	Time (h)	Trap (equiv.)	Trap time (h)	Detection by MS ESI
1	1	1	0.2	24	2	24	No
2	1	1	1.0	24	2	24	No
3	2	1	0.2	24	2	24	No
4	20	1	0.2	24	21	24	No
5	1	1	0.2	72	2	24	No
6	2	1	0.2	24	3	24	No
7	1	1	0.2	72	2	72	Yes
8	1	1	0.2	24	2	72	Yes

Table 2.2. Initial experiments on the synthesis of trapped 2-deoxy-D-ribose **D-132**.

Table 2.2 Entry 1 shows that no product was detected from the initial run, instead only trapped starting materials **24** and **D-115** and dimerised acetaldehyde were observed. When the number of equivalents of acetaldehyde **24** were increased to 2 and then to 20 (Entries 3 and 4) the deoxypentose products were still not observed. The length of time was also

increased to 72 hours and no products were detected (Entry 5). However, increasing the duration of the trapping reaction did seem to have an effect (Entries 7 and 8). When the trapping time length was tripled from 24 to 72 hours the correct mass for 2-deoxy-D-ribose **132**/2-deoxy-D-threopentose **133** was detected in the ESI mass spectrum (**Figure 2.3**). Unfortunately the product could not be detected by ^{1}H NMR or HPLC and hence could not be isolated.

The main aims of the experiment were to obtain an isolated yield of 2-deoxy-D-ribose **D-130** and 2-deoxy-D-threopentose **D-131** and also to determine the diastereomeric ratio of **D-130** and **D-131**. To achieve these aims it was evident that the current reaction had to be modified. The trapping stage was identified as the weakest part of the two-steps. Trapping authentic 2-deoxy-D-ribose **D-131** with 2,4-dinitrophenyl hydrazine **114** only gave an 8 % yield of the hydrazone product. The aldol reaction itself may not be particularly high yielding due to the reasons discussed in Chapter 5, therefore it was imperative to have an efficient trap to maximise the yield of hydrazone products obtained to allow a better opportunity to isolate and analyse these compounds.

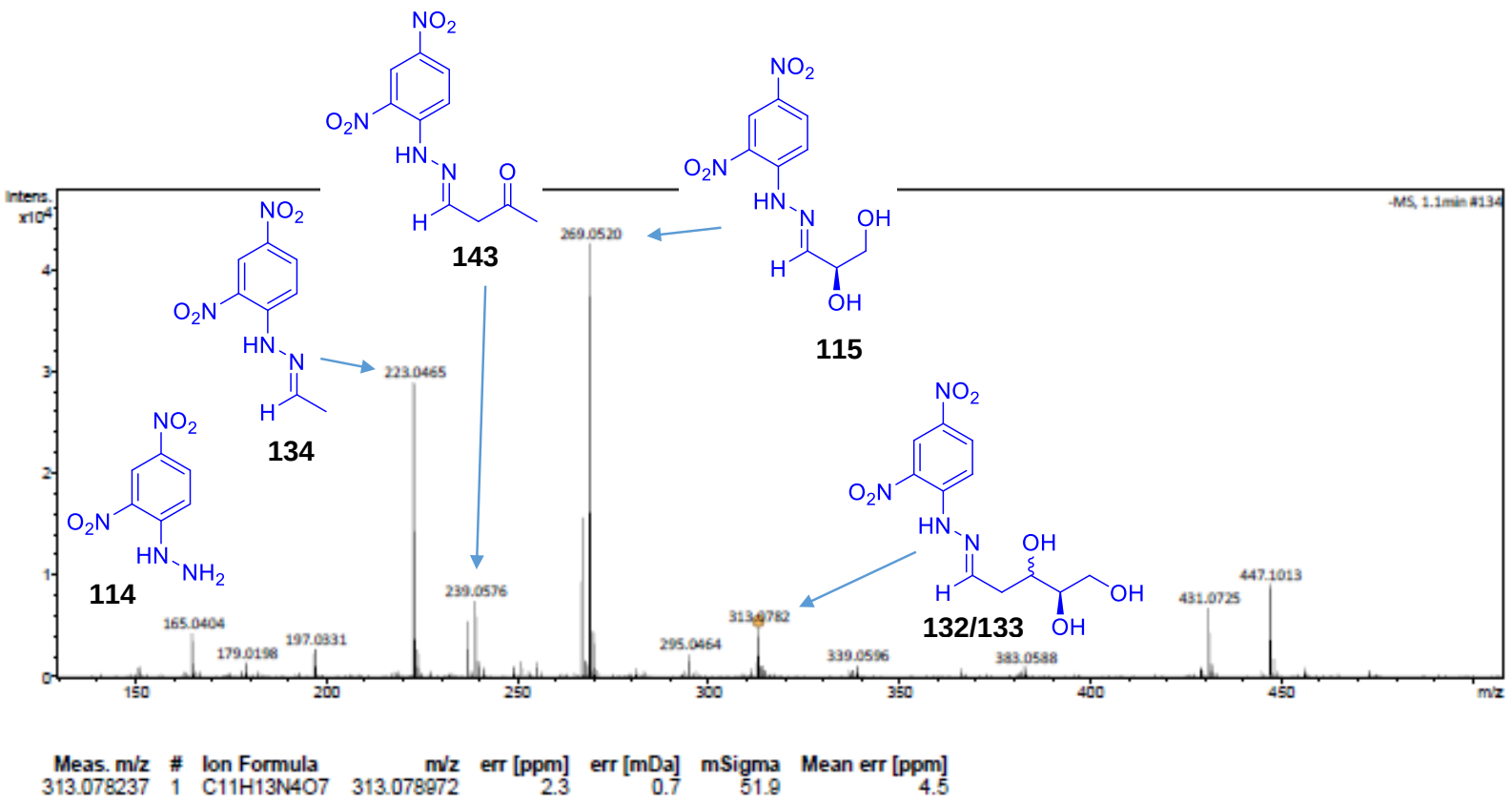
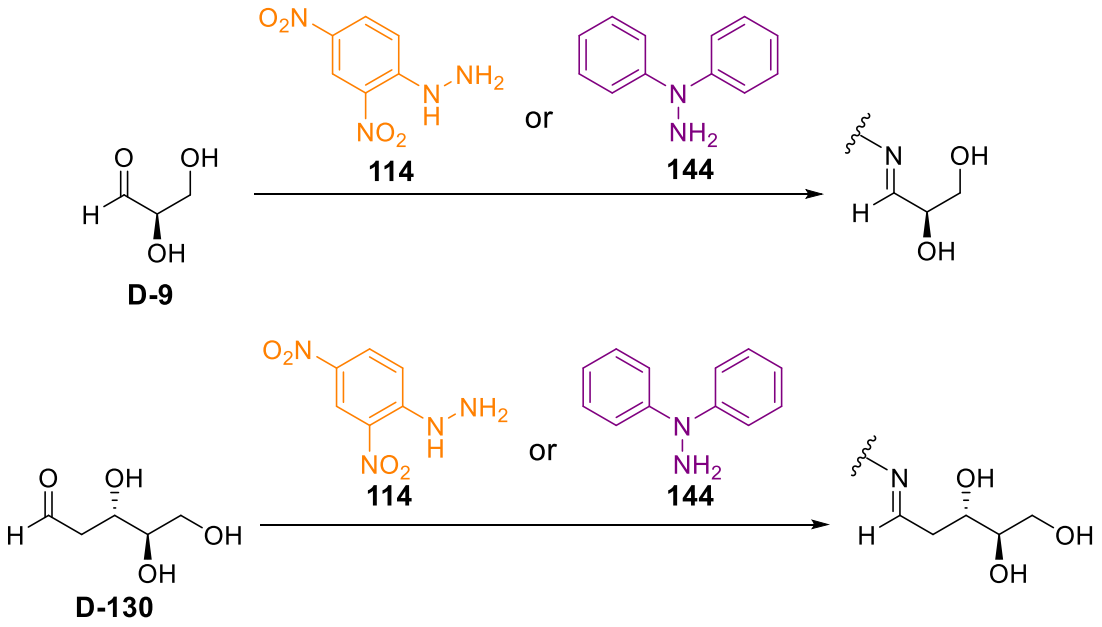


Figure 2.3. ESI mass spectrum for the 2-deoxyribose-forming reaction after 72 hours of trapping. The correct mass of trapped 2-deoxyribose **132** is present.

3. *N, N*-Diphenyl hydrazine trap

In order for the reaction to be successfully analysed an efficient trap needed to be developed to ensure as close to 100 % of the sugar molecules can be accounted for. The trapping ability of 2,4 dinitrophenyl hydrazine **114** was tested against other hydrazines and a comparison of **114** with the best candidate, *N,N*-diphenyl hydrazine **144**, as a trap for 2-deoxy-D-ribose **D-130** and starting material D-glyceraldehyde **D-9**, are shown in **Table 2.3**. Hydrazine **144** was found to be a more efficient trapping agent. This may be due to the reduced solubility of **114** in methanol or the addition of the electron withdrawing groups on **114** reducing the nucleophilicity of the hydrazine nitrogen. Based on this study hydrazine **144** was chosen as the replacement trapping agent of choice for the study.



Entry	Starting Material	Mass of SM (mg)	Trap	Solvent	Time (h)	Isolated Yield (%)
1	**D-9**	50	**114**	MeOH	20	45
2	**D-9**	50	**144**	MeOH AcOH (cat.)	1	91
3	**D-9**	200	**144**	MeOH	1	70
4	**130**	50	**114**	MeOH, AcOH (cat).	72	8
5	**130**	50	**144**	MeOH, AcOH (cat.)	1	98
6	**130**	29	**144**	MeOH, AcOH (cat.)	1	84

Table 2.3. Comparison of the efficiency of 2,4-dinitrophenylhydrazine **114** and diphenyl hydrazine **144** as trapping agents.

With an improved trapping candidate to hand, new trapped hydrazone standards of starting materials and products were synthesised for the purpose of analysis. The trapped starting materials and products, in hydrazone form, are shown in **Figure 2.4** along with percentage isolated yields obtained from the reaction conditions in **Table 2.3**. *N,N*-Diphenyl hydrazine **144** is commercially available as the HCl salt and was first neutralised before the trapping reactions were attempted. The salt was washed with saturated sodium bicarbonate solution and extracted with DCM which gave the free hydrazine as a deep purple oil.

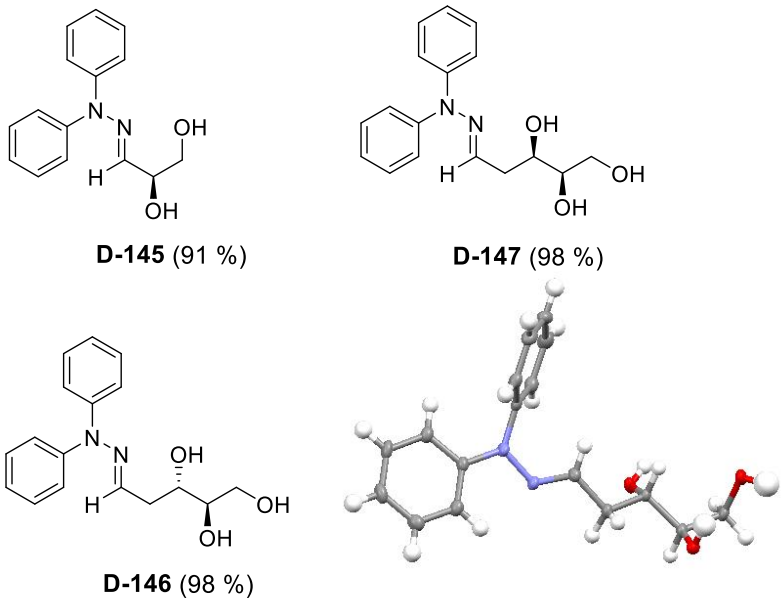


Figure 2.4. Trapped hydrazone standards and their yields using N,N-diphenyl hydrazine **144** after stirring for 1 hour in methanol and catalytic acetic acid. The crystal structure of **146** is also shown as 50 % ellipsoid.

With all standards at hand a method of product analysis needed to be identified. Unfortunately the two diastereomers, **D-146** and **D-147**, had identical R_F values in a range of different solvent systems therefore isolation of each diastereomer was not possible. However, analysis of the two standards on a 500 MHz NMR spectrometer showed differences in the ^{1}H NMR spectra of the two hydrazone products. **Figure 2.5** shows a 500 MHz ^{1}H NMR spectrum of a mixture of 2-deoxy-D-ribose and 2-deoxy-D-threopentose trapped as diphenyl hydrazones (**146** and **147**). **Figure 2.6** and **Figure 2.7** show the separate ^{1}H NMR spectra of 2-deoxy-D-ribose hydrazone **D-146** and 2-deoxy-D-threopentose hydrazone **D-147** respectively. An expanded image of the region of 6.61-6.55 ppm is also shown (inset). The 2 peaks in this region correspond to the azomethine peaks of **D-146** and **D-147** respectively, as highlighted using blue circles on the spectrum. As

these two triplet signals were resolved the diastereoselectivity of the reaction could be calculated based on the integration of the two signals.

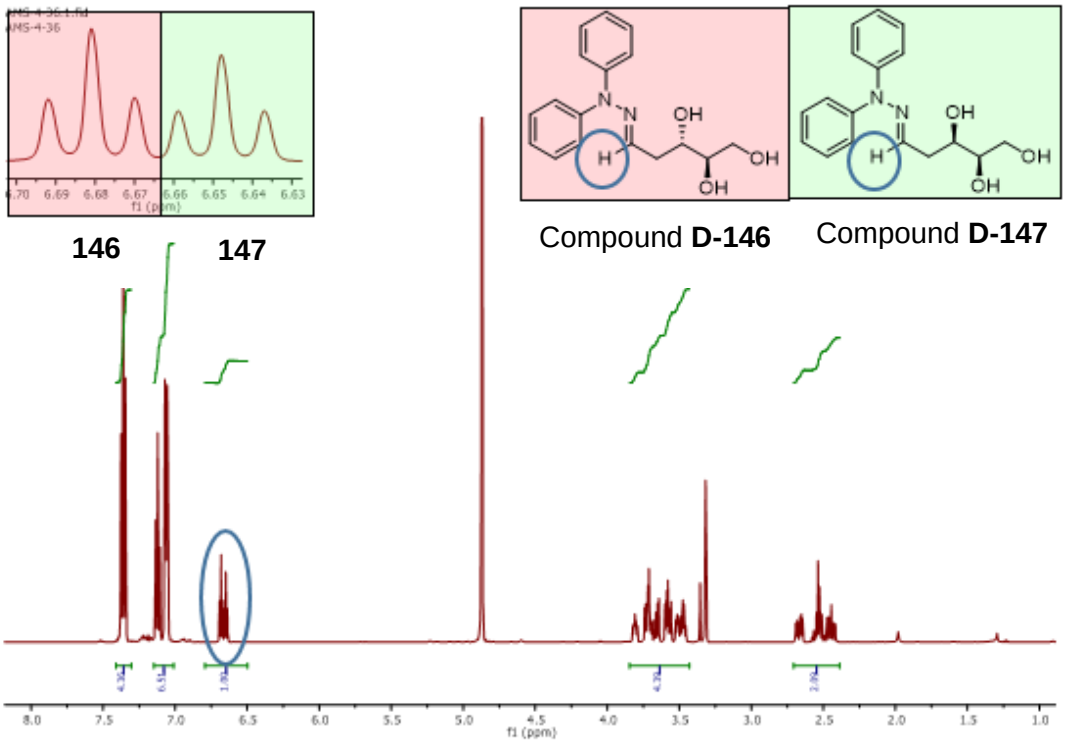


Figure 2.5. 500 MHz ^{1}H NMR spectrum of a mixture of hydrazones **D-146** and **D-147** in methanol d^{4}**.**

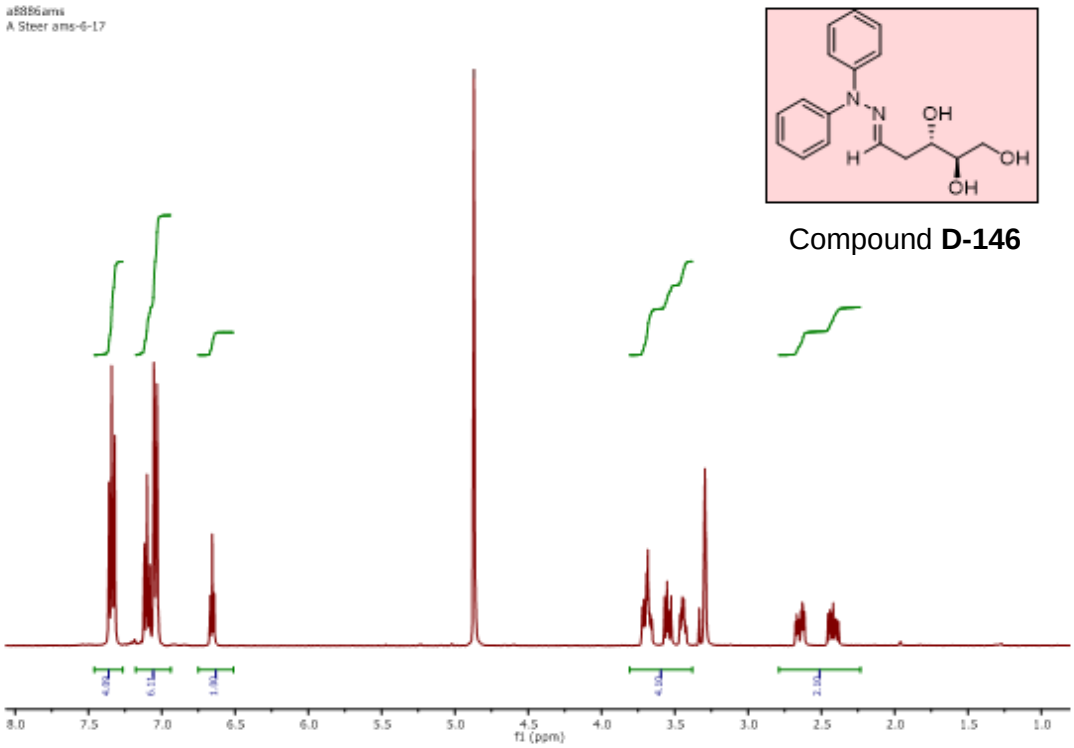


Figure 2.6. 500 MHz ^{1}H NMR spectrum of 2-deoxy-D-ribose hydrazone **D-146** in methanol d^{4}.

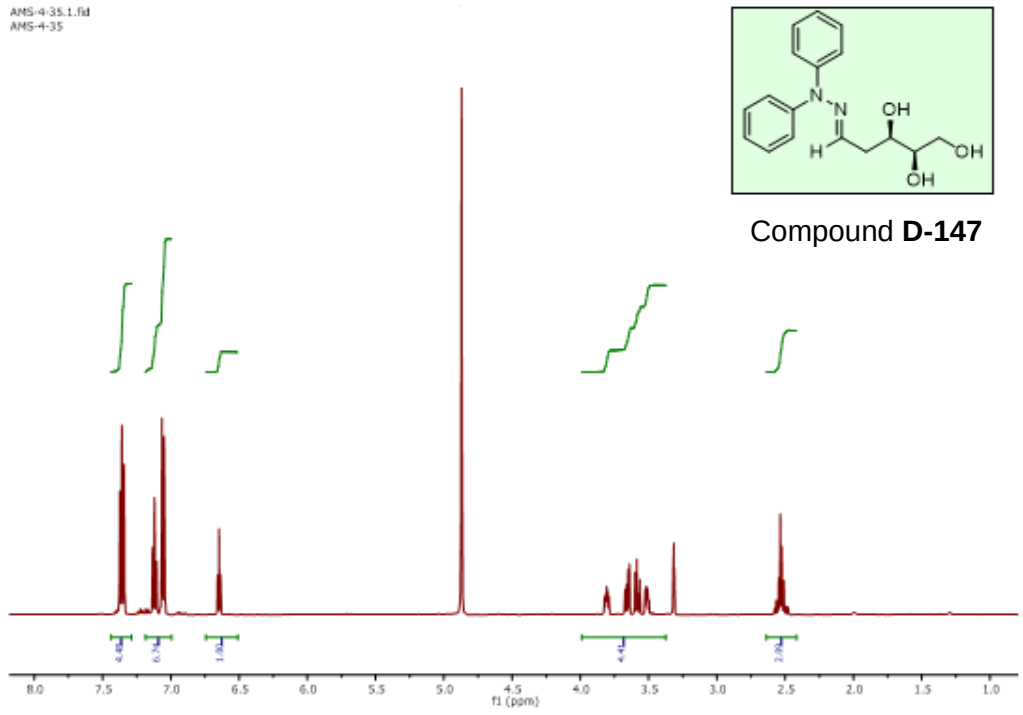
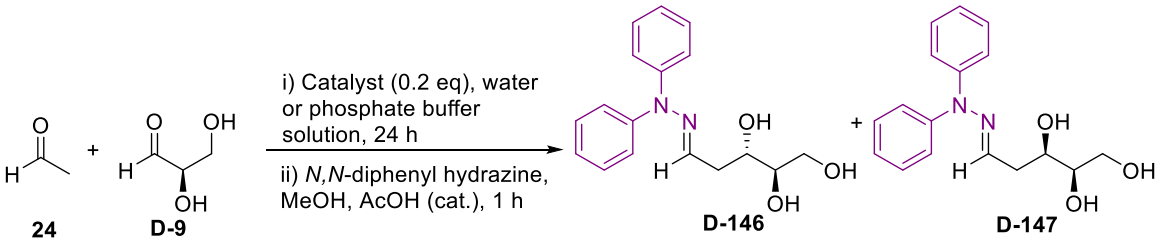


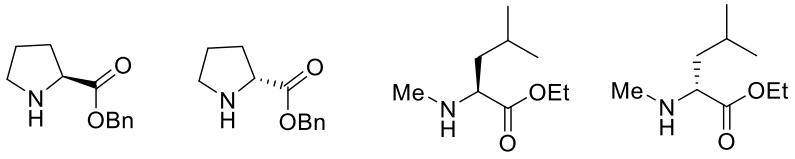
Figure 2.7. 500 MHz ^{1}H NMR spectrum of 2-deoxy-D-threopentose hydrazone **D-147** in methanol d^{4}.

The 2-deoxy-D-ribose **D-130** forming reaction was attempt using the modified trapping conditions. As previously stated by Clarke *et al.* in unbuffered water the enamine catalysed aldol reaction could be competing with general acid or general base promoted reactions which can lead to a reduction of enantioselectvity.[81] To gain a true understanding of the diastereoselectivity arising from the enamine promoted aldol reaction the 2-deoxy-D-ribose **D-130** forming reaction was attempted at pH 7 (which should reduce competing general acid or general base catalysis) and pH 6 (which should have reduce competing general base catalysis) using phosphate buffers. Background reactions were again carried out without a catalyst present at the various pHs and no product formation was observed. Separately, starting material and authentic products were submitted to the reaction conditions and no erosion of enantio-integrity was observed.

One equivalent of each starting material (**24** and **D-9**) with 20 mol % catalyst (**108** or **111**) in an aqueous medium was stirred at room temperature for 24 hours. The solvent was then removed *in vacuo* and the new trapping conditions applied. After 1 hour the reaction mixture was concentrated *in vacuo* and the crude mixture of products was subjected to flash column chromatography (5:95 methanol:DCM) followed by preparative thin layer chromatography (10:90 ethyl acetate:hexane) which isolated hydrazone products **D-146** and **D-147** as a mixture of diastereomers. The results of the reaction using both enantiomers of catalysts **108** and **111** are shown in **Table 2.4** below. The yields and diastereoselectivities of each entry were based on three runs of each experiment and an average of the three was taken.



Catalysts:



Entry	Catalyst	pH	Isolated Yield (%)	Ratio (113:114)
1	None	7	0	N/A
2	**L-111**	7	2	2.0 : 1
3	**L-111**	6	1	1.6 : 1
4	**L-111**	Unbuffered	0	N/A
5	**D-111**	7	2	1.6 : 1
6	**L-108**	7	2	1.8 : 1
7	**L-108**	6	2	1.7 : 1
8	**L-108**	Unbuffered	2	1.5 : 1
9	**D-108**	7	2	1.2 : 1

Table 2.4. Results of the 2-deoxy-D-ribose **D-130** forming reaction. Each of the entries was based on three runs at the same pH (either unbuffered water, pH 6 phosphate buffer or pH 7 phosphate buffer) and the yield and d.r. was the average of the three. The dr was determined using 500 MHz ^{1}H NMR spectroscopy.

From **Table 2.4** it is clear that there was some selectivity for 2-deoxy-D-ribose **146** over 2-deoxy-D-threopentose **147**, albeit quite small. The pH of the reaction did not seem to have an effect on the selectivity or the yield of the reaction. The isolated yields were between (1 and 2 %). This is the first time isolated yields have been reported in this field for carbohydrate formation of 2-deoxy-D-ribose **D-130**. The low yields obtained, and the relatively low drs, could be attributed to the loss of the key hydrogen-bonded chair-like transition state. As described in Chapter 5, this key intermediate transition state is thought to occur in enamine catalysed aldol reactions in organic solvents, however, in aqueous solvents the electrophile and enamine intermediate will most likely form hydrogen bonds with the water solvent instead.

Changing the chirality of the catalyst (compare Entries **6** and **9**) reduced the diastereoselctivity but still showed some selectivity for the formation of 2-deoxy-D-ribose **D-130** over 2-deoxy-D-threopentose **D-131**. It can be implied from this observation that the reaction was controlled to some degree by the catalyst as reversing the chirality of the catalyst reduced the dr of the products. However, the chirality of the starting material, D-glyceraldehyde **D-9**, had a larger impact as aldol reactions tend to favour *anti* products.

An example ^{1}H NMR spectrum of the reaction ran in pH 7 buffer after purification is shown in **Figure 2.8**. This demonstrates that the deoxypentose products formed in the reaction were isolated and purified to a high degree.

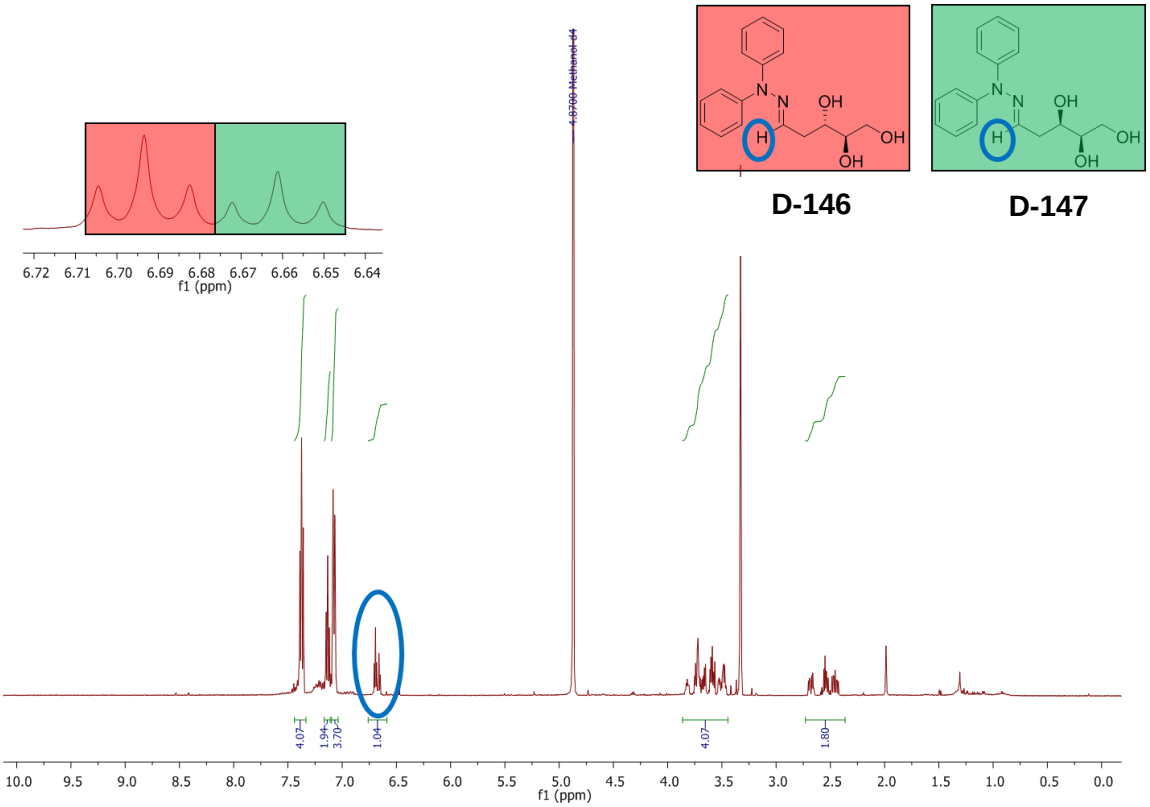


Figure 2.8. An authentic ^{1}H NMR spectrum of the reaction assay after column chromatography using the N-methyl-D-leucine ethyl ester catalyst **D-111**.

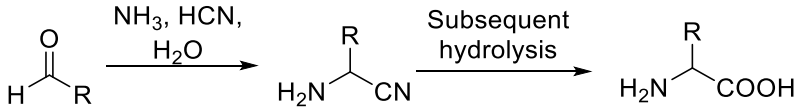
4. Conclusion

This chapter has shown a refined study into the potential prebiotic synthesis of 2-deoxy-D-ribose **130**, the sugar of DNA, from two smaller prebiotic building blocks acetaldehyde **24** and D-glyceraldehyde **D-9**. Diastereoselectivities of 1.5:1 - 2:1 (2-deoxy-D-ribose **146** : 2-deoxy-D-threopentose **147**) with yields of 1 - 2 % were observed using amino esters as catalsyst. The study suggests that the stereochemistry of the amino ester catalysts have limited effect on the chirality of the newly formed stereocentre and the stereoselectivity is primarily due to the chirality of the glyceraldehyde **9** starting material. The low yields and relatively low drs of the reaction are typical for aldol reactions in aqueous solvents as shown in the literature examples presented in Chapter 5 Sections 5.5 and 5.7.

3. Amino nitriles – possible progenitors to amino acids

3.1. Evidence for amino nitriles as prebiotic molecules

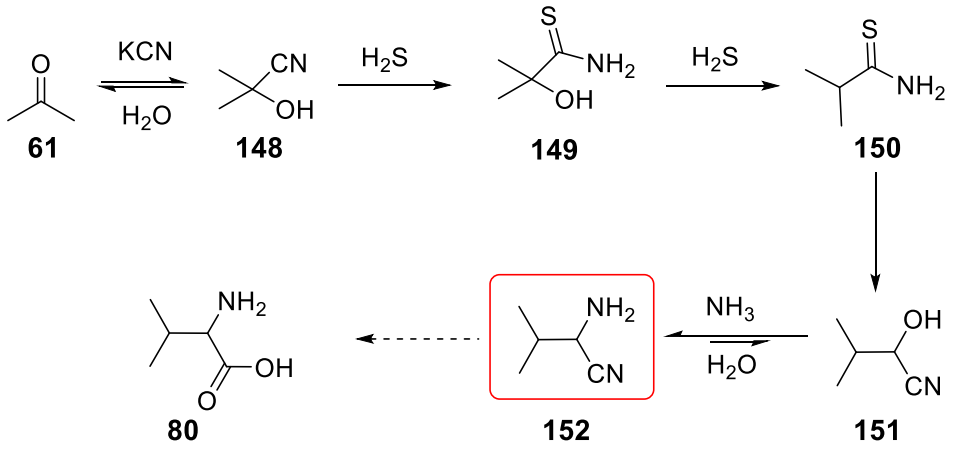
A route to 2-deoxy-D-ribose **D-130** *via* an aldol reaction using two smaller sugar building blocks, acetaldehyde **24** and D-glyceraldehyde **D-9**, catalysed by amino esters has been demonstrated in the previous chapter. The next aim of the research was to make this reaction as prebiotically plausible as possible. The prebiotic nature of amino esters is debatable. To develop this study further more plausibly prebiotic catalysts were investigated. Chapter **2.3** outlined several possible origins of amino acids. One of these theories involved the formation of non-terrestrial amino acids through Strecker reactions in space. The reaction of an aldehyde, ammonia and HCN could form an amino nitrile and subsequent hydrolysis of the nitrile to forms amino acids (**Scheme 3.1**). [47,48]



Scheme 3.1. General Strecker synthesis of an amino nitrile and subsequent hydrolysis to an amino acid.

Another possible prebiotic route to amino nitrile compounds comes from Sutherland and co-workers' prebiotic systems chemistry which explored the origins of a range of biomolecular building blocks including ribonucleotides, lipids and amino acids in reducing environments driven by hydrogen sulphide and the use of copper photoredox cycling.[40,102] In this pathway HCN was added to acetone **61** generating an equilibrium with cyanohydrin **148** (**Scheme 3.2**). Addition of hydrogen sulfide to the reaction mixture gave α-hydroxythioamide **149** *via* nucleophilic attack of HS^- which also pushed the equilibrium towards **148**. Irradiation of **149** caused α-dehydroxylation to give **150** where upon reduction to the aldehyde in the presence

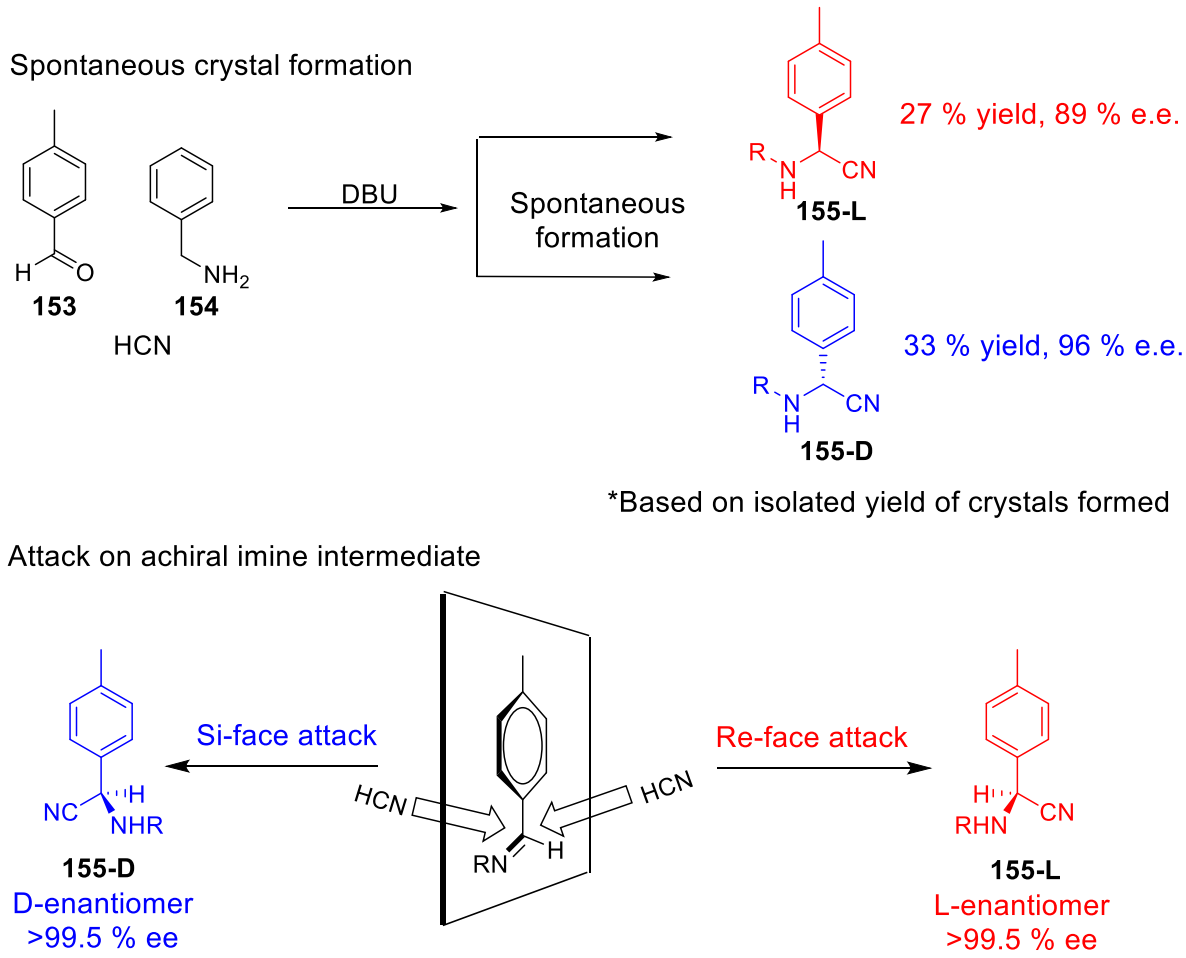
of HCN gave cyanohydrin **151**. Compound **151**, in the presence of ammonia, was found to form amino nitrile **152** which upon hydrolysis of the nitrile would form valine **80**.[40]



Scheme 3.2. Prebiotic route to racemic valine **80** from acetone using Sutherland's systems chemistry.[40]

An issue with both routes to amino nitrile formation is that they are both racemic and of course the naturally occurring amino acids all possess L-stereochemistry. A recent publication by Kawasaki and co-workers addressed this problem through spontaneous crystallisation. By performing a Strecker reaction with the achiral substrate *p*-tolualdehyde **153**, benzhydrylamine **154** and HCN in the presence of DBU led to the spontaneous crystallisation of amino nitriles **155** in up to 96 % ee.[106] In this paper the group also showed how amplification of the crystals in the solid state could occur through processes such as attrition-enhanced ripening and Viedma ripening. Attrition-enhanced ripening is the process whereby a small enantiomeric excess of a crystal solution can lead to enantiopure crystals, as the enantiomeric crystal grows the imbalance in the racemic solution is addressed by converting to more of the depleted enantiomer.[107,108,109] Viedma Ripening involves a near racemic slurry of crystals in solution whereby the crystals are given enough energy through continuous grinding to spontaneously form one enantiomer.[110,111] Furthermore, the group later showed addition of HCN to the intermediate imine crystal face led to the asymmetric

synthesis of amino nitriles in greater than 99.5 % ee.[112] The amino nitriles were then hydrolysed to the amino acids.



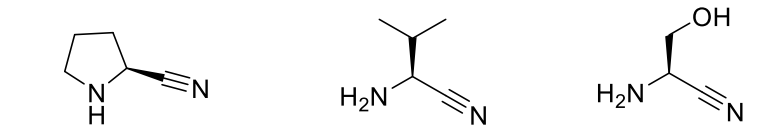
Scheme 3.3. Formation of chiral amino nitriles via Strecker synthesis by Kawasaki et al.[106]

It seems plausible that amino nitrile molecules are possible progenitors of amino acids. These molecules are realistically “more prebiotic” than amino esters and with some evidence of prebiotic asymmetric synthesis of amino nitriles the decision to test chiral amino nitriles as potential catalysts in the 2-deoxy-D-ribose **D-130** forming reaction was taken. A

detailed preliminary literature search found no evidence of amino nitriles as catalysts in any type of organic reaction.

3.2. Synthesis of amino nitriles

Three amino nitriles were synthesised; L-proline nitrile **156**, L-valine nitrile **157** and L-serine nitrile **158** (**Figure 3.1**). L-Proline nitrile **156** was chosen as this is the only secondary amine, of the amino acids, which may be more efficient at catalysing the enamine formation with acetaldehyde over primary amines. L-Valine nitrile **157** was chosen as an aliphatic comparison and L-serine nitrile **158** had the added importance as a common and imperative residue in many enzyme active sites.[113]

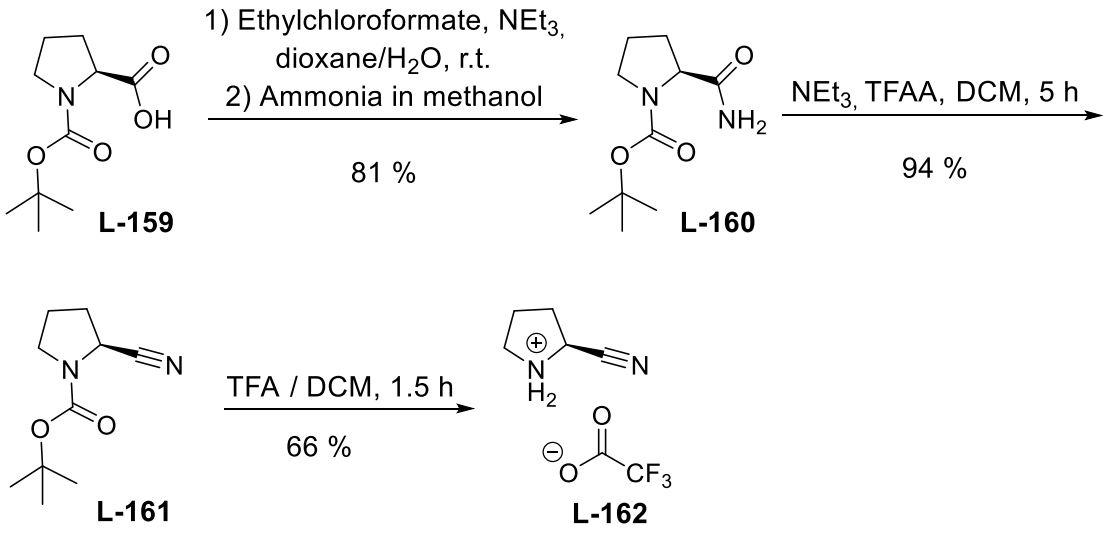


L-156 L-Proline nitrile | **L-157** L-Valine nitrile | **L-158** L-Serine nitrile

Figure 3.1. Amino nitrile targets to be synthesised.

The synthesis of L-proline nitrile **156** and the subsequent use of this compound as a catalyst in the 2-deoxy-D-ribose **D-130** forming reaction were carried out by Nicolas Bia, an Erasmus student, during a three month placement in the Clarke group under the supervision of Andrew Steer. The synthesis of **L-156** began with *N*-Boc-L-proline **L-159** and was readily converted to amide **L-160** in a good 81 % yield by using ethylchloroformate to form a mixed anhydride followed by addition of ammonia (**Scheme 3.4**). Conversion of amide **L-160** to *N*-Boc-L-proline nitrile **L-161** was possible with triflouroacetic anhydride (TFAA) in an excellent 94 % yield. A common problem with this reaction was the condition of the TFAA.

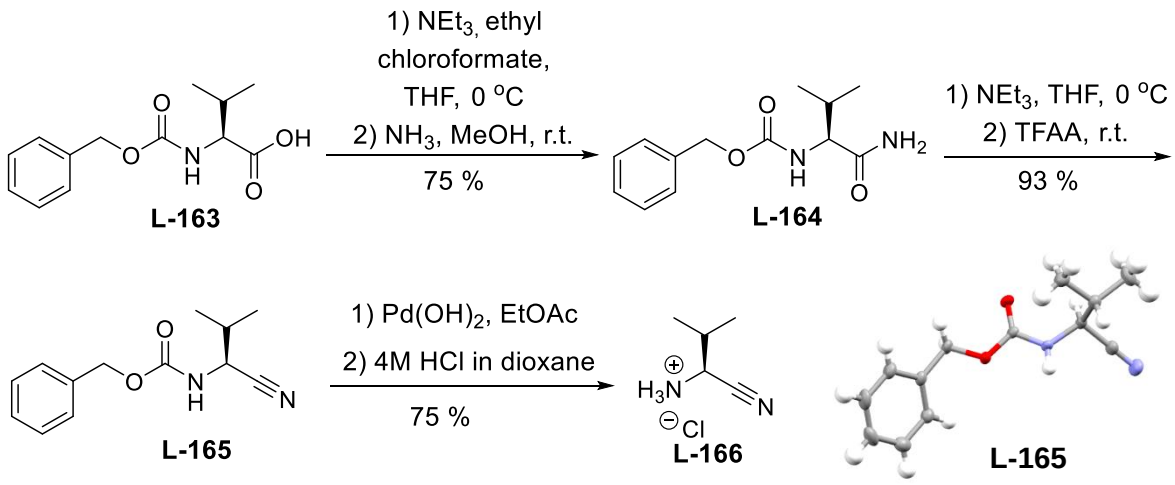
Over time TFAA gradually breaks down to form trifluoroacetic acid and hence fresh reagent was used to ensure a high yield was obtained. The final step of the reaction was Boc deprotection which was carried out using TFA. Upon evaporation of solvent the nitrile product was isolated as the TFA salt **L-162** in a 66 % yield. Unfortunately it was difficult to isolate the free amino nitrile. Washing with saturated sodium bicarbonate and extraction with dichloromethane gave a low (< 50 %) return of the neutral amino nitrile as **L-156** was very water soluble. However, dissolving **L-162** in DCM and stirring over solid sodium bicarbonate did increase the yield of free amino nitrile **L-156** to 85 %.



Scheme 3.4. Synthesis of L-proline amino nitrile **L-156**.

The synthesis of L-valine nitrile **L-157** was next attempted. The protecting group strategy was changed from Boc to carboxybenzyl (Cbz). Cbz-L-Valine **L-163** or (*Z*-L-valine) was commercially available and had the added benefit of obtaining the neutral amine by hydrogenation of the Cbz group in the final step of the synthesis. The same route for the synthesis of L-proline nitrile **L-156** was applied to *Z*-L-valine **L-157** (**Scheme 3.5**). Amide formation proceeded in a good 75 % yield and subsequent dehydration to nitrile **L-165** proceeded in an excellent 93 % yield, after column chromatography. Nitrile **L-165** readily formed long colourless crystals allowing an x-ray crystal structure to be obtained. The X-

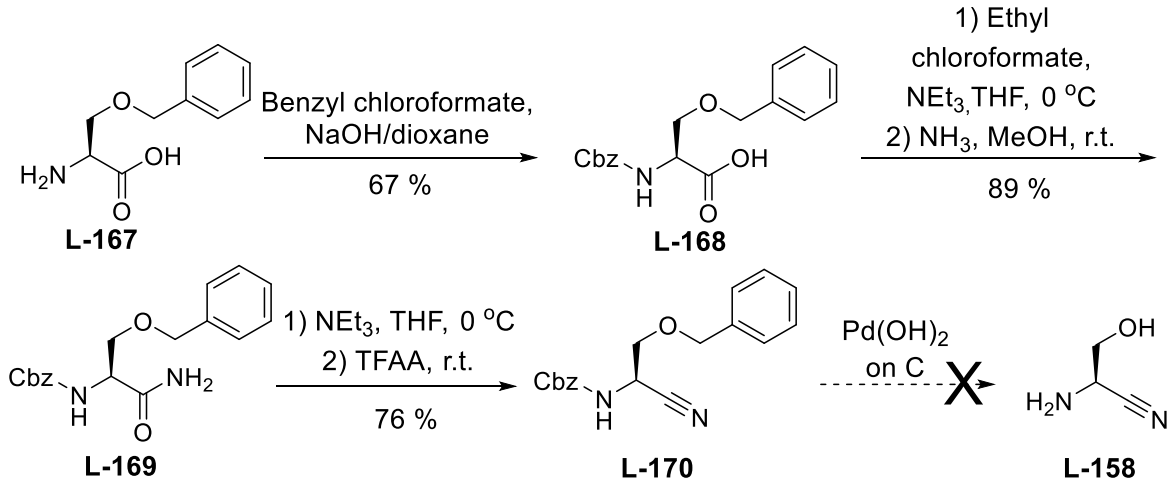
ray data and the polarimetry data (-37.3 (c = 0.97 g cm^{-3} in methanol) compare with literature -55 (c = 1.13 g cm^{-3} in chloroform))[114] confirmed that racemisation had not occurred to any great extent during amino nitrile formation. The final step of the synthesis, deprotection of the Cbz group, was conducted using Pearlman's reagent ($Pd(OH)_2$ on carbon). The reaction proceeded cleanly and quickly to give deprotected amino nitrile **L-166**. Unfortunately, due to the low boiling point of **L-166** (70 °C at 0.5 Torr)[115] isolation by evaporation of the solvent proved difficult. Alternatively a 4M solution of HCl in dioxane was added to the amide following filtration of the reaction mixture. This caused amine salt **L-166** to precipitate out of solution and a 75 % isolated yield was achieved following filtration.



Scheme 3.5. Synthesis of L-valine nitrile **L-157**. The X-ray crystal structure of **L-165** as 50 % ellipsoid is also shown.

Potentially the synthesis of L-serine nitrile **L-158** could be achieved through a similar procedure to L-valine nitrile **L-157**. The initial idea was to protect the alcohol of the serine side chain with a benzyl group to allow a double deprotection in the final step of the reaction to give the free nitrile. This initial reaction scheme is shown in **Scheme 3.6**. The synthesis began with the Cbz protection of *O*-benzyl-L-serine **L-167** with benzyl chloroformate which produced the doubly protected amino acid **L-168** in a moderate 67 % yield. Standard amide formation and subsequent dehydration to nitrile **L-170** proceeded in good yields of 89 and

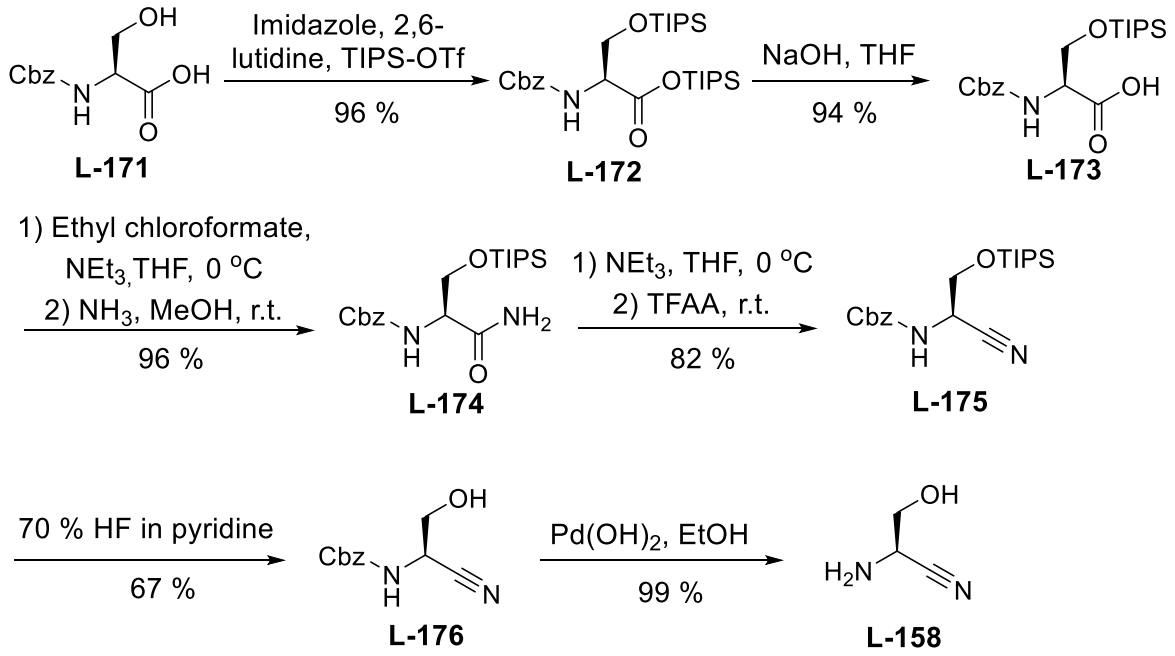
76 % respectively. The final double deprotection step was challenging. Nitrile **L-170** was stirred in a variety of solvents (such as ethyl acetate, isopropanol, ethanol, methanol) with Pearlman's reagent but only deprotection of the Cbz at the N-terminus occurred. The catalyst was changed to palladium on carbon but this still did not remove the benzyl protecting group on the alcohol, nor did changing the solvent. Longer exposure of **L-170** to palladium on carbon (> 24 h) resulted in decomposition of the compound indicated by a complex ^{1}H NMR spectrum and multiple spots on the TLC plate.



Scheme 3.6. Initial synthetic route to L-serine **L-158**. The final step of the reaction did not remove the benzyl protecting group.

It was clear that the synthetic strategy was flawed and alternative, longer, protecting group strategy was employed instead. The new route (**Scheme 3.7**) began from commercially available *N*-Z-L-serine **L-171**. The first step of the sequence was tri*iso*propyl silyl (TIPS) protection of the hydroxyl side chain. Compound **L-171** was dissolved in DMF and stirred with TIPS triflate, imidazole and 2,6-lutidine for 22 hours. This led to, not only, protection of the alcohol but also formation of the TIPS ester **L-172** in a very good 96 % yield. Hydrolysis of the TIPS-ester in 1M NaOH provided the free carboxylic acid **L-173** in an

excellent 94 % yield. Standard amide formation, *via* a two-step process of mixed anhydride formation and subsequent attack with ammonia, and dehydration, with TFAA, gave nitrile **L-175** in a 79 % yield (over 2 steps). The final steps involved the removal of the protecting groups. The TIPS group was removed using a commercially available 70% solution of hydrogen fluoride in pyridine which gave the free alcohol **L-176** in a moderate 67 % yield. The Cbz group was removed, as before, using Pearlman's reagent with full deprotection achieved after just 10 minutes. This gave amino nitrile **L-158** as the free amine in a 99 % yield.

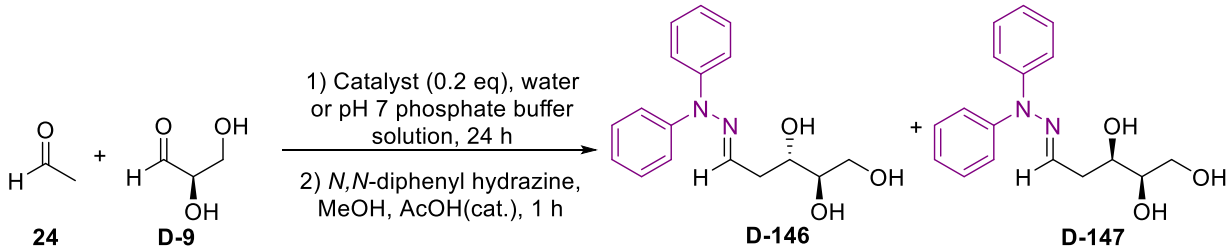


Scheme 3.7. Modified synthesis of L-serine amino nitrile **L-158**.

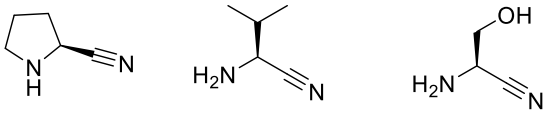
3.3. Amino nitriles as potential catalysts

With prebiotically plausible amino nitriles **L-156 –** L-**158** in hand the 2-deoxy-D-ribose **D-130** forming reaction was attempted using these molecules as the catalysts. All other variables remained the same i.e. acetaldehyde **24** (1 mmol) and D-glyceraldehyde **D-9** (1

mmol) were stirred in an aqueous solvent (3 mL) with an amino nitrile (0.2 mmol) for 24 hours before the solvent was removed *in vacuo*. *N*-*N*-Diphenyl hydrazine **144** (3 mmol) was then added in methanol (5 mL) and acetic acid (catalytic) for 1 hour before the solvent was removed *in vacuo* and crude material was purified by chromatography. **Table 3.1** shows the results of these reactions.



Catalysts:

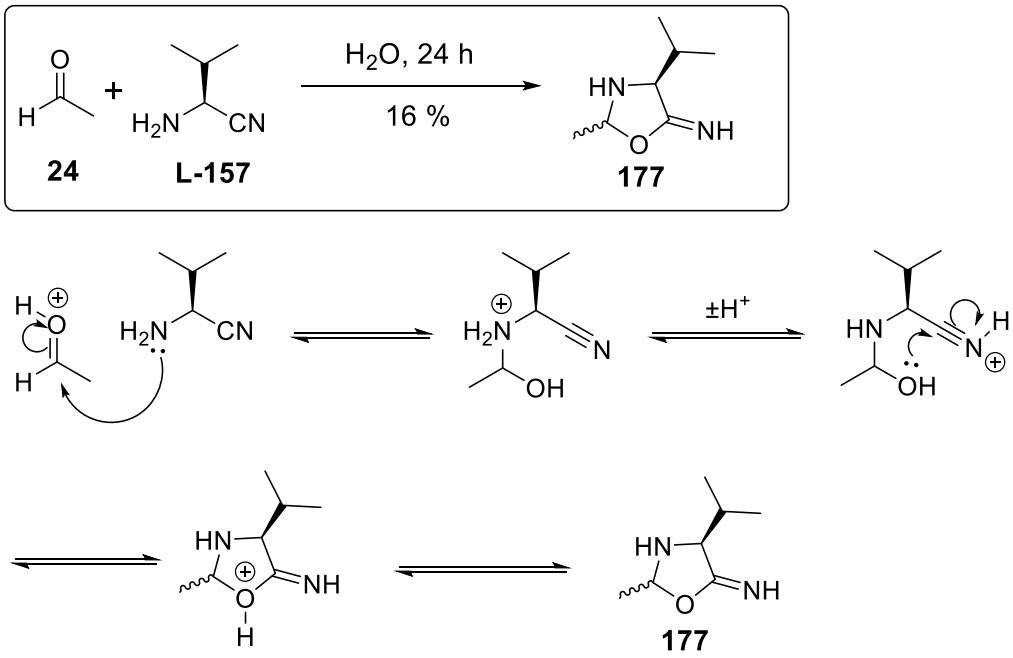


Entry	Catalyst	pH	Isolated Yield (%)	Ratio (146:147)
1	**156**	7	2	1.5 : 1
2	**156**	Unbuffered	0	-
3	**157**	7	5	1.7 : 1
4	**157**	Unbuffered	2	1.7 : 1
5	**158**	7	0	-
6	**158**	Unbuffered	0	-

Table 3.1. Results of the 2-deoxy-D-ribose **D-130** forming reaction. Each of the entries is based on three runs at the same pH and the dr is the average of the three. The dr was determined using 500 MHz ^{1}H NMR spectroscopy.

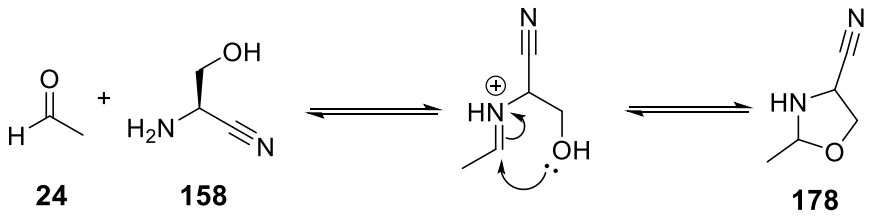
Table 3.1 shows amino nitriles **L-156** and **L-157** were successful at promoting the formation of 2-deoxy-D-ribose **D-130** but L-serine **L-158** was not. L-Proline nitrile **L-156** generated a ratio of products in very similar yield and ratio to L-proline benzyl ester **L-108**, however, it was unable to catalyse the reaction in unbuffered water (Entry 2) possibly due to protonation of the amine. L-valine nitrile **L-157** was the most successful catalyst and gave a slightly higher selectivity for 2-deoxy-D-ribose over 2-deoxy-D-threopentose and a much higher yield in pH 7 phosphate buffer. This yield of 5 % is the largest yield of all the nitriles and esters tested and shows that a primary amino nitrile is capable at catalysing this aldol reaction. However, L-serine nitrile **L-158** was not successful at catalysing the reaction in neither unbuffered water nor pH 7 phosphate buffer.

Further investigation of the amino nitriles were undertaken. A control experiment was conducted which involved stirring one equivalent of L-valine nitrile **L-157** with one equivalent of acetaldehyde **24** for 24 hours. Upon purification a cyclic by-product **177** (**Scheme 3.8**) was obtained in a 16 % yield along with unreacted amino nitrile **L-157**. This showed that during the 2-deoxy-D-ribose **D-130** forming reaction, the acetaldehyde **24** starting material could be sequestering the amino nitrile through the alternative mechanism shown in **Scheme 3.8**, reducing the yield of the reaction. Therefore it may not be correct to call the amino nitriles “catalysts” as not all of the nitrile is recovered at the end of the reaction, however, despite this they are as efficient if not more so at promoting the reaction than amino esters. For example amino nitrile **L-157** gave a 5 % yield of 2-deoxypentose products (Entry 3, **Table 3.1**) compared to amino ester **L-108** and **L-111** which gave yields of 2 % in pH 7 buffer (Entry 2 & 6, **Table 6.4**).



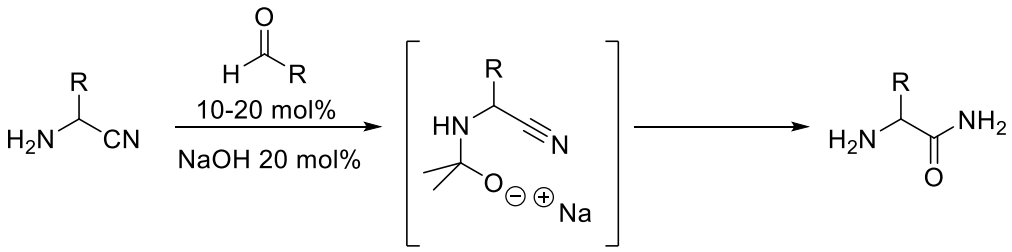
Scheme 3.8. Alternative reaction pathway for the reaction of acetaldehyde with L-valine nitrile **L-157**.

With the discovery of the amino nitrile by-product **177**, a possible explanation for the lack of product formation when using L-serine nitrile **158** could be explained in a similar way through further reactions upon the formation of the enamine intermediate (**Scheme 3.9**). Once L-serine **158** has reacted with acetaldehyde **24** to form an iminium ion, the internal attack of the serine oxygen could form a cyclic by-product **178** which maybe catalytically unreactive. This is purely speculative as there was not enough time to carry out this reaction in the laboratory.



Scheme 3.9. Possible alternate pathway to from a potential L-serine by-product L-**178**.

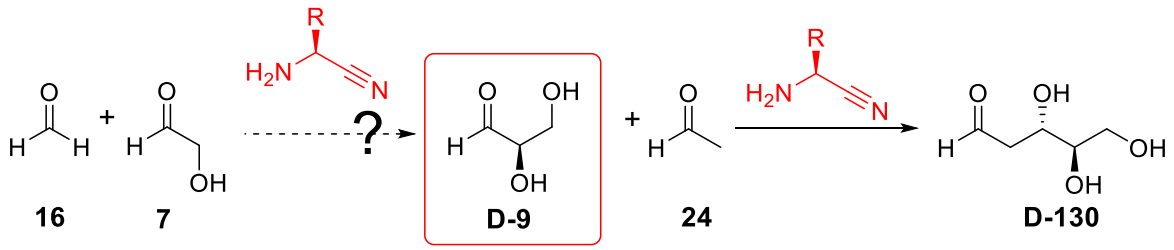
Another possible fate of the amino nitirles is hydrolysis to the amino amide and the amino acid as demonstarted by Beauchemin *et al* who showed this was possible with catalytic amounts of aldehyde in basic conditions including formaldehyde **16**, glycolaldehyde **7** and glycerlaldehyde **9** (**Scheme 3.10**).[116] This is further demonstarted in the recent work of Coggins and Powner with amino nitrile phosphates.[117]



Scheme 3.10. Hydrolysis of amino nitriles to amino amides using catalytic aldehydes.[116]

3.4. Prebiotic formation of glyceraldehyde

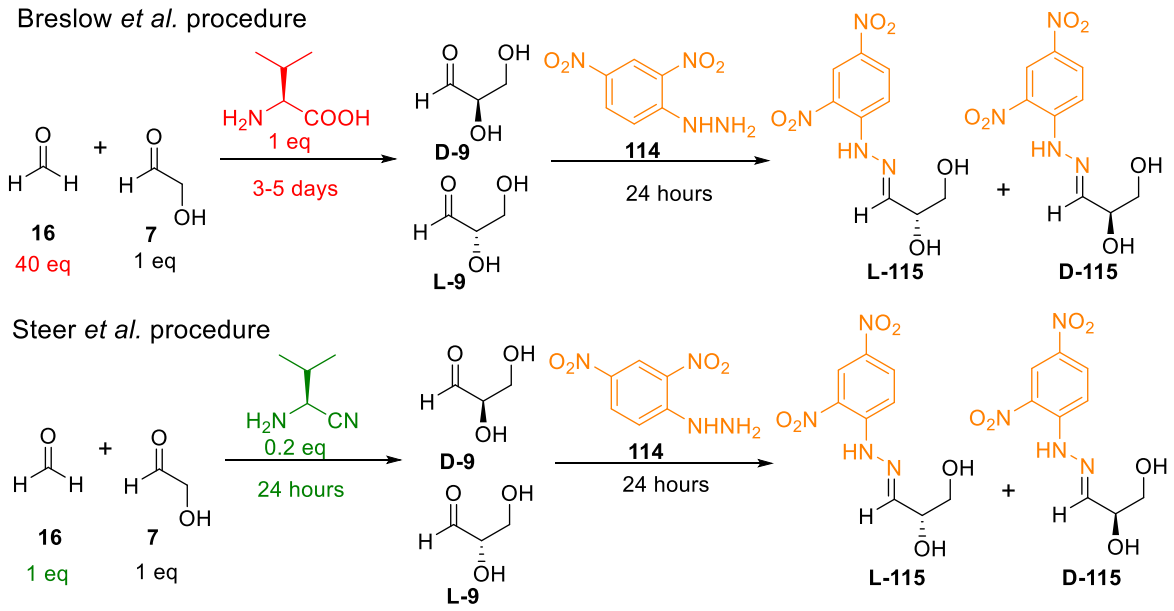
In an attempt to further the prebiotic validity of the 2-deoxy-D-ribose **D-130** forming reaction attention was turned to the starting materials; acetaldehyde **24** and D-glyceraldehyde **D-9**. Acetaldehyde **24** is known as an interstellar building block, however, D-glyceraldehyde **D-9** is not.[118] A further retro-analysis of glyceraldehyde **9** breaks the 3-carbon sugar down further into two carbon building blocks, formaldehyde **16** and glycolaldehyde **7**. An amino nitrile promoted aldol reaction of **16** and **7** to **D-9** would provide a two-step formal synthesis from glycolaldehyde **7** to 2-deoxy-D-ribose **D-130** (**Scheme 3.11**).



Scheme 3.11. Proposed prebiotically plausible synthetic pathway from glycolaldehyde **7** to 2-deoxy-D-ribose **D-130**.

The prebiotic formation of D-glyceraldehyde **D-9** from formaldehyde **16** and glycolaldehyde **7** has been attempted previously as discussed earlier in Chapter 5.[85,86,87] Breslow and Cheng have previously reported the aldol reaction of formaldehyde **16** and glycolaldehyde **7** in a prebiotic setting.[85] In their procedure the authors dissolved formaldehyde **16** (40 mmol), glycolaldehyde **7** (1 mmol) and an amino acid (1 mmol) in deionised water (20 mL). The solution was stirred for 3-5 days before 2,4-dinitrophenyl hydrazine **114** (41 mmol) was added. The mixture of L and D-glyceraldehyde **9** was trapped in hydrazone form and the enantioselectivity of the isolated products determined by HPLC.

Breslow's experiment was modified to include 1 mmol each of formaldehyde **16** and glycolaldehyde **7**, compared to Brelsow's 40:1 ratio, as the relative abundances of formaldehyde **16** and glycolaldehyde **7** on the early Earth is unknown. The length of the experiment was also reduced from 3-5 days to 1 day in line with the previous studies on the formation of 2-deoxy-D-ribose **D-130**. Below, **Scheme 3.12**, shows a comparison of Breslow's experiment using 1 equivalent of L-valine **L-80** compared to the Steer *et al.* procedure which used 0.2 equivalents of L-valine amino nitrile **L-157**.



Scheme 3.12. Comparison of Breslow et al.'s synthesis of D-glyceraldehyde **D-9** (above) with Steer et al.'s (below).

Breslow's experiment led to a 4 % ee of D-glyceraldehyde **D-9** using L-valine **L-80** as the catalyst although a yield for this particular example was not stated. Whereas the Steer *et al.* procedure gave a 6 % ee in favour of D-glyceraldehyde **D-9** in a 1 % overall yield. The results with the amino nitrile promoted reaction compared favourably to those of Breslow's early results and suggest that amino nitriles could be viable promoters for the two step synthesis of 2-deoxy-D-ribose **D-130**. The ee and yield for the reaction was calculated based on three runs of the reaction. An authentic HPLC trace from the reaction is shown in **Figure 3.2**. It should be noted that this reaction was used to demonstrate a formal route from glycolaldehyde **7** to 2-deoxy-D-ribose **D-130**. The D-glyceraldehyde **D-9** forming reaction was not optimised any further.

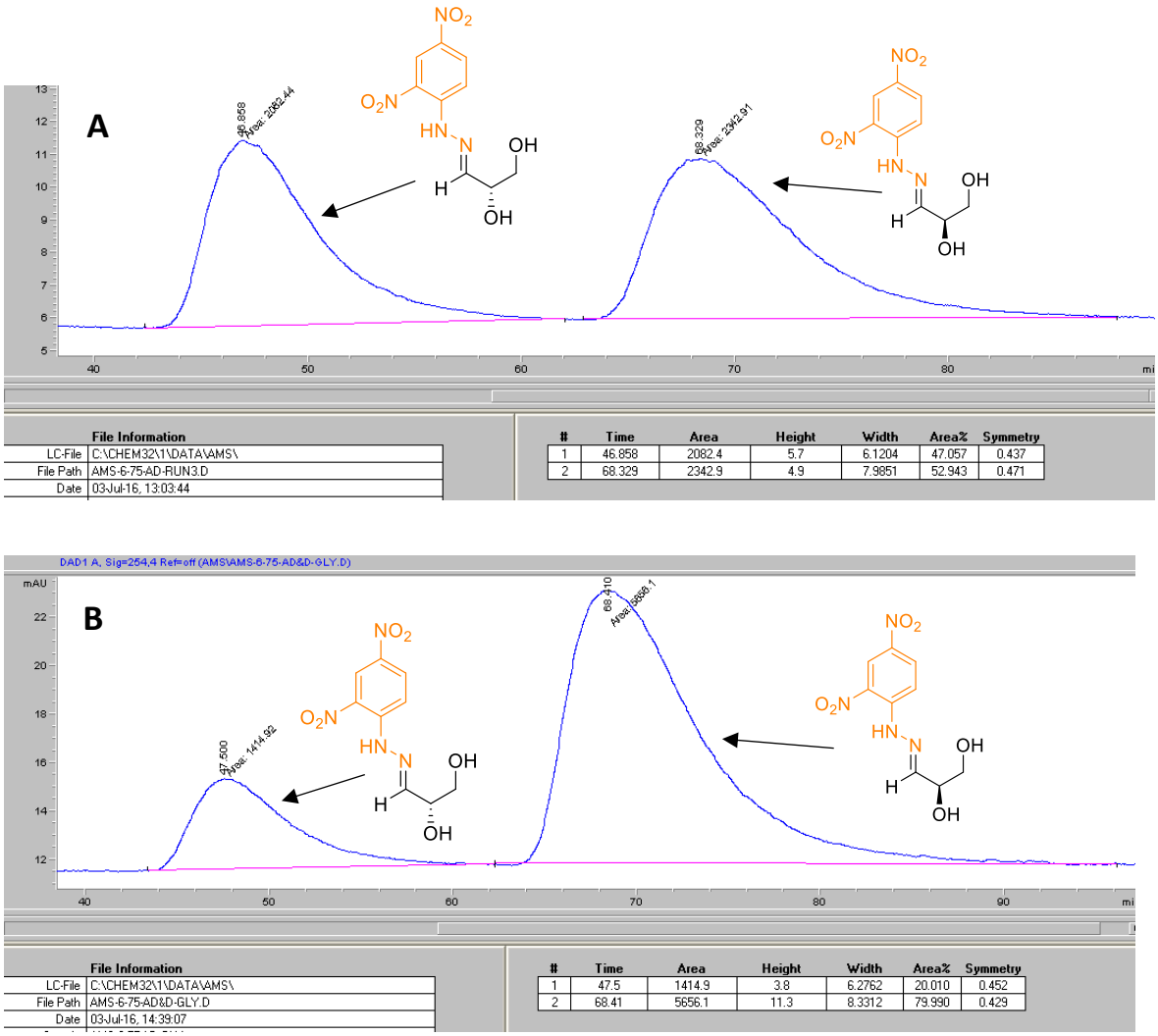


Figure 3.2. (A) Authentic HPLC chromatogram of hydrazone product from the reaction. (B) Hydrazone product spiked with D-glyceraldehyde **D-9** hydrazone to identify the peaks.

Using the results so far, a picture could be built whereby during the early bombardment period, comets brought interstellar sugar building blocks to Earth along with amino nitriles that had been previously synthesised in space through Strecker reactions or evolved on the early Earth through pathways suggested by Sutherland *et al.*[47,40] Over millions of years these promoters could first be used to catalyse the formation of D-glyceraldehyde **D-9** in a slight excess which through a number of chiral amplification pathways such as eutectic concentrations, Viedma ripening or attrition enhance ripening, could lead to enhanced ees. In a further step, D-glyceraldehyde **D-9** could undergo an aldol reaction with acetaldehyde

24 with some selectivity for 2-deoxy-D-ribose **D-130** over 2-deoxy-D-threopentose **D-131**. This scenario is shown graphically in **Figure 3.3** below.

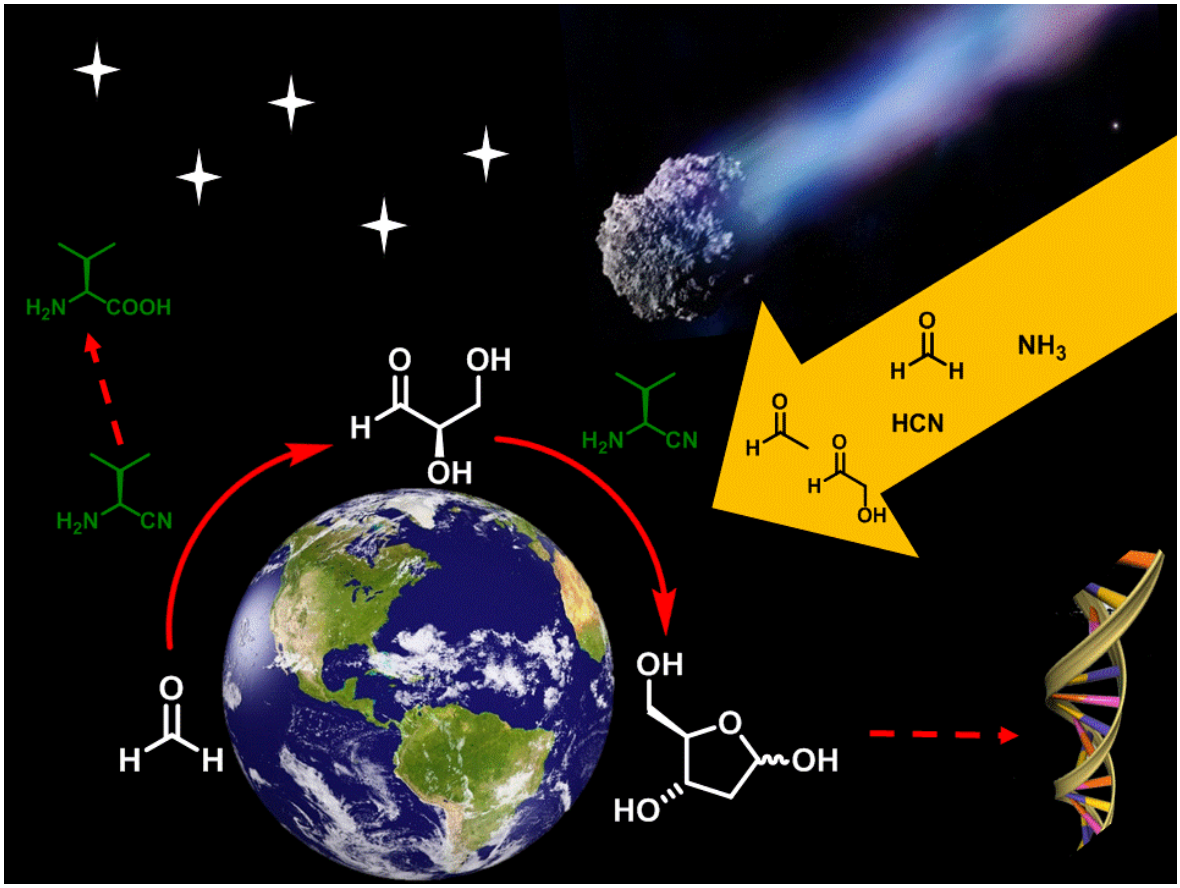


Figure 3.3. Vision of the synthesis of 2-deoxy-D-ribose **D-130** on the early Earth.

3.5. Racemic glyceraldehyde

The next logical step was to investigate the stereochemical outcome of using racemic glyceraldehyde **rac-9** in place of D-glyceraldehyde **D-9** in the 2-deoxy-D-ribose **D-130** forming reaction. As racemic glyceraldehyde contains molecules of both the L and D enantiomers four possible pentose products could potentially arise: 2-Deoxy-D-ribose **D-130**, 2-deoxy-D-threopentose **D-131**, 2-deoxy-L-ribose **L-130** and 2-deoxy-L-threopentose

L-131. The ratio of 2-deoxy-L-ribose **L-130** to 2-deoxy-D-ribose **D-130** was to be analysed by HPLC. A HPLC trace of authentic 2-deoxy-D-ribose and 2-deoxy-L-ribose in their hydrazone forms (**L-146** and **D-147**) is shown in **Figure 3.4**.

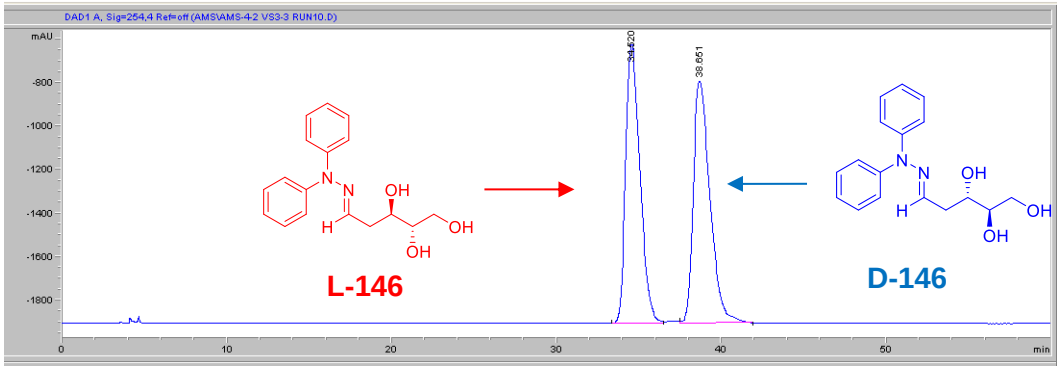
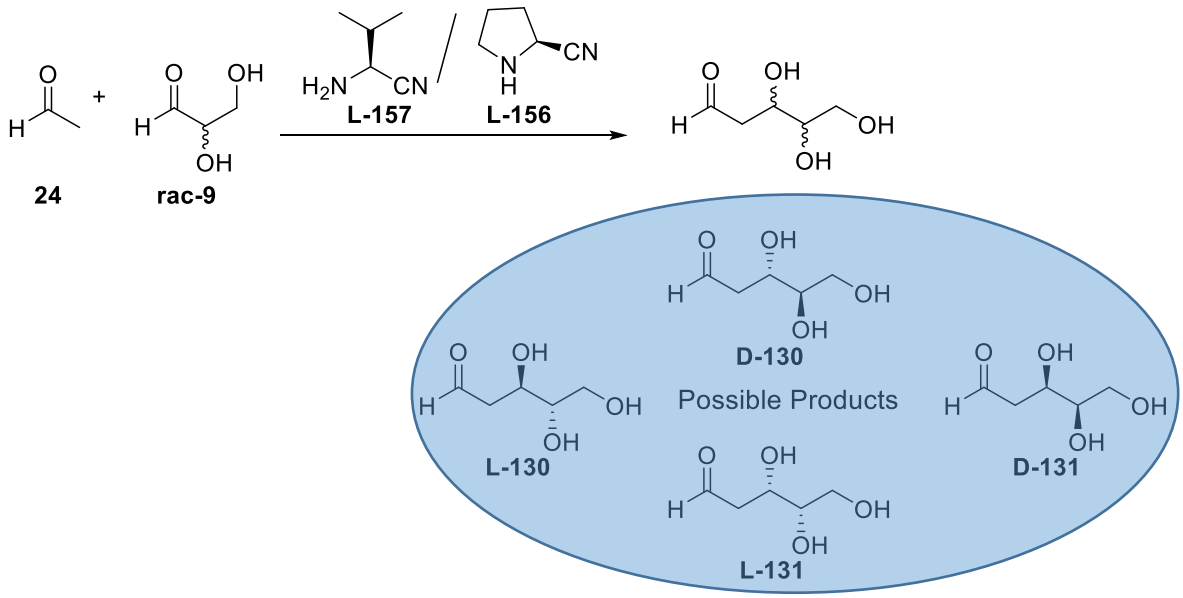


Figure 3.4. HPLC trace of authentic standards of 2-deoxy-L-ribose **L-130** and 2-deoxy-D-ribose **D-130** trapped in hydrazone form.

The reaction was carried out exactly the same as before; acetaldehyde **24** (1 mmol), glyceraldehyde **rac-9** (1 mmol) and L-amino nitrile (0.2 mmol) were dissolved in pH 7 phosphate buffer (3 mL) and stirred for 24 hours. The solvent was then removed *in vacuo* and the crude material redissolved in methanol (5 mL). The trapping agent, *N,N*-diphenyl hydrazine **144** (3 mmol) and acetic acid (2 drops) were added and stirred for 1 hour before concentrating *in vacuo* and purification *via* chromatography. The results are shown in **Table 3.2** below.



Entry	Nitrile	pH	Isolated Yield (%)	Ratio (146:47)	ee of 2-deoxyribose
1	**L-157**	7	4	1.6 : 1	8 % ee (L)
2	**L-156**	7	5	1.6 : 1	Racemic

Table 3.2. Results of the 2-deoxyribose **D-130** forming assay with racemic glyceraldehyde **rac-9** and amino nitrile promtoers. The dr of 2-deoxyribose **146** : 2-deoxythreopentose 1**47** was based on the integration of the azomethine peaks in the ^{1}H NMR spectra and confirmed through HPLC. The ee of deoxyribose was calculated by comparing the peak area of **L-146** and **D-146** from HPLC traces.

The HPLC chromatograms of the reactions are shown in **Figure 3.5** and **Figure 3.6**. A sample of products from the reaction with L-valine nitrile **L-157** was doped with the hydrazone of 2-deoxy-D-ribose **D-146** to clarify the peaks. The peak at 36 minutes represented the hydrazone of 2-deoxy-L-ribose **L-146** and the peak at 40 minutes was the hydrazone of 2-deoxy-D-ribose **D-146**. The other two inseparable peaks at approximately 45 minutes belonged to those of the two enantiomers of 2-deoxythreopentose hydrazones (**D-147 and L-147**). The dr of trapped 2-deoxyribose **146** to 2-deoxythreopentose **147** in

the HPLC was 1.6:1 (2-deoxyribose:2-deoxythreopentose) in agreement with that of the ^{1}H NMR spectra.

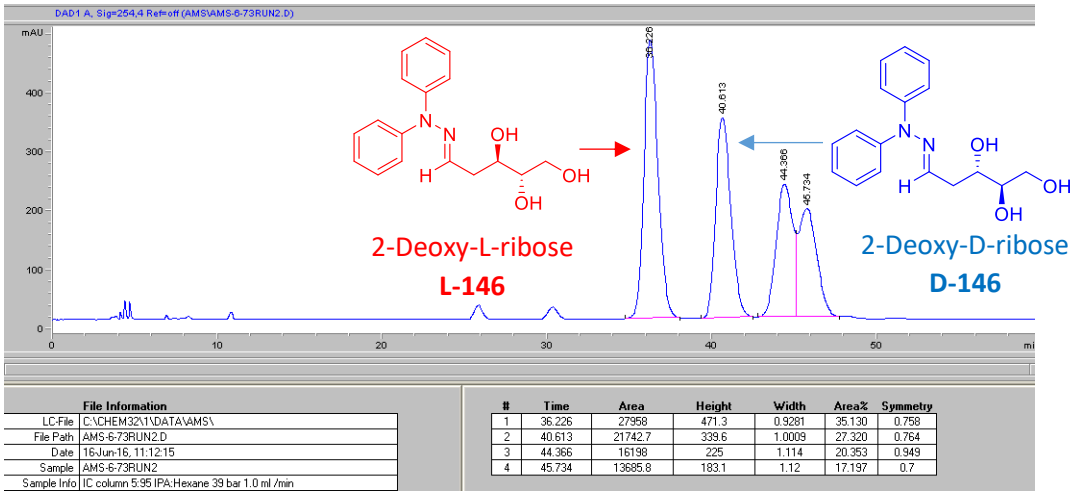


Figure 3.5. HPLC chromatogram from L-valine nitrile **L-157** run.

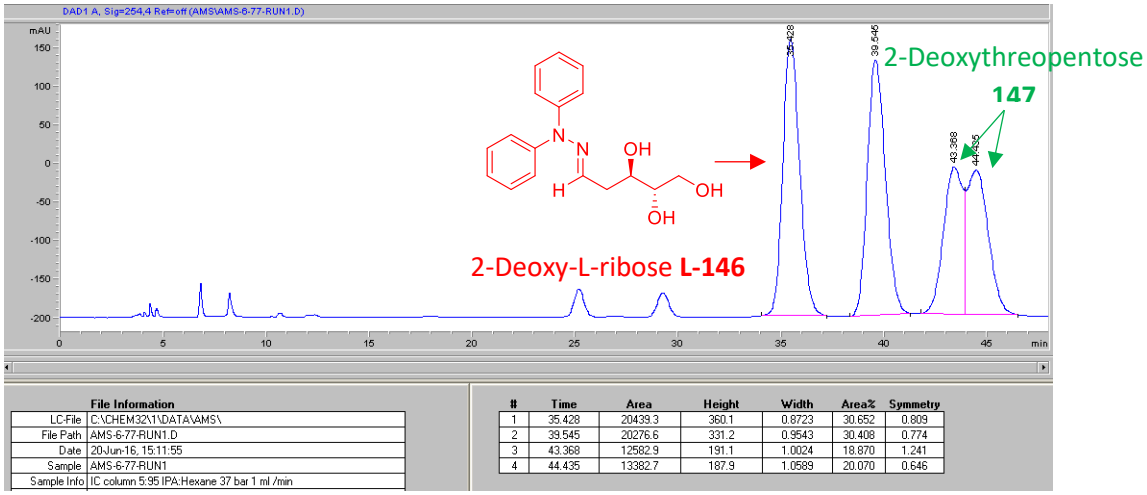
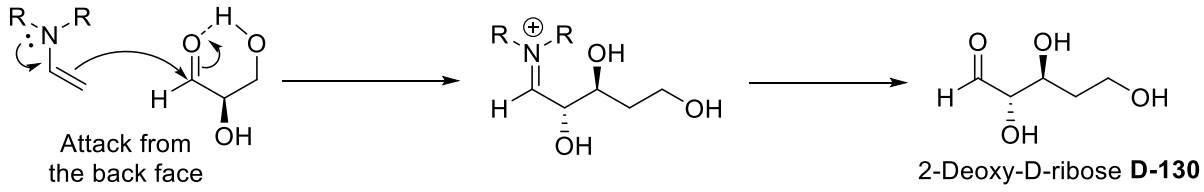


Figure 3.6. HPLC chromatogram from L-proline nitrile **L-156** run.

Interestingly both reactions with racemic glyceraldehyde **rac-9** gave the same ratio of products, 1.6:1 (**146**:**147**), as using D-glyceraldehyde **D-9** as the starting material. From this we inferred that the *anti*-configuration of the two hydroxyl groups is preferred to the *syn*-configuration when forming pentose sugars. This may be due to an intramolecular

hydrogen bond between the hydroxyl group and carbonyl of the glyceraldehyde **9** electrophile (**Scheme 3.13**). This would prevent rotation of the C2 bond and provide two possible trajectorys for attack of the nucleophile. The nucleophile will prefer to attack from the back of the molecule to avoid steric clash with the OH group, hence showing selectivity for the resulting *anti* configuration. The OH group is not sterically very bulky and so this would account for the slight preference for *anti* over *syn* geometry.



Scheme 3.13. Rationale for the preferred formation of 2-deoxy-D-ribose **D-130** over 2-deoxy-D-threopentose **D-131**.

However, in the racemic glyceraldehyde experiment when using L-valine nitrile **L-157** as the promoter gave a small preference for the formation of 2-deoxy-L-ribose **L-130** over 2-deoxy-D-ribose **D-130** which was not seen when using L-proline nitrile **L-156** as the promoter. By comparing the two promoters it is possible that L-valine **L-157** is of significant steric bulk to create a matched-mismatched effect between the promoter-nucleophile molecule and the glcyerladehyde **9** starting material. Hence the chiral enamine complex of L-valine nitrile **L-157** may react faster with L-glycerladehyde **L-9** than D-glyceraldehyde **D-9** leading to a small excess of 2-deoxy-L-ribose **L-131**. Unfortunately no further experiments were conducted to test this hypothesis and future work should focus on computational studies to test the hypothesis as well as experiments involving the reaction of racemic glyceraldehyde **rac-3** using D-valine nitrile **D-157**. If the hypothesis is correct an excess of 2-deoxy-D-ribose **D-130** should be observed. Further experiments should also

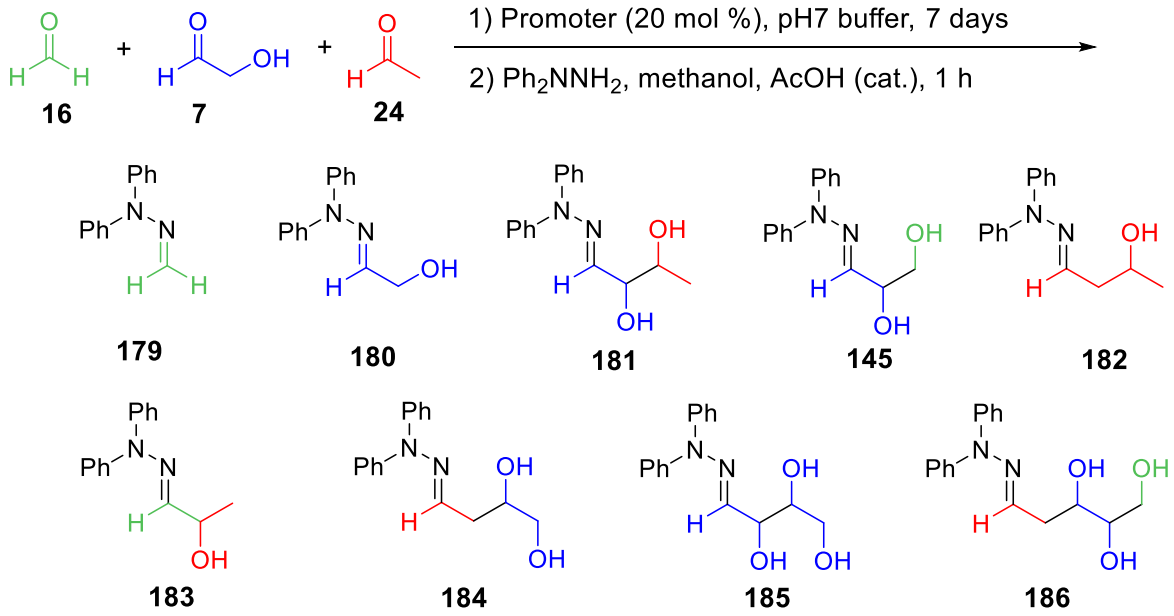
include other bulky amino nitrile promoters to determine if increasing steric bulk could influence the selectivity further.

3.6. One pot synthesis of 2-deoxyribose

Having demonstrated the organocatalytic synthesis of glyceraldehyde **9** and 2-deoxy-D-ribose **D-130** separately, in an effort to enhance the prebiotic relevance of the study, the next experiments combined all three starting materials, formaldehyde **16**, glycolaldehyde **7** and acetaldehyde **24** in a one pot system to potentially form 2-deoxyribose **130** and other pentose products. For this to be achieved formaldehyde **16** and glycolaldehyde **7** first needed to react to form glyceraldehyde **9**. Next, a subsequent aldol reaction needed to occur between newly formed glyceraldehyde **9** and acetaldehyde **24 to** form 2-deoxyribose **130**. Considering the yields for the latter reaction alone provided pentose sugars in only 2-5 %, and the complex number of products that could be produced from this reaction, the quantity of pentose sugars, if any, was expected to be very low (< 1 %).

The one-pot reaction was first attempted using 0.2 equivalents of L-proline benzyl ester **L-108** with a 1:1:1 ratio of aldehyde starting materials (formaldehyde **16**, glycolaldehyde **7** and acetaldehyde **24**) in pH 7 buffer. After 24 hours the solvent was evaporated and the usual trapping conditions were applied. Only hydrazine-trapped starting materials were recovered. The reaction was repeated with an increased reaction time of 7 days, again, the solvent was removed *in vacuo* and the trapping conditions applied. After 24 hours the solvent was removed *in vacuo* once more and the different hydrazone species separated *via* column chromatography. Each compound was then characterised by ESI mass spectrometry and ^{1}H NMR spectroscopy and compared to literature compounds. In the event that the product had not been documented in the literature an authentic standard of

the compound was made to verify its identity. The results of the experiment are shown in **Table 3.3** using (a) no promoter (b) L-proline benzyl ester **L-108** and (c) L-valine nitrile **L-157**.



Product	No Promoter	L-ProBn (L-108)	L-Valine Nitrile (157)
179	0.78 mmol (17 %)	0.25 mmol (5%)	0.77 mmol (17 %)
180	0.87 mmol (22%)	0.70 (18%)	0.76 mmol (19%)
181	0.20 mmol (6 %)	0.056 mmol (1.7%)	0.065 mmol (0.8 %)
145	0.026 mmol (0.8%)	0.071 mmol (1.3 %)	0.050 mmol (1.5%)
184	-	0.037 mmol (1.1%)	0.041 mmol (1.3%)
185	0.016 mmol (0.5%)	0.035 mmol (1.1%)	0.041 mmol (1.3%)
186	-	Trace	-
Total	1.89 mmol (47 %)	1.15 mmol (29 %)	1.73 mmol (42 %)

Table 3.3. Products and starting materials recovered from one-pot synthesis reaction in hydrazone form. The quantity of each product is given in mmols and % yield based on the combined mmols of the three starting materials.

From **Table 3.3** it can be seen that the majority of recovered material is trapped starting material **179** and **180**. In the absence of a promoter a larger amount of **181** (glycolaldehyde **7** + acetaldehyde **24**) is formed as well as a smaller amount of products **145** and **185**. There was no recovered acetaldehyde **24**, nor acetaldehyde dimer **182**. This is most likely because **182** could further eliminate water to from an enone which, along with acetaldehyde **24** have low boiling points, therefore when removing the solvent after the reaction and before the trap, the molecules will be removed *in vacuo*. None of product **184** (acetaldehyde **24** + glycolaldehyde **7**) or product **186** was isolated. As predicted the quantity of compounds **181**, **145**, **184** and **185** increased in the presence of a promoter. In the case of L-proline benzyl ester **L-108** the crude mass spectrum showed a peak corresponding to the mass of trapped 2-deoxyribose **146** to be present. Unfortunately, isolation of 2-deoxyribose **146** was not possible due to the small quantity of the pentose product. However, after column chromatography, a fraction containing the mass of the tetrose product and the mass of 2-deoxyribose was isolated. Further HPLC studies that involved this fraction and this fraction doped with authentic 2-deoxy-D-ribose hydrazone **D-146** revealed 2-deoxy-D-ribose **D-146** to be present (**Figure 3.7**). Intriguingly, when the fraction was doped with a racemic mixture of 2-deoxyribose hydrazones it was revealed that 2-deoxy-L-ribose **L-146** was not present in the sample.

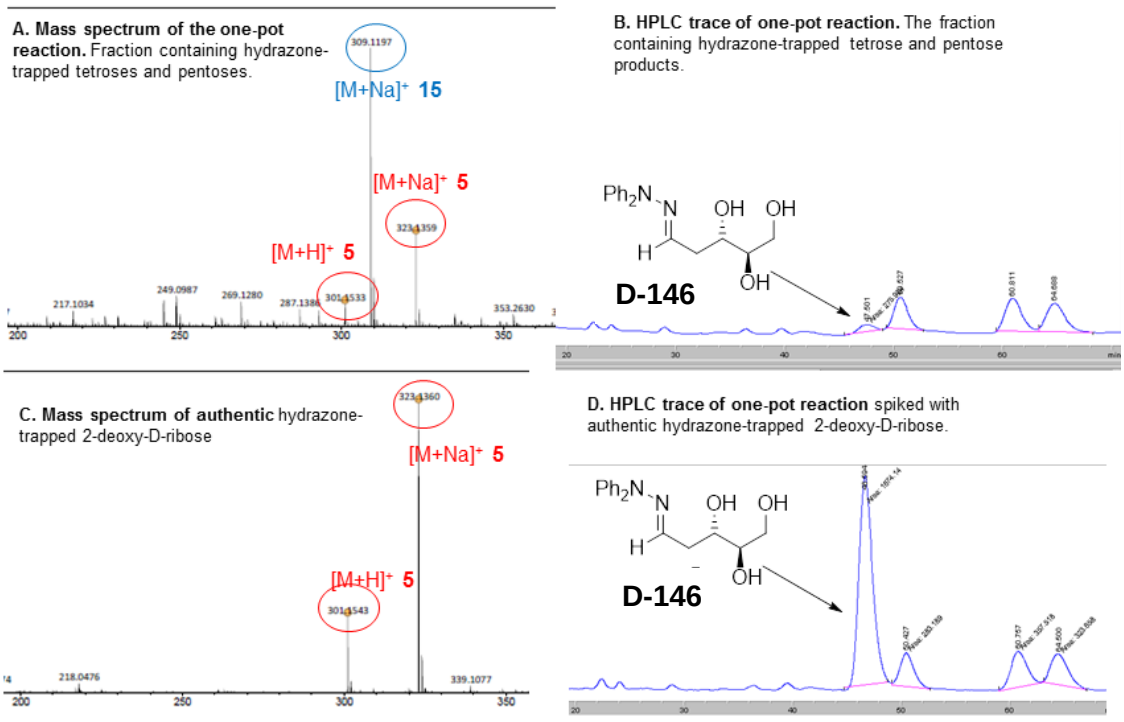


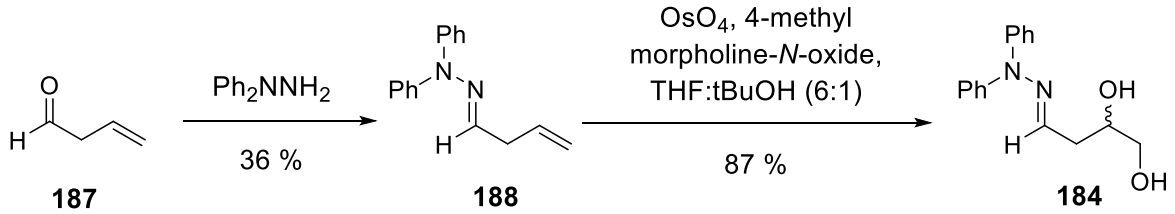
Figure 3.7. Mass spectra and HPLC chromatograms of tetrose fraction containing 2-deoxy-D-ribose hydrazone **D-146** and HPLC trace with pure 2-deoxy-D-ribose hydrazone **D-146**.

Interestingly switching from L-proline benzyl ester **L-108** to L-valine nitrile **L-157** gave similar quantities of products **181**, **145**, **184** and **185** but did not produce any 2-deoxy-D-ribose **146** detectable by ESI mass spectrometry. This may have been due to removal of the L-valine nitrile **L-157** promoter from the reaction over time through the pathway of imine formation and subsequent cyclisation. In all cases the recovered mass was less than 50 % of the total mass. A large proportion of this may have been due to the loss of unreacted starting materials with low boiling points (such as formaldehyde and acetaldehyde) during rotary evaporation before the trapping step of the assay or indeed from less efficient trapping of the sugar products and starting materials.

In order to confirm the assignment of products **181**, **145**, **184** and **185** were correct the products were to be compared to pure authentic standards. Compound **181** is known in the

literature and comparison of the ^{1}H NMR data to the literature confirmed the identity of **181**.[119] Compound **145** had been synthesised previously as an authentic standard in Chapter 6 and the data for **145** was in agreement. Compound **184** and **185** had not been reported in the literature and therefore needed to be synthesised.

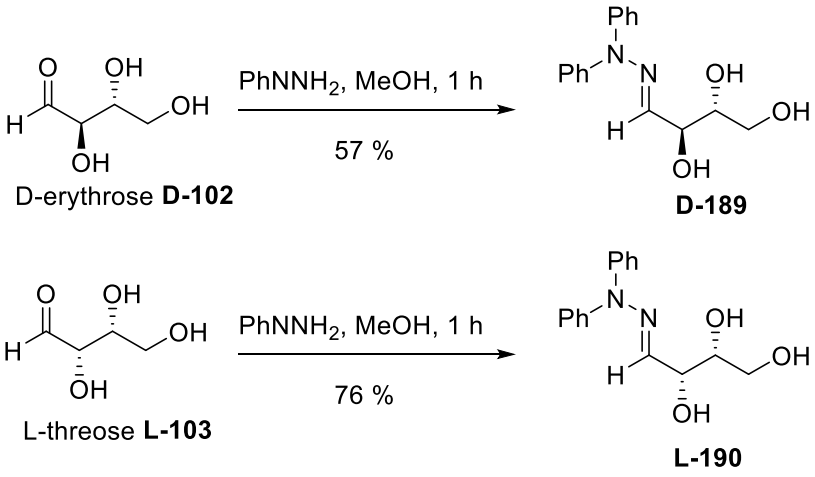
Compound **184** was synthesised from 3-butenal **187** in a 2-step process (**Scheme 3.14**). Firstly 3-butenal **187** was dissolved in DCM and trapped with *N,N*-diphenyl hydrazine **144** which gave the resulting hydrazone **188** in a low 36 % yield. Hydroxylation of the alkene with catalytic osmium tetroxide in the presence of 4-methyl morpholine *N*-oxide gave the diol product **184** in a good 87 % yield. Comparison of ^{1}H NMR and ESI MS data of the authentic standard and compound **184** isolated from the one-pot reaction revealed them to be the same compound.



Scheme 3.14. Synthesis of **184** from butenal **187**.

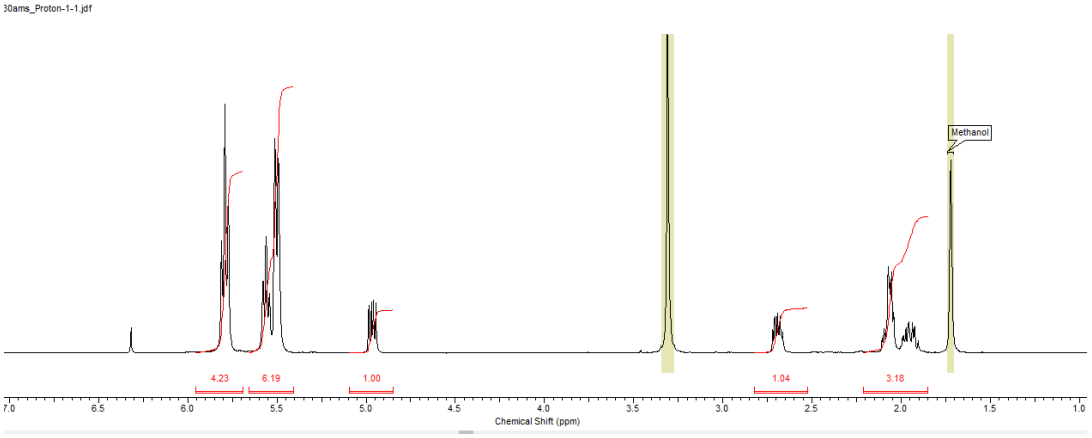
The final compound **185** contained two stereocentres and therefore existed in up to four diastereomers: D/L threose **103** and D/L erythrose **102**. As enantiomers have the same physical and chemical properties, the ^{1}H NMR spectrum showed this as a mixture of threose/erythose diastereomers. To confirm **185** was the tetrose products it was only necessary two synthesise to standards; L or D threose **103** and L or D erythrose **102**. Commercially available D-erythrose **D-102** and L-threose **L-103** were trapped with *N,N*-

diphenyl hydrazine **144** to produce the hydrazones **D-189** and **L-190** (**Scheme 3.15**). A ^{1}H NMR spectrum containing a 1:1 mixture of erythose:threose proved identical to that of the ^{1}H NMR spectrum of **185** (**Figure 3.8**).



Scheme 3.15. Synthesis of D-erythose and L-threose diphenyl hydrazones **189** and **190** for comparison to compound **185**.

A) Mixture of standards **189** and **190**



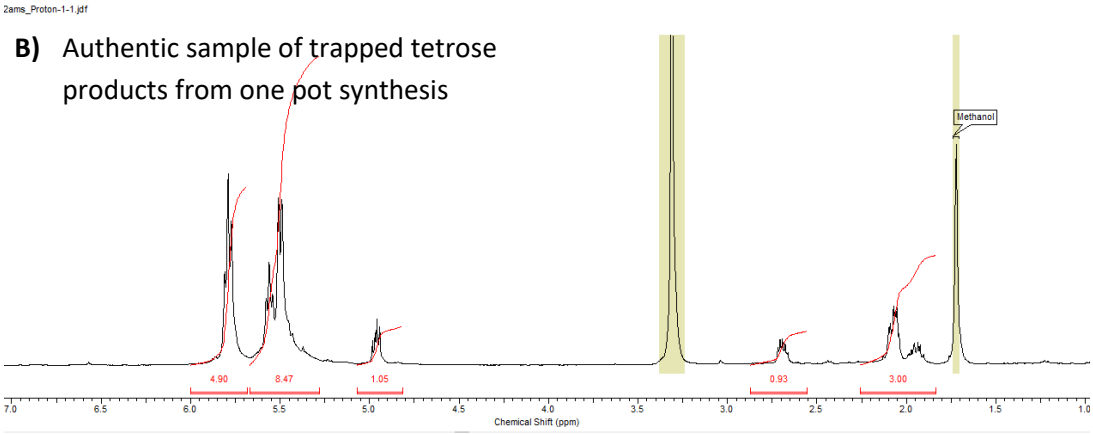


Figure 3.8. ^{1}H NMR spectra of; (A) a mixture of authentic threose and erythrose tetrose standards. (B) Authentic sample of tetrose products from one-pot synthesis after column chromatography.

This study proved that 2-deoxy-D-ribose **D-130**, as well as many other small simple sugars, can be produced in a one-pot synthesis from interstellar building blocks, acetaldehyde **24**, formaldehyde **16** and glycolaldehyde **7**. Although L-valine nitrile **L-157** (arguably the more prebiotically plausible catalyst) could not catalyse the formation of 2-deoxyribose **130** the problem may lie in the synthesis of the cyclic by-product and could be overcome by increasing the equivalents of amino nitrile or adding further portions of amino nitrile to the reaction over the course of the 7 days or increasing the reaction time further.

3.7. Conclusions

Within this chapter amino nitriles have been shown to promote the synthesis of 2-deoxy-D-ribose **D-130** and 2-deoxy-D-threopentose **D-131** in an approximate 2:1 ratio from D-glyceraldehyde **D-9** and acetaldehyde **24** in the highest yields yet reported for this reaction (5 %). This is the first example of an amino nitrile promoted reaction. L-Valine nitrile **L-157** has been shown to promote the formation of D-glyceraldehyde **D-9** from formaldehyde **16**

and glycolaldehyde **7** and hence a formal 2-step amino nitrile promoted synthesis of glycolaldehyde to 2-deoxy-D-ribose **D-130** has been demonstrated.

Amino nitriles were also used to promote the reaction of racemic glyceraldehyde **rac-9** with acetaldehyde **24** but L-proline nitrile **L-156** gave no selectivity for 2-deoxy-D-ribose **D-130** over 2-deoxy-L-ribose **L-130** whereas L-valine nitrile **L-157** showed an 8 % ee in favour of 2-deoxy-L-ribose **L-130**. The results suggested that glyceraldehyde **9** needs to be somewhat enantiomerically enriched towards the D-enantiomer in order to selectively form 2-deoxy-D-ribose **D-130**. The final part of the study showed that 2-deoxy-D-ribose **D-130** could be formed in a one-pot synthesis from interstellar building blocks using L-proline benzyl ester **L-108** as a catalyst.

4. Prebiotic Protocells

4.1. A Protocell Environment

Having designed a successful experiment for the potential prebiotic formation of 2-deoxy-D-ribose **D-130** in an aqueous environment, the next step was to develop this chemistry further within a primitive cell-like structure, a so-called "protocell". The concept of a protocell is that of a primitive ancestor of the biological cell. Researchers in the field of Origins of Life have different ideas on what defines a protocell, for detailed reviews see references.[120,121,122] In its simplest form a protocell could be described as a self-assembled compartmentalised system whereby a chemical reaction can take place using energy from the environment and containing some type of information system.[122] This could include a range of supramolecular forms from vesicles to gels systems.

The modern biological cell contains a phospholipid bilayer which is thought to have been a later stage evolution of a simpler monomer. Vesicles are good supramolecular candidates as primitive cells and may have been composed of simple single chain fatty acids. Scientists have confirmed the presence of simple fatty acids on extra-terrestrial meteorites and form a good case for their use as the monomers for simple vesicle structures.[123,124] Fatty acids are amphiphilic molecules composed of a long hydrophobic tail-group and a hydrophilic head-group. When exposed to aqueous medium the monomers arrange themselves to minimise the number of unfavourable interactions between hydrocarbon tail and aqueous solvent, and maximise the number of interactions between polar-head group and solvent. Above a certain concentration, the monomers will arrange into supramolecular structures to help minimise unfavourable interactions.

There has been a lot of research into the mechanism of vesicle formation and the dependency to form vesicles as opposed to micelles is in large due to pH. The pH of the solution should ideally be approximately the same as the pK_a of the fatty acid in the aggregated state to allow for a balance between protonated and deprotonated molecules.[125] This allows for hydrogen bonding between pairs of fatty acids and fatty acid salts.[126,127] As well as this it has been demonstrated that vesicles are sensitive to changes in ionic concentrations.[128] Obviously a pH specific supramolecular structure is not a very robust model for a protocell, however, researchers have overcome this pH dependency by mixing fatty acid monomers with alcohols. The addition of alcohols, such as glycerol, also had the added benefit of reducing the critical vesicle concentration (CVC) of the system.[129,130,131] Work from the Szostak lab has developed vesicle research further showing how model protocell vessels of oleic acid in the presence of glycerol monooleate can become stabilised, and permeable to magnesium ions, which is required for ribozyme activity and RNA synthesis.[132] Moreover, the group has shown that in the presence of magnesium ions, ribozymes can be activated and encapsulated within vesicles.[133] Moving towards synthesis in vesicle "protocells", Szostak demonstrated the peptide bond formation of a dipeptide **195** from lysine amide **192** and phenylalanine amide ester **193** using the dipeptide Ser-His **194** as a catalyst. The product would then localise to the bilayer of the vesicle membrane. The group found that vesicles containing **195** in the bilayer would grow in size at the detriment of micelles which did not contain **195**, which would shrink (**Figure 4.1**).[134]

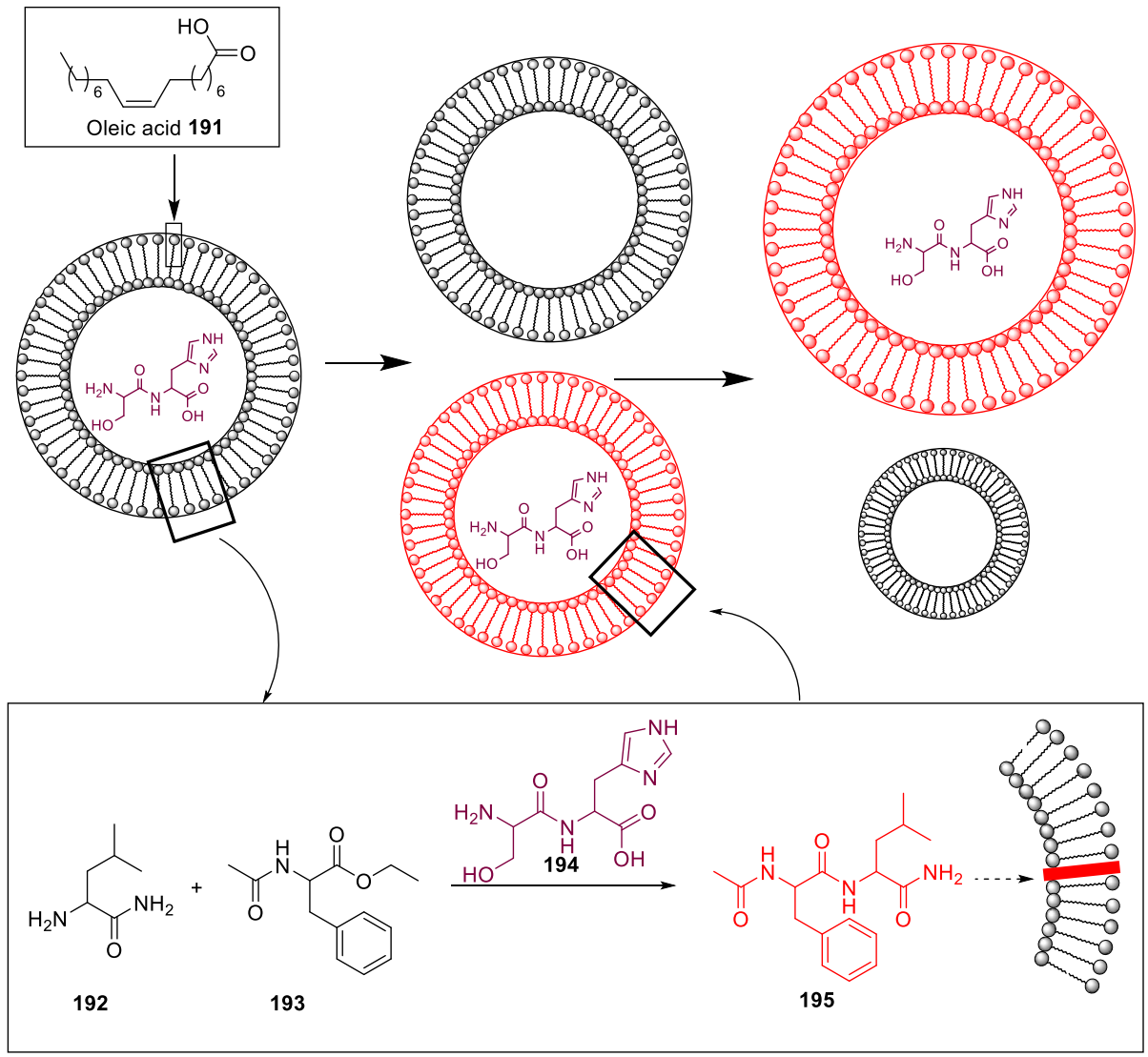


Figure 4.1. Synthesis of dipeptide derivative **195** inside a vesicle by the catalysis of Ser-His **194**. Once synthesised the product moved to the bilayer of the vesicle. The group found vesicles with the product would grow at the detriment to vesicles without the product.[134]

An alternative supramolecular protocell environment may be a hydrogel. Supramolecular mimetic hydrogels are two-component colloidal systems made up of a network of gelator molecules that form nanoscale fibres through weak, non-covalent, intermolecular interactions, that become entangled and trap the bulk solvent usually due to some sort of physical change (i.e. temperature, pH). In the case of hydrogels the solvent is water, and

in this case hydrogen bonds between gelator molecules tend to be less prevalent and the spontaneous nano-fibre assembly often relies more on the hydrophobic effect. A gel possesses characteristics of both a solid and a liquid as it retains a fixed shape which is usually demonstrated by inverting the gel sample with the gel being observed to hold its form, whilst the liquid molecules within the gel, although mobile on the molecular scale, do not exhibit the ability to flow on the experimental timescale. A gel can be considered to be made up of a primary structure, the "molecular recognition aggregate", which is the way in which the monomers assemble – usually as a nanofibril aggregate.[135] The secondary structure can be defined as the morphology in which the fibres assemble further into nanofibres. The tertiary structure is then the interaction of these fibres to yield a 3-D microscale network. This is summarised in **Figure 4.2**.

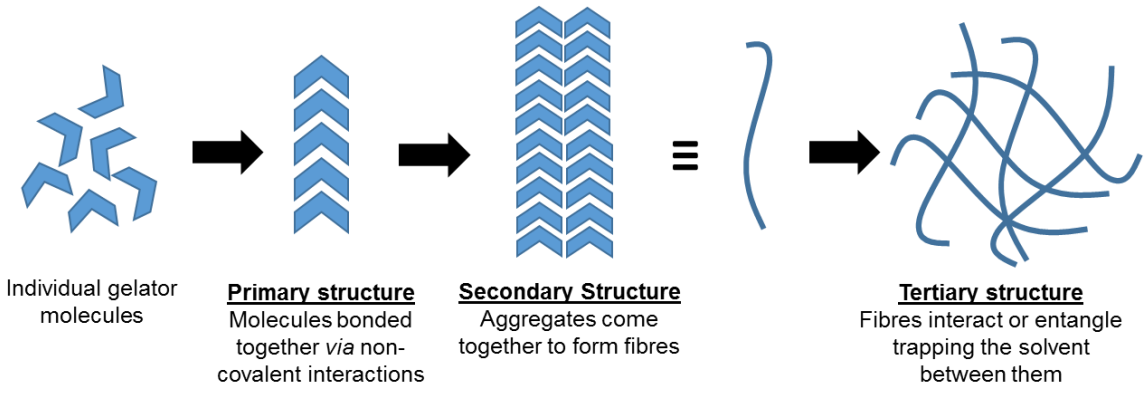


Figure 4.2. Primary, secondary and tertiary structure of a self-assembled gel.

Over the last 20 years a lot of advances have been made in the area of hydrogels with a lot of the research focusing around enzyme mimics, involving long chain peptides. For the purpose of this chapter, a brief introduction into plausible "prebiotic" hydrogels will be given. For more detailed reviews of hydrogels in general please see references 136 and 137.[136,137] A number of hydrogelators involving amino acids rely on large aromatic groups, such as the Fmoc protecting group (**Figure 4.3**, compound **201**), for gelation. The aromatic groups

provide key π-π stacking interactions to help order the monomers into fibres. However, the Fmoc group cannot be considered prebiotic. Compounds **196** and **197** (**Figure 4.3**) are bolaamphiles with polar head groups at either end of a non-polar hydrophobic chain. This allows the monomers to stack up one on top of each other with hydrogen bonding dominating between the polar regions and hydrophobic interactions between the greasy alkyl chains. The prebiotic nature of these peptides is up for debate and, for example **196**, would have been formed from a diacid and two molecules of glutamic acid.

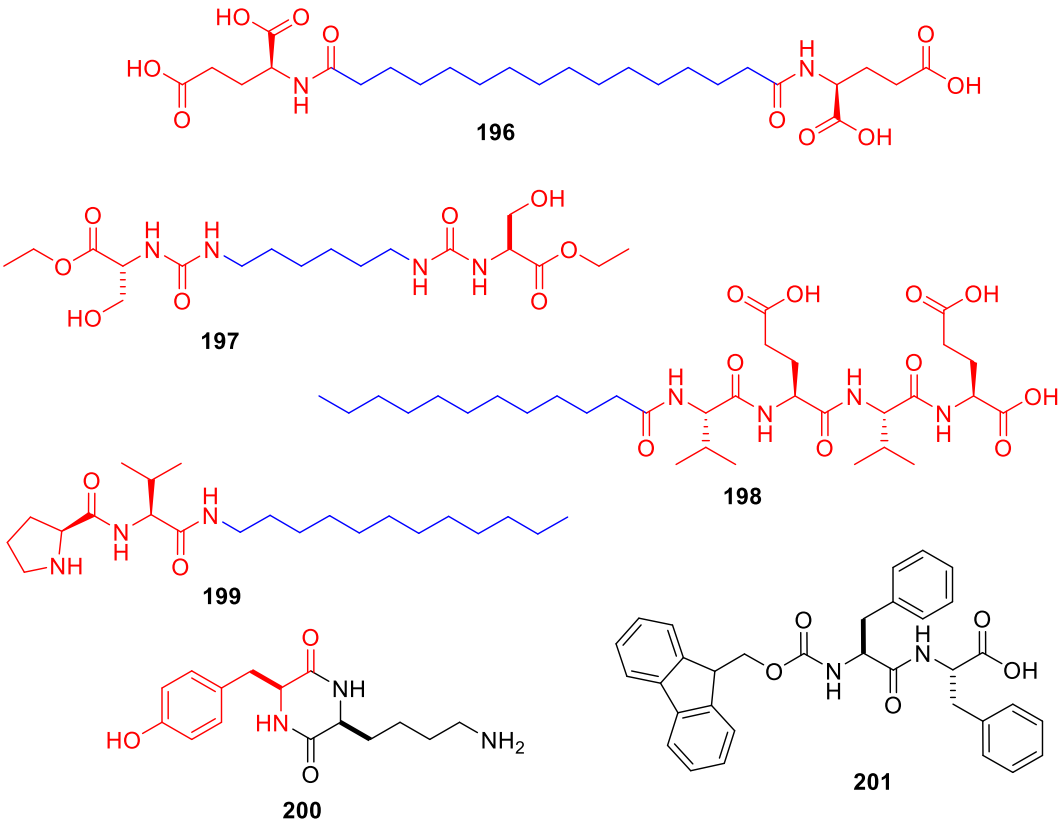


Figure 4.3. Previously reported peptide-based hydrogelators.

Perhaps gelators such as **198** and **199**, composed of a long alkyl chain connected to a peptide core, are more prebiotically relevant. In the case of **199** a simple dipeptide is connected to a long alkyl chain *via* an amide bond. This molecules can form hydrogels in water, a big driving force of this being the need to minimize the number of interactions between the hydrophobic tails and the aqueous solvent.[138] Compound **198** has been used

as a catalyst in the aldol reaction of cyclohexanone and *para*-nitrobenzaldehyde in excellent yields and selectivities albeit using an organic solvent to dissolve the starting materials.[139] Monomer **200** is a good prebiotic hydrogel candidate consisting of only two amino acids. However, Feng *et al*. reported that, although the monomer can form a gel in water, it is very unreliable and would usually precipitate out of solution.[140]

Studies are now moving towards combining supramolecular chemistries to design protocell systems that demonstrate compartmentalization within a membrane. The pioneering work of Sapara *et al* showed an aqueous droplet could be stabilized in an oil/lipid mixture, all encapsulated within an agarose hydrogel.[141]

There are many similarities between hydrogels and cells, in fact the cytoplasm itself can be considered to be a gel.[142] Therefore on the early Earth, it is possible to speculate that, before primitive cells developed membranes there could have been isolated gel-like objects. Other similarities of primitive cells and hydrogels include:[143]

- The ability to retain integrity even in the absence of a membrane.
- Some control over what enters and exits *via* solute exclusion; for example, if the molecule is too large it will not be able to enter the matrix or will become trapped.
- Capability to do “work” within it through small changes in environment that can lead to phase transitions.
- A closed environment which holds everything in close proximity, enabling metabolites from organelles in cells as little distance as possible to travel whilst still being isolated from one another.

In this concept starting material could represent “food” entering the protocell, a chemical reaction could then take place within the gel and then the subsequent products or “waste” removed (**Figure 4.4**). The next aim of the research project was to apply this concept to the formation of 2-deoxy-D-ribose as a possible pre-metabolic pathway.

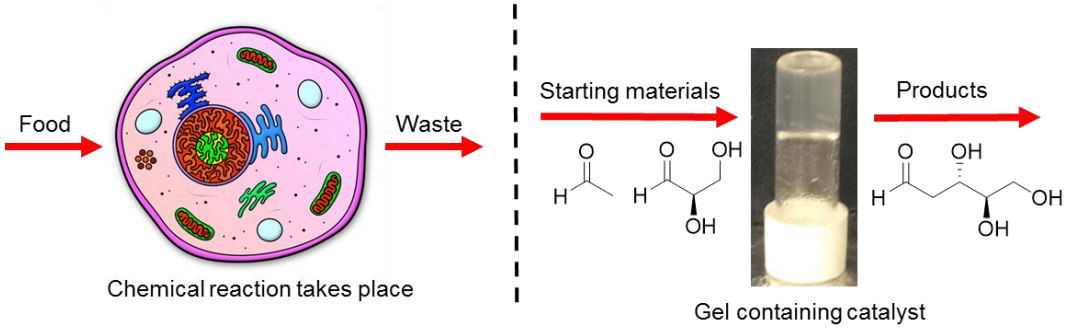


Figure 4.4. Comparison of a chemical reaction taking place in a mammalian cell and a hydrogel.

4.2. Proof of concept using agarose

Before investing time in designing and creating gelator candidates that look prebiotically plausible, the theory of using gels as containers for the aldol reaction to take place was tested. The commercially available, but non-prebiotic, gelator agarose was chosen (**Figure 4.6**). Agarose is a polysaccharide well documented in the literature for its gelation chemistry.[144,145,146] Agarose itself, should not catalyse the aldol reaction thus it is necessary to encapsulate an active catalyst within the gel matrix. Gelation of agarose is trivial and achieved by applying a simple heat-cool cycle. A schematic for the gelation of agarose with L-Pro-OBn **L-108** trapped inside the gel is shown in **Figure 4.5**. Firstly water (1 mL) was added to the agarose powder (5 mg), which was then heated until a homogeneous solution formed. This hot solution was transferred to a vial containing the catalyst **L-108** (23 mg). The mixture was reheated then allowed to rest. Once the vial had cooled below the T_{gel}

(temperature of gelation), the network of agarose fibres assembled around the solvent and catalyst and the gel was formed.

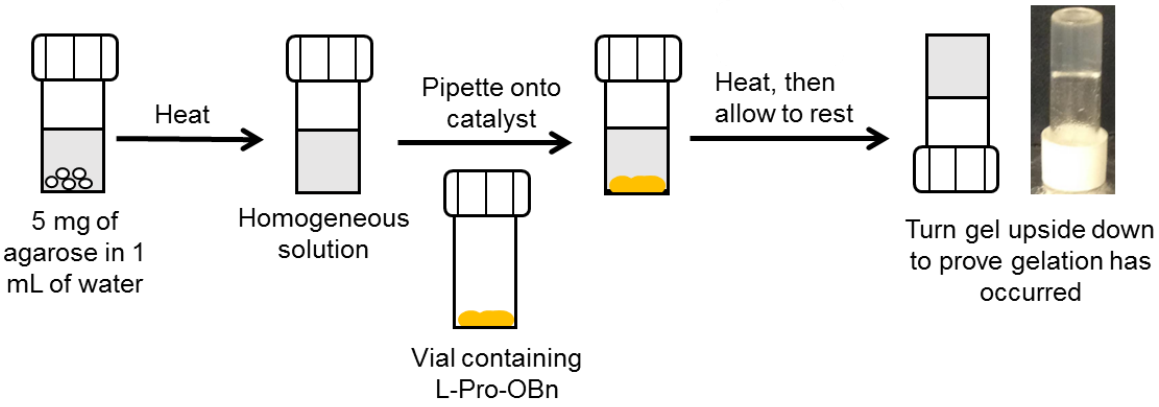


Figure 4.5. Schematic to show the gelation for encapsulation of L-Pro-OBn **L-108** in an agarose hydrogel gel.

With the catalyst within the agarose gel the 2-deoxy-D-ribose **D-130** forming experiment was then attempted. The experimental set-up was the same as before with the two reagents, acetaldehyde **24** (1 mmol) and D-glyceraldehyde **D-9** (1 mmol) dissolved in 3 mL of aqueous solvent. The hydrogel was then carefully levered out of the vial it was formed in and into the reaction vessel with the reagents. This set-up is shown in **Figure 4.6**.

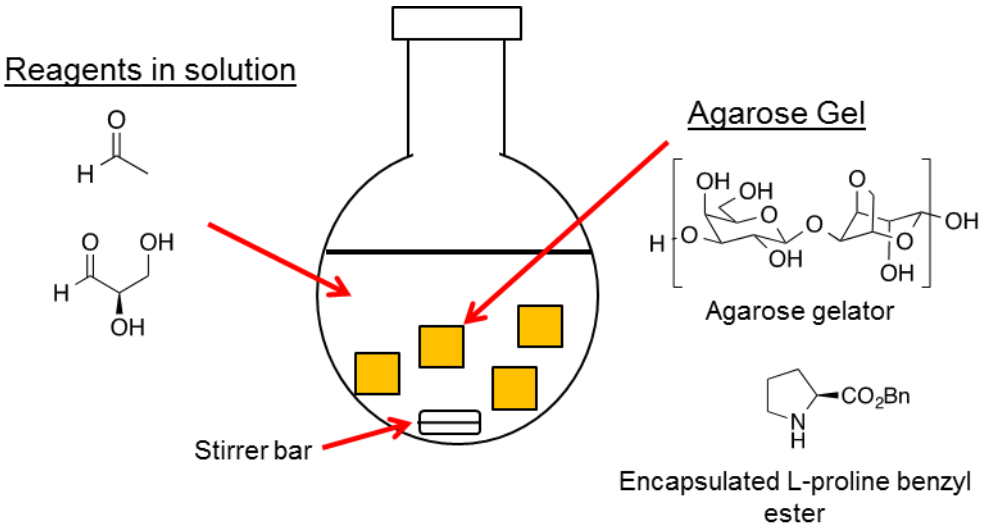
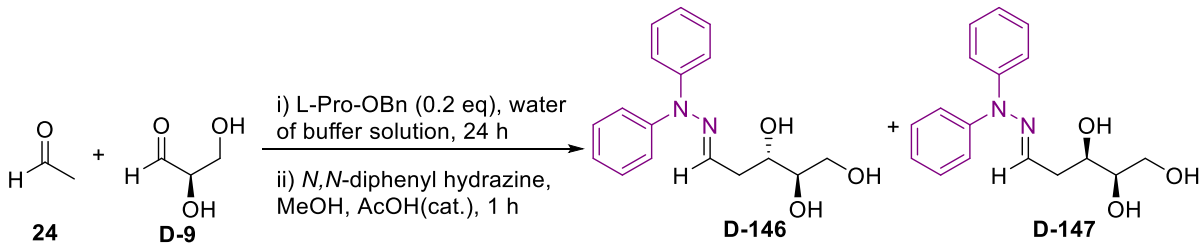


Figure 4.6 Agarose gel reaction set-up.

Due to the way the gel was added to the reaction, the hydrogel was in a number of pieces in the reaction vessel. After stirring slowly for 24 hours the gel was filtered through a sinter, the filtrate concentrated *in vacuo*, and the sugar trapping process and purification were carried out as previously reported. 2-Deoxy-D-ribose **D-146** was indeed formed but in a lower yield and similar dr compared with solution (**See Table 4.1**).



Entry	Catalyst	Solution or gel?	pH	Yield (%)	Ratio (146:147)
1	L-Pro-OBn	Solution	7	2	1.8 : 1
3	None	Gel	7	0	-
4	L-Pro-OBn	Gel	7	1	1.5 : 1

Table 4.1. Results of the 2-deoxy-D-ribose **D-130** forming reaction comparing L-Pro-OBn in gel and solution to an agarose hydrogel. Ratios of **146:147** were calculated as usual from the 500 MH_z ^1H NMR spectrum.

The method of encapsulation of L-Pro-OBn **L-108** within the agarose gel was next examined. It was possible that heating the hot agarose solution with the catalyst inside may have hydrolysed the L-proline ester. A summary of this investigation is shown in a flow chart in **Figure 4.7**. Some important control experiments were conducted in an effort to develop a new method of gelation. In the first set of conditions (Method 1) hot deionised water (1 mL) was poured onto L-Pro-OBn **L-108** (23 mg) followed by subsequent heating of the mixture for 20 minutes – imitating the current gelation conditions but with no agarose present. This was compared with Method 2, wherein hot deionised water (1 mL) was added to the same amount of L-Pro-OBn **L-108** (23 mg) but this time instead of additional heating the mixture was shaken for 30 seconds. The solvent in both experiments was then removed *in vacuo* over 45 minutes with the temperature of the water bath set to 45 °C. Comparison of the ^1H NMR spectra showed that 44 % of the ester was hydrolysed in Method 1 compared with 15 % in Method 2. A further comparison with a control experiment involving removal of water (1 mL) from L-Pro-OBn **L-108** (23 mg) at 45 °C for 45 minutes by rotary evaporation

revealed that 15 % of the ester had been hydrolysed. Therefore we could assign this proportion of hydrolysis as being due to the rotary evaporation conditions. Therefore a true estimate of proline-ester hydrolysis during the gelation method is 29 % for Method 1 and 0 % for Method 2.

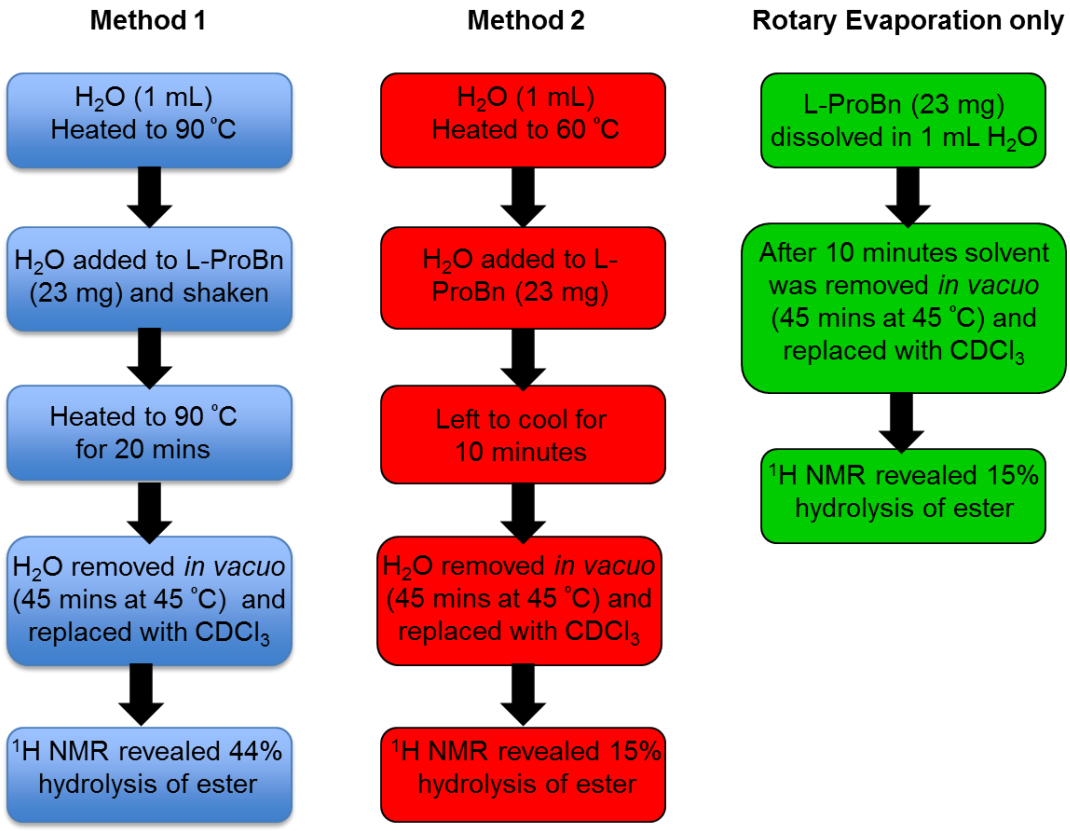


Figure 4.7. Three separate ^1H NMR experiments for the hydrolysis of L-proline benzyl ester in the absence of the agarose gel.

The 2-deoxy-D-ribose **D-130** forming experiment was run again with gels prepared using Method 2. The same d.r of 1.5:1 was obtained, in a similar yield to the previous examples. As the yields of products from both methods were approximately the same it could be possible that the hydrolysed L-proline present in gels prepared through Method 1 were also catalytically proficient. To test this theory, agarose gel was prepared again, only

encapsulating 20 mol% of L-proline, and the assay was run as previously reported in pH 7 buffer. The pentose sugars were synthesised but in a yield of < 1%, with a similar d.r. of 1.5:1.

Having shown some evidence of L-proline catalysis in a gel the original solution phase reaction was run again with L-proline as the catalyst in pH 7 buffer. L-Proline was indeed able to catalyse the formation of 2-deoxy-D-ribose **D-130** in similar yields and drs to previous solution phase reactions (**Table 4.2** Entry 2). The same reaction was attempted in unbuffered water only a trace of the product was formed, interestingly with no selectivity for either diastereomer, this was probably the background reaction which did not occur *via* enamine catalysis and hence had no product selectivity. The difference in reactivity may be due to L-proline changing the pH of unbuffered water leading to a higher concentration of the zwitterionic form of L-proline. With the nitrogen protonated the catalyst is inactive and enamine formation cannot occur. The yield is slightly better in solution than in the agarose gel possibly due to the catalyst being more accessible to the reagents in solution rather than trapped inside the gel.

Entry	Catalyst	Solution or Gel?	pH	Yield (%)	Ratio (146:147)
1	L-proline	gel	7	1	1.5:1
2	L-proline	solution	7	2	1.5:1
3	L-proline	solution	unbuffered	trace	1:1

Table 4.2 Using L-proline as the catalyst for the aldol reaction.

4.3. Prebiotic protocells based on dipeptide amphiphile hydrogels

The agarose gel experiments demonstrated that it was indeed possible to encapsulate amino acids and esters in a gel matrix and indicated that the aldol reaction could still occur in similar drs and yields to the reactions in solution. The next step was to attempt the same reaction, the formation of 2-deoxy-D-ribose **D-130**, in a more prebiotically plausible hydrogel. There are only a few examples of small, simple, hydrogel molecules that could credibly have existed on the early Earth. However one recent example, by Escuder *et al.,* which consisted of a dipeptide connected to a long alkyl chain *via* an amide bond was found to not only form a hydrogel but could successfully catalyse the aldol reaction of *p*-nitrobenzaldehyde **62** with cyclohexanone **75** in excellent yields and drs when the reagents were added to the top of the gel predissolved in toluene.[139] This molecule could be considered prebiotic as it consisted of just two amide bonds. The catalysis stems from the free N-terminus of the L-proline residue (Gelator **A**, **Figure 4.8**) allowing the formation of the key enamine intermediate of the aldol reaction. It was decided that this gelator would be synthesised and tested as a potential hydrogel catalyst in the 2-deoxy-D-ribose **D-130** forming reaction. As a control, a second non-catalytic gelator (Gelator **B**) with the amine site blocked by a methyl group, was also proposed. However, it was then envisaged that catalysts, such as L-Pro-OBn **L-108**, could be trapped inside the gel network of **B** to provide a comparison to the agarose system. This led to the proposal for two different types of catalytic gel (**Figure 4.8**) one in which the catalytic sites are an integral part of the nanofibres (Gelator **A**) and the other in which they are simply encapsulated within the gel (Gelator **B**).

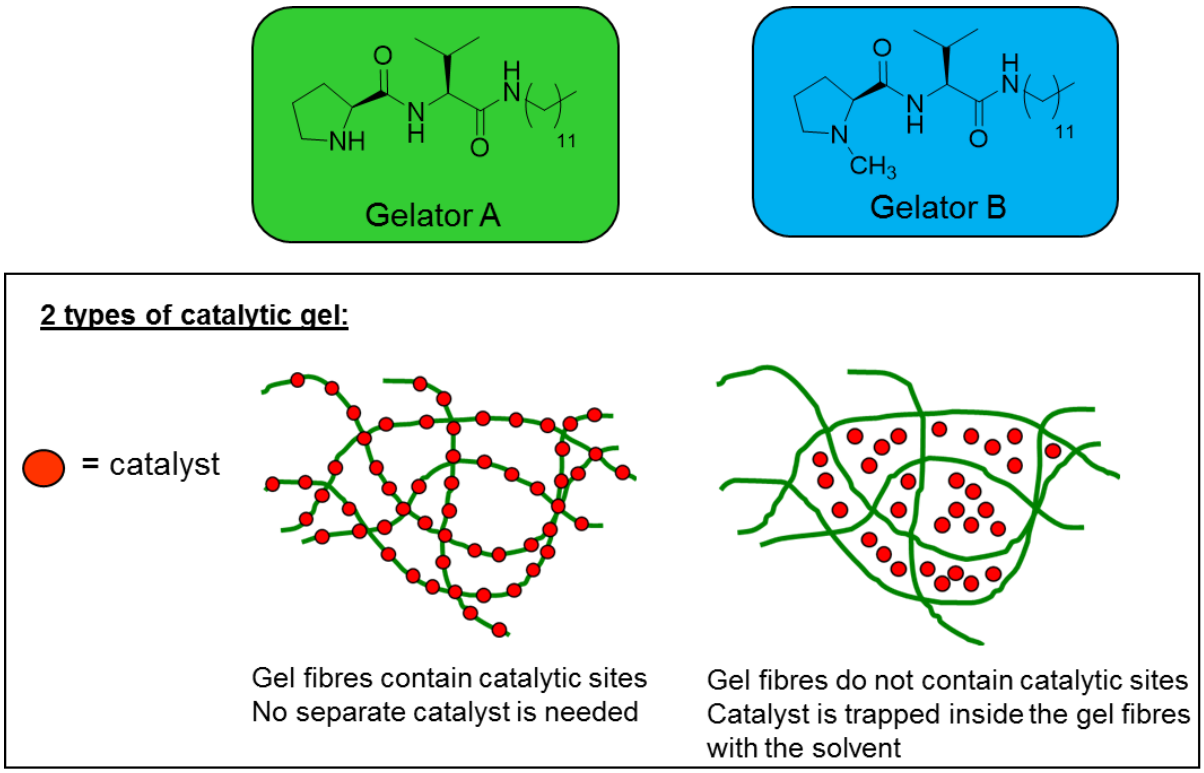
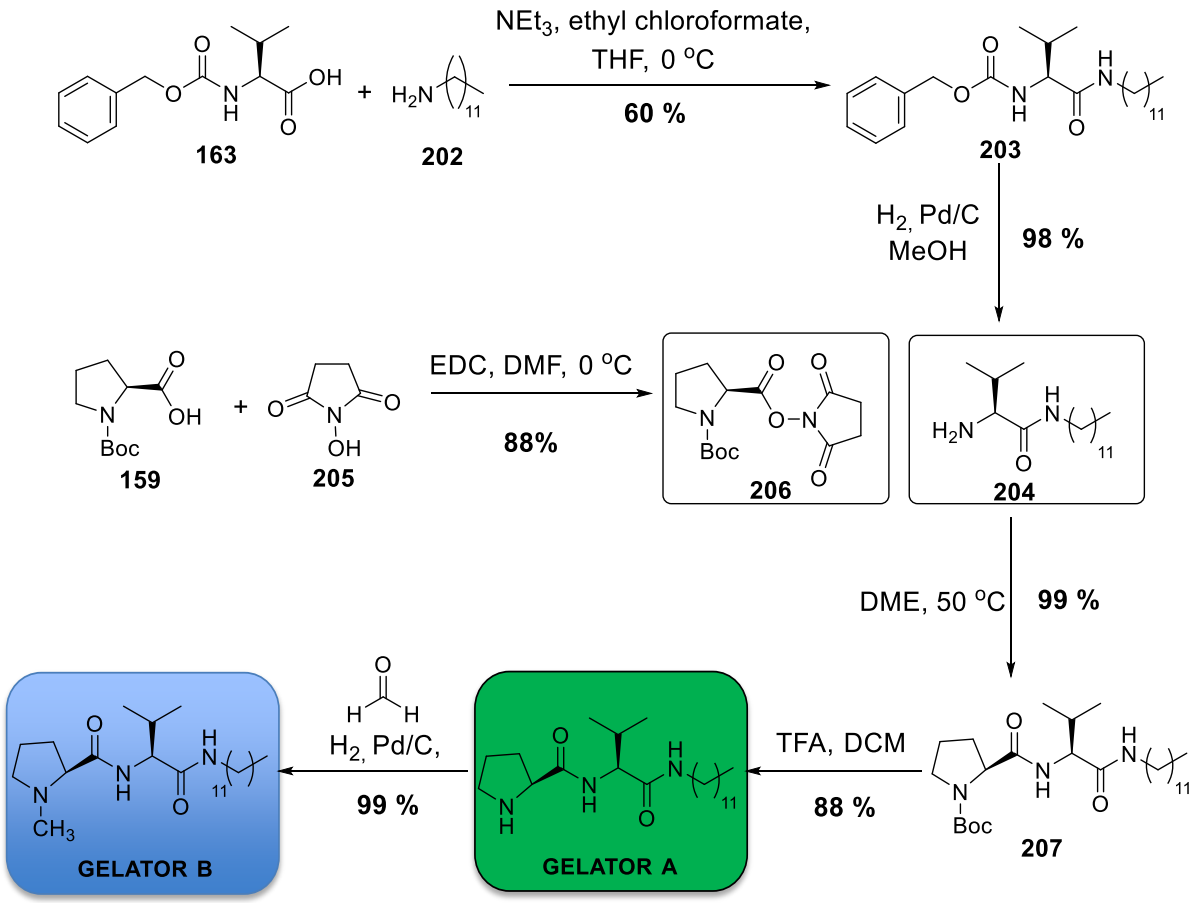


Figure 4.8. The chemical structures of **A** and **B** and the resulting gel fibres they should form.

Escuder's synthetic approach to Gelator **A** was followed to give the product as a white powder.[139] The synthesis involved the amide coupling of *Z*-L-valine **163** to dodecylamine **202** followed by hydrogenation of the Cbz group. This fragment, **204**, was then coupled to L-Pro-OSuc **206** by heating in DME (1,2-dimethoxyethane), with succinimide acting as a good leaving group, to give the dipeptide product **207** in an excellent 99 % yield. Standard Boc deprotection with TFA gave Gelator **A**. Reductive amination of Gelator **A** with formaldehyde gave Gelator **B** in an excellent 99 % yield (**Scheme 4.1**). Both Gelator **A** and Gelator **B** were obtained as white flowing powders.



Scheme 4.1. Synthesis of Gelators **A** and **B**.

Gelation of gelator **A** was attempted first. This gel has already been studied in previous papers.[147] Gelation was attempted through a heat-cool cycle, however, gelation did not occur. After several failed attempts a literature search revealed a recent publication by Escuder *et al.* that detailed the problems the group had encountered regarding the gelation of **A**.[148] The paper discussed the factors affecting the gelation and gave a table of all the conditions the group attempted. Unfortunately, in the paper, they only succeeded in gelling the hydrochloride salt of the Gelator **A,** and gelation only occurred in pH 8 buffer. Undeterred by this, a detailed investigation into the gelation of **A** was attempted taking into

account all of the variables that may affect gelation such as concentration of gelator, pH, sonication time and heating. A selection of the results is shown in

Table 4.3.

Entry	Solvent (mL)	Mass of gelator (mg)	pH	Results
1	2	2.5	8	Dispersion
2	2	5.0	8	Dispersion
3	2	8.4	8	Solid
4	4	2.5	8	Dispersion
5	4	5.0	8	Dispersion
6	4	8.4	8	Weak Gel
7	2	2.5	7	Dispersion
8	2	5.0	7	Solid
9	2	8.4	7	Solid
10	4	2.5	7	Dispersion
11	4	5.0	7	Solid
12	4	8.4	7	Solid

Table 4.3. A selection of the conditions used to attempt gelation of the HCl salt of Gelator **A**.

Having had no success, and running out of gelator, a second batch of Gelator **A** was synthesised. Gelation was successfully attempted in water with the free gelator molecule. (Gelator **A** (5 mg) in deionised water (1 mL)). Previous attempts under these exact conditions had failed. Escuder *et al.* elude to the fact that this molecule can form different polymorphs and that many of these polymorphs do not gel.[148] Once a system has formed one polymorph it is difficult to convert it to another. Possibly, this underpins the capricious nature of this particular gelator.

The reaction was then conducted under two separate sets of conditions (**Figure 4.9**). In condition **1**, the same procedure as for the agarose gel was used in which the gel was

dispersed in the sample. Again, this used a pH 7 buffer, but it was found after 24 hours the gel had completely broken apart to form a dispersion – small blobs of gel in the pH 7 phosphate buffer, and the solution had changed colour as before from colourless to yellow. The second set of conditions, **2**, removed any form of agitation from the assay and was inspired by the gel reactions of Escuder *et al.* The experiment involved forming a gel between the gelator and water on a bigger scale than previously used and then adding the reagents; acetaldehyde **24** (1 mmol) and D-glyceraldehyde **D-9** (1 mmol), in the minimum solvent (1 mL) to the top of the gel (15 mg in 3 mL water). A larger volume of gel is used here to increase the surface area in contact between gel and reagents, hence to increase the rate of diffusion into and out of the gel. The idea was that the reagent would diffuse into the gel, catalysis could then occur, and then the products diffuse back out into the phosphate buffer. After 24 hours the phosphate buffer layer was decanted and the trapping and purification repeated as normal.

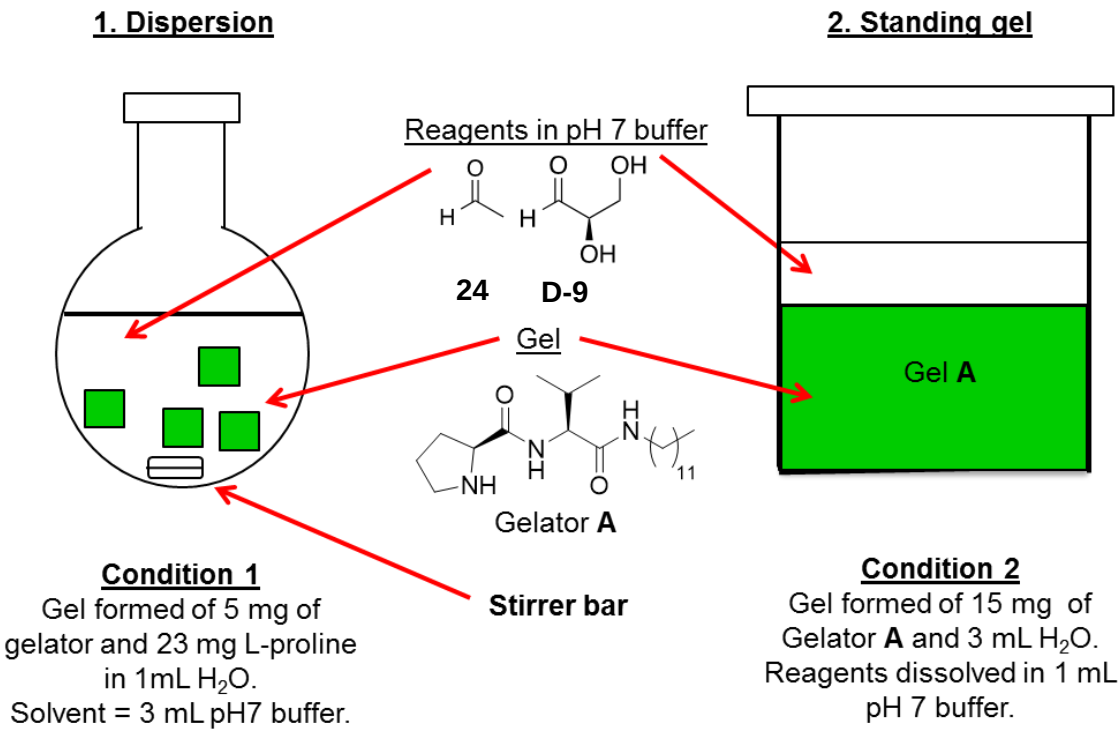


Figure 4.9. The two sets of conditions used to catalyse the aldol reaction using Gelator **A**.

Unfortunately, under either set of conditions, no pentose products were isolated in pH 7 phosphate buffer or unbuffered water (**Table 4.4**). It should be noted that in Escuder's hands the aldol reaction carried out in Gelator **A** involved dissolving the reagents in toluene, as opposed to water.[139,147] In Escuder's case the aldol reaction may be occurring at the surface of the gel where both reagents are in their favoured organic solvent. In this work no organic solvents were used.

Entry	Conditions	Gelator (%)	Vol of water in gel (mL)	pH	Yield (%)
1	1	4	3	7	0
2	2	4	3	7	0
3	2	7	3	unbuffered	0
4	2	20	8	7	0
5	2	20	8	unbuffered	0
6	Solid gelator	20	-	7	0

Table 4.4. Conditions used for the 2-deoxy-D-ribose **D-130** forming reaction using Gelator **A** as the catalyst. Conditions 1 and 2 refer to the gel set-ups shown in **Figure** 4.9.

One explanation for the failure of entries 1 and 2 is the catalyst loading. In the solution 20 mol % of catalyst was used compared to just 4 mol % in the gel examples. The gelation was therefore attempted at higher loadings, however, this proved very difficult. The gelator was very insoluble in water and needed to be heated above 100 °C for the compound to dissolve. Upon leaving the hot solution to cool the compound would usually fall out of solution as a precipitate rather than gel. Even in entries 4 and 5 (**Table 4.4**) in 8 mL of water it was possible to force gelation to occur at higher loadings but not with the entire 20 ml % of compound; there were still clumps of white solid around the edges of the gel (**Figure 4.10 A**). Therefore the gel still did not contain the full 20 mol% of the catalyst. In some

cases, after 24 hours, the gel had completely broken down (**Figure 4.10 B(ii)**) which again shows how unpredictable this gelator was. Finally, entry 6 used the powdered gelator and was stirred for 5 days in pH 7 buffer with one equivalent of each starting material and also gave no product. We can conclude from this study that gelator **A** does not catalyse the reaction.

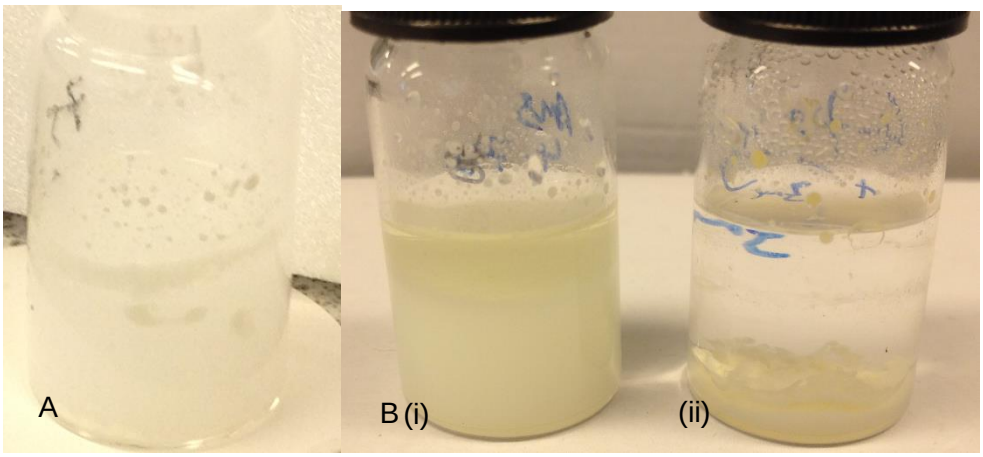


Figure 4.10 (A) Attempted gelation of Gelator A – not all of the gelator actually gels. (B) Shows the gelation reactions after 24 hours with (i) gelator **B** and (ii) gelator **A**.

Gelation of **B** was carried out by adding L-proline **L-58** (23 mg) (20 mol % in the aldol reaction) to Gelator **B** (15 mg) in deionized water (1 mL). Method 2 (**Figure 4.7**) gelation conditions were applied to give an opaque white gel upon resting (**Figure 4.10A**). Transition electron microscopy (TEM) images were obtained for Gel A and B. The images (**Figure 4.11**) show both gels consisted of similar disordered fibrous networks. The similarity of these networks shows that the free N-H of the L-proline residue in Gel A cannot be very important in terms of forming non-covalent interactions responsible for self-assembly of the gel nanostructure. This was a positive feature, as if this amine formed a vital interaction, such as a hydrogen bond within the gel network, the gel would weaken, and possibly fall apart during enamine formation and reaction catalysis.

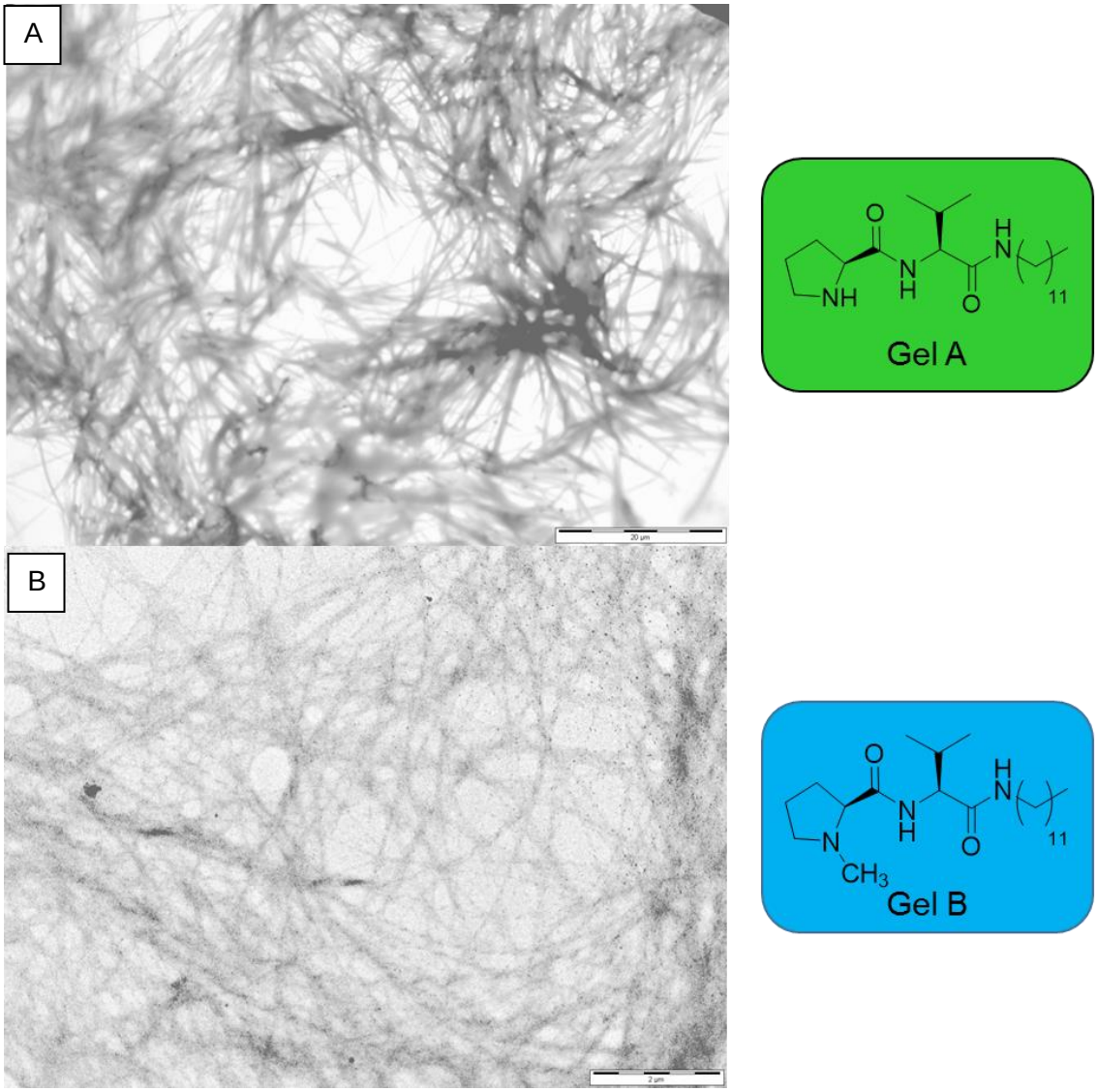


Figure 4.11. (A) TEM image of Gel **A**. Scale bar = 20μm. (B) TEM image of Gel **B**. Scale bar = 2 μm.

Escuder *et al*. postulate that the gel fibres ordered themselves with the long alkyl chains stacked on top of each other forming many weak hydrophobic interactions.[139] The hydrophilic head components were believed to orient at alternate ends, keeping the hydrophobic tails as protected as possible from the water solvent in an interdigitated bilayer assembly mode. The group also suggested key hydrogen bond interactions between amide groups may have helped strengthen the interactions between residues (**Figure 4.12**).[139]

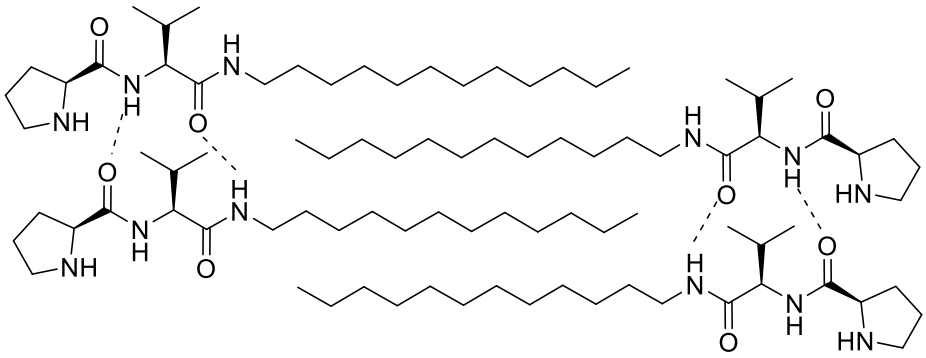


Figure 4.12. Proposed interactions of Gelator **A** by Escuder *et al.*[139]

To determine whether the L-proline **L-58** inside the novel gel was trapped within the gel fibres or directly bound to the gel fibres ^{1}H NMR experiments were conducted. The gel was formed again in an NMR tube using D_2O instead of water and spiked with a known amount of DMSO. If the catalyst was directly bound to the gel fibres there would have been no NMR signal for proline, or for the gelator in the ^{1}H NMR spectrum, as both would be immobile on the NMR timescale. The amount of L-proline **L-58** bound to the gel fibres could be calculated due to the known amount of DMSO added to the sample which was mobile within the gel nanostructure and hence visible. The ^{1}H NMR spectrum revealed the presence of L-proline **L-58** (**Figure 4.13**). Integration of the proline additive peak compared with the DMSO peak revealed 90% of L-proline was mobile within the gel. This showed that L-proline **L-58** is not to any great extent involved in the solid-like self-assembled gel network.

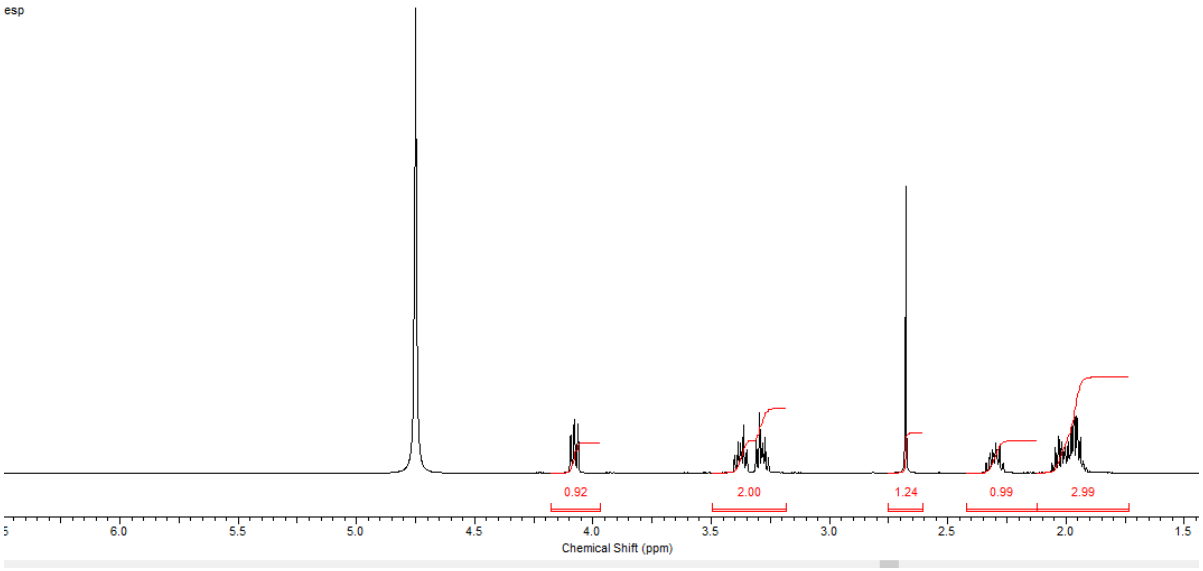


Figure 4.13. The ^{1}H NMR spectrum of the gel containing L-proline **L-58**. As the ^{1}H NMR of L-proline is visible, it is not bound to the gel fibres.

Before attempting the 2-deoxy-D-ribose **D-130** forming reaction with Gelator **B**, a blank gel control was attempted where no L-proline **L-58** was added. This revealed that the gel was indeed catalytically inactive and no 2-deoxy-D-ribose **D-130** nor its diastereomer **D-131** were formed. The reaction was then attempted gain, with L-proline **L-58** (20 mol %), using conditions 1 and 2 (Figure 4.9). The desired product was indeed formed (**Table 4.5**). The reactions were repeated with unbuffered water and no product was observed.

Entry	Gel conditions	pH	Yield (%)	Ratio (146:147)
1	1	7	< 2	1.5 : 1
2	2	7	< 1	1.5 : 1
3	1	Unbuffered	0	-
4	2	Unbuffered	0	-

Table 4.5. Reactions using Gelator **B** containing 20 mol % of L-proline. Each entry is based on 3 runs and an average of the three has been taken. Conditions **A** and **B** refer to **Figure 4.9.**

In the case of condition **2** (Entry 2), the gel changed colour from white to yellow over the course of the reaction. The yield was lower using condition **2** but this was probably because only the solvent layer was decanted off and further product may be expected to have been trapped within the gel fibres. The ratio and yield of the reactions were the same as using L-proline in solution at pH 7 therefore it can be deduced that Gel **B** does not influence yield or selectivity. One possible explanation for the similarity in results is that the catalyst leaches out of the gel into solution. This theory was easy to test with Gelator **B**. A control experiment was set up where L-proline (10 mg) was dissolved in D_2O (0.6 mL) containing DMSO (1.2 µL). The ^1H NMR is shown in **Figure 4.14** below. The ratio between the α-proton of L-proline and the DMSO signal was 1.0:1.09 (**Figure 4.14**). A gel containing Gelator **B** (15 mg), water (3 mL) and L-proline **L-58** (23 mg) was made and water (1 mL) placed on top of the gel. These conditions are identical to condition **2** (minus the reagents). After 24 hours the water layer was pipetted off, the solvent removed *in vacuo* and replaced with D_2O (0.6 mL) containing DMSO (1.2 µL). If 100 % leaching had occurred a ratio between the α-proton of L-proline and DMSO to be 1.0:0.47 would be expected, however, it was found to be 1.0:3.11 (**Figure 4.15**). This means only 15 % of L-proline had leached into the water layer. Therefore even though leaching was occurring in the gels it was of such a small amount that catalysis by L-proline inside the gel must be contributing to the synthesis of pentose sugars.

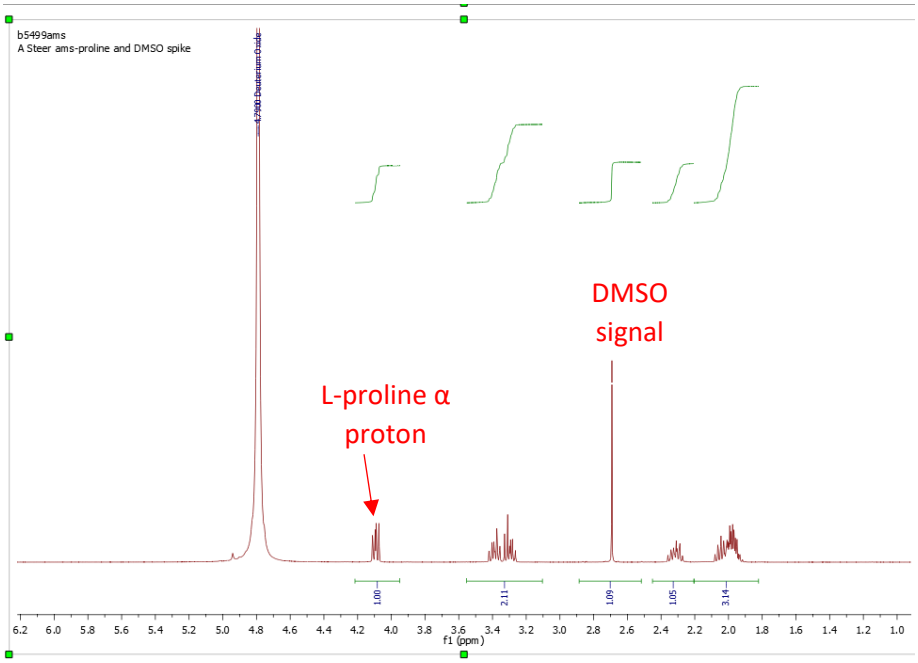


Figure 4.14. ^{1}H NMR spectrum of control experiment comparing ratios of a known amount of L-proline and DMSO.

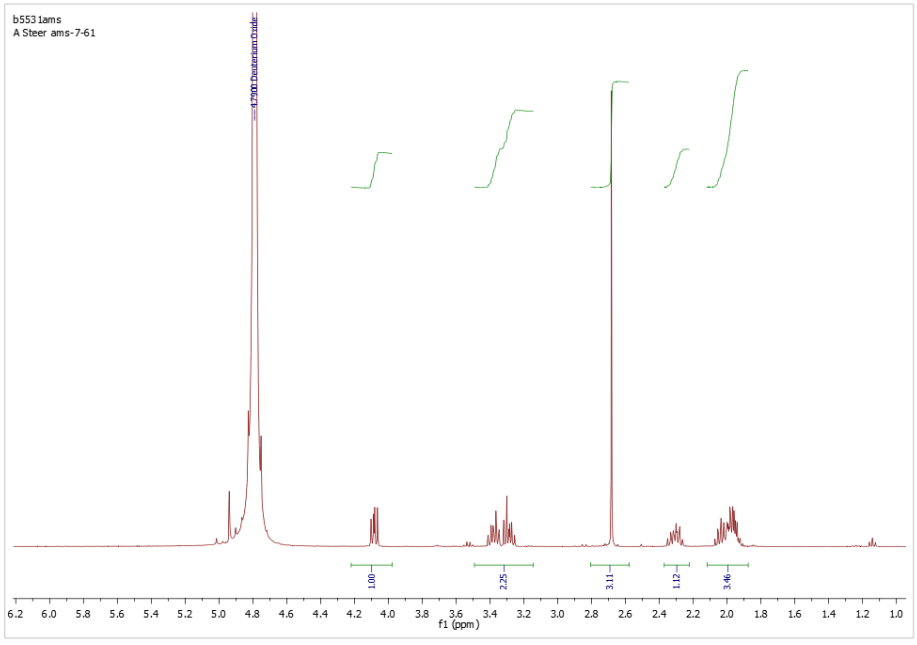


Figure 4.15. ^{1}H NMR spectrum of the leaching experiment. After 24 hours a ratio of α-proton of L-proline : DMSO was 1 : 3.11.

In an attempt to induce catalysis of the 2-deoxy-D-ribose **D-130** forming reaction on the gel fibres themselves, a third dipeptide hydrogel was constructed, Gelator **C**, containing L-proline **L-58** and L-phenylalanine. It was hypothesised that with the addition of the aromatic rings of L-phenylalanine the gelation process would be much more reliable due to π-π stacking interactions between phenyl rings. Indeed this seemed to be the case, and gelation of Gelator **C** occurred instantly after a heat-cool cycle. The synthesis of L-proline-L-phenylalanine dodecylamide (Gelator **C**) was carried out as for Gelator **A** (**Scheme 4.1**). TEM images (**Figure 4.16**) showed the fibril network formed by Gelator **C**. As with Gelators **A** and **B**, Gelator **C** formed an opaque hydrogel in deionized water, the gelation being significantly more reliable than that of **A**. The 2-deoxy-D-ribose **D-130** forming reaction was attempt using a 20 mol % loading, as before, however, Gelator **C** failed to catalyse the aldol reaction.

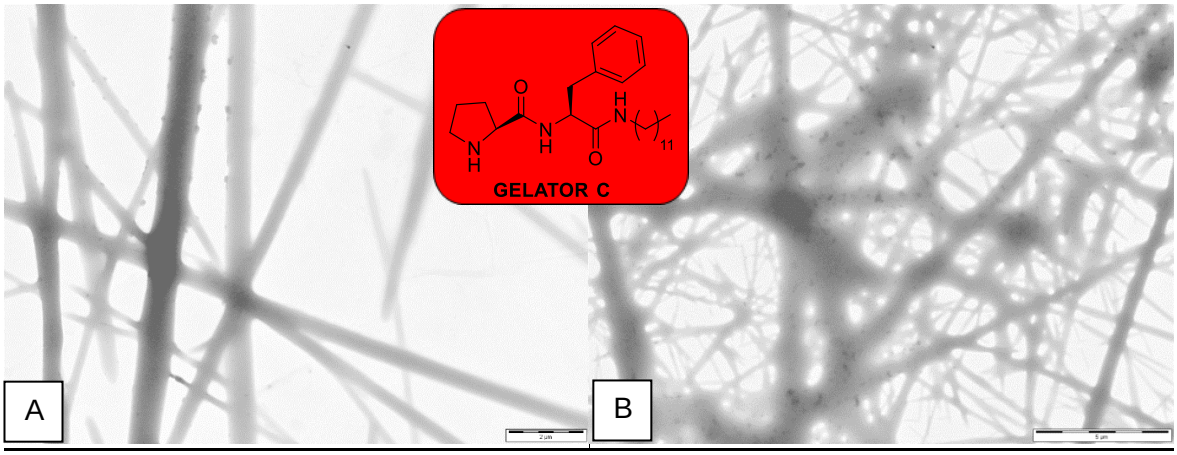


Figure 4.16. TEM images of Gelator **C**. Scale bars = (A) 2 μm (B) 5 μm.

As neither gel **A** nor **C** could catalyse the aldol reaction the problem may lie with the nature of the amino acids. Both L-valine and L-phenylalanine contain sterically bulky, hydrophobic amino acids. It could be suggested that the R-groups block or repel the hydrophilic starting

material from getting close enough to the amine of the L-proline residue for enamine catalysis to occur. Phenylalanine and valine are both classed as relatively hydrophobic amino acids with k_d values of 4.2 and 2.8 respectively, based on Kyte and Doolittle's scale of hydrophobicity.[149] If this is the case switching amino acid residues to more hydrophilic residues, providing gel formation occurs, such as lysine or arginine may facilitate enamine formation.

4.4. Conclusions

In conclusion, in this chapter vesicles and hydrogels have been discussed as potential primitive cells. The hydrogel model implemented, based on that reported by Escuder *et al*., consisted of plausibly prebiotic dipeptide hydrogels with long alkyl chains connected by an amide bond. The two distinct regions; a hydrophilic and hydrophobic region, allow the monomers to gel. Catalysis of 2-deoxy-D-ribose **D-130** formation has been shown to occur provided the catalyst is trapped within a catalytically inactive gel. Attempts to use the hydrogel network itself as the catalyst has proven unsuccessful thus far. The reason for this could have been the hydrophobic nature of the amino acids used in the dipeptide chain. This may have resulted in the hydrophilic D-glyceraldehyde **D-9** being repelled from the imine site preventing the electrophile from getting close enough for the enamine to attack. In order to succeed in the gel promoted catalysis of 2-deoxy-D-ribose **D-130** more hydrophilic amino acids, such as arginine and lysine, should be used in place of valine. These amino acids would have the added benefit of another amine functional group for a second potential site of catalysis.

5. Further investigations into hydrogel catalysis

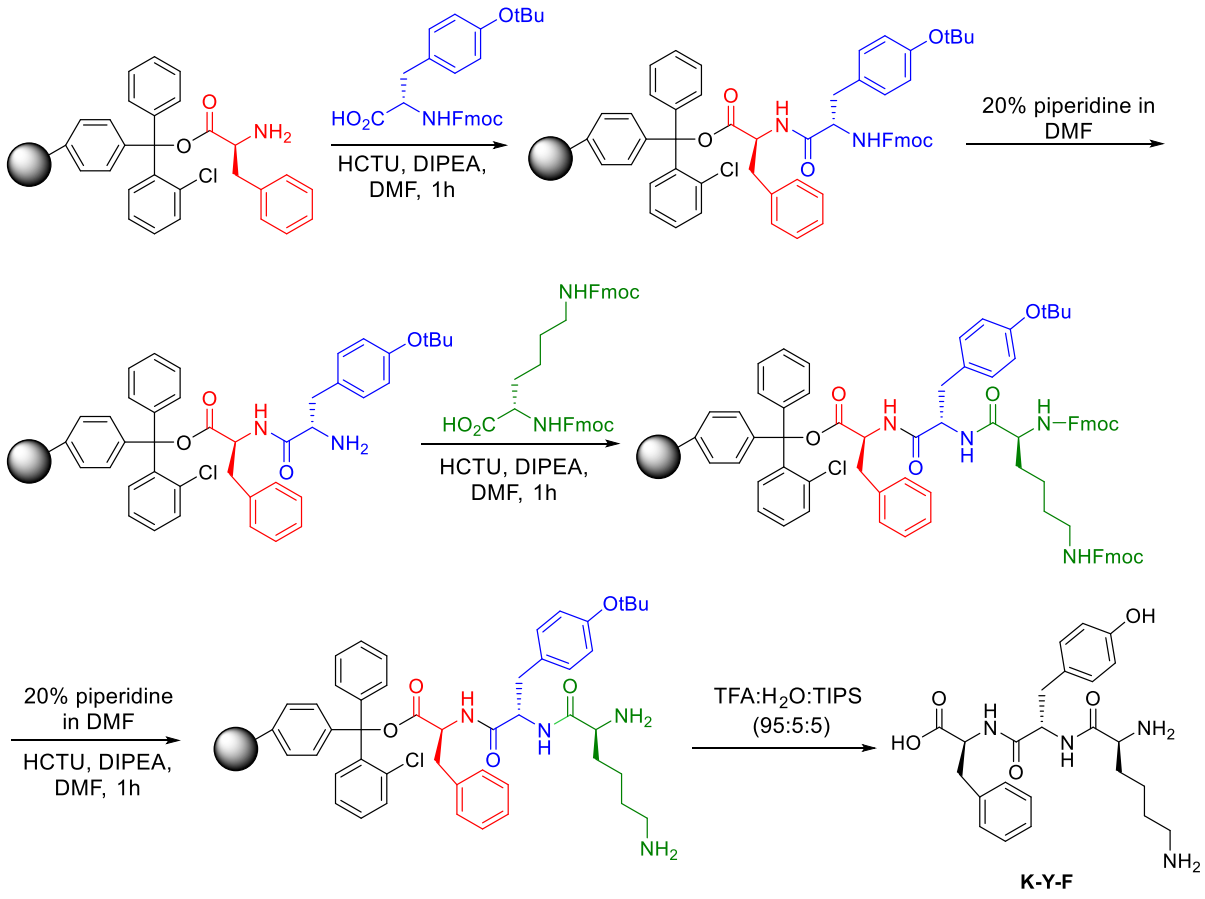
5.1. A tripeptide hydrogel candidate

Chapter 4 detailed the degree of success in using Gelator **B** as a non-catalytic gel within which starting material and catalyst could be mixed together carry out the aldol reaction to form 2-deoxy-D-ribose **D-130**. Unfortunately attempts to use catalytically active gel fibres to catalyse the reaction had failed. This chapter concerns further efforts to develop a catalytically active "prebiotic protocell" environment for the 2-deoxy-D-ribose **D-130** forming reaction.

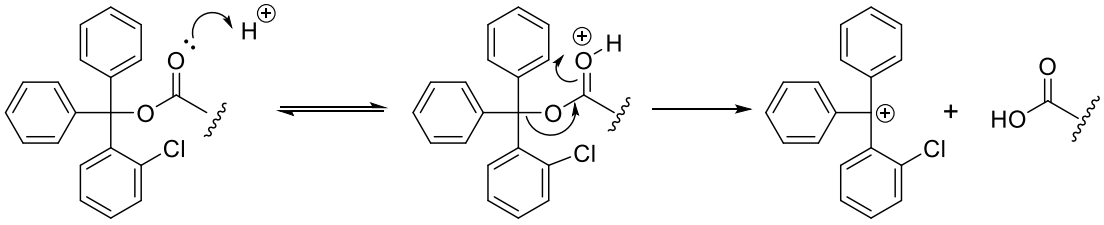
Continuing the theme of small peptide gelators, a study was published in 2015 by Frederix *et al.* detailing an extensive computational investigation of the gelling potential of a wide range of tripeptide molecules based on a number of criteria.[150] The highest scoring tripeptides were then synthesised and their gelling ability tested. The lead candidate from this study was NH_2-L-lysine-L-tyrosine-L-phenyalanine-OH (K-Y-F) with the authors suggesting that the π-stacking ability of the tyrosine and phenylalanine aromatic rings plays a large part in enabling gelation. From a prebiotic perspective tripeptides, with just two amide linkages, are plausible prebiotic molecules on the early Earth.[151] K-Y-F also contained two primary amines on the lysine residue and hence two potential catalytic sites.

The synthesis of K-Y-F was carried out *via* solid phase peptide synthesis (SPPS). This method had the benefit of growing the peptide chain on resin by attaching an Fmoc-protected amino acid to the resin followed by deptrotection of the Fmoc-group using a 20 % solution of piperidine in DMF. This procedure could be repeated to rapidly build up the peptide chain, with the key benefit of simplicity of purification. As the peptide remains bound

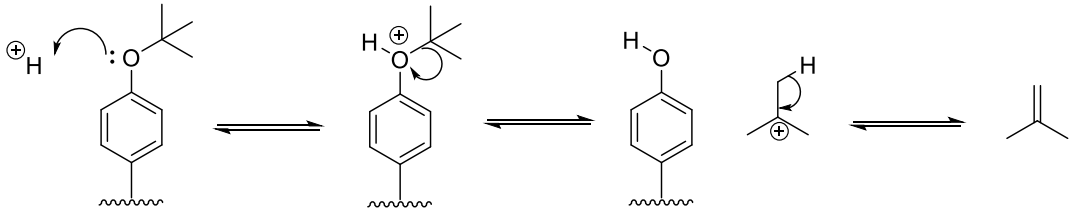
to the resin beads, purification could be achieved through washing with copious amounts of solvent after each synthetic step. For the synthesis of K-Y-F, commercially available chlorotrityl resin was used as the resin of choice. The peptide was easily cleaved from the resin in the final step of the reaction using a strong acid such as TFA. This allowed for deprotection of the *t*-butyl group on tyrosine and cleavage to occur in the same step. A standard solution of 95:5:5 TFA : DCM : TIPS was used for cleavage from resin, and the peptide product was filtered into ice cold ether where upon the peptide precipitated out of solution as a white powder. The synthesis of the tripeptide is shown in **Scheme 5.1** below and the mechanism of cleavage and deprotection is shown in **Scheme 5.2**.



Scheme 5.1. Solid phase peptide synthesis of the tripeptide K-Y-F.



(A) Cleavage from chlorotrityl resin



(B) Cleavage of *t*-butyl protecting group.

Scheme 5.2**.** Dual usage of TFA for t-butyl deprotection and cleavage from chlorotrityl resin.

As the tripeptide had various ionisable groups the pK_a of each group needed to be investigated to determine whether at approximately pH 7 either of the lysine N-termini would be in the neutral non-protonated "NH_2" form and thus able to act as possible catalytic sites for the 2-deoxy-D-ribose **D-130** forming reaction. A pH titration was therefore carried out using KYF (10 mg) in 12mM NaOH (5 mL). This was titrated against 0.01 mM HCl and the resulting graph of volume of acid added vs pH is shown **Figure 5.1**.

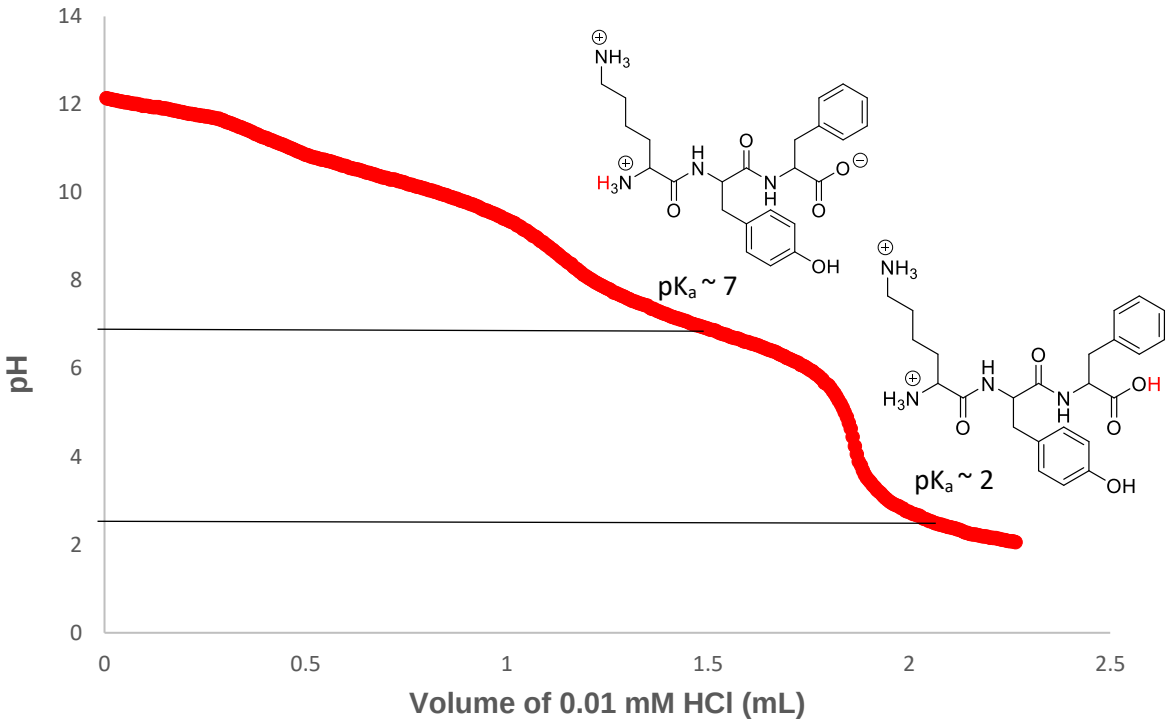


Figure 5.1. A pH titration curve of 10 mg of K-Y-F in 12 mM NaOH titrated with 0.01 mM HCl at a rate of 0.1 M per min.

Four ionisable groups were expected from K-Y-F; the C-terminus, the two NH_2 groups and the tyrosine OH. This is summarised in **Figure 5.2** below. The initial phase of the titration was difficult to interpret with no distinct vertical drops in pH. In this region, between 0 and 1 mL of acid added, the protonation events of the lysine R group and the tyrosine R group occur with relatively similar pK_a values. The next protonation event occured at the N-terminus which, from the graph, has a pK_a of approximately 7. The final protonation event was the protonation of the carboxylate anion represented by the steep drop in pH after the addition of approximately 1.8 mL of acid which corresponds to a pK_a of around 2. From this data it was proposed that if the gelation conditions were around pH 7 then there should be at least one NH_2 in the neutral state with the potential to carry out enamine catalysis.

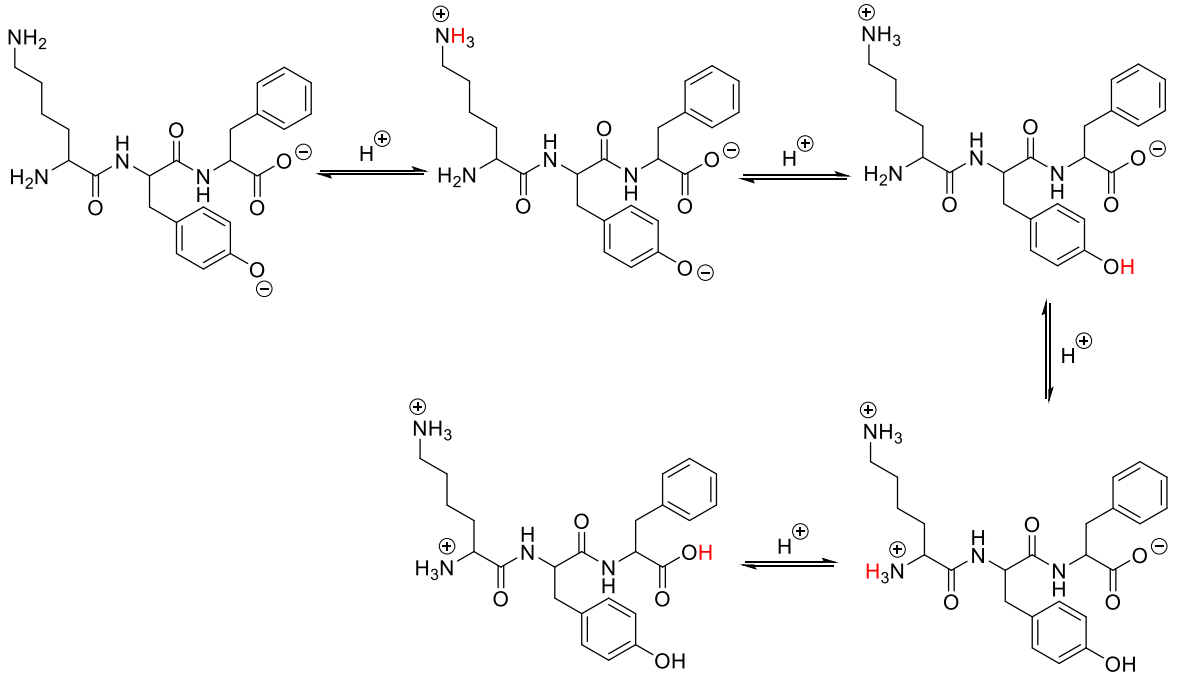


Figure 5.2. Protonation events of KYF with increasing acidity.

Gelation of K-Y-F was attempted, as described in the Frederix *et al.* paper, *via* a method of gelation known as pH switching[150], unlike previous gels formed in Chapter 8 where an insoluble solid was heated in a solvent until homogeneous and gel formation occurred upon cooling. Tripeptide K-Y-F dissolved readily in pH 7 phosphate buffer, where gelation did not occur. Dropwise addition of 0.1M sodium hydroxide to this solution caused instant gelation. Gelation occurred so quickly that if the sodium hydroxide was applied directly to the surface of the solution, gelation would only take place on the surface creating a layer of hydrogel with free-flowing solvent trapped underneath (**Figure 5.3 B**). In order to achieve full gelation of the solution, simultaneous addition of both the gelator in pH7 buffer and 0.1M sodium hydroxide was attempted which created a more uniform, translucent gel (**Figure 5.3 A**). Gelation of K-Y-F, therefore, occurred in basic conditions but not at pH 7.

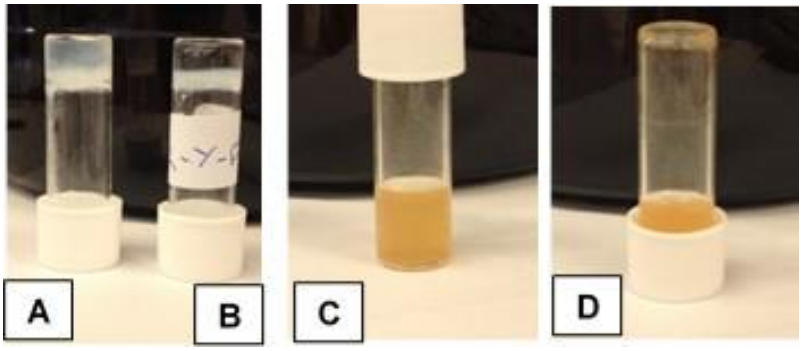
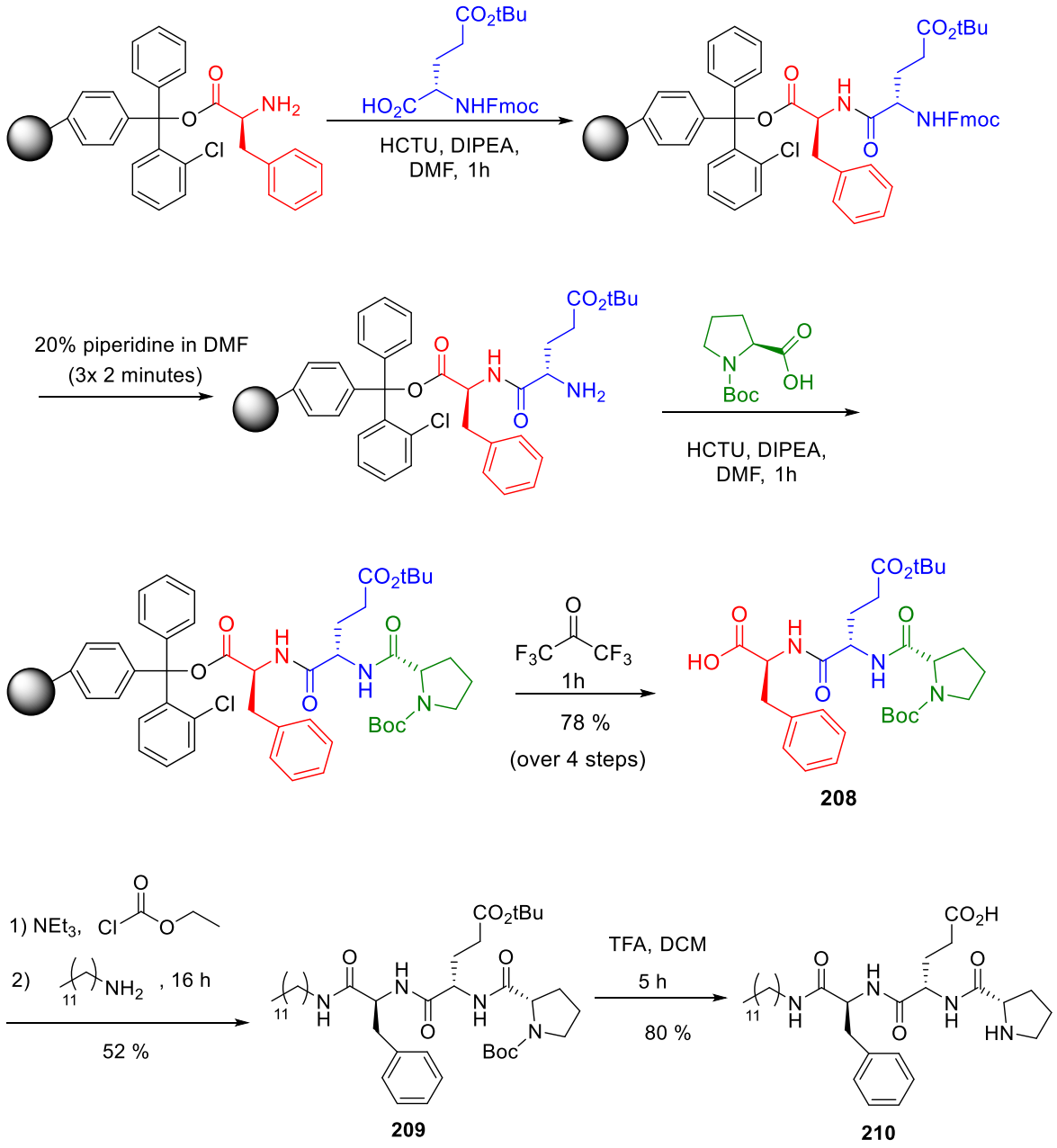


Figure 5.3. (A) Gelation of K-Y-F via simultaneous addition of K-Y-F in pH 7 buffer and 0.1M NaOH solution. (B) Gelation of K-Y-F by addition of 0.1M NaOH solution into a solution of K-Y-F in pH 7 phosphate buffer. (C) K-Y-F hydrogel immediately after addition of starting materials to surface of gel. (D) K-Y-F no longer a hydrogel after 1 minute since addition of starting materials.

Unfortunately due to the sensitivity of the hydrogel to pH change, the gel broke down within one minute of adding the reagents (acetaldehyde **24** and D-glyceraldehyde **D-9**) in the minimum amount of solvent (0.1 mL) to the surface of the gel (**Figure 5.3 C** and **D**). This is probably because the polar reagents interacting with the gelator disturbing key interactions between gelator monomers causing the gel to collapse, or lowering the pH of the gel and hence reversing the gelation process. It appeared that the reaction was actually catalysing a trace amount of 2-deoxy-D-ribose **D-130** formation, however upon completing a control experiment with just sodium hydroxide and no gelator, we saw that the same level of catalysis occurred indicating that the sodium hydroxide base was catalysing the formation of the trace amounts of product and not the tripeptide. A possible reason for the failure of the reaction could be the amount of gelator used. Only experiments using up to 10 % gelator were used. This is because 0.1 mmol (10 % catalyst loading) was the equivalent of 40 mg and it was impractical to gelate the usual 0.2 mmol (80 mg) in 3 mL solvent.

5.2. Supramolecular structures of a tripeptide amide

During the project Escuder *et al.* published a demonstrating som e tripeptide hydrogelators as catalysts in aldol reactions.[152] These gelators had 12 carbon chains attached by an amide bond to the C-terminus of a tripeptide. This seemed like a logical progression from the tripeptide K-Y-F candidate and so the synthesis of these molecules was conducted. The first candidate; L-proline-L-glutamic acid-L-phenylalanine dodecylamide **210** (P-E-F dodecylamide) (**Scheme 5.3**) was synthesised by a mixture of solid phase and solution phase peptide synthesis. Firstly, the tripeptide fragment was made, as in the case of K-Y-F, by using commercially available chlorotrityl resin preloaded with Fmoc-phenylalanine using a series of deprotections and couplings (**Scheme 5.3**). The next step was to remove the tripeptide from the resin. This time instead of using TFA, which would also remove the *t*-butyl protecting group of glutamic acid and the Boc group of proline, hexafluoroisopropanol (HFIP) was used. HFIP is a much milder acid than TFA with a pK_a of 9.8 (compared to -0.3 for TFA) and therefore was only strong enough to cleave the tripeptide from the resin leaving the protecting groups intact. Purification of the tripeptide was the same as for K-Y-F obtaining the compound **208** as a white solid. The final stage of the synthesis was to attach dodecylamine to the peptide C-terminus. Amide coupling was carried out, as before, using ethyl chloroformate to form an anhydride and then subsequent attack with dodecylamine to give **209**. This was a rather low yielding step (52 %) due to the difficulty of forming the anhydride, with the rest of the starting material recovered. The final step, double deprotection with TFA, gave the desired product as the TFA salt **210**.



Scheme 5.3. Synthesis of P-E-F dodecylamide **210**.

With the hydrogelator candidate in hand a series of gelation experiments were conducted varying the mass of gelator, volume of solvent and pH of solvent. The results of these attempts are shown in **Table 5.1**. Unfortunately it was not possible to form gels with **210** in aqueous media. At this point Escuder *et al.* published their gel studies and also found that

P-E-F dodecylamide **210** did not gel, but instead formed an opaque solution in phosphate buffer.[152]

Entry	Mass (mg)	Volume of solvent (mL)	Solvent	Result
1	10 mg	1	Water	Upon heating compound dissolved to give a colourless solution
2	20 mg	1	Water	Upon heating compound dissolved to give a colourless solution
3	10 mg	1	pH 7 phosphate	Upon heating compound dissolved. Upon cooling gave an opaque solution
4	10 mg	2	pH 7 phosphate	Upon cooling compound dropped out of solution as a white precipitate.
5	10 mg	1	pH 8 phosphate	Upon cooling compound dropped out of solution as a white precipitate.
6	2.8 mg	1	pH 7 phosphate	Upon heating compound dissolved. Upon cooling gave an opaque solution, eventually solid precipitated out.
7	2.8 mg	0.5	pH 7 phosphate	Upon heating compound dissolved. Upon cooling gave an opaque solution, eventually solid precipitated out.

Table 5.1. Gelation attempts of P-E-F dodecylamide **210**.

Interestingly, in Entries 1 and 2, when deionised water was used as the solvent the TFA peptide salt dissolves readily, with 20 mg of the compound dissolving in 1 mL of water. As gelation was not occurring an investigation into possible supramolecular structure alternatives was conducted.

5.3. Evidence of Micelle Aggregation

It is a logical hypothesis that P-E-F dodecylamide **210** molecules could arrange themselves in aqueous media to form self-assembled micellar structures. The amide consists of two distinctly different sections; a hydrophilic “water-loving” section made up of the amino acid residues and a hydrophobic “water-hating” greasy dodecyl tail. In aqueous medium the monomers would arrange in such a way as to minimise the number of interactions between the aqueous solvent and the hydrophobic section of the molecule. One way to achieve this is to form a micelle with all of the polar heads pointing out at the solvent as shown in **Figure 5.4**.

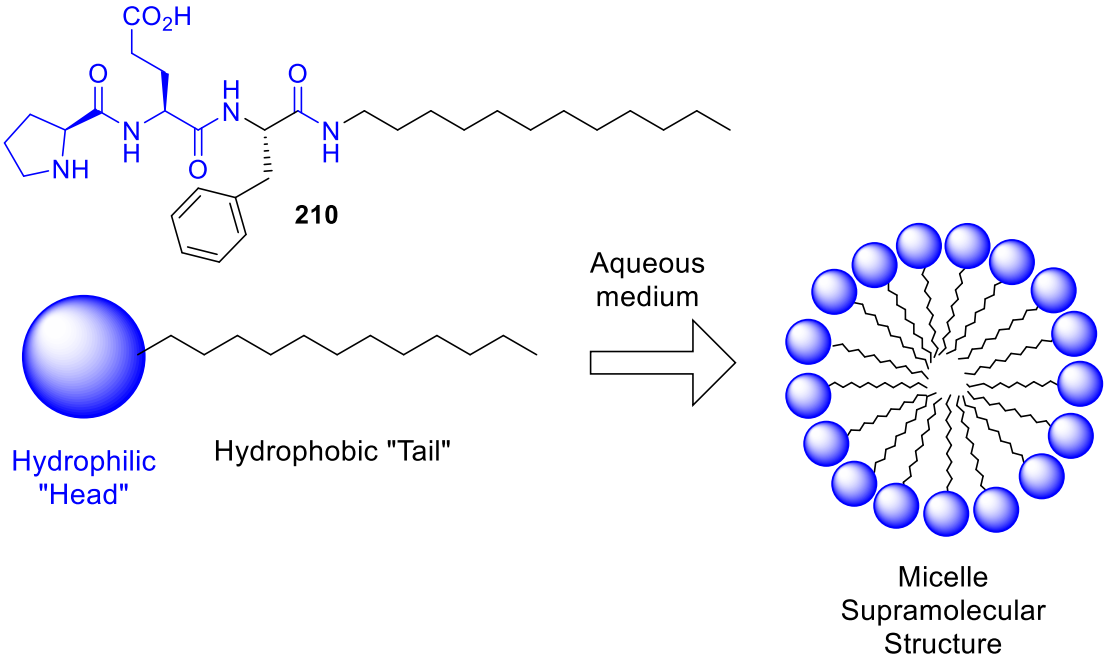


Figure 5.4. Likening of P-E-F dodecylamide **210** to an amphiphile and possible arrangement to a micelle in aqueous medium.

To probe this hypothesis Dynamic Light Scattering (DLS) was initially used. This is a light scattering technique used to measure the size of supramolecular structures from proteins and polymers to micelles and colloids. In simple terms, a sample, in solution, is illuminated with a laser beam and the fluctuations of the scattered light are detected at a specific angle, the scattering angle. This technique can be used to measure structures with diameters as small as 1 nm. The results of the experiment are depicted as two different distribution charts. The intensity distribution (**Figure 5.5**) is absolute and measures the amount of light scattering. In this case the largest intensity is around 53 nm. However, this was not a true reflection of the sample as any larger supramolecular structures scatter more light and hence will have a greater intensity. The second chart (**Figure 5.6**) shows the volume distribution which takes the scattering of larger structures into account to give a representation of how much of each structure is in the sample. In this case a maximum intensity at around 6 nm was seen with a small shoulder showing a limited number of larger supramolecular structures.

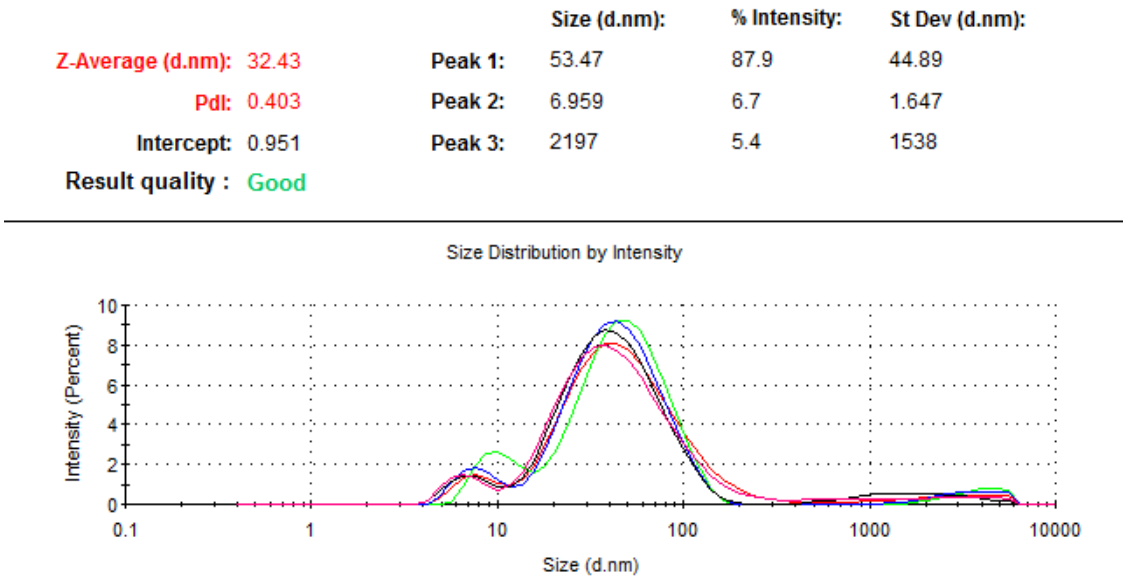


Figure 5.5. Size distribution by intensity graph created from DLS experiments of P-E-F dodecylamide **210** in deionised water. This graph is based on 5 runs of the experiment.

		Size (d.nm):	% Volume:	St Dev (d.nm):
Z-Average (d.nm): 32.43	Peak 1:	6.084	86.1	1.664
PdI: 0.403	Peak 2:	21.56	13.9	11.95
Intercept: 0.951	Peak 3:	1255	0.0	481.4
Result quality : Good				

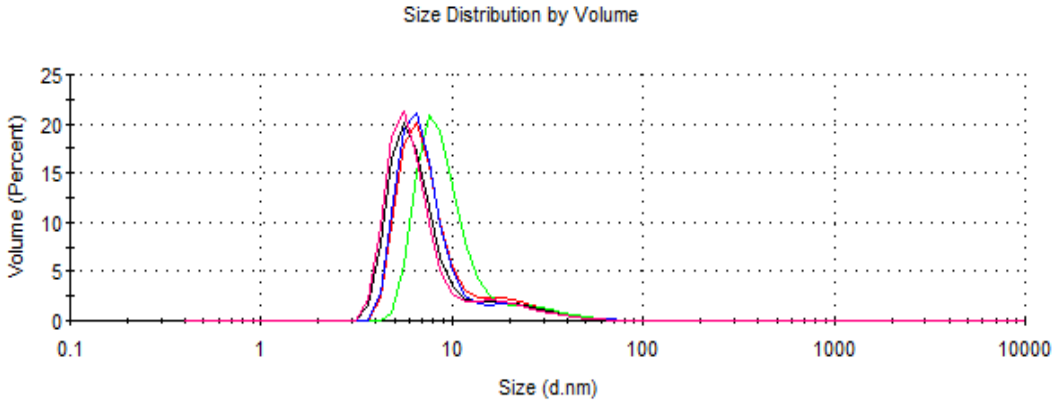


Figure 5.6. Size distribution by volume graph created from DLS experiments of P-E-F dodecylamide **210** in deionised water. This graph is based on 5 runs of the experiment.

The average diameter of 6 nm shown in **Figure 5.6** is a good fit for the P-E-F dodecylamide **210** amphiphiles. As a rough calculation, if we assumed each bond length takes up approximately 0.1 nm of space then the distance from the end of the alkyl chain to the top of the L-proline residue would be 23 bonds. As the diameter of the micelle includes two molecules this is 4.6 nm. Factoring in the solvent shell of the micelle and extra space for inefficient packing, 6 nm seemed a reasonable diameter for micelle formation. The peak at larger diameter in the size distribution (**Figure 5.5**) could arise from larger supramolecular structures. For example, it could be that in solution some cylindrical gel fibres form, however this was not the dominant supramolecular structure and hence gelation of the solvent did not occur but some long cylindrical fibres may have existed (**Figure 5.7**). The ratio of these structures would also be expected to depend on concentration.

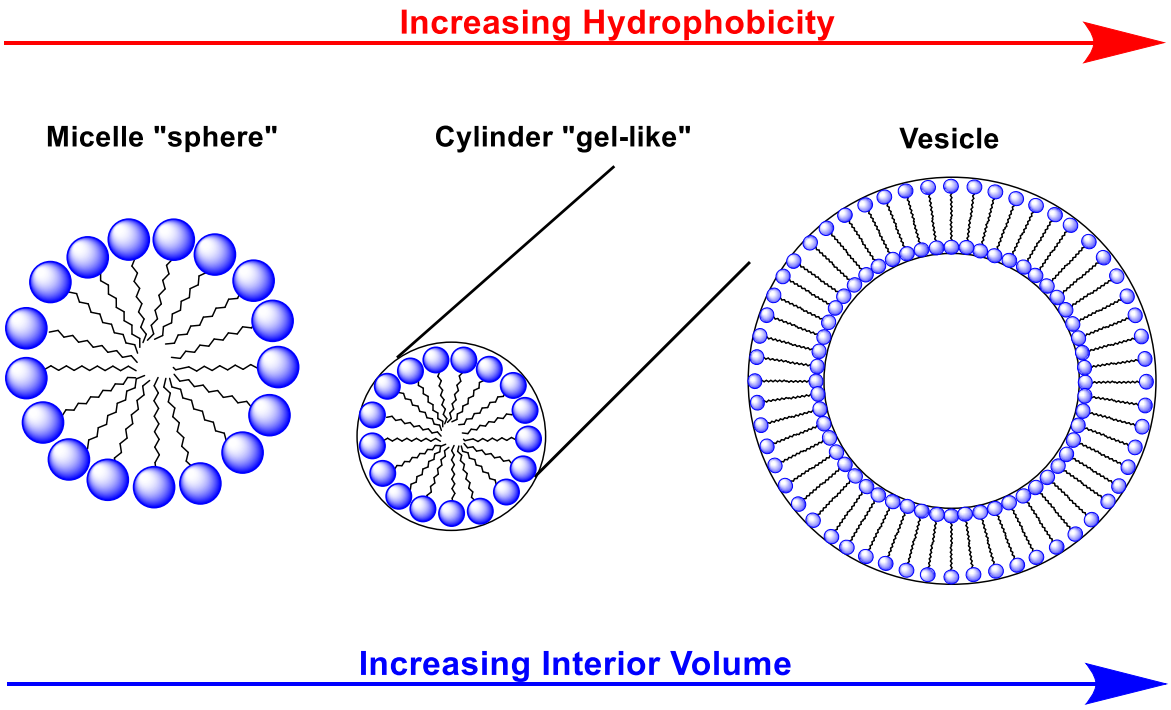
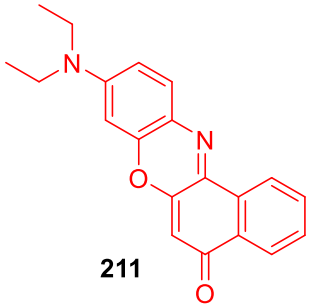


Figure 5.7. Different supramolecular structures and their link between increasing hydrophobicity and increasing interior volume.

In order to gain more evidence for formation of self-assembled nanostructures, a Nile Red **211** (**Figure 5.8**) encapsulation assay was performed. This technique has been widely used previously by other groups to determine the critical aggregation concentration of amphiphiles in aqueous medium – that is the concentration at which amphiphiles spontaneously self-assemble to form micelles.[153,154] Nile Red **211** is an uncharged heterocycle, the fluorescence of which depends on its environment. In an aqueous, hydrophilic, environment the fluorescence is quenched. In a hydrophobic environment, such as the interior of a micelle the fluorescence is enhanced at 635 nm. Therefore by mixing with various concentrations of the monomer in aqueous solution allows determination of the concentration in which micelle aggregation takes place.



211

Figure 5.8. Nile Red **211**.

The encapsulation study used phosphate buffered saline (PBS) at pH 7.5 as the aqueous medium. A starting stock solution containing **210** was made and aliquots of this were taken and diluted to known concentrations. Nile Red **211** (1 μL) was dissolved in ethanol (2.5 mM) and added to the solution of **210** and the fluorescence intensity at 635 nm measured. This assay was performed over a range of P-E-F dodecylamide **210** concentrations and a plot of the log of the concentration against absorbance was constructed and the critical micelle concentration (CMC) determined as the point of inflection. This is the point at which micelles starting forming, encapsulating Nile Red **211**, and hence increasing the fluorescence intensity observed. (**Figure 5.9**). The grey point on the graph was omitted from both trend lines as this was the point with maximum curvature; including this in either trend line would alter the CMC by approximately 5 μM. By solving the simultaneous equations of the two tend lines the CMC was determined as 27 μM ± 2 μM. The relatively small value showed that P-E-F dodecylamide **210** could efficiently aggregate and forms micelles in aqueous media.

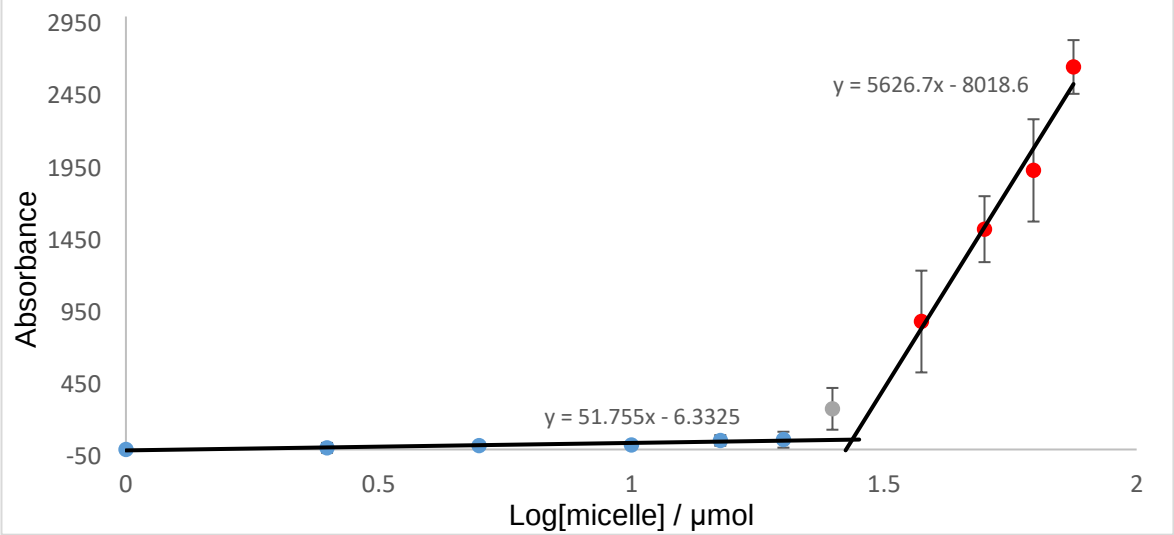


Figure 5.9. CMC plot of Nile Red **211** assay using P-E-F dodecylamide **210**. The CMC is determined as the point of inflection.

The final piece of evidence for micelle formation came from transition electron microscopy (TEM) imaging. P-E-F dodecylamide **210** (0.75 mg) was dissolved in deionised water (1 mL). The sample was dried and then exposed to uranyl acetate as a means of negative staining, and TEM images were taken (**Figure 5.10**). **Figure 5.10A** shows multiple dark regions indicating that a supramolecular structure had been stained. **Figure 5.10B** shows a zoomed in image of one of these structures. These shapes are larger than the diameter of the micelles calculated from DLS and it was assumed that the drying procedure had caused the micelles to clump together into large conglomerates. Therefore the size shown by the scale bars in **Figure 5.10** does not reflect the true diameter of the micelles. Careful inspection of these agglomerates does suggest some sub-structuring consistent with this view.

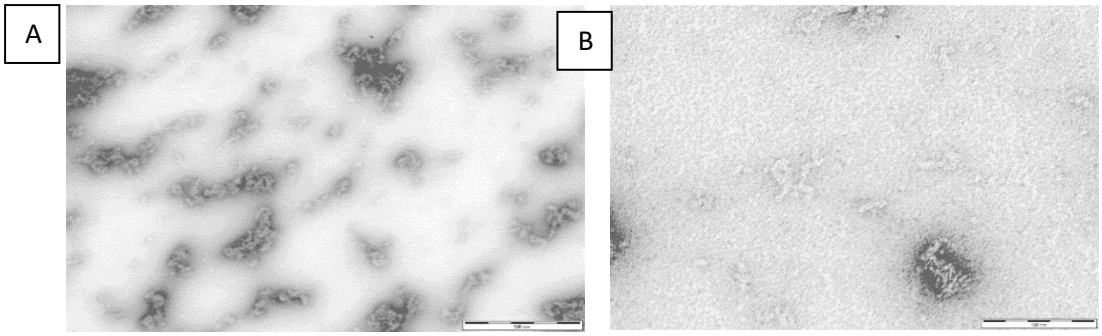
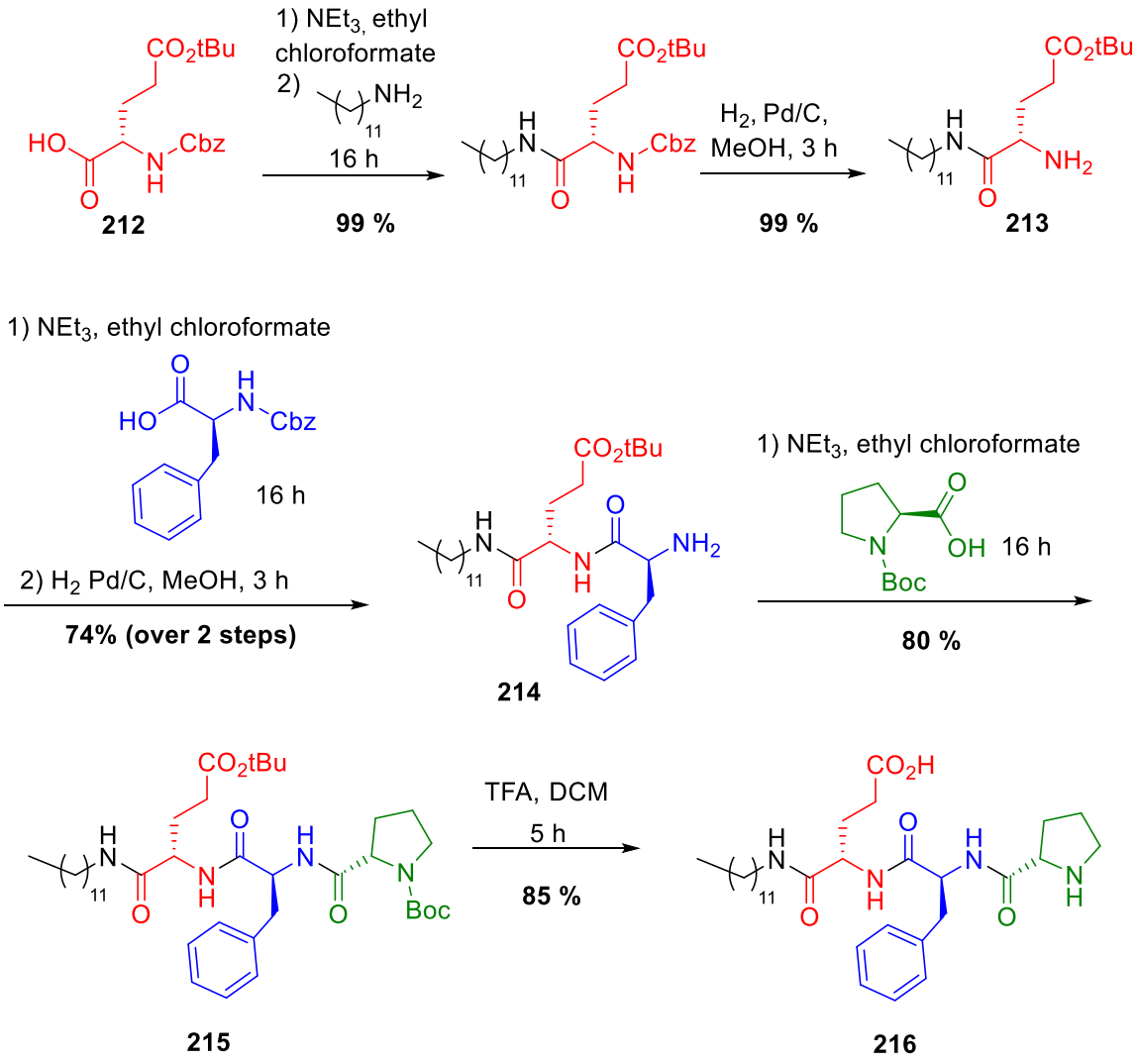


Figure 5.10. (A) TEM image of PEF-amide **210** in water at a concentration of 0.75 mg/mL stained with uranyl acetate. Scale bar is 500 μm. (B) A zoomed in version of the TEM image, the scale bar is 100 μm.

With evidence to support the view that P-E-F dodecylamide **210** spontaneously forms micelles in aqueous solution, the next step was to test if these micelles could catalyse the formation of 2-deoxy-D-ribose **D-130** from D-glyceraldehyde **D-9** and acetaldehyde **24**. The reaction was performed as usual but the catalyst was replaced with P-E-F dodecylamide **210**. Due to the large molecular weight of the hydrochloride salt (595 g / mol), only a 10 % catalyst loading was used. This was equivalent to 60 mg of **210**, well above the CMC. Following 5 minutes of sonication the majority of the salt dissolved in deionised water (3 mL). The reagents were added to the saturated peptidic solution and the reaction stirred slowly for 24 hours. At the end of the experiment the reaction vessel contained an opaque orange mixture. The trapping and purification procedures were carried out as usual but unfortunately no product was found. The reason for this may simply be that 10 % catalyst loading was not sufficient to promote the reaction after 24 hours or possibly the proline residue was protonated in the deionised water medium preventing formation of the crucial imine intermediate, or the site of catalysis may have been sterically blocked by the neighbouring R-group of glutamic acid.

5.4. Changing the amino acid sequence

At the time of the micelle work, Escuder *et al.* did show that reversing the glutamic acid and phenylalanine residues, to give P-F-E-dodecylamide **216**, did form a gel network in pH 7 buffer. In an attempt to replicate this, **216** was synthesised. One limitation of solid phase synthesis was the difficulties in scalability to gram quantities with the equipment at hand and so a solution phase approach was used instead. This is shown in **Scheme 5.4**.



Scheme 5.4. Solution phase synthesis of P-F-E dodecylamide **216**.

The synthesis of P-F-E dodecylamide involved simple repetition of amide coupling and deprotection. For the amide coupling steps ethyl chloroformate was used to form the reactive anhydride intermediate. Each of the steps of the synthesis was high yielding and 900 mg of the Boc-protected tripeptide amide **215** was made with ease.

When attempting the gelation of **216** a range of conditions were used, varying aqueous media (pH 7 buffer and water), mass of gelator and volume of solvent. Unfortunately gelation was not possible under any of these conditions. A pH titration was carried out (**Figure 5.11**) using the same conditions as Escuder *et al*. who reported pK_a values of 9.72 and 6.85 compared to the experimental values here of approximately 10 and 7 which showed no noticeable difference in pK_a values.[152] It could be that, in the laboratory, a different polymorph of the product had formed which did not favour gelation. This would have some precedent as previously noted in a key paper by Escuder.[148] The synthesis of **216** was towards the end of the project and due to lack of time this synthesis was unable to be repeated.

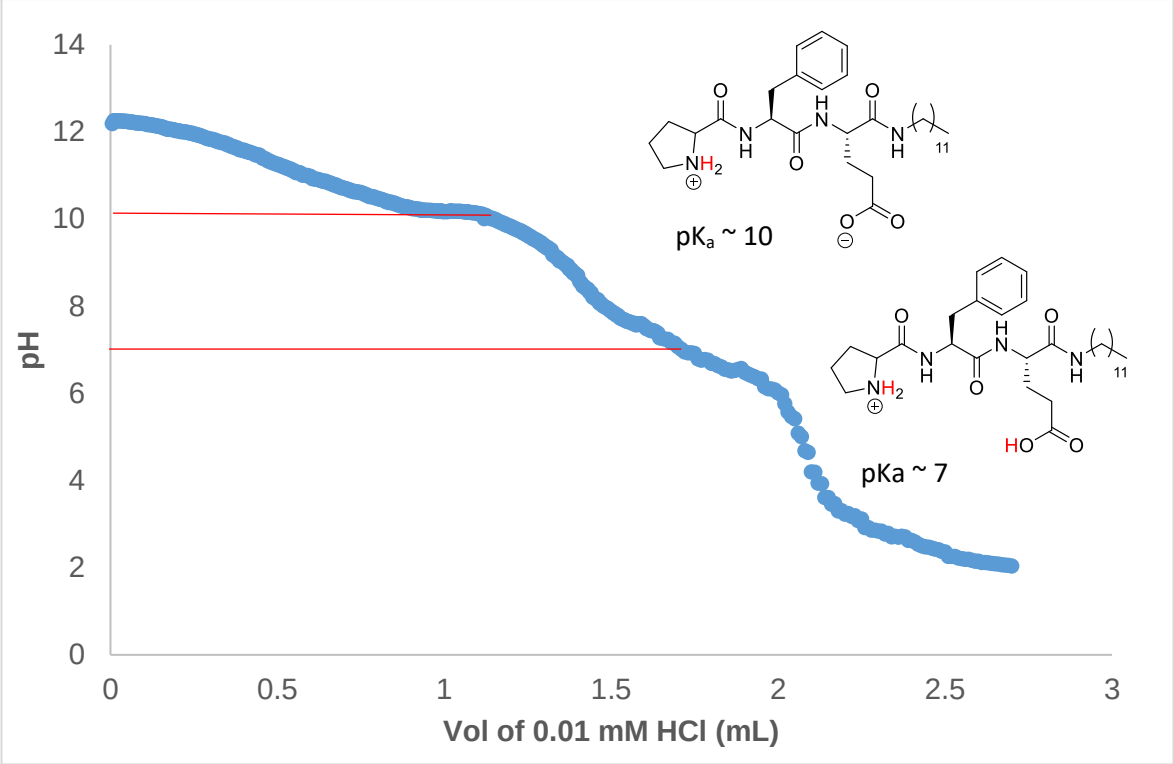
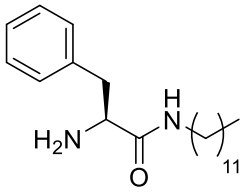


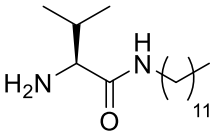
Figure 5.11. A pH titration curve of 10 mg of PFE-amide **216** in 12 mM NaOH titrated with 0.01 mM HCl at a rate of 0.1 M per min.

5.5. Single amino amides for supramolecular catalysis

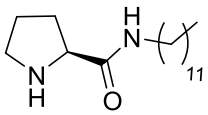
Whilst attempting the gelation of the tripeptide compounds a serendipitous gelation discovery was unearthed. Three simple amino amides were found to spontaneously assemble into supramolecular structures. The benefit of these systems is that a 20 mol % catalyst loading (for the 2-deoxy-D-ribose-forming reaction) could be achieved with a much smaller mass of compound, compared to the tripeptide candidates. **Table 9.2** summarises the supramolecular properties of L-phenylalanine dodecylamide **217**, L-valine dodecylamide **218** and L-proline dodecylamide **219**.



217



218



219

Species	Mass (mg)	mmol	Vol of water (mL)	Structure
217	33 mg	0.1	3 mL	Opaque gel
218	30 mg	0.1	3 mL	Opaque Gel
219	56 mg	0.2	3 mL	Micelle aggregates

Table 5.2. Overview of monopeptidic amides and their supramolecular structures in deionised water.

The gelation of **217** and **218** was originally attempted with 0.2 mmol of gelator and 3 mL of water, as this would be the ideal conditions to directly test the catalytic ability of **217** and **218** against the previous conditions used in the 2-deoxy-D-ribose **D-130** forming reaction. Gelation was attempted using a cycle of heating, sonicating and cooling. A concentration of 0.2 mmol in water (3 mL) proved too concentrated and all of the material could not be dissolved. Also upon cooling **217** and **118** would precipitate out of solution very quickly and so gel fibres did not have time to form. Successful gelation occurred at 0.1 mmol of gelator in solvent (3 mL) through a continuous heating, sonicating, cooling and resting cycle that could take anywhere between 1 and 7 cycles for gelation to occur. **Figure 5.12A** shows **182** (33 mg) gelled in deionised water (3 mL). **Figure 5.12B** is the TEM image of the gelled structure. This TEM image showed that the gel looked rather crystalline with larger micron sized aggregates. As such the gel may be more microcrystalline in nature than true nano-gels.

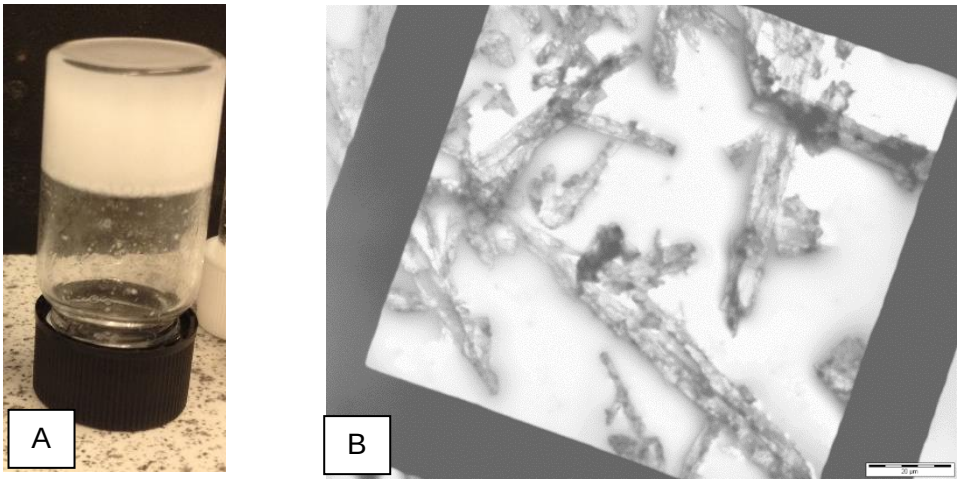


Figure 5.12 (A) L-Phenylalanine-dodecylamide **217** gelled in deionised water and (B) The TEM image of the gel, scale bar = 20 μm.

Figure 5.13A shows the many attempts to gelate **217** at various concentrations in deionised water. Compound **217** formed an opaque gel that was particularly strong and sticky. The TEM image of the gel (**Figure 5.13B**) showed the presence of gel fibres. These fibres were particularly susceptible to the ion beam of the microscope; the fibres on the left hand side of the image look blurred as the ion beam had already destroyed them. The fibres appeared significantly narrower than those of **217**, although they were again relatively needle-like and crystalline in nature.

The 2-deoxy-D-ribose **D-130** forming reaction was attempted with both of these gels using Condition 2 from Chapter 8 i.e. the gel was made in deionised water (3 mL) and the reagents; acetaldehyde **24** (1 mmol) and D-glyceraldehyde **D-9** (1 mmol) were dissolved in deionised water (1 mL) and added dropwise to the top of the gel. Unfortunately, in both cases, within 5 minutes the gel broke down. **Figure 5.14** shows images of the reaction with the L-valine gel at various time points.

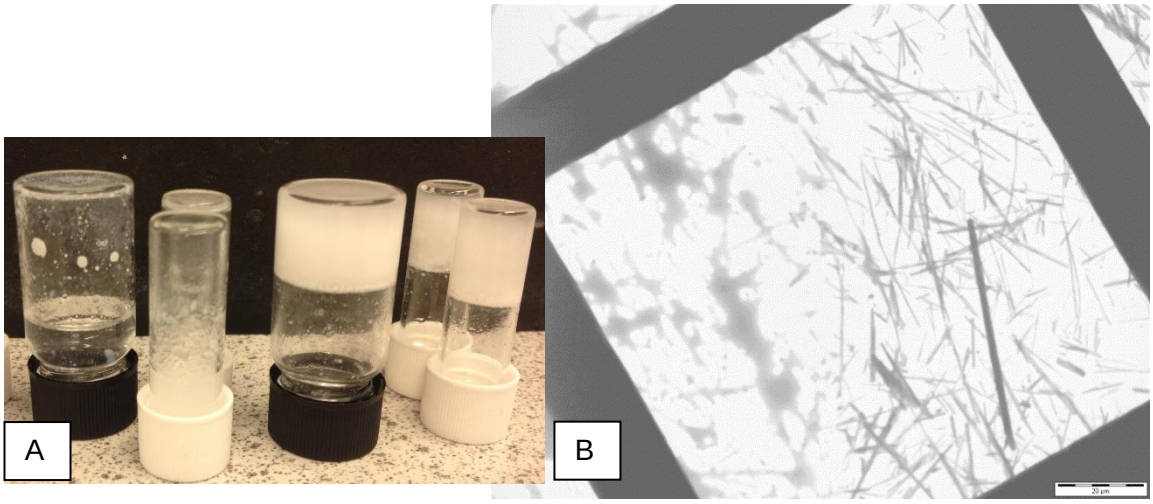


Figure 5.13 (A) L-Valine dodecylamide **218** gelation attempts in deionised water. (B) TEM image of gel, scale bar = 20 μm.

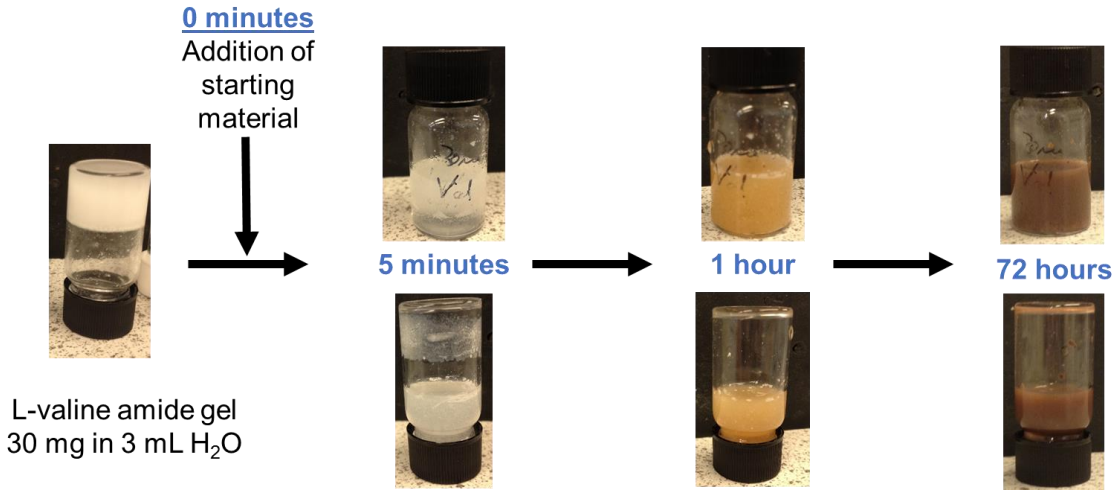


Figure 5.14. L-Valine amide **218** gel at various time points after addition of D-glyceraldehyde **D-9** and acetaldehyde **24** to the surface of the gel.

Over the course of an hour, the opaque white gel had broken down to give a colourless solution and a white precipitate which then turned yellow, and when left for 72 hours turned brown. The contents of the vial were concentrated *in vacuo* and the trapping and purification

procedure was carried out as before. No 2-deoxy-D-ribose **D-130** or 2-deoxy-D-threopentose **D-131** was formed over the course of the reaction. The reason for the breakdown of the gel can be perhaps attributed to the size of the gelator. Referring back to Escuder's hypothesis of the key hydrogen bonding interactions of Gelator **A** (L-proline-L-valine dodecylamide), these mono-aminoamides lose one of the key hydrogen bond interactions leaving just one hydrogen bond between molecules, as shown in the L-valine dodecylamide **218** example in **Figure 5.15**.[152,155] As well as this, the site of imine formation has been moved closer to the key hydrogen bond interaction and therefore it is possible when the primary amine interacts with the aldehyde starting material, the weak hydrogen bond may break leading to the disassembly of the gel fibres.

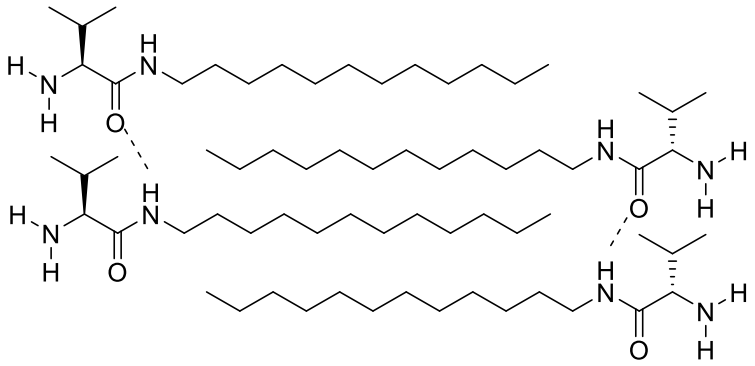


Figure 5.15. Possible assembly of gel fibres in L-valine dodecylamide **218**.

The L-proline dodecylamide **219** did not form gel fibres and instead remained dissolved in deionised water. As in the case of P-E-F dodecylamide **210** we assumed supramolecular aggregation had occured. To prove this, DLS experiments were conducted in deionised water. **Figure 5.16** and **Figure 5.17** show the intensity and volume distributions. In both cases a single scattering peak was observed indicating an average diameter of 143 nm by intesntisy and 112 nm by volume. This was much larger than the 6 nm diameter of P-E-F dodecylamide **210** micelles and suggests that simple micelles had not formed here. In this case, either elongated micelles or fibres may have formed, that are much more cylindrical in shape leading to a larger dimeter of the supramolecular structure. It should be noted that

DLS does not provide accurate sizes for such structures as they are not spherical – DLS reports the average size of an equivalent spherical system. However, the size distributions are relatively narrow which would suggest a relatively well, controlled assembly process giving rise to similarly sized nanoscale objects.

		Size (d.nm):	% Intensity:	St Dev (d.nm):
Z-Average (d.nm): 116.0	Peak 1:	143.0	100.0	61.81
PdI: 0.223	Peak 2:	0.000	0.0	0.000
Intercept: 0.963	Peak 3:	0.000	0.0	0.000
Result quality : Good				

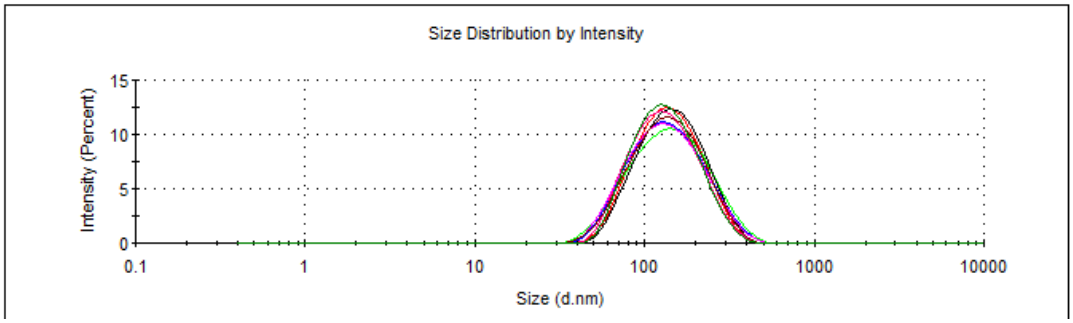


Figure 5.16. Size distribution by intensity graph created from DLS experiments of L-proline dodecylamide **219** in deionised water. This graph is based on 8 runs of the experiment.

		Size (d.nm):	% Volume:	St Dev (d.nm):
Z-Average (d.nm): 116.0	Peak 1:	112.0	100.0	59.18
PdI: 0.223	Peak 2:	0.000	0.0	0.000
Intercept: 0.963	Peak 3:	0.000	0.0	0.000
Result quality : Good				

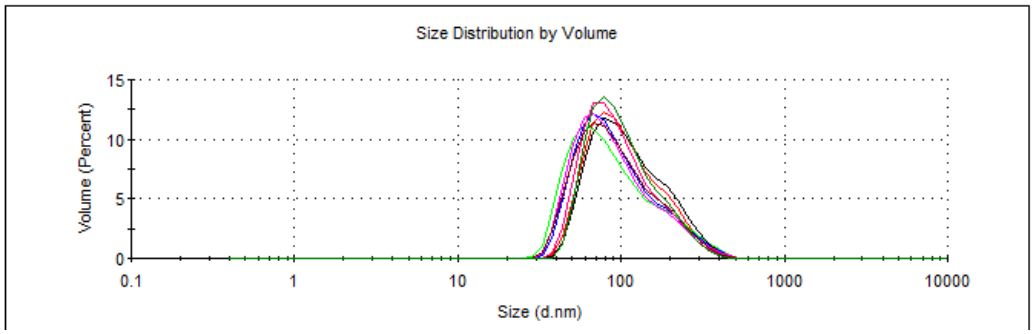


Figure 5.17. Size distribution by volume graph created from DLS experiments of **219** in deionised water. This graph is based on 8 runs of the experiment.

The CMC was again calculated using the Nile Red assay. A plot of log(concentration) vs absorbance was made and the point of inflection of the two trend lines identified as the CMC (**Figure 5.18**). This was an order of magnitude higher than that of PEF-dodecylamide **210** with a value of 172 μmol ± 2 μmol. This suggests more L-proline dodecylamide **219** was needed in solution to form an elongated micelle or fibre indicative of a less effective assembly process.

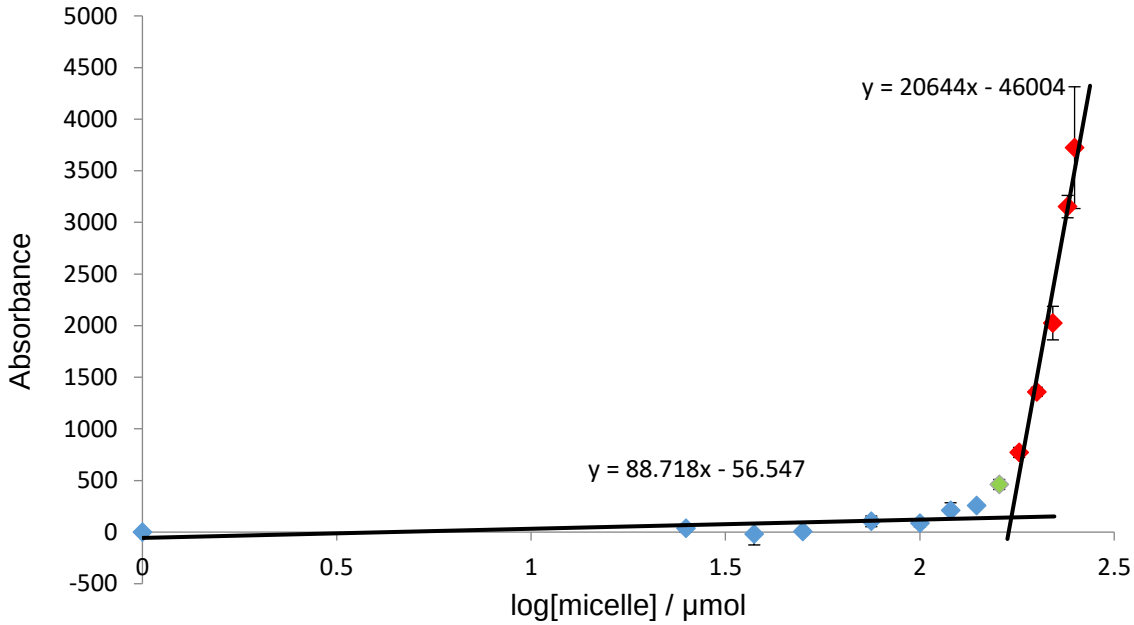


Figure 5.18. Nile Red **211** assay performed on L-proline amide **219**. The CMC is determined as the point of inflection calculated as 172 μmol ± 2 μmol.

Unfortunately when tested as a catalyst for the 2-deoxy-D-ribose **D-130** forming reaction at a concentration of 0.2 mol (20 mol %) in deionised water no pentose sugars were formed. Due to time constraints these reactions were not repeated in pH 7 phosphate buffer, which may have more active catalyst present and hence turnover product in the 24 hour time period.

5.6. Conclusions

The synthesis and characterisation of a number of supramolecular structures has been performed. The hydrogels that have been formed have not been able to catalyse the aldol reaction of acetaldehyde **24** and D-glyceraldehyde **D-9** and collapsed upon addition of starting material to the surface of the gel. This indicates that the polar starting materials are disrupting key interactions that stabilise the gel fibres or that the reaction of the aldehydes with the catalytic amine to form the intermediate iminium species disrupts the gelator and breaks down the self-assembled nanostructure. In the cases of P-E-F dodecylamide **210** and L-proline dodecylamide **219** these molecules preferred to form other nanoscale supramolecular aggregates (e.g. micelles) in water. Unfortunately these molecules could not catalyse the formation of 2-deoxy-D-ribose **D-130** either. Further research in this area should aim to develop the novel supramolecular structures of the single amino amides. Amino acid candidates with R-groups capable of forming additional hydrogen bonds, such as lysine, glutamic acid, arginine etc, should be used as gelators. Having the extra hydrogen bonding potential may stabilise the gel fibres whilst enamine formation with the aldehyde starting material is occurring.

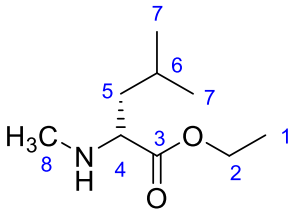
6. Experimental

6.1. General Experimental

Unless otherwise noted all compounds were bought from commercial suppliers and used without further purification. Where a solvent is described as “dry” it was purified by PureSolv alumina columns from Innovative Technologies. Melting points were determined using a Stuart SMP3 apparatus. Optical rotations were carried out using a JASCO-DIP370 polarimeter and $[\alpha]_D$ values are given in deg. $cm^3\ g^{-1}\ dm^{-1}$. Infra-red spectra were acquired on a ThermoNicolet Avatar 370 FT-IR spectrometer. Nuclear magnetic resonance spectra were recorded on a Jeol ECS-400, a Jeol 500 Avance III HD 500 or a Jeol AV500 at ambient temperature. Coupling constants (*J*) are quoted in Hertz. Mass spectrometry was performed by the University of York mass spectrometry service using electron spray ionisation (ESI) technique. Thin layer chromatography was performed on glass-backed plates coated with Merck Silica gel 60 F_{254}. The plates were developed using ultraviolet light, acidic aqueous ceric ammonium molybdate or basic aqueous potassium permanganate. Liquid chromatography was performed using forced flow (flash column) with the solvent systems indicated. The stationary phase was silica gel 60 (220–240 mesh) supplied by Sigma-Aldrich. Preparative Thin Layer Chromatography (PTLC) was carried out on 20x20 2000 micron silica plates with UV_{254} purchased from Uniplate. High Performance Liquid Chromatography (HPLC) was performed using an Agilent 1100 series instrument using the chiral columns indicated and a range of wavelengths from 210-280 nm for detection. Buffer solutions, pH 6 and pH 7 phosphate buffers, were purchased as ready-made solutions from Fisher Scientific.

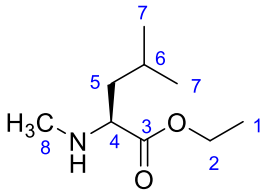
6.2. Methods and Characterisation of Compounds

N-Methyl-D-leucine ethylester (D-111)



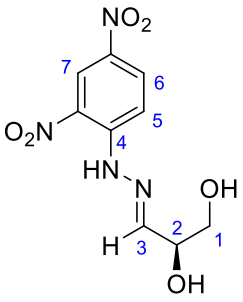
Trifluoroacetic acid (1.12 mL, 14.6 mmol) was added to a solution of Boc-*N*-methyl-D-leucine ethyl ester **D-142** (200 mg, 0.73 mmol) in DCM (20 mL) and was stirred for 14 hours under a nitrogen atmosphere. The solution was then concentrated *in vacuo* and partitioned between DCM (4 mL) and a saturated solution of sodium bicarbonate (4 mL) and extracted two more times with DCM. The organic layers were combined, dried over magnesium sulfate and concentrated *in vaco* to give **D-111** as a colourless oil in a 75 % yield (115 mg, 0.55 mmol). **IR** (ATR): ν_{max} 2956 , 1731, 1469, 1368, 1178, 1026 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -1.3 (*c* = 1.0, CH_3Cl); **^{1}H NMR** (400 MHz, $CDCl_3$) δ 4.17 (2H, q, J = 7.0 Hz, H-2), 3.81 (1H dd, J = 7.1, 7.1 Hz, H-4), 2.77 (3H, s, H-8), 1.73-1.65 (1H, m, H-6), 1.47-1.41 (2H, m, H-5), 1.26 (3H, t, J = 7.0 Hz, H-1), 0.91 (3H, d, J = 6.5 Hz, H-7), 0.88 (3H, d, J = 6.5 Hz, H-7); **^{13}C NMR** (400 MHz, $CDCl_3$): δ 175.9 (C-3), 61.8 (C-2), 60.6 (C-4), 42.7 (C-5), 34.8 (C-8), 25.0 (C-6), 22.7 (C-7), 22.5 (C-7), 14.4 (C-1); **HRMS** (ESI): $[M+H]^+$ HRMS found 174.1491, $C_9H_{20}NO_2$ required 174.1489.

***N*-Methyl-L-leucine ethylester (L-111)**



N-Methyl-L-leucine ethyl ester (**L-111**) was prepared in the same way as **D-111** from **D-142**. **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) +1.0 (*c* = 1.0, chloroform) literature +0.57 (*c* = 0.74, chloroform);[81] **^{1}H NMR** (400 MHz, $CDCl_3$) δ 4.28 (2H, q, J = 7.3 Hz, H-2), 3.79 (1H, dd, J = 8.6, 4.9 Hz, H-4), 2.75 (3H, s, H-8), 1.82 (1H, dd, J = 17.1, 8.6 Hz, H-5), 1.77-1.70 (2H, m, H-5, H-6), 1.30 (3H, t, J = 7.3 Hz, H-1), 0.96 (3H, d, J = 5.8 Hz, H-7), 0.95 (3H, d, J = 5.8 Hz, H-7); **^{13}C NMR** (400 MHz, $CDCl_3$): δ 168.9 (C-3), 62.7 (C-2), 59.6 (C-4), 38.2 (C-5), 31.5 (C-8), 24.8 (C-6), 22.9 (C-7), 21.4 (C-7), 14.0 (C-1); **HRMS** (ESI): $[M+H]^+$ HRMS found 174.1489, $C_9H_{20}NO_2$ required 174.1496. Spectroscopic data was in agreement with the literature.[81]

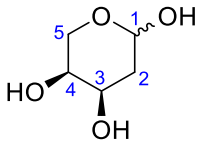
***Trans*-3-[(2, 4-Dinitrophenyl)-hydrazono]-propane-1,2-diol (D-115)**



D-Glyceraldehyde **D-9** (50 mg, 0.56 mmol) and 2,4-dinitrophenylhydrazine **114** (132 mg, 0.67 mmol) were dissolved in water (2.5 mL) and stirred at room temperature for 24 hours. The reaction mixture was concentrated *in vacuo* to give the crude product as an orange

solid. Purification *via* flash column chromotography (1:1 petroleum ether:ethyl acetate) gave the desired product, **D-115**, as an orange solid in a 45 % yield (67 mg, 0.25 mmol); **Decomposition temperature:** 150-152 °C; **IR** (ATR): ν_{max} 3301, 3106, 3094, 2929, 1614, 1586, 1504, 1419, 1320, 1222, 1090, cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) +32.0 (*c* = 0.1, chloroform) literature +36.9 (*c* = 0.07, chloroform)[156]; **^{1}H NMR** (400 MHz DMSO-d^6) δ 11.40 (1H, br s, N-H), 8.84 (1H, d, 2.7 Hz, H-7), 8.36 (1H, dd, 9.6 Hz, 2.7 Hz, H-6), 7.95 (1H, d, 6.0 Hz, H-3), 7.91 (1H, d, 9.6 Hz, H-5), 5.35 (1H, br, s, OH), 4.79 (1H, br s, OH), 4.17 (1H,dd, 6.0, 6.0 Hz, H-2), 3.54 (2H, d, 6.0 Hz, H-1); **^{13}C NMR** (400 MHz DMSO-d^6) δ 155.2 (C-3), 145.4 (CNO_2), 137.3 (C-4), 130.3 (C-6, 129.7 (CNO_2), 123.5 (C-7), 117.1 (C-5), 72.2 (C-1), 64.5 (C-2); **HRMS** (ESI): $[M-H]^-$ HRMS found 269.0535, $C_9H_9N_4O_6$ required 269.0528.

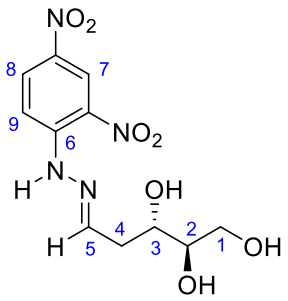
2-Deoxy-D-threopentose (D-131)



To a flask containing 2-bromo-2-deoxy-D-threo-pentofuranose **139** (75 mg, 0.35 mmol) in dry THF (5 mL), tin tributyl hydride (0.11 mL, 0.43 mmol) and AIBN (9.3 mg, 0.057 mmol) were added and the resulting mixture was refluxed for 4 hours. At that point the reaction was deemed complete through TLC analysis and the reaction mixture was concentrated *in vacuo.* The crude mixture was partitioned between water (10 mL) and ethyl acetate (10 mL) and the aqueous layer extracted. The organic layer was further extracted with water (3x 10 mL) and the aqueous layers combined and concentrated *in vacuo* to give the title compound **D-131** as a colourless oil in an 86 % yield as a mixture of anomers (47 mg, 0.30 mmol). **IR** (ATR): ν_{max} 3327, 2933, 2886, 1648, 1436, 1359, 1263, 1239, 1129, 1060 cm^{-1};

$[\alpha]_D^{25}$ (deg cm^3 g^{-1} dm^{-1}) -10.0 (*c* = 0.5, water) literature -4.8 (*c* = 0.5, water)[157]; **^{1}H NMR** as a mixture of α and β anomers; **^{13}C NMR** (400 MHz CD_3OD): α-anomer δ 95.7 (C-1), 72.2 (C-4), 69.7 (C-3), 63.9 (C-5), 40.7(C-2); β-anomer 93.1 (C-1), 72.1 (C-4), 66.7 (C-3), 63.9 (C-5), 38.7 (C-2); **HRMS** (ESI): $[M+Na]^+$ HRMS found 157.0470, $C_5H_{10}O_5$ required 157.0471.

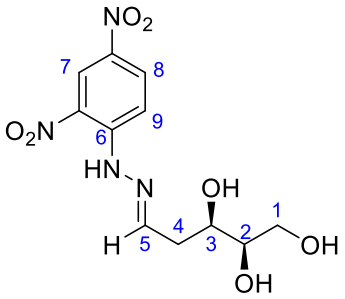
(2-*S*,3-*R*,*E*)-5-[(2,4-Dinitro-phenyl)-hydrazono]-pentane-1,2,3-triol (D-132)



2-Deoxy-D-ribose **D-130** (50 mg, 0.37 mmol) and 2,4-dinitrophenyl hydrazine **114** (110 mg, 0.37 mmol) were partly dissolved in ethanol and stirred at room temperature for 72 hours. Concentration *in vacuo* gave the crude product as an orange solid. Purification *via* preprative thin layer chromotgpahy (50:50 ethyl acetate:petrol) gave the title compound **D-132** as an orange solid in a 7 % yield (8 mg, 0.025 mmol). **IR** (ATR): ν_{max} 3350, 3287, 2928, 1616, 1584, 1420, 1327, 1304, 1262, 1215, 1145, 1071 cm^{-1}; **^{1}H NMR** (400 MHz CD_3OD): δ 9.00 (1H,d, 2.8 Hz, H-7), 8.23 (1H, dd, 10.0, 2.8 Hz, H-8), 7.98 (1H, d, 10.0 Hz, H-9), 7.79 (1H, dd, 5.8, 5.8 Hz, H-5), 3.85 (1H, ddd, J= 10.4, 8.8, 3.5 Hz, H-3), 3.74 (1H, dd, J= 11.4, 4.0 Hz, H-1), 3.60 (1H, dd, J= 11.4, 6.4 Hz,H-1), 3.52 (1H, ddd, J= 10.4, 6.4, 4.0 Hz, H-2), 2.78 (1H, ddd, 14.9, 5.8, 3.5 Hz, H-4), 2.54 (1H, ddd, 14.9, 8.8, 5.8 Hz, H-4);

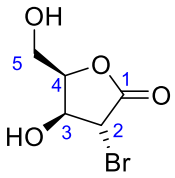
^{13}C NMR (400 MHz CD_3OD): δ 153.0, 146.4, 138.7, 130.6, 130.2, 124.1, 117.6, 76.0, 71.3, 64.6, 37.8; **HRMS** (ESI): $[M-H]^-$ HRMS found 313.0798, $C_{11}H_{13}N_4O_7$ required 313.0790.

(2-R,3-,R,E)-5-[(2,4-Dinitro-phenyl)-hydrazono]-pentane-1,2,3-triol (D-133)



To a solution of 2-deoxy-D-threopentose **D-131** (40 mg, 0.30 mmol) in dry methanol (4 mL) was added 2,4-dinitrophenyl hydrazine **114** (71 mg, 0.36 mmol). The reaction mixture was stirred for 72 hours at room temperature under an atmospehere of nitrogen. The mixture was concentrated *in vacuo* and purified by preparitve thin layer chromatography to yield **D-133** as an orange solid in an 11% yield (10 mg, 0.032 mmol). **IR** (ATR) ν_{max} 3421, 3280, 3158, 1546, 1409 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) +1.04 (*c* = 0.825, methanol); **^{1}H NMR** (400 MHz, CD_3OD): δ 8.96 (1H, d, 2.8 Hz, H-9), 8.26 (1H, dd, 9.6, 2.8 Hz, H-8), 7.93 (1H, d, 9.6 Hz, H-7), 7.75 (1H, dd, 5.5, 5.5 Hz, H-5), 3.96 (1H, ddd, J= 10.1, 10.1, 5.5 Hz, H-2), 3.67-3.60 (3H, m, H-3, H-1), 2.64-2.61 (2H, m, H-4); **^{13}C NMR** (400 MHz CD_3OD): δ 151.4 (C5), 145.0 (CNO_2), 137.4 (C-6), 129.4 (C-7), 129.3 (CNO_2), 122.7 (C-8), 116.2 (C-9), 73.8 (C-2), 69.1 (C-3), 62.9 (C-1), 36.6 (C-4); **HRMS** (ESI): $[M-H]^-$; HRMS found 313.0795, $C_{11}H_{13}N_4O_7$ required 313.0790.

2-Bromo-2-deoxy-D-lyxono-1,4-lactone (136)

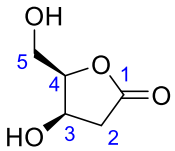


D-Lyxose **121** (3.00 g, 22.2 mmol), and solid sodium bicarbonate (2.52 g, 30.0 mmol) were dissolved in deionised water (25 mL) and stirred at 0 °C for 5 minutes. Bromine was added dropwise to the solution every 20 minutes for 1 hour (3 x 0.38 mL, 14.8 mmol) and the reaction stirred at room temperature for 4 hours. Sodium thiosulfate (20 mL) was added to destroy the excess bromine and the mixture concentrated *in vacuo* to give an off-white precipitate. The crude material was purified by extracting with boiling methanol (3 x 50 mL). The extracts were combined and concentrated *in vacuo* to give D-lyxono-1,4-lactone **135** as an off white solid (4.73 g). This was used in the next step without further purification.

Crude D-lyxono-1,4-lactone **135** (1.00 g, 6.75 mmol), was added to a solution of 33 % hydrogen bromide in acetic acid (10 mL) and stirred for 2 hours at room temperature at which point TLC confirmed that all of the starting material had been consumed. The reaction was quenched by the addition of methanol and the reaction mixture stirred for a further 24 hours. The mixture was concentrated *in vacuo*, the residue dissolved in chloroform (10 mL) and extracted with water (7 x 10 mL). The aqueous extracts were combined and concentrated *in vacuo* to give the crude product as a red oil. Purification *via* flash column chromatography over silica (50:50 cyclohexane:ethyl acetate) gave the title compound **136** as a yellow oil in a 14 % yield over 2 steps (0.20 g, 0.92 mmol). **IR** (ATR): ν_{max} 3315, 2991, 2967, 2949, 1763, 1464, 1372, 1329, 1182, 1145, 1023 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) +20.1 (*c* = 1.0, ethyl acetate) literature +26 (*c* = 0.2, ethyl acetate)[104]; **^{1}H NMR** (400 MHz D_2O): δ 4.91 (1H, ddd, J= 5.2, 5.2, 4.3 Hz, H-4), 4.77 (1H, dd, J= 5.2, 4.3 Hz, H-3), 4.68

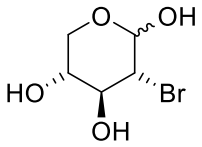
(1H, d, J= 4.3 Hz, H-2), 3.94 (1H, d, J= 5.2 Hz, H-5), 3.93 (1H, d, J= 4.3 Hz, H-5); **^{13}C NMR** (400 MHz D_2O): δ 174.8 (C-1), 83.3 (C-4), 74.3 (C-3), 59.1(C-5), 42.7 (C-2); **HRMS** (ESI) $[M+Na]^+$ found 232.9422 and 234.9397 in a 1:1 ratio, $C_5H_7Br^{79}O_4Na$ required 232.9420 $C_5H_7Br^{81}O_4Na$ required 234.9399. An artefact of ESI MS through methanolic opening of lactone calculated for $C_6H_{11}Br^{79}NaO_5$, 264.9682, found 264.9686.

2-Deoxy-D-lyxono-1,4-lactone (137)



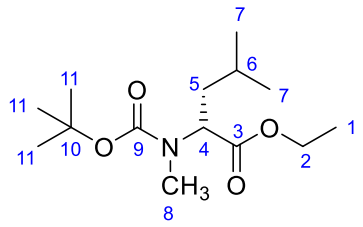
To a flask containing 2-bromo-2-deoxy-D-lyxono-1,4-lactone **136** (200 mg, 0.95 mmol) in dry THF (5 mL) under a nitrogen environment was addedd tributyl tin hydride (0.25 mL, 0.95 mmol) and AIBN (25 mg, 0.15 mmol). The mixture was refluxed for 3 hours. At this point the reaction was deemed finished through TLC and the reaction mixture diluted with water (10 mL) washed with ethyl acetate (10 mL), extracted with water (3 x 10 mL) and concentrated *in vacuo* to give the title compound **137** as a colourless oil (126 mg, 0.95 mmol). **IR** (ATR): ν_{max} 3366, 2939, 1749, 1162 cm^{-1}; **^{1}H NMR** (400 MHz, D_2O); 4.55 (1H, ddd, J= 5.5, 4.0, 1.1 Hz, H-4), 4.53 (1H, ddd,.6.4, 4.0, 4.0 Hz, H-3), 3.78 (1H, d, 4.0 Hz, H-2), 3.77 (1H, d, 6.4 Hz, H-2), 2.91 (1H, dd J= 18.1, 5.5 Hz, H-5), 2.42 (1H, dd, 18.1, 1.1 Hz, H-5); **^{13}C NMR** (400 MHz D_2O): 179.6 (C-1), 85.7 (C-4), 67.8 (C-3), 59.8 (C-5), 38.7 (C-2); **HRMS** (ESI) $[M + H]^+$ HRMS found 133.0497, $C_5H_9O_4$ required 133.0495. $[M+Na]^+$ found 155.0316, $C_5H_8NaO_4$ required 155.0315. Spectroscopic data was in agreement with the literature.[104]

2-Bromo-2-deoxy-D-threo-pyranose (139)



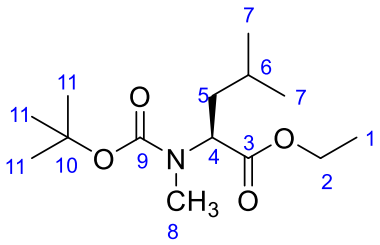
2-Bromo-D-lyxono-1,4-lactone **136** (435 mg, 2.06 mmol) was dissolved in water (7 mL) and stirred at 0 °C, amberlite IR-120-H resin was added to reduce the pH to pH 3. Sodium borohydride (39 mg, 1.03 mmol) was added in portions (~10 mg) along with amberlite resin to keep the pH of the reaction to approximately 6. The reaction was then allowed to stir for 30 minutes at which time there was no starting material visible by TLC. The reaction mixture was filtered to remove the resin and then concentrated *in vacuo* to give the crude residue. Purification *via* flash column chromatography over silica (40:60 petroleum ether:ethyl acetate) gave the pure title compound **139** as a colourless oil in a 45 % yield as a mixture of anomers (199 mg, 0.93 mmol). **IR** (ATR): ν_{max} 3307, 2943, 2836, 1417, 1353, 1110, 1060, 1016 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) +64.4 (*c* = 1.0, methanol) literature +51.5 (*c* = 0.4, water)[104]; **^{1}H NMR** spectrum showed a mixture of anomers and was unassignable, therefore characterisation was based on comparison of carbon NMR data with literature values. **^{13}C NMR** (400 MHz CD_3OD): α-anomer: δ 98.8, 78.9, 72.1, 67.1, 58.3; β-anomer: 94.2, 74.7, 72.8, 62.6, 55.6; **HRMS** (ESI) $[M+Na]^+$ calculated at 232.9420 for $C_5H_9Br^{79}O_4Na$ and 234.9400 for $C_5H_9Br^{81}O_4Na$, HRMS found 232.9475. All data was in agreement with the literature.[104]

Boc-*N*-Methyl-D-leucine ethyl ester (D-142)



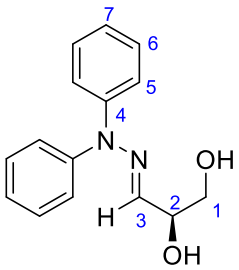
A solution of potassium bis(trimethylsilyl)amide 0.5 M in toluene (20 mL, 9.76 mmol) was added to a stirred solution of *N*-Boc-D-leucine ethyl ester **D-141** (2.3 g, 3.09 mmol) in dry THF (20 mL) at -78 °C. After 30 minutes methyl iodide (0.61 mL, 9.76 mmol) was added dropwise (over 5 minutes) and the reaction stirred for 1 hour at -78 °C and a further 16 hours at room temperature. The reaction mixture was washed with saturated potassium carbonate solution (30 mL) and extracted with DCM (3 x 30 mL). The combined organic extracts were washed with 1 M sodium hydroxide (30 mL), then brine (30 mL). The organic layer was dried over magnesium sulfate, filtered and then concentrated *in vacuo* to give the crude product as a yellow oil. Purification *via* column chromatography (10:90 ethyl acetate:hexane) gave the title produce, **D-142**, as a colourless oil in an 83 % yield (700 mg, 2.56 mmol). **IR** (ATR): ν_{max} 2962, 1741, 1694, 1390, 1363, 1320, 1148, 1031 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) +8.9 (*c* = 1.0, ethanol); **^{1}H NMR** (400 MHz $CDCl_3$): Apparent 1:1 mixture of rotamers δ 4.84 (1H,dd, *J* = 8.0, 8.0 Hz, H-4), 4.55 (1H, dd, J= 10.8, 4.7 Hz, H-4), 4.16 (4H, q, J = 7.1 Hz, H-2), 2.79 (3H, s, H-8), 2.76 (3H, s, H-8), 1.74-1.51 (6H, m, H-5, H-6 (both rotamers)), 1.45 (9H, s, H-11), 1.44 (9H, s, H-11), 1.27-1.23 (6H, m, H-1), 0.94 (6H, d, J= 6.8 Hz, H-7), 0.92 (6H, d, J = 6.8 Hz, H-7); **^{13}C NMR** (400 MHz $CDCl_3$): δ 172.4 (C-3), 156.4 (C-9), 80.0 (C-10), 61.1, 57.3, 56.1, 38.0, 28.5, 24.8, 23.4, 21.5, 21.3, 14.4 (C-1); **HRMS** (ESI): $[M+H]^+$ HRMS found 274.2000, $C_{14}H_{28}NO_4$ required 247.2013. $[M+Na]^+$ HRMS found 296.1822, $C_{14}H_{27}NO_4Na$ required 296.1832.

Boc-*N*-Methyl-L-leucine ethylester (L-142)



Boc-*N*-Methyl-L-leucine ethyl ester **L-142** was prepared in the same way as **D-142** from **D-141**; **[α]$_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -9.0 (*c* = 1.0, methanol) literature -8.2 (*c* = 0.74, chloroform);[81] **^{1}H NMR** (400 MHz, $CDCl_3$): Apparent 1:1 mixture of rotamers δ 4.83 (1H, dd, J = 7.3, 7.3 Hz, H-4), 4.55 (1H, dd, J = 10.8, 4.7 Hz, H-4), 4.15 (4H, q, J = 6.9 Hz, H-2 (both rotamers)), 2.79 (3H, s, H-8), 2.75 (3H, s, H-8), 1.74-1.48 (6H, m, H-5, H-6 (both rotamers)), 1.29-1.28 (18H, m, H-11 (both rotamers)), 1.27-1.24 (6H, m, H-1(both rotamers)), 0.94 (6H, d, J = 6.9 Hz, H-7(both rotamers)), 0.92 (6H, d, J = 6.9 Hz, H-7(both rotamers)). Spectroscopic data was in agreement with the literature.[81]

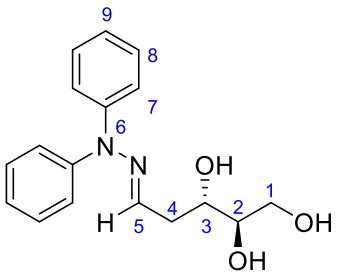
(*N,N*-Diphenyl-hydrazono)-propane,1-2-diol (145)



D-Glyceraldehyde **D-9** (51 mg, 0.56 mmol) was dissolved in methanol (8 mL). *N,N*-Diphenyl hydrazine **144** (172 mg, 0.93 mmol) followed by two drops of acetic acid was added to the solution and stirred for 1 hour before concentrating *in vacuo* to give the crude product as a brown oil. Purification *via* column chromatography (10:90 methanol:DCM) yielded **145** as

a crystalline solid in a 95 % yield (145 mg, 0.54 mmol). **Mp** 91-93 °C; **IR** (ATR) ν_{max} 3281, 2932, 1589, 1492, 1293, 1052 cm^{-1}; **$[\alpha]_D^{25}$** (deg. cm^3 g^{-1} dm^{-1}) -0.083 (*c* = 1.0, chloroform) literature = -3.3 (c = 3.23, chloroform)[158]; **^{1}H NMR** (500 MHz, CD_3OD): δ 7.40-7.34 (2H dd, J= 8.5, 7.4 Hz, H-7), 7.14 (4H**,** tt, J= 7.4, 1.2 Hz, H-6), 7.07-7.04 (4H dd, J= 8.5, 1.2 Hz, H-5), 6.45 (1H, d, 5.2 Hz, H-3), 4.31 (1H, dt, 6.5, 5.2 Hz, H-2), 3.66 (1H, dd, J= 11.3, 5.2 Hz, H-1), 3.60 (1H, dd, J= 11.3, 6.5 Hz, H-1); **^{13}C NMR** (500 MHz, CD_3OD): δ 145.1 (C-4), 138.6 (C-3), 130.8 (C-6), 125.5 (C-7), 123.5 (C-5), 73.6 (C-2), 65.8 (C-1); **HRMS** (ESI): $[M+H]^+$ found 257.1279, $C_9H_{11}N_4O_6$ required 257.1285. $[M+Na]^+$ found 279.1100, $C_9H_{10}N_4O_6Na$ required 279.1104.

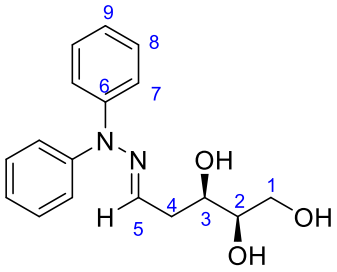
(2-*S*,3-*R*,*E*)-5-[(*N*,*N*-Diphenyl)-hydrazono]-pentane-1,2,3-triol (D-146)



2-Deoxy-D-ribose **D-130** (50 mg, 0.37 mmol) and *N,N*-diphenyl hydrazine **144** (125 mg, 0.68 mmol) were dissolved in methanol (5 mL). Two drops of acetic acid were added and the reaction stirred at room temperature for 1 hour. The reaction mixture was concentrated *in vacuo* to give a brown oil. Upon purification *via* preparative thin layer chromatography (5:95 methanol:DCM) the pure compound, **D-146**, was obtained as a white crystalline solid in a 98 % yield (100 mg, 0.34 mmol). **Mp** 114-117 °C; **IR** (ATR) ν_{max} 3221, 2926, 2875, 1586, 1487, 1298 cm^{-1}; **$[\alpha]_D^{25}$** (deg. cm^3 g^{-1} dm^{-1}) = -5.1 (*c* = 0.1, methanol); **^{1}H NMR** (500 MHz CD_3OD): δ 7.33 (4H, dd, J= 8.6 Hz, 7.4 Hz, H-9), 7.09 (2H, tt, J= 7.4, , 1,2 Hz, H-8),

7.03 (4H, dd, J= 8.6, 1.2 Hz, H-7), 6.64 (1H, dd, J= 5.5, 5.5 Hz, H-5), 3.69 (1H, dd, J= 11.3, 3.7 Hz, H-1), 3.67 (1H, ddd, J= 10.4, 8.6, 3.8 Hz, H-3), 3.53 (1H, dd, J= 11.3, 6.5 Hz, H-1), 3.43 (1H, ddd, 10.4, 6.5, 3.7 Hz, H-2), 2.64 (1H, ddd, J= 14.9, 5.5, 3.8 Hz, H-4), 2.41 (1H, ddd, J= 14.9, 8.6, 5.5 Hz, H-4); **^{13}C NMR** (500 MHz CD_3OD): δ 145.7 (C-6), 139.1 (C-5), 130.7 (C-8), 125.1 (C-9), 123.6 (C-7), 76.0 (C-2), 72.0 (C-3), 64.6 (C-1), 37.5 (C-4); **HRMS** (ESI): $[M+H]^+$ found 301.1543, $C_{17}H_{21}N_2O_3$ required 301.1547. $[M+Na]^+$ found 323.1360, $C_{17}H_{20}N_2O_3Na$ required 323.1366.

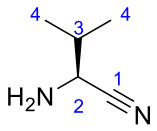
(2-R, 3-*R*)-5-[(*N,N*-Diphenyl)-hydrazono]-pentane-1,2,3-triol (D-147)



2-Deoxy-D-threopentose **D-131** (29 mg, 0.21 mmol) and diphenyl hydrazine **144** (77 mg, 0.42 mmol) were dissolved in methanol (5 mL). Two drops of acetic acid were added and the reaction stirred at room temperature for 1 hour. The reaction mixture was concentrated *in vacuo* to give a brown oil. Upon purification *via* preparative thin layer chromatography (5:95 methanol:DCM) the pure compound **D-147** was obtained as a white colourless oil in a 84 % yield. (53 mg, 0.18 mmol); **IR** (ATR): ν_{max} 3351, 3059, 2898, 1588, 1493, 1297 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) +8.0 (*c* = 0.1, methanol); **^{1}H NMR** (500 MHz CD_3OD); δ 7.36 (4H, dd, 8.4, 7.4 Hz, H-9), 7.12 (2H, tt, 7.4, 1.1 Hz, H-8), 7.06 (4H, dd, 8.4, 1.1 Hz, H-7), 6.65 (1H, dd, 5.5, 5.5 Hz, H-5), 3.81 (1H, ddd J= 8.2, 5.1, 3.5 Hz, H-3), 3.66 (1H, dd, J= 11.1, 5.1 Hz, H-1), 3.58 (1H, dd, J= 11.1, 6.4 Hz, H-1), 3.52 (1H, ddd, J= 6.4, 5.1, 3.5 Hz,

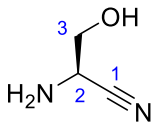
H-2), 2.58-2.48 (2H, m, H-4); **^{13}C NMR** (500 MHz CD_3OD): δ 145.6 (C-6), 138.7 (C-5), 130.7 (C-8), 125.1 (C-9), 12.5 (C-7), 75.1 (C-2), 71.2 (C-3), 64.3 (C-1), 37.6 (C-4); **HRMS** (ESI): $[M+Na]^+$ HRMS found 323.1366, $C_{17}H_{20}N_2O_3Na$ required 323.1362,

L-Valine nitrile (L-157)



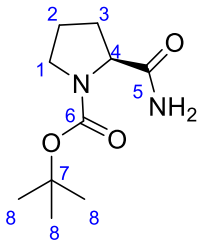
A flask containing *Z*-L-valine nitrile **L-165** (400 mg, 1.72 mmol), Pearlman's reagent (20% b.w., 119 mg) and ethyl acetate (15 mL) were evacuated and a hydrogen balloon added to the top of the flask. After 1 hour the mixture was filtered through a pad of celite and the celite washed thoroughly with ethyl acetate (50 mL). A solution of 4M HCl in dioxane (1.0 mL) was added and stirred for 10 minutes which turned the solution cloudy. Upon evaporation the chloride salt **L-166** was isolated as an off-white solid in a 75 % yield (173 mg, 1.29 mmol). The free, neutral amine was isolated by dissolving in DCM and stirring over solid sodium bicarbonate for 10 minutes. The mixture was then filtered and concentrated *in vacuo* to give the neutral amine **L-157** as a pale yellow oil**.** **IR** (ATR): ν_{max} 2960, 2866, 1727, 1707, 1160 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -8.3 (*c* = 0.83, DCM); **^{1}H NMR** ($CDCl_3$ 400 MHz): δ 3.52 (1H, d, J= 5.6 Hz, H-2), 1.93, (1H, dspt, 6.8, 5.6 Hz, H-3), 1.66 (2H, br s, NH_2), 1.07 (3H, d, J= 6.8 Hz, H-4), 1.06 (3H, d, J= 6.8 Hz, H-4); **^{13}C NMR** ($CDCl_3$ 400 MHz): δ 121.4 (C-1), 49.9 (C-2), 33.0 (C-3), 19.0 (C-4), 17.7 (C-4); **HRMS** (ESI): $[M+H]^+$ HRMS found 99.0917, $C_5H_{11}N_2$ required 99.0917. Spectroscopic data agrees with literature.[159]

L-Serine nitrile (L-158)



Z-L-Serine nitrile **L-176** (60 mg, 0.27 mmol) was dissolved in ethanol (5 mL). Pearlman's reagent (20% b.w., 18 mg) was added and the flask evacuated. A hydrogen balloon was added to the top of the flask and the reaction stirred for 10 minutes. The reaction mixture was filtered through a pad of celite and the pad washed with copious amounts of ethanol. The solvent was removed *in vacuo* to give the pure title compound **L-158** as a colourless oil in a 99 % yield (23 mg, 0.27 mmol). **IR** (ATR): ν_{max} 3288, 2935, 2867, 2240, 1595, 1454, 1361, 1051 cm^{-1}; **^{1}H NMR** (CD_3OD, 400 MHz): δ 3.81 (1H, dd, J = 5.5, 5.5 Hz, H-2), 3.70 (1H, dd, J= 10.5, 5.5 Hz, H-3), 3.66 (1H, dd, J= 10.5, 5.5 Hz, H-3); **^{13}C NMR** ($CDCl_3$, 400 MHz): δ 122.0 (C-1), 64.6 (C-3), 46.5 (C-2); **HRMS** (ESI): $[M+H]^+$ HRMS found 87.0552, $C_3H_7N_2O$ required 87.0553.

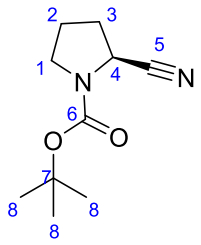
N-Boc-L-proline amide (L-160)



To a stirred solution of Boc-L-proline **L-159** (100 mg, 0.46 mmol), and triethylamine (46.5 mg, 0.46 mmol) in dioxane (2 mL) was added ethylchloroformate (0.1 mL, 1.06 mmol) at

room temperature. A 35% solution of $NH_3.H_2O$ (1.53 mL, 2.3 mmol), was added and the mixture stirred for 4 hours. Dioxane was removed *in vacuo* and the crude residue was redissolved in water (9 mL). The aqueous layer was partitioned against DCM (8 mL) and the organic layer extracted. The aqueous layer was extracted twice more with DCM (2 x 8mL), the combined organic layers dried over magnesium sulfate and the solvent removed *in vacuo* to give the title compound **L-160** as a white powder in an 81 % yield (78 mg, 0.37 mmol). **Mp** 94-97 °C; **IR** (ATR): ν_{max} 3373, 3201, 1669 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -20.1 (*c* = 1.0, chloroforn) literature -42.4 (*c* = 1.0, methanol);[160] **^{1}H NMR** (400 MHz, $CDCl_3$) δ 6.83 (1H, s, N-H), 5.56-6.13 (1H, m, N-H), 4.36-4.14 (1H, m, H-4), 3.55-3.17 (2H, m, H-1), 2.40-1.74 (4H, m, H-2, H-3), 1.45 (9H, s, H-8), **HRMS** (ESI): $[M+Na]^+$ HRMS found 237.1209, $C_{10}H_{18}N_2O_3Na$ required 237.1215.

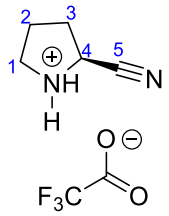
***N*-Boc-L-proline nitirle (L-161)**



To a stirred solution of **L-160** (1.5 g, 7.0 mmol) and triethylamine (4.85 mL, 35 mmol) in DCM (3 mL) at 0 °C was added trifluoroacetic anhydride (2.0 mL, 14.0 mmol) under a nitrogen atmosphere. After 30 minutes the reaction was warmed to room temperature. After a further 5 hours the mixture was washed with saturated sodium bicarbonate solution and extracted with DCM (3 x 8 mL). The combined organic extracts were washed with water (8 mL), then brine (8 mL), dried over magnesium sulfate and concentrated in *vacuo*.

The crude residue was purified by flash column chromatography (50:50 ethyl acetate:hexane) to give the title compound **L-161** as a colourless oil in a 94 % yield (1.29 g, 6.6 mmol). **IR** (ATR): ν_{max} 2976, 2239, 1797, 1692 cm^{-1}; **$[\alpha]_D$** (deg cm^3 g^{-1} dm^{-1}) -92.2 (*c* = 0.4, chloroform) literature -95.5 (*c* = 1.3, methanol)[160]; **^{1}H NMR** (400 MHz, $CDCl_3$) δ 4.39-4.60 (1H, m, H-4), 3.24−3.58 (2H, m, H-1), 1.93−2.30 (4H, m, H-2, H-3), 1.50 (9H, s, H-8); **^{13}C NMR** (400 MHz, $CDCl_3$) δ 153.1 (C-6), 119.3 (C-5), 81.6 (C-7), 47.3 (C-4), 45.8 (C-1), 31.8 (C-3), 28.4 (C-8), 23.9 (C-2); **HRMS** (ESI): $[M+Na]^+$ HRMS found 219.1107, $C_{10}H_{16}N_2O_2Na$ required 219.11040. Spectroscopic data is in agreement with the literature.[161]

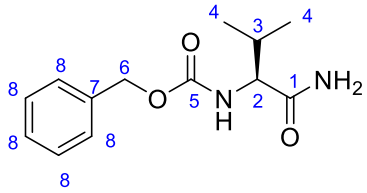
L-Proline nitrile trifluoracetic acid salt (L-162)



To a solution of **L-161** (580 mg, 2.96 mmol) in DCM (10 mL) at 0 °C was added trifluoroacetic acid (5.7 mL, 73.9 mmol). The solution was stirred for 16 hours and then concentrated *in vacuo* to give the compound **L-162** as the TFA salt in a 37 % yield (230 mg, 1.10 mmol). **Mp** 92-94 °C; **IR** (ATR): ν_{max} 2992, 2789, 2393, 1674 cm^{-1}, **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -16.7 (*c* = 1.0, methanol); **^{1}H NMR** (400 MHz, CD_3OD) δ 4.66 (1H, t, J = 7.4 Hz, H-4), 3.50-3.34 (2H, m, H-1), 2.54-2.44 (1H, m, H-3), 2.36-2.05 (3H, m, H-2, H-3); **^{13}C NMR** (400 MHz, $CDCl_3$) δ 116.5 (C-5), 47.9 (C-4), 47.0 (C-1), 31.2 (C-3), 24.5 (C-2); **HRMS** (ESI): $[M+H]^+$ found 97.0757, $C_5H_9N_2$ required 97.07602. Spectroscopic data is in agreement with the literature.[162,,163] The free amine was liberated through dissolution in DCM and the resultant

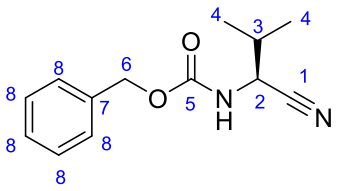
solution stirred over solid sodium bicarbonate for 10 mins. The mixture was then filtered and the filtrate concentrated *in vacuo*.

***Z*-L-Valine amide (L-164)**



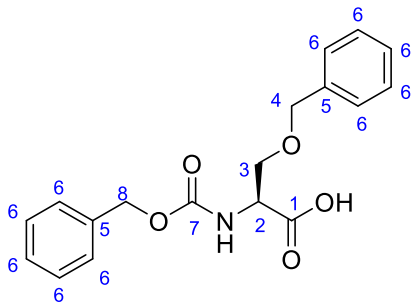
To a stirred solution of *Z*-L-valine **L-163** (2.0 g, 7.96 mmol) and triethylamine (1.2 mL, 8.8 mmol) in dry THF (40 mL) at 0 °C was added ethyl chloroformate (0.76 mL, 7.96 mmol). After 30 minutes 7N ammonia in methanol (1.66 mL, 11.9 mmol) was added and stirred at 0 °C for 1 hour and a further 19 hours at room temperature. The resultant white precipitate was filtered and washed with ice cold water to give the title compound **L-164** as a white solid in a 75 % yield (1.5 g, 6.0 mmol). **Mp** 205-208 °C; **IR** (ATR): ν_{max} 3380, 3316, 3063, 3027, 2957, 2873, 1683, 1645, 1536, 1455, 1305, 1246, 1040 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) +25.0 (*c* = 1.0, dimethyl formamide) literature +24.2 (*c* = 1.0, dimethyl formamide)[164]; **^{1}H NMR** (400 MHz DMSO d^6) δ 7.37-7.30 (6H, m, H-8, N-H), 7.16 (1H, d, J= 9.0 Hz, N-H), 7.04 (1H, s, N-H), 5.03 (2H, s, H-6), 3.80 (1H, dd, J= 9.0, 6.8 Hz, H-2), 1.95 (1H, apparent oct, J= 6.8 Hz, H-3), 0.86 (3H, d, J= 6.8 Hz, H-4), 0.83 (3H, d, J= 6.8 Hz, H-4); **^{13}C NMR** (DMSO d^6 400 MHz) δ 173.3 (C-1), 156.2 (C-5), 131.2 (C-7), 128.4 (C-8), 127.0 (C-8), 65.4 (C-6), 60.1 (C-2), 30.2 (C-3), 19.4, (C-4) 18.0 (C-4); **HRMS** (ESI): $[M+H]^+$ HRMS found 251.1387, $C_9H_{20}NO_2$ required 251.1390. $[M+Na]^+$ HRMS found 273.1217, $C_{13}H_{18}N_2O_3Na$, required 273.1210.

***Z*-L-Valine nitrile (L-165)**



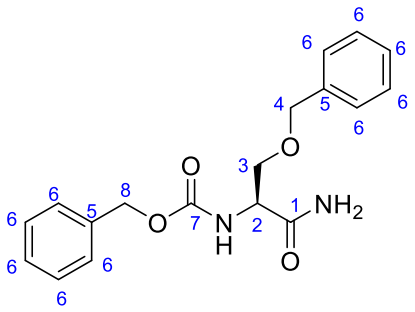
Z-L-Valine amide **L-164** (1.5 g, 6 mmol) and triethylamine (1.83 mL, 13.2 mmol) were dissolved in dry THF (20 mL) at 0 °C. After 30 minutes trifluoroacetic anhydride (1.26 mL, 9 mmol) was added and stirred at 0 °C for 1 hour and a further 14 hours at room temperature. The solvent was removed *in vacuo* and the crude oil redissolved in ethyl acetate. The organic layer was washed three times with 2M HCl, then once with brine, dried over magnesium sulfate, filtered and concentrated *in vacuo* to give the crude product as a red translucent oil. Purification *via* flash column chromatography (5:95 methanol:DCM) gave the crude product **L-165** as a colourless oil in a 93 % yield (1.3 g, 5.6 mmol), upon trituration a colourless crystalline solid was formed. **Mp** 53-56 °C; **IR** (ATR): ν_{max} 3294, 3062, 2980, 2929, 2243, 1690, 1535, 1467, 1455, 1321, 1303, 1253, 1136, 1028, 1049 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -37.3 (*c* = 0.97, methanol) literature -55 (*c* = 1.13, chloroform)[165]; **^{1}H NMR** (400 MHz DMSO d^6): δ 8.22 (1H, br d, J= 8.0 Hz, N-H), 7.39-7.33 (5H, m, H-8), 5.09 (2H, s, H-6), 4.40 (1H, dd, J= 8.0, 8.0 Hz, H-2), 1.98, (1H, m, H-3), 1.00 (3H, d, J= 6.8 Hz, H-4), 0.94 (3H, d, J= 6.8 Hz, H-4); **^{13}C NMR** (400 MHz $CDCl_3$): δ 155.5 (C-5), 135.7 (C-7), 128.8 (C-8), 128.6 (C-8), 128.4 (C-8),117.8 (C-1), 67.9 (C-6), 49.1 (C-2), 31.9 (C-3), 18.6 (C-4), 18.0 (C-4); **HRMS** (ESI): $[M+Na]^+$ HRMS found 255.1113, $C_{13}H_{16}N_2O_2Na$ required 255.1104.

***Z*-(O-Benzyl)-L-Serine (L-168)**



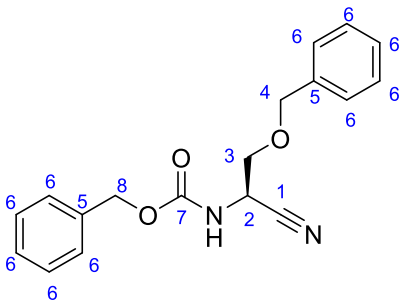
To a solution of *O*-benzyl-L-serine **L-167** (1.0 g, 5.1 mmol) in dioxane (5 mL) and 2M NaOH solution (5 mL) was added benzyl chloroformate (1.15 mL, 8.2 mmol) dropwise over 5 minutes at 0 °C. The solution was warmed to room temperature and stirred for a further 2 hours before concentration *in vacuo*. The crude reaction mixture was redissolved in diethyl ether (30 mL) and washed with water (30 mL). The aqueous layer was extracted twice more with ether (2 x 30 mL). The aqueous layer was then acidified to pH 1-2 with 2M HCl and extracted with DCM (3 x 30 mL). The organic layers were combined, dried over sodium sulfate, filtered and concentrated *in vacuo* to give the title compound **L-168** in a 67 % yield as an off-white solid (1.1 g, 3.4 mmol). **Mp** 98–99 °C; **IR** (ATR): ν_{max} 3368, 3164, 2949, 2879, 1748, 1731, 1665, 1539, 1258, 1200, 1108, 1054, 1027 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) +11.6 (*c* = 1.0, methanol) literature +11.3 (c = 1.0 g, methanol);[166] **^{1}H NMR** (400 MHz, DMSO d^6) δ 12.79 (1H, br s, O-H), 7.59 (1H, d, J = 8.2, N-H), 7.33-7.24 (10H, m, H-6), 5.00 (2H, s, H-8), 4.99 (1H, d, J = 12.8, H-8), 4.47 (1H, d, J = 12.1, H-4), 4.43 (1H, d, J = 12.1, H-4), 4.25 (1H, ddd, J = 4.6, 6.2, 8.2, H-2), 3.67 (1H,dd, J= 10.1, 4.6 Hz, H-3) 3.63 (1H, dd, J= 10.1, 6.2 Hz, H-3); **^{13}C NMR** ($CDCl_3$ 400 MHz): δ 171.7 (C-1), 156.1 (C-7), 138.1 (Ar), 137.0 (Ar), 128.4 (Ar), 128.2 (Ar), 127.9 (Ar), 127.8 (Ar), 127.5 (Ar), 127.5 (Ar), 127.5 (Ar), 72.1 (CH_2), 69.3 (CH_2), 65.5 (CH_2), 54.2 (C-2); **HRMS** (ESI): $[M+Na]^+$ HRMS found 352.1145, $C_{18}H_{19}NNaO_5$ required 352.1155. Physical and spectroscopic data is in agreement with the literature.[167]

***Z*-(O-benzyl)-L-Serine amide (L-169)**



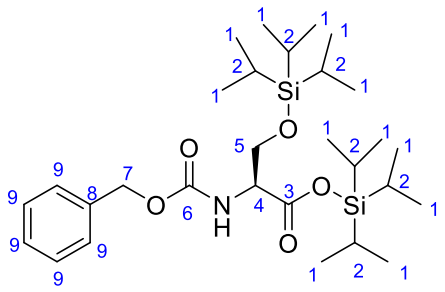
Ethyl chloroformate (0.29 mL, 3.04 mmol) was added dropwise (0.5 mL per min) to a solution of *Z*-(*O*-benzyl)-L-serine **L-168** (1.0 g, 3.04 mmol) and triethylamine (0.46 mL, 3.34 mmol) in dry THF (20 mL) at 0 °C. After 4 hours the solvent was removed *in vacuo* and the crude material redissolved in ethyl acetate (30 mL) and washed with saturated sodium bicarbonate solution (30 mL). The aqueous layer was extracted twice more with ethyl acetate (2 x 30 mL), the organic layers were combined and washed with brine (50 mL), dried over sodium sulfate, filtered and concentrated *in vacuo* to give the pure title compound **L-169** as a white powder in a 89 % yield (0.89 g, 2.70 mmol). **Mp** 129-131 °C; **IR** (ATR): ν_{max} 3446, 3313, 3204, 2864, 1654, 1530, 1240 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) +5.7 (*c* = 1.0, methanol); **^{1}H NMR** (400 MHz, DMSO d^6) δ 7.4-7.16 (13 H, m, , H-6, N-H), 5.06 (1H, d, J= 12.7, H-8), 5.02 (1H, d, J = 12.7, H-8), 4.48 (2H, s, H-4), 4.23 (1H, ddd, J= 8.5, 7.0, 5.0 Hz, H-2), 3.59 (1H, dd, J= 10.1, 5.0 Hz, H-3), 3.54 (1H, dd, J= 10.1, 7.0 Hz); **^{13}C NMR** ($CDCl_3$, 400 MHz): δ 171.6 (C-1), 156.0 (C-7), 138.2 (Ar), 137.1 (Ar), 128.4 (Ar), 128.2 (Ar), 128.1 (Ar), 127.7 (Ar), 127.5 (Ar), 127.5 (Ar), 72.0 (C-8), 70.0 (C-4), 65.5 (C-3), 54.7 (C-2); **HRMS** (ESI): $[M+Na]^+$ HRMS calculated for $C_{18}H_{20}N_2NaO_4$, 351.1315, found 351.1305.

***Z*-(O-benzyl)-L-Serine nitrile (L-170)**



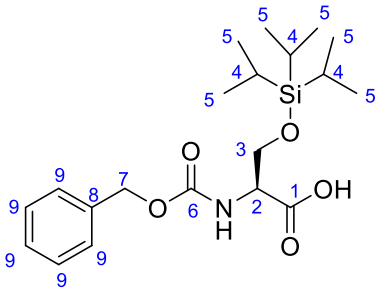
Triethylamine (0.68 mL, 4.90 mmol) was added to a suspension of *Z*-(O-benzyl)-L-serine amide **L-169** (735 mg, 2.24 mmol) in dry THF (20 mL) and stirred at 0 °C for 10 minutes before trifluoracetic anhydride (0.47 mL, 3.36 mmol) was added dropwise (0.1 mL / min) to the suspension. After a further 6 hours at 0 °C the solvent was removed *in vacuo*. The crude material was redissolved in DCM (20 mL) and washed with 2M HCl (20 mL). The aqueous layer was extracted twice more with DCM (2 x 20 mL). The organic extracts were combined, washed with brine, dried over sodium sulfate, filtered and concentrated *in vacuo* to give the crude product as a yellow oil. Purification *via* flash column chromatography (5:95 methanol:DCM) gave the title compound **L-170** as a white solid in a 76 % yield (530 mg, 1.71 mmol). **Mp** 51–53 °C; **IR** (ATR): ν_{max} 3292, 3036, 2871, 2250, 1694, 1684, 1527 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -19.9 (*c* = 1.0, methanol); **^{1}H NMR** (400 MHz, DMSO d^6) δ 8.30 (1H, d, J = 7.8 Hz, NH), 7.40-7.29 (10 H, m, H-6), 5.09 (2H, s, H-8), 4.85 (apparent dt, J = 7.8, 6.3 Hz, H-2), 4.56 (2H, s, H-4), 3.67 (2H, d, J = 6.3 Hz, H-3); **^{13}C NMR** ($CDCl_3$ 400 MHz): δ 155.2 (C-7), 136.7 (Ar), 135.6 (Ar), 128.8 (Ar), 128.6 (Ar), 128.5 (Ar), 128.4 (Ar), 128.0 (Ar), 117.4 (C-1), 73.8 (C-8), 68.9 (C-4), 67.9 (C-3), 43.1 (C-2); **HRMS** (ESI): $[M+Na]^+$ HRMS found 333.1201, $C_{18}H_{18}N_2NaO_3$ required 333.1210.

Z-(*O*-tri*iso*propylsilyl)-L-Serine-tri*iso*propyl silate (L-172)



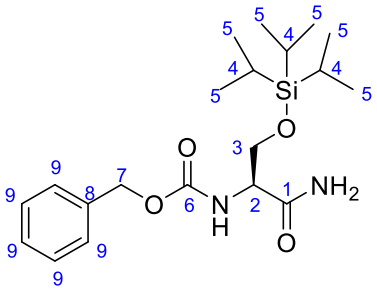
To a solution of *Z*-L-serine **L-171** (500 mg, 2.09 mmol) in dimethyl formamide (20 mL) was added 2,6-lutidine (0.49 mL, 4.18 mmol) and tri*iso*propylsilyl triflate (0.60 mL, 2.09 mmol). After 22 hours the solvent was removed *in vacuo* and the crude mixture acidified with 2M HCl to pH 2 (ca. 20 mL) and extracted with DCM (3 x 50 mL). The organic extracts were combined, washed with brine (20 mL), dried over magnesium sulfate, filtered and concentrated *in vacuo* to give the title compound **L-172** as a colourless oil in a 96 % yield (1.11 g, 2.01 mmol). **IR** (ATR): ν_{max} 3446, 2944, 2867, 1719, 1501, 1464, 1343, 1254, 1116, 1053 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) +4.51 (*c* = 1.38, DCM); **^{1}H NMR** ($CDCl_3$, 400 MHz): δ 7.36-7.30 (5H, m, H-9), 5.63 (1H, d, 8.3 Hz, N-H), 4.40 (ddd, 1H, 8.3, 2.8, 2.2 Hz, H-4), 4.22 (1H, dd, 9.9, 2.2 Hz, H-5), 4.00 (1H, dd, 9.9, 2.8 H-5), 1.34-1.01 (42H, m, H-1); **^{13}C NMR** ($CDCl_3$, 400 MHz): 170.3 (C-3), 156.0 (C-6), 136.6 (Ar), 128.6 (Ar), 128.2 (Ar), 66.9 (C-7), 64.1 (C-5), 57.1 (C-4), 18.0 (TIPS), 17.9 (TIPS), 17.8 (TIPS), 12.0 (TIPS); **HRMS** (ESI): $[M+H]^+$ HRMS found 574.3363, $C_{29}H_{53}NNaO_5Si_2$ required 574.3354.

***Z*-(Tri*iso*propylsilyl)-L-Serine (L-173)**



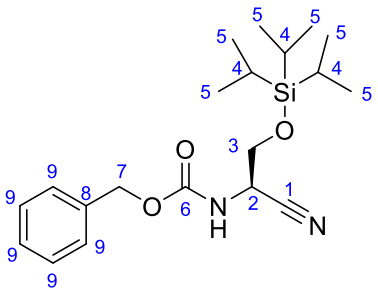
Z-(Tri*iso*propylsilyl)-L-serine-tri*iso*propyl silate **L-172** (948 mg, 1.72 mmol) was dissolved in a 1:1 mixture of THF and 1M sodium hydroxide (20 mL) and stirred vigorously. After 3 hours the pH was adjusted to pH 2 using 2M HCl and the THF removed *in vacuo*. The remaining aqueous solution was extracted with DCM (3 x 10 mL) and the combined organic extracts washed with brine (20 mL). The organic layer was dried over magnesium sulfate, filtered and concentrated *in vacuo* to give the crude product as a colourless oil. Purification *via* flash column chromatography (97:3 hexane:ethyl acetate) was used to remove tri*iso*propylsilanol and then adjusted (90:10 DCM:methanol) to elute the pure product **L-173** as yellow oil in a 94 % yield (636 mg, 1.61 mmol). **IR** (ATR): ν_{max} 3445, 2943, 2866, 1721, 1512, 1463, 1246, 1211, 1116, 1063 cm^{-1}, **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) +21.0 (*c* = 1.0, DCM); **^{1}H NMR** ($CDCl_3$, 400 MHz): δ 7.37-7.32 (5H, m, H-9), 5.61 (1H, d, 7.6 Hz, N-H), 5.13 (2H, s, H-7), 4.46 (1H, ddd, J= 7.6, 4.6, 2.5 Hz, H-2), 4.24 (1H, dd, J= 9.7, 2.5 Hz, H-3), 3.92 (1H, dd, J= 9.7, 4.6 Hz, H-3), 1.12-1.03 (21H, m, H-5); **^{13}C NMR** ($CDCl_3$, 400 MHz): δ 175.1 (C-1), 155.9 (C-6), 136.0 (Ar), 128.4 (Ar), 128.0 (Ar), 127.9 (Ar), 67.0 (C-7), 63.6 (C-3), 55.85 (C-2), 17.6 (TIPS), 11.6 (TIPS); **HRMS** (ESI): $[M+Na]^+$ HRMS found 418.2034, $C_{20}H_{33}NNaO_5Si$ required 418.2020. Spectroscopic data was in agreement with the literature.[168]

***Z*-(*O*-tri*iso*propylsilyl)-L-serine amide (L-174)**



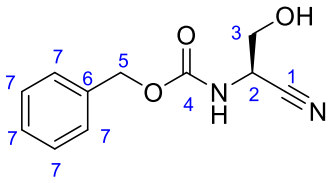
To a solution of *Z*-(*O*-tri*iso*propylsilyl)-L-serine **L-173** (4.50 g, 11.3 mmol) and triethylamine (1.90 mL, 13.5 mmol) in THF (50 mL) was added ethyl chloroformate (1.20 mL, 12.4 mmol) at 0 °C. After 60 minutes ammonia in methanol (7N, 1 mL) was added and the solution warmed to room temperature. After 3 hours the solvent was removed *in vacuo* and the crude mixture redissolved in DCM (50 mL). The organic layer was washed with saturated sodium bicarbonate solution (50 mL) then brine (50 mL) extracting each time with DCM (3 x 50 mL). The combined organic extracts were dried over magnesium sulfate, filtered and concentrated *in vacuo* to give the pure product **L-174** as a white solid in a 96 % yield (4.27 g, 10.8 mmol). **Mp** 70-74 °C; **IR** (ATR): ν_{max} 3063, 1656, 1515, 1440, 1182, 1132 cm^{-1}; **^{1}H NMR** ($CDCl_3$, 400 MHz): δ 7.36-7.31 (5H, m, H-9), 6.59 (1H, br s, NH), 5.70 (2H, m, NH), 5.12 (2H, s, H-7), 4.25-4.19 (1H, m, H-3), 4.17-4.12 (1H, m, H-3), 3.74-3.71 (1H, m, H-2), 1.07-1.05 (21H, m, H-5); **^{13}C NMR** ($CDCl_3$, 400 MHz): δ 172.7 (C-1), 156.0 (C-6), 136.2 (Ar), 128.7 (Ar), 128.4 (Ar), 128.3 (Ar), 67.2 (C-7), 63.5 (C-3), 55.6 (C-2), 18.0 (TIPS), 11.8 (TIPS); **HRMS** (ESI): $[M+Na]^+$ HRMS found 417.2163, $C_{20}H_{34}N_2NaO_4Si$ required 417.2180. Spectroscopic data was in agreement with the literature.[168]

***Z*-L-(Tri*iso*propylsilyl) serine nitrile (L-175)**



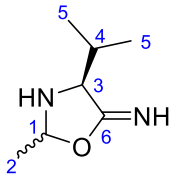
Triethylamine (0.90 mL, 6.36 mmol) was added to a solution of *Z*-L-(tri*iso*propylsilyl) serine amide **L-174** (1.14 g, 2.89 mmol) in THF (8 mL) at 0 °C. After 30 minutes trifluoroacetic anhydride (0.61 mL, 4.34 mmol) was added dropwise (0.1 mL / min) and stirred for a further 15 hours at room temperature. The solvent was removed *in vacuo* and the crude material redissolved in DCM (10 mL). The organic layer was washed with 2M HCl (10 mL), then brine (10 mL) extracting each time with DCM (3 x 10 mL). The combined organic extracts were dried over magnesium sulfate, filtered and concentrated *in vacuo* to give the pure title compound **L-175** as a light yellow oil in a 82 % yield (0.90 g, 2.38 mmol). **IR** (ATR): ν_{max} 3320, 2944, 2867, 2249, 1806, 1707, 1499, 1463, 1384, 1328, 1250, 1218, 1124, 1027 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -18.4 (*c* = 1.0, DCM); **^{1}H NMR** ($CDCl_3$, 400 MHz): δ 7.40-7.33 (5H, m, H-9), 5.43 (1H, d, J= 7.6 Hz, N-H), 5.16 (2H, s, H-7), 4.73-4.71 (1H, ddd, J= 7.6, 4.0, 2.9 Hz, H-2), 4.01 (1H, dd, J= 10.1, 2.9 Hz, H-3), 3.90 (1H, 10.1, 4.0 Hz, H-3), 1.08-1.05 (21H, m, H-5); **^{13}C NMR** ($CDCl_3$ 400 MHz): δ 155.3 (C-6), 135.7 (Ar), 128.8 (Ar), 128.7 (Ar), 128.5 (Ar), 117.6 (C-1), 67.9 (C-7), 63.7 (C-3), 45.1 (C-2), 18.0 (TIPS), 11.9 (TIPS); **HRMS** (ESI): $[M+Na]^+$ found 399.2062, $C_{20}H_{32}N_2NaO_3Si$ required 399.2074.

***Z*-L-serine nitrile (L-176)**



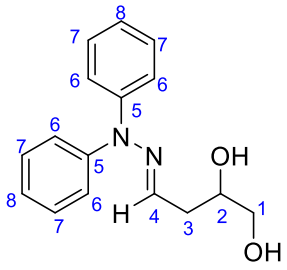
To a solution of *Z*-L-(tri*iso*propylsilyl) serine nitrile **L-175** (700 mg, 1.86 mmol), in pyridine (4 mL) was added a solution of hydrogen flouride (70 % bwt in pyridine) (0.50 mL, 19.2 mmol). After 5 hours the reaction was washed with saturated sodium bicarbonate solution (5 mL) and extracted with ethyl acetate (3 x 5mL). The organic extracts were combined and washed with 1M HCl (2 x 5 mL), then brine (5 mL) and dried over sodium sulfate, filtered and concentrated *in vacuo* to give the pure product **L-176** as a colourless oil in a 67 % yield (274 mg, 1.25 mmol). **IR** (ATR): ν_{max} 3423, 3281, 3064, 2960, 2257, 1698, 1541, 1252, 1069 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -45.3 (*c* = 1.0, DCM); **^{1}H NMR** ($CDCl_3$, 400 MHz): δ 7.38-7.31 (5H, m, H-7), 5.87 (1H, d, J= 8.5 Hz, N-H), 5.13 (2H, s, H-5), 4.70-4.68 (1H, ddd, J = 8.5, 4.2, 3.6 Hz, H-2), 3.89 (1H, dd, J= 11.5, 3.6 Hz, H-3), 3.79 (1H, dd, J= 11.5, 4.2 Hz, H-3), 2.74 (1H, br s, O-H); **^{13}C NMR** ($CDCl_3$, 400 MHz): δ 155.5 (C-4), 135.5 (Ar), 128.8 (Ar), 128.7 (Ar), 1284 (Ar), 117.5 (C-1), 68.0 (C-5), 62.6 (C-3), 45.1 (C-2); **HRMS** (ESI): $[M+Na]^+$ HRMS found 243.0738, $C_{11}H_{12}N_2NaO_3$ required 243.0740.

4-Isopropyl-2-methyl-oxazolidin-5-ylidene amine (177)



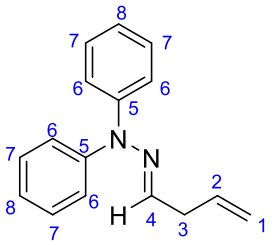
L-valine nitrile, **L-157** (43 mg, 0.44 mmol) was dissolved in water (1 mL) and added to a solution of acetaldehyde **24** (19 mg, 0.44 mmol) in water (1 mL). The solution was stirred at room temperature for 24 hours before the solvent was removed *in vacuo*. Purification *via* flash column chromatography (95:5 DCM:methanol) gave the title compound **177**, as a colourless oil in a 16 % yield (10 mg, 0.07 mmol) as a 3:1 mixture of diastereomers. **IR** (ATR): ν_{max} 3239, 2962, 2931, 2872, 1688, 1464, 1432, 1380, 1346, 1292, 1099 cm^{-1}; **^{1}H NMR** ($CDCl_3$, 400 MHz) major: δ 6.67 (1H, br s, N-H), 4.66 (1H, q, J = 5.6 Hz, H-1), 3.41 (1H, d, J = 4.2 Hz, H-3), 2.15 (1H, dsept, J = 6.9, 4.2 Hz, H-4), 1.85 (1H, br s, N-H), 1.34 (3H, d, J = 5.6 Hz, H-2), 1.04 (3H, d, J = 6.9 Hz, H-5), 0.92 (3H, d, J = 6.9 Hz, H-5); Minor diastereomer δ 6.76 (1H, br s, N-H), 4.68 (1H, q, J = 5.6 Hz, H-1), 3.42 (1H, d, J = 4.2 Hz, H-3), 2.11 (1H, dsept, J = 6.9, 4.2 Hz, H-4), 1.85 (1H, br s, N-H), 1.34 (3H, d, J = 5.6 Hz, H-2), 1.02 (3H, d, J = 6.9 Hz, H-5), 0.93 (3H, d, J = 6.9 Hz, H-5) **^{13}C NMR** ($CDCl_3$ 400 MHz): As a mixture of diastereomers. Major δ 178.4 (C-6), 65.6 (C-1), 65.2 (C-3), 29.0 (C-4), 23.2 (C-2), 20.0 (C-5), 17.1 (C-5); Minor δ 178.4 (C-6), 67.1 (C-1), 64.3 (C-3), 30.1 (C-4), 24.0 (C-2), 20.0 (C-5), 17.4 (C-5); **HRMS** (ESI): $[M+H]^+$ HRMS found 143.1171, $C_7H_{15}N_2O$ required 143.1179. $[M+Na]^+$ HRMS found 165.0998, required $C_7H_{14}N_2ONa$, 165.0998.

4-[(*N*,*N*-diphenyl)-hydrazono]-butane-1,2,-triol (184)



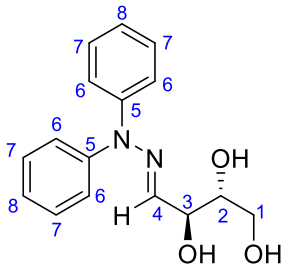
To a flask containing hydrazone **188** (20 mg, 0.085 mmol), 4-methyl morpholine-N-oxide (20 mg, 0.17 mmol) and *t*-butyl alcohol (0.25 mL) was added a solution of osmium tetroxide (0.43 mg, 0.0017 mmol) in THF (1.75 mL). The reaction was stirred for 23 hours before a solution of saturated sodium thiosulfate (2 mL) was added to quench the reaction. After 40 minutes the product was extracted with ethyl acetate (3 x 3 mL) and the combined organic extracts washed with water (1 x 5 mL) then brine (1 x 5 mL), dried over magnesium sulfate, filtered and concentrated *in vacuo*. Purification *via* silica filtration gave the title compound **184** as a colourless oil in an 87 % yield (20 mg, 0.074 mmol). **IR** (ATR): ν_{max} 3359, 2923, 1589, 1494, 1298, 1210 cm^{-1}; **^{1}H NMR** (400 MHz, CD_3OD) δ 7.33 (4H, dd, J = 7.8, 7.3 Hz, H-7), 7.09 (2H, t, J = 7.3 Hz, H-8), 7.02 (4H, d, J = 7.8 Hz, H-6), 6.59 (1H, dd, J = 5.5, 5.5 Hz, H-4), 3.75 (1H, dddd, J = 7.3, 6.0, 5.5, 4.6 Hz, H-2), 3.48 (1H, dd, J = 11.0, 4.6 Hz, H-1), 3.42 (1H, dd, J = 11.0, 6.0, H-1), 2.47 (1H, apparent ddd, J = 14.7, 5.5, 5.5 Hz, H-3), 2.36 (1H, ddd, J = 14.7, 7.3, 5.5 Hz, H-3); **^{13}C NMR** (400 MHz, CD_3OD): δ 145.6 (C-5), 138.2 (C-4), 130.7(C-7), 125.2 (C-8), 123.5 (C-6), 71.8 (C-2), 66.9 (C-1), 37.7 (C-3); **HRMS** (ESI): $[M+H]^+$ HRMS found 271.1431, $C_{16}H_{19}N_2O_2$ required 271.1441. $[M+Na]^+$ HRMS found 293.1247, $C_{16}H_{18}N_2NaO_2$ required 293.1260.

<u>*N*-But-3-enylidene-*N*,*N*-diphenylhydrazone</u> (188)



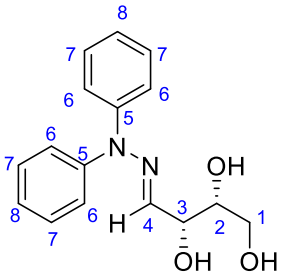
3-Butenal **187** (71 mg, 1.01 mmol) and *N*-*N*-diphenyl hydrazine **144** (223 mg, 1.21 mmol) were stirred in DCM (1 mL) at room temperature for 17 hours before the solvent was removed *in vacuo* to give a brown solid. Purification *via* flash column chromatography (95:5 petroleum ether:ethyl acetate) gave the pure title compound, **188**, as a yellow oil in a 36 % yield (84 mg, 0.35 mmol). **IR** (ATR): ν_{max} 3061, 1589, 1494, 1298, 1208 cm^{-1}; **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.40-7.35 (4H, m, H-7), 7.15-7.09 (6H, m, H-6, H-8), 6.49 (1H, t, J = 5.5 H.z, H-4), 5.88 (1H, ddt, J = 17.9, 9.5, 6.4 Hz, H-2), 5.08-5.07-5.03 (2H, m, H-1), 3.06 (2H, dd, J = 6.4, 5.5 Hz, H-3); **^{13}C NMR** (400 MHz, $CDCl_3$) δ 143.9 (C-5), 136.9 (C-4), 134.2 (C-2), 129.5 (C-7), 123.8 (C-8), 122.2 (C-6), 116.4 (C-1), 36.9 (C-3); **HRMS** (ESI): $[M+H]^+$ HRMS found 237.1379, $C_{16}H_{17}N_2$ required 237.1386. $[M+Na]^+$ HRMS found 259.1196, $C_{16}H_{16}N_2Na$ required 259.1206.

(2-*R*-3-*S*,*E*)-4-[(*N*,*N*-Diphenyl)-hydrazono]-butane-1,2,3-triol (D-189)



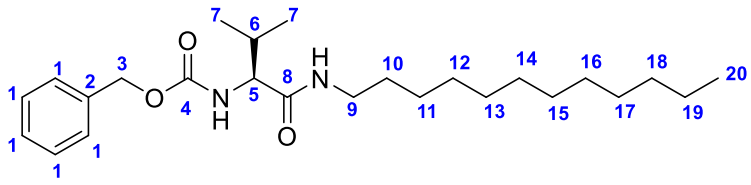
D-Erythrose **D-102** (20 mg, 0.16 mmol) and *N*,*N*-diphenyl hydrazine **144** (35 mg, 0.19 mmol) were dissolved in methanol (3 mL) and stirred for 45 minutes at room temperature before concentration *in vacuo*. Purification *via* flash column chromatography (95:5 DCM:methanol) gave the title compound **D-189** as a brown oil in a 57 % yield (26 mg, 0.091 mmol). **IR** (ATR): ν_{max} 3352, 3061, 2925, 1590, 1494, 1297, 1212, 1040 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -10.9 (*c* = 0.85, chloroform); **^{1}H NMR** (400 MHz, CD_3OD) δ 7.37-7.34 (4H, m, H-7), 7.12 (2H, t, J = 7.4, H-8), 7.06 (4H, m, H-6), 6.54 (1H, d, J = 6.0, H-4), 4.24 (1H, dd, J = 6.0, 5.9, H-3), 3.67-3.60 (2H, m, Ha, H-2), 3.55-3.50 (1H, m, H-1); **^{13}C NMR** (400 MHz, CD_3OD): δ 145.2 (C-5), 139.2 (C-4), HRMS 130.7 (C-7), 125.4 (C-8), 123.5 (C-6), 75.5 (C-2), 73.7 (C-3), 64.3 (C-1); **HRMS** (ESI): $[M+H]^+$ HRMS found 287, $C_{16}H_{19}N_2O_3$ required 287.1390. $[M+Na]^+$ found 309.1212, $C_{16}H_{18}N_2NaO_3$ required 309.1210.

(2-*S*-3-*S*,*E*)-4-[(*N*,*N*-Diphenyl)-hydrazono]-butane-1,2,3-triol (L-190)



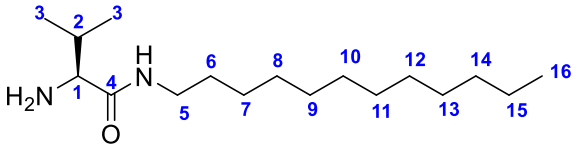
L-threose **L-103** (20 mg, 0.16 mmol) and *N*,*N*-diphenyl hydrazine **144** (35 mg, 0.19 mmol) were dissolved in methanol (3 mL) and stirred at room temperature for 1 hour before concentration *in vacuo*. Purification *via* flash column chromatography (95:5 DCM:methanol) gave the title compound **L190** as a brown oil in a 79 % yield (36 mg, 0.13 mmol). **IR** (ATR): ν_{max} 3353, 3060, 2927, 1590, 1494, 1296, 1212, 1037 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -5.2 (*c* = 1.1, chloroform); **^{1}H NMR** (400 MHz, CD_3OD) δ 7.35 (4H, dd, J = 8.4, 7.5 Hz, H-7), 7.18 (2H, t, J = 7.5, H-8), 7.05 (4H, d, J = 8.4 Hz, H-6), 6.51 (1H, d, J = 5.7, H-4), 4.27 (1H, dd, J = 5.7, 4.4 Hz, H-3) 3.63 (1H, ddd, J= 8.0, 5.0, 4.4 Hz, H-2), 3.63 (1H, dd, J= 12.5, 5.0 Hz, H-1), 3.50 (1H, dd, J = 12.5, 8.0 Hz, H-1); **^{13}C NMR** (400 MHz, CD_3OD): δ 145.1 (C-5), 139.0 (C-4), 130.8 (C-7), 125.5 (C-8), 123.5 (C-6), 75.2 (C-2), 73.4 (C-3), 64.0 (C-1); **HRMS** (ESI): $[M+H]^+$ HRMS found 287.1395, $C_{16}H_{19}N_2O_3$, required 287.1390. $[M+Na]^+$ HRMS found 309.1210, $C_{16}H_{18}N_2NaO_3$ required 309.1210.

***Z*-L-Valine dodecylamide (L-203)**



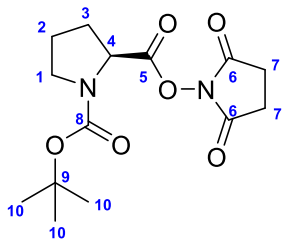
Triethylamine (2.8 mL, 20 mmol) was added to a stirred solution of *Z*-L-valine **L-163** (5.00 g, 20 mmol) in THF (50 mL) at 0 ˚C. After 5 minutes ethyl chloroformate (1.90 mL, 20 mmol) was added dropwise (0.2 mL/min) to the solution and then stirred for a further 30 minutes. Dodecylamine (3.69 g, 20 mmol) was dissolved in THF (10 mL) and added to the mixture causing a white precipitate to form. This was dissolved upon the addition of more THF (50 mL). The reaction was warmed to room temperature and stirred for a further 14 hours. The resulting white solid was filtered and washed with ice cold 1M NaOH (15 mL), followed by ice cold methanol (15 mL) and finally water (15 mL) to give the title compound **L-203** as a crystalline white solid in a 99 % yield (8.70 g, 20.7 mmol). **Mp** 125-127 ºC; **IR** (ATR): ν_{max} 3291, 2919, 2851, 1687, 1645, 1538, 1293, 1246, 1041 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -11.5 (c= 1.0 g cm^{-3}, chloroform); **^{1}H NMR** (400 MHz $CDCl_3$); δ 7.38-7.29 (5H, m, H-1), 5.85-5.83 (1H, br t, J = 5.5 Hz, N-H amide), 5.35 (1H,br d, J = 9.0, N-H carbamate), 5.10 (2H, s, H-3), 3.90 (1H, dd, J= 9.0, 6.3 Hz, H-5), 3.30-3.16 (2H, m, H-9), 2.14-2.09 (1H, sptd, J= 6.9, 6.3 Hz, H-6), 1.51-1.44 (2H, m, H-10), 1.34-1.22 (18H, m, H-11 - H-19), 0.95 (3H, d, J= 6.9 Hz, H-7), 0.92 (3H, d, J= 6.9 Hz, H-7), 0.88 (3H, t, J= 6.9, H-20), **^{13}C NMR** (400 MHz, $CDCl_3$); 171.1 (C-8), 156.2 (C-4), 136.0 (Ar), 128.7 (Ar), 128.4 (Ar), 128.2 (Ar), 66.9 (C-3), 60.5 (C-5), 39.4 (C-9), 31.7 (CH_2), 30.8 (C-6), 29.5 (CH_2), 29.4 (CH_2), 29.3 (CH_2), 29.2 (CH_2), 29.1 (CH_2), 26.7 (CH_2), 19.1 (C-7), 17.7 (C-7), 13.9 (C-20); **HRMS** (ESI): $[M+Na]^+$ HRMS found 441.3074 $C_{25}H_{42}N_2NaO_3$ required 441.3088.

L-Valine dodecylamide (L-204)



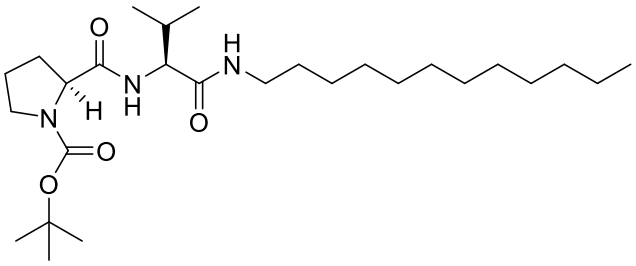
Methanol (25 mL) was added to a mixture of *Z*-L-valine dodecylamide **L-203** (0.50 g, 1.19 mmol) and palladium on carbon (0.15 g, 10% b.w.) and the flask was evacuated. A hydrogen balloon was added to the top of the flask and the reaction stirred for 1 hour. The reaction mixture was filtered through a pad of celite and the pad washed with copious amounts of methanol and the solvent reduced *in vacuo* to give compound **L-204** as a white solid in a 99 % yield (0.33 g, 1.17 mmol). **Mp** 132-133 $^{\circ}$C; **IR** (ATR): ν_{max} 3295, 2917, 2845, 1641, 1553, 1472, 1463, 1234, 1160 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -27.7 (*c* = 1.0, chloroform); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.29 (1H, br s, NH amide), 3.29-3.16 (2H, m, H-5), 3.20 (1H, d, J= 3.7 Hz, H-1), 2.29 (1H, sptd, J= 6.9, 3.7 Hz, H-2), 1.51-1.46 (2H, m, H-6), 1.26 (18H, m, H-7 – H-15), 0.97 (3H, d, J= 6.9 Hz, H-3), 0.87 (3H, t, J= 6.9 Hz, H-16), 0.80 (3H, d, J= 6.9 Hz, H-3); **^{13}C NMR** (400 MHz $CDCl_3$) δ 174.3 (C-4), 60.3 (C-1), 39.1 (C-5), 32.0 (CH_2), 30.9 (C-2), 29.8 (CH_2), 29.7(CH_2), 29.7(CH_2), 29.7(CH_2), 29.6 (CH_2), 29.4 (CH_2), 29.4 (CH_2), 27.1(CH_2), 22.8 (CH_2), 19.8 (C-3), 16.04 (C-3), 14.19 (C-16); **HRMS** (ESI) $[M+H]^+$ HRMS found 482.3936 $C_{27}H_{52}N_3O_4$ required 482.3952; $[M+Na]^+$ HRMS found 504.3752 $C_{27}H_{51}N_3O_4Na$ required 504.3772.

***N*-Boc-L-proline-*O*-succinimide (L-206)**



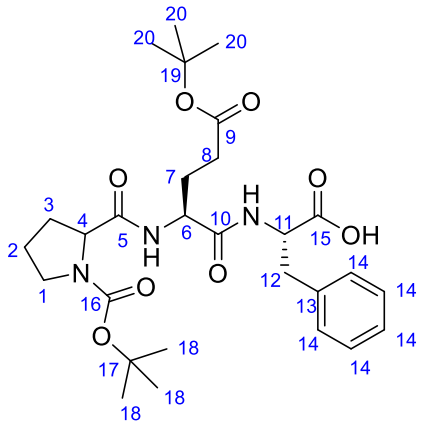
Boc-L-Proline **L-159** (1.01 g, 4.46 mmol) and *N*-hydroxysuccinimide **205** (0.55 g, 4.91 mmol) were dissolved in dry DMF (20 mL) and stirred at 0 °C for 30 minutes. 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (1.02 g, 5.35 mmol) was dissolved in dry DMF (10 mL) and added dropwise (0.1 mL / min) to the stirred solution. The reaction was stirred for 1 hour at 0 °C and a further 14 hours at room temperature before the solvent was removed *in vacuo*. The resulting oil was partitioned between ethyl acetate (20 mL) and 2M HCl (20 mL) and extracted twice more with ethyl acetate (2 x 20 mL). The combined organic layers were washed with brine (20 mL) then water (20 mL), extracting each time with ethyl acetate (20 mL). The organic layers were combined, dried over magnesium sulfate and concentrated *in vacuo* to give the title compound **L-206** as a white solid in an 88% yield (1.23 g, 3.94 mmol). **Mp** 132-133 °C; **IR** (ATR): $\mathbf{\nu_{max}}$ 2938, 1817, 1789, 1738, 1698, 1387, 1365, 1199, 1157, 1079 cm^{-1}; $\mathbf{[\alpha]_D^{25}}$ (deg cm^3 g^{-1} dm^{-1}) -57.6 (*c* = 1.0, chloroform) literature -55 (c = 2.1, dioxane);[169] **^{1}H NMR** (400 MHz $CDCl_3$); δ 4.54 (1H, dd, *J*= 8.8, 3.7 Hz, H-4), 3.59 (1H, ddd, J= 10.1, 7.3, 4.6 Hz, H-1), 3.46 (1H, ddd, J= 10.1, 7.3, 7.3 Hz, H-1), 2.87-2.81 (4H, m, H-7), 2.43-2.36 (1H, m, H-3), 2.34-2.25 (1H, m, H-3), 2.13-1.84 (2H, m, H-2), 1.46 (9H, s, H-10); **^{13}C NMR** (400 MHz $CDCl_3$); 168.9 (CO), 168.8 (CO), 153.6 (C-8), 81.2 (C-9), 57.3 (C-4), 46.4 (C-1), 31.5 (C-2), 28.2 (C-10), 25.7 (C-7), 23.6 (C-3); **HRMS** (ESI); $[M+K]^+$; HRMS found 351.0937 $C_{14}H_{20}KN_2O_6$ required 351.0953; $[M+Na]^+$; HRMS found 335.1200 $C_{14}H_{20}N_2NaO_6$ required 335.1214. Spectroscopic data was in agreement with the literature.[170]

<u>*N*-Boc-L-Proline-L-valine dodecylamide (207)</u>



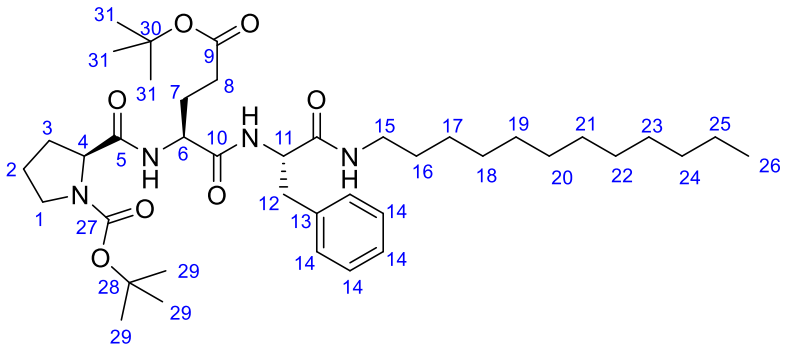
N-Boc-L-proline-*O*-succinimide **L-206** (3.00 g, 9.7 mmol) and L-valine dodecylamide **L-204** (3.30 g, 11.6 mmol) were dissolved in DME (85 mL) and stirred at 50 °C for 2 hours upon which the reaction was deemed complete by TLC. The solvent was removed *in vacuo* and the resulting colourless oil partitioned between DCM (50 mL) and saturated sodium bicarbonate solution (50 mL). The organic layer was washed with brine (50 mL) and extracted three times with DCM (3 x 50 mL). The organic layers were combined, dried over magnesium sulfate and concentrated *in vacuo* to give **207** as a colourless oil in a 99 % yield (4.70 g, 9.7 mmol). **IR** (ATR): ν_{max} 3288, 2923, 2853, 1702, 1641, 1547, 1466, 1392, 1365, 1235, 1162 cm^{-1}; **HRMS** (ESI) $[M+H]^+$ HRMS found 482.3933, $C_{27}H_{52}N_3O_4$ required 482.3952. $[M+Na]^+$ HRMS found 504.3755, $C_{27}H_{51}N_3O_4Na$ required 504.3772. The ^1H NMR spectrum was very broad with no splitting, probably due to rotamers, so the product was used in the next step without full characterisation.

***N*-Boc-L-Proline-L- (*O-tert*butyl)glutamic acid-L-phenylalanine (208)**



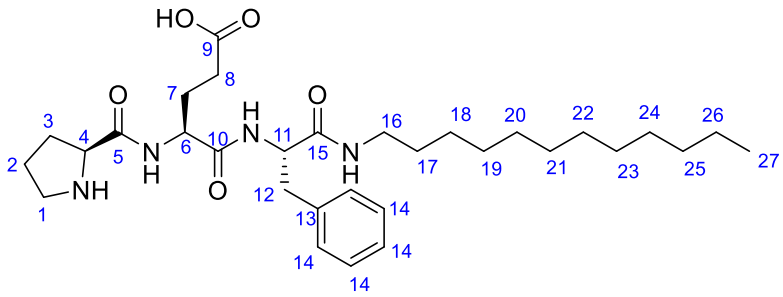
N-Boc-L-Proline-L-glutamic acid(*O-tert*butyl)-L-phenylalanine **208** was prepared using the general procedure for one pot peptide synthesis using cleavage method B to remove the peptide from resin (**Section 10.3.3**). The tripeptide was acquired as a white powder (as the TFA salt). **^{1}H NMR** (400 MHz, $CDCl_3$); δ 7.81 (1H, br d, J= 7.1 Hz, N-H), 7.34 (1H, br d, J= 7.3 Hz, N-H), 7.25-7.15 (5H, m, H-14), 4.78-4.69 (1H, m, H-11), 4.53-4.32 (1H, m, H-6), 4.24-4.14 (1H, m, H-4), 3.50-3.31 (2H, m, H-1), 3.26-3.12 (1H, m, H-12), 3.05-2.96 (1H, m, H-12), 2.25-2.21 (2H, m, H-8), 2.09-1.78 (6H, m, H-2, H-3, H-7), 1.44-1.37 (18H, m, H-18 & H-20); **^{13}C NMR** (400 MHz, CD_3OD) δ 173.6 (CO), 173.4 (CO), 173.2 (CO), 171.3 (C=), 155.8 (C-16), 136.6 (C-13), 129.4 (C-14), 128.5 (C-14), 126.9 (C-14), 87.2 (C-19), 81.0 (C-17), 60.4 (C-4), 53.7 (C-11), 53.3 (C-6), 43.7 (C-1), 37.4 (C-12), 31.6 (CH_2) 29.1 (CH_2), 28.4 (CH_3), 28.2 (CH_2), 28.1 (CH_3), 27.0 (CH_2), 24.7 (CH_2).

_N_-Boc-L-Proline-(_O-tert_butyl)L-glutamicacid-L-pheneylalanine dodecylamide (209)



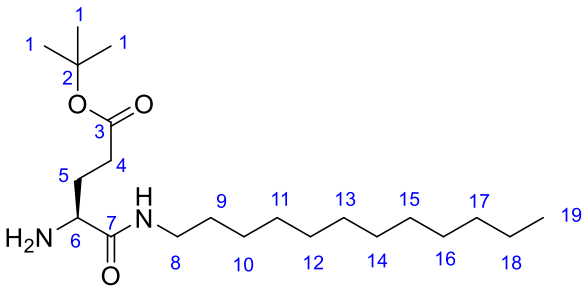
Triethylamine (0.15 mL, 0.91 mL) was added to a stirred solution of *N*-Boc-L-proline-(*O-tert*butyl)L-glutamic acid-L-phenylalanine **208** (500 mg, 0.91 mmol) in THF (10 mL) at 0 ˚C followed by ethylchlorformate (0.12 mL, 0.91 mmol). After 30 minutes dodecylamine (180 mg, 0.97 mmol) in THF (4 mL) was added and stirred at room temperature for 14 hours. The resulting white solid was filtered and washed with ice cold water (10 mL). The crude product was purified *via* column chromatography (30:70 to 50:50 ethyl acetate:hexane) to give the title compound **209** as a white solid in a 52 % yield (340 mg, 0.48 mmol). **Mp** 96.5-98.1 ˚C; **IR** (ATR): ν_{max} 3279, 2926, 2855, 1731, 1702, 1635, 1538, 1391, 1365, 1246, 1151 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -39.7 (*c* = 1.0, DCM); **^{1}H NMR** (400 MHz, $CDCl_3$) δ 8.33 (1H, br d, J= 4.6 Hz, NH amide), 7.32 (1H, br d, J= 9.2 Hz, NH amide), 7.25-7.12 (5H, m, H-14), 6.86 (1H, br t, J= 5.5 Hz, N-H amide), 4.90-4.80 (1H, m, H-11), 4.19-4.03 (2H, m, H-4, H-6), 3.65-3.09 (5H, m, H-1, H-15, H-12), 2.84 (1H, dd, J = 14.7, 11.0 Hz, H-12), 2.25-1.72 (8H, m, H-2, H-3, H-7, H-8), 1.55-1.37 (20 H, m, H-16 – H-25), 1.30-1.15 (18 H, m, H-29 & H-31), 0.85 (3H, t, J= 6.6 Hz, H-26); **^{13}C NMR** (400 MHz, $CDCl_3$); δ 175.0 (CO), 174.1 (CO), 170.9 (CO), 155.9 (CO), 138.3 (Ar), 129.4 (Ar), 128.8 (Ar), 128.5 (Ar), 128.3 (Ar), 126.3 (Ar), 81.6, 81.2, 81.0, 61.5 (CH), 56.8 (CH), 53.6 (CH), 47.4 (CH_2), 39.9 (CH_2), 37.1 (CH_2), 32.0 (CH_2), 32.0 (CH_2), 29.8 (CH_2), 29.8 (CH_2), 29.5 (CH_2), 29.4 (CH_2), 29.4 (CH_2), 28.5 (CH_3), 28.2 (CH_3), 27.0 (CH_2), 27.0 (CH_2), 25.3 (CH_2), 24.7 (CH_2), 22.8 (CH_2), 14.2 (CH_3); **HRMS** (ESI): $[M+Na]^+$ HRMS found 737.4795 $C_{40}H_{66}N_4NaO_7$ required 737.4824. Spectroscopic data was in agreement with the literature.[152]

L-Proline-L-glutamic acid-L-phenyalanine dodecylamide (210)



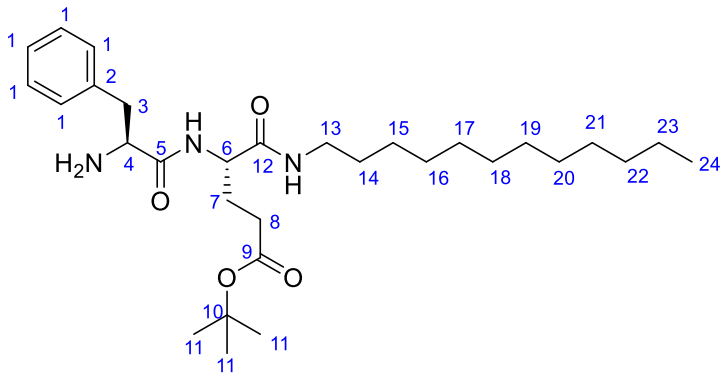
TFA (0.8 mL, 10.9 mmol) was added to a stirred solution of **209** (260 mg, 0.36 mmol) in DCM (2 mL). After 4 hours the solvent was removed *in vacuo*, and the residue redissolved in DCM (3 mL) and washed with 2M HCl (3 mL). The organic layer was dried over magnesium sulfate, filtered and concentrated *in vacuo* to give the title compound **210** as a white solid in a 47 % yield (95 mg, 0.17 mmol). **Mp** 116-118 °C; **IR** (ATR): ν_{max} 3281, 3069, 2923, 2853, 1645, 1556, 1177, 1135 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -14.5 (c= 1.0, methanol); **^{1}H NMR** (400 MHz, CD_3OD) δ 7.95-7.91 (1H, NH amide), 7.24-7.17 (5H, m, H-14), 4.60-4.49 (1H, m, H-11), 4.38-4.24 (2H, m, H-4, H-6), 3.35-2.77 (6H, m, H-1, H-16, H-12), 2.38-1.81 (8H, m, H-2, H-3, H-7, H-8), 1.36-1.15 (20 H, m, H-17 – H-26), 0.87 (3H, t, J= 6.6 Hz, H-27); **^{13}C NMR** (400 MHz, $CDCl_3$); δ 176.4 (CO), 172.7 (CO), 169.8 (CO), 138.2 (Ar), 130.4 (Ar), 130.3 (Ar), 129.6 (Ar), 129.5 (Ar), 127.8 (Ar), 60.9 (CH), 56.1 (CH), 54.4 (CH), 47.5 (CH_2), 40.4 (CH_2), 39.3 (CH_2), 31.0 (CH_2), 30.8 (CH_2), 30.7 (CH_2), 30.5 (CH_2), 30.4 (CH_2), 30.2 (CH_2), 28.3 (CH_2), 27.9 (CH_2), 25.0 (CH_2), 23.7 (CH_2), 14.5 (C-27); **HRMS** (ESI): $[M+H]^+$ HRMS found 559.3863, $C_{31}H_{51}N_4O_5$ required 559.3854. $[M+Na]^+$; HRMS found 581.3682, $C_{31}H_{50}N_4NaO_5$ required 581.3673

L-Glutamic acid (*O-tert*butyl)-dodeclyamide (213)



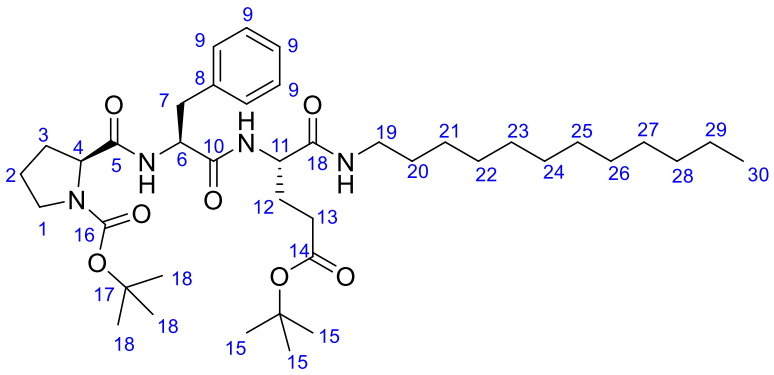
Methanol (30 mL) was added to a mixture of *Z*-(*O-tert*butyl)-Glutamic acid dodecylamide **223** (1.2 g, 2.4 mmol) and palladium on carbon (10% b.w., 225 mg, 0.24 mmol) and the flask evacuated. A hydrogen balloon was added to the top of the flask and the reaction stirred for 90 minutes. The reaction mixture was filtered through a pad of celite and the pad washed with copious amounts of methanol. The solvent was removed *in vacuo* to give the title compound **213** as an opaque oil in a 99 % yield (889 mg, 2.4 mmol). **IR** (ATR): ν_{max} 3301, 2922, 2853, 1727, 1656, 1457, 1367, 1151 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) +2.5 (*c* = 1.0, methanol); **^{1}H NMR** (400 MHz, CD_3OD); δ 7.16 (1H, apparent br t, J = 6.9 Hz, N-H amide), 3.35 (1H, dd, J = 7.3, 5.0 Hz, H-6), 3.22 (2H, apparent dt, J= 6.9, 6.9 Hz, H-8), 2.35 (1H, ddd, 16.0, 7.3, 7.3 Hz, H-4), 2.31 (1H, ddd, 16.0, 7.3, 7.3 Hz, H-4), 2.07 (1H, dddd, J = 14.0, 7.3, 7.3, 5.0 Hz, H-5), 1.79 (1H, dddd J = 14.0, 7.3, 7.3, 7.3 Hz, H-5), 1.50-1.40 (11H, m, H-1, N-H amine), 1.27-1.24 (18 H, m, H-10 - H-18), 0.86 (3H, t, J = 6.9 Hz, H-19); **^{13}C NMR** (400 MHz, $CDCl_3$); δ 173.3 (C-7), 171.3 (C-3), 81.0 (C-2), 52.0 (C-6), 40.8 (CH_2), 39.3 (CH_2), 32.0 (CH_2), 29.5 (CH_2), 28.0 (C-1), 26.8 (CH_2), 22.8 (CH_2), 14.0 (C-19); **HRMS** (ESI) $[M+H]^+$ HRMS found, 371.3262 $C_{21}H_{43}N_2O_3$ required 371.3268.

L-Phenylalanine-(*O-tert*butyl)-L-glutamic acid dodecylamide (214)



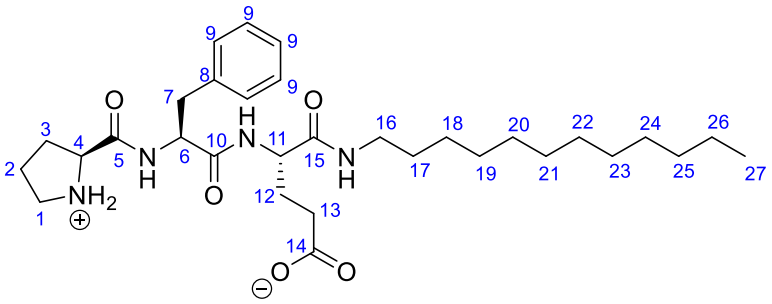
Methanol (20 mL) was added to a mixture of *Z*-L-phenylalanine-(*O-tert*butyl)-L-glutamic acid dodecylamide **224** (1.10 g, 1.69 mmol) and palladium on carbon (10 % b.w., 149 mg) and the mixture evacuated. A hydrogen balloon was added to the top of the flask and the reaction stirred for 3.5 hours. The reaction mixture was filtered through a pad of celite and the pad washed with copious amounts of methanol. The filtrate was concentrated *in vacuo* to give the pure title compound **214** as a colourless oil in a 97 % yield (0.85 g, 1.64 mmol). **IR** (ATR): ν_{max} 3313, 2924, 2853, 1728, 1650, 1513, 1455, 1367, 1153 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -8.5 (c= 1.0, methanol); **^{1}H NMR** (400 MHz, $CDCl_3$); δ 7.86 (1H, d, J= 8.2 Hz, N-H amide), 7.32-7.18 (5H, m, H-1), 6.47 (1H, br t, J= 5.5 Hz, N-H amide), 4.36 (1H, ddd, J= 8.2, 7.0, 6.4 Hz, H-6), 3.60 (1H, dd, J= 9.2, 4.1 Hz, H-4), 3.21-3.16 (2H, m, H-13), 3.18 (1H, dd, J= 13.7, 4.1 Hz, H-3), 2.74 (1H, dd, J = 13.7, 9.2, H-3), 2.32 (1H, ddd, J = 16.5, 7.0, 6.4 Hz, H-8), 2.19 (1H, ddd, J = 16.5, 7.0, 6.4 Hz, H-8), 2.05 (1H, dddd, J = 13.5, 6.4, 6.4, 6.4 Hz, H-7), 1.85 (1H, dddd, J = 13.5, 7.0, 7.0, 7.0 Hz, H-7), 1.52-1.43 (13H, m, H-11, H-14, NH_2), 1.25-1.23 (18H, m, H-15 - H-23), 0.86 (3H, t, 6.9 Hz, H-24); **HRMS** (ESI): $[M+H]^+$ HRMS found 518.3943, $C_{30}H_{52}N_3O_4$ requires 518.3952. $[M+Na]^+$ HRMS found 540.3740, $C_{30}H_{51}N_3NaO_4$ required 517.3880.

***N*-Boc-L-Pro-L-Phe-L(*O*-*t*ertbutyl)-Glutamic acid dodecylamide (215)**



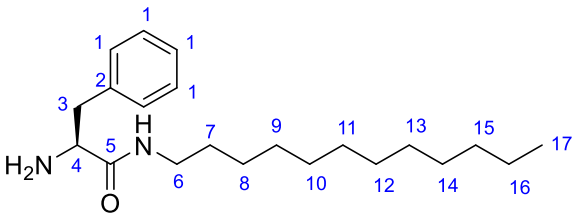
Triethylamine (0.22 mL, 1.55 mmol) was added to a stirred solution of *N*-Boc-L-proline **L-159** (334 mg, 1.55 mmol) in DCM (12 mL) at 0 ºC. After 30 minutes ethyl chloroformate (0.15 mL, 1.55 mmol) was added dropwise (0.1 mL / min). After a further 1.5 hours a solution of L-phenylalanine-(*O*-*tert*butyl)-L-glutamic acid dodecylamide **214** (800 mg, 1.55 mmol) in DCM (4 mL) was added. After 19 hours the crude mixture was washed with 2M HCl (20 mL) and the aqueous layer extracted with DCM (3 x 20 mL). The organic layers were combined, dried over magnesium sulfate, filtered and concentrated *in vacuo* to give the title compound **180** as a white solid in a 75 % yield (834 mg, 1.17 mmol). **IR** (ATR): ν_{max} 3269, 3088, 2926, 2855, 1735, 1705, 1633, 1546, 1390, 1365, 1254, 1152 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -35.2 (c= 1.0, methanol); **^{1}H NMR** (400 MHz, $CDCl_3$); 7.36-7.16 (6H, m, H-9, N-H amide), 6.61-6.57 (2H, m, N-H amide), 4.59 (1H, dd, J = 6.0, 6.0 Hz, H-6), 4.47 (1H, dd, J = 7.8, 7.8 Hz, H-4), 4.2 (1H, dd, J = 8.5, 5.3 Hz, H-11), 3.39-3.22 (3H, m, H-1, H-7), 3.18 (2H, m, H-19), 3.05 (1H, dd, J = 14.0, 6.0 Hz, H-7), 2.77-2.14 (4H, m, H-3, H-13), 1.89-1.86 (1H, m, H-12), 1.81-1.74 (3H, m, H-2, H-12), 1.50-1.20 (38H, m, H-15, H-18, H-20 – H-29) 0.86 (3H, t, J = 6.9 Hz, H-30); **^{13}C NMR** (400 MHz, $CDCl_3$); 173.1 (CO), 172.2 (CO), 170.7 (CO), 170.6 (CO), 155.7 (C-15), 135.9 (C-8), 129.1 (C-9), 127.6 (C-9), 81.2 (C-16), 80.4 (C-16), 61.2 (C-4) 54.3 (C-6), 52.9 (C-11), 47.3 (C-1), 39.7 (C-19), 36.5 (C-7), 31.9 (CH_2), 29.6 (C-17), 28.3 (C-17), 26.9 (CH_2), 24.6(CH_2), 22.7(CH_2), 14.1 (C-30); **HRMS** (ESI): $[M+H]^+$ HRMS found 715.4980, $C_{40}H_{67}N_4O_7$ required 715.5004. $[M+Na]^+$ HRMS found 737.4805, $C_{40}H_{66}N_4NaO_7$ required 737.4824.

L-Proline-L-phenylalanine-L-glutamic acid dodecylamide trifluoroacetic acid salt (216)



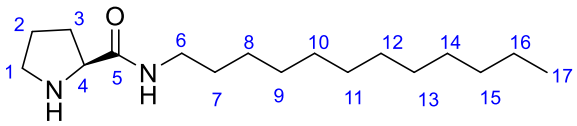
N-Boc-L-proline-L-phenylalanine-L(*O*-*t*ertbutyl)-glutamic acid dodecylamide **215** (350 mg, 0.490 mmol) was dissolved in DCM (4 mL). Trifluoroacetic acid (1.2 mL, 15.68 mmol) was added and stirred for 2 hours at room temperature before the solvent was removed *in vacuo*. Co-distillation with diethyl ether (5 x 5 mL) gave the title compound **216** as a white solid as the TFA salt in an 85 % yield (280 mg, 0.416 mmol). **Mp** 154-158 °C; **IR** (ATR): ν_{max} 3296, 2919, 2851, 1728, 1634, 1552, 1394, 1178, 1135 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -27.3 (c= 1.0, methanoo); **^{1}H NMR** (400 MHz, $CDCl_3$); 9.14 (1H, br s, N-H), 8.69 (1H, d, J= 8.2 Hz, NH amide), 8.41 (br s, N-H), 8.22 (1H, d, J= 8.5 Hz, NH-amide), 7.78 (1H, br t, J= 5.5, N-H amide), 7.26-7.13 (5H, m, H-9), 4.58 (1H, ddd, J= 10.1, 8.2, 4.2 Hz, H-6), 4.18 (1H, ddd, J= 8.5, 8.0, 5.5 Hz, H-11), 4.06-4.03 (1H, m, H-4), 3.14-2.95 (4H, m, H-1, H-16), 3.00 (1H, dd, J= 13.7, 4.2 Hz, H-7), 2.73 (1H, dd, J= 13.7, 10.1 Hz, H-7), 2.26-2.13 (3H, m, H-3, H-13),1.86-1.66 (5H, m, H-2, H-3, H-12), 1.34-1.31 (2H, m, H-17), 1.20-1.18 (18H, m, H-18 – H-26), 0.81 (3H, t, J = 6.9 Hz, H-27); **^{13}C NMR** (400 MHz, $CDCl_3$); 173.9 (C-14), 171.0 (CO amide), 171.0 (CO amide), 168.1 (CO amide), 137.5 (C-8), 129.2 (C-9), 128.1 (C-9) 126.5 (C-9), 58.7(C-4), 54.4 (C-6), 52.0 (C-11), 45.7 (C-1), 38.9 (C-7), 37.7 (C-16), 30.0 (CH_2), 29.5 (CH_2), 29.1 (CH_2), 28.7 (CH_2), 27.7 (CH_2), 26.3 (CH_2), 23.4 (CH_2), 22.1 (CH_2), 14.0 (C-27); **HRMS** (ESI) $[M+H]^+$; HRMS found 559.3846, $C_{31}H_{51}N_4O_5$ required 559.3854. Spectroscopic data is in agreement with the literature.[152]

L-Phenylalanine dodecylamide (217)



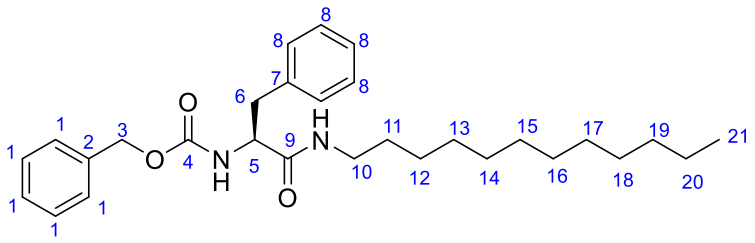
Methanol (40 mL) was added to a mixture of *Z*-L-phenylalanine dodecylamide **220** (1.22 g, 2.62 mmol) and palladium on carbon (10 % b.w., 0.28 g, 0.26 mmol,) and the flask evacuated. A hydrogen balloon was added to the top of the flask and stirred for 16 hours. The mixture was filtered through a pad of celite and the pad washed with copious amounts of methanol. The filtrate was concentrated *in vacuo* to give the title compound **220** as a white powder in an 83 % yield (0.72 g, 2.17 mmol). **Mp** 61-63 °C; **IR** (ATR) 3287, 2916, 2850, 1633, 1549, 1523, 1470 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -46.1 (*c* = 1.0, chloroform); **^{1}H NMR** (400 MHz, CD_3OD) δ 7.34-7.19 (5H, m, H-1), 3.51 (1H, dd, J= 7.0, 7.0 Hz, H-4), 3.18-3.11 (1H, m, H-6), 3.07-3.00 (1H, m, H-6), 2.95 (1H, dd, J= 13.3, 7.0 Hz, H-3), 2.82 (1H, dd, J= 13.3, 7.0 Hz, H-3), 1.38-1.18 (20H, m, H-7-H16), 0.89 (3H, t, J= 6.9 Hz, H-17); **^{13}C NMR** (400 MHz, CD_3OD) δ 176.0 (C-5), 138.7 (Ar), 130.4 (Ar), 129.5 (Ar), 127.7 (Ar), 57.8 (C-4), 42.6 (C-3), 40.3 (C-6), 33.1 (CH_2), 30.8 (CH_2), 30.8 (CH_2), 30.7 (CH_2), 30.6 (CH_2), 30.5 (CH_2), 30.4 (CH_2), 30.3 (CH_2), 27.9 (CH_2), 23.7 (CH_2), 14.5 (C-17); **HRMS** (ESI): $[M+H]^+$ found 333.2901, $C_{21}H_{37}N_2O$ required 333.2900. $[M+Na]^+$ found 355.2728, $C_{21}H_{36}N_2NaN_2O$ requires 355.21720.

L-Proline dodecylamide (219)



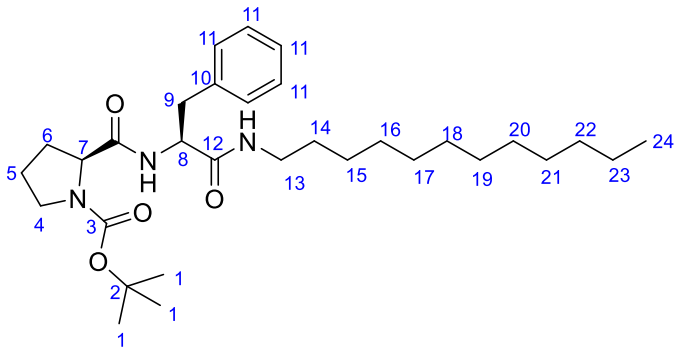
TFA (4.8 mL, 62.8 mmol) was added to a stirred solution of *N*-Boc-L-proline dodecylamine (1.2 g, 3.14 mmol) in DCM (20 mL). After 2 hours the reaction was neutralised with saturated sodium bicarbonate solution (10 mL) and extracted with DCM (3 x 20 mL). The organic extracts were combined and washed with brine (50 mL), dried over magnesium sulfate, flittered and concentrated *in vacuo* to give the title compound **219** as a white solid in a 96 % yield (803 mg, 3.00 mmol). **Mp** 47–50 °C; **IR** (ATR): ν_{max} 3301, 2918, 2850, 1666, 1632, 1553, 1523, 1471 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -39.3 (*c* = 1.0, chloroform); **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.77 (1H, br t, J= 6.9 Hz, N-H amide), 3.89 (1H, dd, J= 8.7, 5.5 Hz, H-4), 3.34 (1H br s, N-H), 3.20 (2H, dt, J = 6.9, 6.9 Hz, H-6), 3.12-3.06 (1H, dt, J = 10.5, 6.9 Hz, H-1), 3.00-2.94 (1H, dt, J = 10.5, 6.9 Hz, H-1), 2.25-2.15 (1H, ddt, J = 13.0, 8.7, 6.9 Hz, H-3), 1.96-1.87 (1H, dtd, J= 13.0, 6.9, 5.5 Hz, H-3), 1.75 (2H, apparent quintet, J = 6.9 Hz, H-2), 1.50-1.45 (2H, m, H-7), 1.30-1.24 (18H, m, H-8-H16), 0.87 (3H, t, J = 6.6 Hz, H-17); **^{13}C NMR** (400 MHz, $CDCl_3$); δ 173.8 (C-5), 60.5 (C-4), 47.3 (C-1), 39.3 (C-6), 32.0 (CH_2), 29.8 (CH_2), 29.7 (CH_2), 29.7 (CH_2), 29.5 (CH_2), 29.4 (CH_2), 27.1 (CH_2), 26.1 (CH_2), 22.8 (CH_2), 14.3 (C-17); **HRMS** (ESI) $[M+H]^+$; HRMS found 283.2748 $C_{17}H_{35}N_2O$ required 283.2744; $[M+Na]^+$ HRMS found 305.2567 $C_{17}H_{34}N_2NaN_2O$ required 305.2563.

***Z*-L-Pheneylalanine dodecylamide (220)**



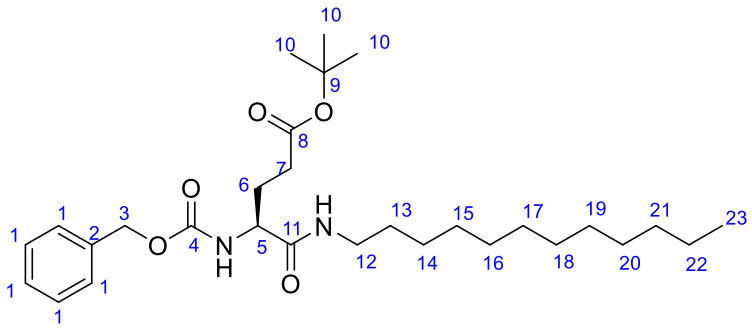
Triethylamine (0.46 mL, 3.34 mmol) was added to a stirred solution of *Z*-L-phenylalanine **221** (1.00 g, 3.34 mmol) in THF (20 mL) at 0 °C. After 15 minutes ethyl chloroformate was added (0.32 mL, 3.34 mmol) followed after 45 minutes by dodecylamine (619 mg, 3.34 mmol). The mixture was stirred for 1 hour at 0 °C before being warmed to room temperature. After 16 hours the resulting white solid was filtered, washed with ice cold 1M NaOH (15 mL), then ice cold deionised water (15 mL). The product **220** was collected a white powder in a 96 % yield (1.50 g, 3.21 mmol). **Mp** 112-114 °C; **IR** (ATR): ν_{max} 3301, 2919, 2851, 1686, 1525, 1285, 1235, 1038 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -1.5 (*c* = 1.0, dimethylsulfoxide); **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.36-7.14 (10 H, m, H-1 & H-8), 5.64 (1H, t, J = 4.8 Hz, N-H), 5.48 (1H, br d, J = 7.5 Hz, N-H), 5.06 (2H, s, H-3), 4.32 (1H, ddd, J= 7.5, 7.5, 7.5 Hz, H-5), 3.16-3.07 (2H, m, H-10), 3.12 (1H, dd, J= 13.7, 7.5 Hz, H-6), 2.98 (1H, dd, J = 13.7, 7.5 Hz, H-6), 1.32-1.12 (20H, m, H-11 – H-20), 0.87 (3H, t, 6.9, H-21); **^{13}C NMR** δ 170.6 (C-7), 136.2 (C-2), 129.4 (Ar), 128.8 (Ar), 128.6 (Ar), 128.3 (Ar), 128.1 (Ar), 127.1 (Ar), 67.1 (C-3), 56.6 (C-5), 39.6 (C-10), 39.0 (C-6), 32.0 (CH_2), 29.7 (CH_2), 29.7 (CH_2), 29.6 (CH_2), 29.6 (CH_2), 29.4 (CH_2), 29.3 (CH_2), 29.3 (CH_2), 26.8 (CH_2), 22.8 (CH_2), 14.2 (C-10); **HRMS** (ESI); $[M+H]^+$ HRMS found 467.3280, $C_{29}H_{43}N_2O_3$ requires 467.3268. $[M+Na]^+$ HRMS found 489.3109, $C_{29}H_{42}N_2NaN_2O_3$ requires 489.3088.

***N*-Boc-L-Proline-L-phenylalanine dodecylamide (222)**



To a stirred solution of L-phenylalanine dodecylamide **217** (1.2 g, 3.4 mmol) in DME (40 mL) was added *N*-Boc-L-proline-*O*-succinimide **206** (1.06 g, 3.4 mmol). After 17 hours at room temperature the solution was heated to 50 °C for a further 2 hours. Upon cooling the solvent was removed *in vacuo* and redissolved in DCM (30 mL) and washed with saturated sodium bicarbonate solution (30 mL), then brine (30 mL). The organic layer was dried over magnesium sulfate, filtered and concentrated *in vacuo* to give the title compound **222** as a colourless oil in a 94 % yield (1.70 g, 3.21 mmol). **IR** (ATR): ν_{max} 3301, 2924, 2854, 1645, 1548, 1391, 1241, 1162 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -56.0 (*c* = 1.0, chloroform); **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.29-7.13 (5H, m, H-11), 6.63-6.53 (1H, m, N-H), 6.34 (1H, br d, J = 7.3 Hz, N-H), 4.73-4.62 (1H, m, H-8), 4.12-4.16 (1H, m, H-7), 3.39-3.23 (3H, m, H-4, H-9), 3.20-3.10 (2H, m, H-13), 3.05-2.97 (1H, m, H-9), 2.10-1.95 (2H, m, H-6), 1.88-1.62 (2H, m, H-5), 1.45-1.10 (29H, m, H-1, H-14 - H-23), 0.86 (3H, t, J = 6.6 Hz, H-24); **HRMS** (ESI) $[M+H]^+$ HRMS found 530.393738, $C_{31}H_{52}N_3O_4$ required 530.395234. $[M+Na]^+$ HRMS found 552.375035, $C_{31}H_{51}N_2NaN_3O_4$ required 552.377178.

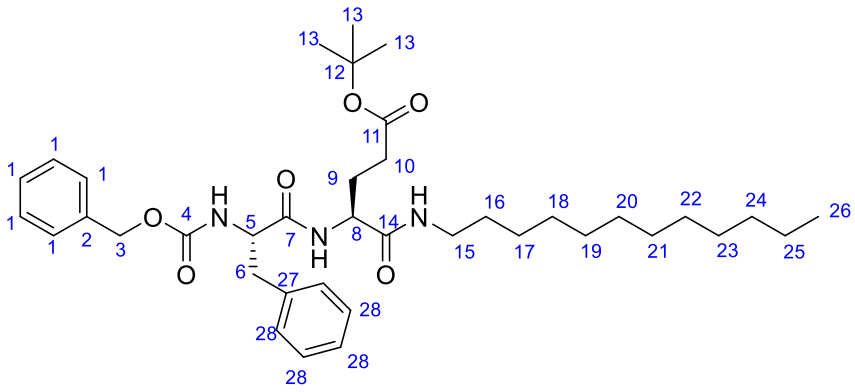
***Z*-(*O*-*tert*butyl)-Glutamic acid dodecylamide (223)**



Triethylamine (0.41 mL, 2.96 mmol) was added to a stirred solution of *Z*-(*O*-*tert*butyl)-glutamic acid **212** (1.0 g, 2.96 mmol) in dry THF (40 mL) at 0 °C. After 15 minutes ethyl chloroformate (0.28 mL, 2.96 mmol) was added dropwise (0.1 mL/min) followed by dodecylamine (548 mg, 2.96 mmol) after a further 30 minutes of stirring. After 3 hours the solvent was removed *in vacuo* and the crude material redissolved in DCM (40 mL). The organic layer was washed with a saturated solution of sodium bicarbonate (40 mL) and the aqueous layer extracted with DCM (2 x 40 mL). The combined organic extracts were washed with brine (40 mL), dried over magnesium sulfate, filtered and concentrated *in vacuo* to give the pure compound **223** as a white solid in a 99 % yield (1.48 g, 2.93 mmol). **Mp** 58-60 °C; **IR** (ATR): ν_{max} 3293, 2920, 2851, 1720, 1691, 1645, 1534, 1250, 1153, 1049 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -8.4 (*c* = 1.0, methanol); **^{1}H NMR** (400 MHz, $CDCl_3$); δ 7.37-7.29 (5H, m, H-1), 6.32 (1H, br t, J= 6.5 Hz, N-H amide), 5.72 (1H, d, J= 7.8 Hz, N-H carbamate), 5.01 (1H, s, H-3), 4.19-4.14 (1H, ddd, J= 7.8, 6.4, 6.4 Hz, H-5), 3.22 (2H, dt, J= 6.5, 6.5 Hz H-12), 2.42 (1H, ddd, J= 16.5, 6.4, 6.4 Hz, H-7), 2.28 (1H, ddd, J= 16.5, 6.4, 6.4 Hz, H-7), 2.1-2.0 (1H, dddd, J = 13.0, 6.4, 6.4, 6.4 Hz, H-6), 1.97-1.87 (1H, dddd, J= 13.0, 6.4, 6.4, 6.4 Hz, H-6), 1.46-1.38 (11H, m, H-10, H-13), 1.28-1.25 (18H, m, H-14 – H-22), 0.87 (3H, t, J= 6.8 Hz, H-23); **^{13}C NMR** (400 MHz, $CDCl_3$); δ 173.1 (CO), 171.2 (CO), 156.4 (C-4), 136.3 (C-2), 128.7 (C-1), 128.3 (C-1), 128.1 (C-1), 81.2 (C-9), 67.1 (C-3), 54.5 (C-5), 39.7 (C-12), 32.0 (CH_2), 29.8 (CH_2), 29.7 (CH_2), 29.7 (CH_2), 29.5 (CH_2), 29.4 (CH_2), 29.2 (CH_2), 28.2 (CH_2), 27.0 (CH_2), 22.8 (CH_2), 14.2 (C-23); **HRMS** (ESI) $[M+H]^+$ HRMS

found 505.3621, $C_{29}H_{49}N_2O_5$ required 505.3636. $[M+Na]^+$ HRMS found 527.3437, $C_{29}H_{48}N_2NaO_5$ required 527.3455.

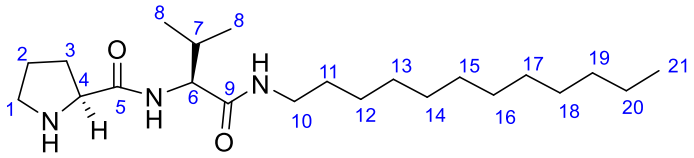
<u>*Z*-L-Phenylalanine-(*O-tert*butyl)-L-glutamic acid dodecylamide (224)</u>



Triethylamine (0.30 mL, 2.16 mmol) was added to a stirred solution of *Z*-L-phenylalanine **221** (646 mg, 2.16 mmol) in 30 mL THF (30 mL) at 0 °C. After 15 minutes ethyl chloroformate (0.20 mL, 2.16 mmol) was added dropwise (0.1 mL / min). After a further 50 minutes a solution of L-glutamic acid (*O-tert*butyl) dodecylamide **213** (800 mg, 2.16 mmol) in THF (5 mL) was added to the reaction and stirred for 19 hours before the solvent was removed *in vacuo*. The crude white solid was redissolved in DCM (30 mL), washed with a saturated sodium bicarbonate solution (30 mL), 0.1 M HCl (30 mL) then brine (30 mL). The organic layer was dried over magnesium sulfate, filtered and concentrated *in vacuo* to give the crude product as a white solid. Purification *via* flash column chromatography (98:2 DCM:methanol to 95:5 DCM:methanol) gave the title compound **224** in a 77% yield as a white powder (1.09 g, 1.67 mmol). **Mp** 100-103 °C; **IR** (ATR): ν_{max} 3286, 2921, 2851, 1726, 1690, 1633, 1535, 1260, 1151 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -12.3 (*c* = 1.0, methanol); **^{1}H NMR** (400 MHz, $CDCl_3$); δ 7.35-7.08 (11H, m, H-1, H-28, NH amide), 6.39 (1H, br t, J= 6.0 Hz, N-H amide), 5.24 (1H, br d, J= 6.4 Hz, N-H amide), 5.08 (1H, d, J= 12.4 Hz, H-3), 5.04 (1H, d, J= 12.4 Hz, H-3), 4.41 (1H, dt, J= 6.4, 6.4 Hz, H-5), 4.33 (1H, ddd, J= 6.9, 7.0, 7.0 Hz, H-8), 3.25-3.14 (2H, m, H-15), 3.07 (2H, d, J= 6.4 Hz, H-6), 2.34 (1H, ddd, J= 16.9, 7.0,

7.0 Hz, H-10), 2.19 (1H, ddd, 16.9, 7.0, 7.0 Hz, H-10), 2.05-1.98 (1H, m, H-9), 1.91-1.83 (1H, m, H-9), 1.45-1.42 (2H, m, H-16), 1.41 (9H, s, H-13), 1.26-1.23 (18H, m, H-17 - H-25), 0.86 (3H, t, J = 6.6 Hz, H-26), **^{13}C NMR** (400 MHz, $CDCl_3$); δ 173.6 (C-11), 171.1 (CO), 170.5 (CO), 156.2 (C-4), 136.1 (C-2), 129.3 (C-1), 128.9 (C-Ar), 128.7 (C-Ar), 128.4 (C-Ar), 128.2 (C-Ar), 127.3 (C-Ar), 81.3 (C-12), 67.3 (C-3), 56.5 (C-5), 53.0 (C-8), 39.8 (C-15), 38.2 (C-9), 32.0 (C-10), 29.8 (CH_2), 29.7 (CH_2), 29.5 (CH_2), 29.5 (CH_2), 28.2 (C-13), 27.0 (CH_2), 22.8 (CH_2), 14.3 (C-26); **HRMS** (ESI) $[M+H]^+$ HRMS found 652.4320, $C_{38}H_{58}N_3O_6$ required 652.4320. $[M+Na]^+$ HRMS found 674.4146, $C_{38}H_{57}N_3NaO_6$ required 674.4140.

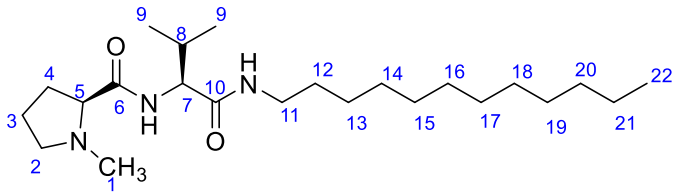
L-Proline-L-valine dodecylamide (Gelator A)



Trifluoroacetic acid (15.6 mL, 0.2 mol) was added to a stirred solution of **207** (4.92 g, 10.2 mmol) in DCM (15 mL). After 2 hours the reaction was deemed complete by TLC and the mixture concentrated *in vacuo*. The resulting yellow oil was redissolved in 20 mL DCM and washed with saturated sodium bicarbonate solution (20 mL). The organic layer was washed with brine (20 mL) and extracted three times with DCM (3 x 20 mL). The organic layers were combined, dried over magnesium sulfate, filtered and concentrated *in vacuo* to give the title compound, **Gelator A**, as a white solid in an 88 % yield (3.43 g, 9.0 mmol). **Mp** 98-102 °C; **IR** (ATR): ν_{max} 3292, 2918, 2850, 1637, 1560, 1539, 1463, 1371, 1224 cm^{-1} ; **$[\alpha]_D^{25}$** -53.14 (*c* = 1.0, chloroform); **^{1}H NMR** (400 MHz, $CDCl_3$) δ 8.17 (1H, br d, J = 9.2 Hz, N-H), 6.28 (1H, br t, J = 7 Hz, N-H), 4.06 (1H, dd, 9.2, 7.3, H-6), 3.74 (1H, dd, 9.2, 5.0, H-4), 3.25 (1H, ddt, J = 12.8, 7.0, 6.0 Hz, H-10), 3.14 (1H, ddt, J = 12.8, 7.0, 6.0 Hz, H-10), 3.02 (1H,

dt, J = 10.1, 7.0 Hz, H-1), 2.92 (1H, dt, J = 10.1, 7.0 Hz, H-1), 2.18-2.13 (1H, m, N-H), 2.16 (1H, ddt, J= 12.1, 9.2, 7 Hz, H-3), 2.14 (1H, dspt, J = 7.3, 6.9 Hz, H-7), 1.89 (1H, dtd, J = 12.1, 7.0, 5.0 Hz, H-3), 1.70 (2H, quintet, J = 7.0 Hz, H-2), 1.50-1.42 (2H, m, H-11), 1.28-1.20 (18H, m, H-12 – H-20), 0.92 (3H, d, J = 6.9, H-8), 0.89 (3H, d, 6.9 Hz, H-8), 0.85 (3H, t, J = 6.9 Hz, H-21); **^{13}C NMR** (400 MHz, DMSO d_6) δ 173.5 (CO), 170.2 (CO), 60.1 (CH), 56.6 (CH), 46.7 (C-1), 38.3 (C-10), 31.5 (CH_2), 31.3 (CH_2), 30.6 (C-7), 26.3 (CH_2), 25.8 (CH_2), 22.1 (CH_2), 19.2 (C-8), 17.8 (C-8), 14.0 (C-21); **HRMS** (ESI): $[M+H]^+$ HRMS found 382.3416, $C_{22}H_{44}N_3O_2$ required 382.3428. $[M+Na]^+$ HRMS found 404.3237, $C_{22}H_{43}N_3NaNO_2$ required 404.3237. Spectroscopic data was in agreement with the literature.[139]

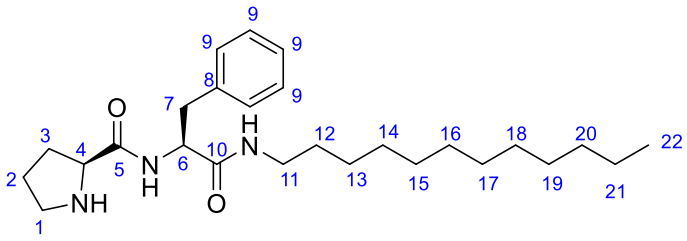
<u>*N*-Methyl-L-proline-L-valine dodecylamide (Gelator B)</u>



Methanol (5 mL) was added to a mixture of L-proline-L-valine dodecylamide (**Gelator A**) (1.0 g, 2.39 mmol,) and palladium on carbon (10 % b.w., 100 mg,). The flask was evacuated and a 40% solution of formaldehyde in water (0.2 mL, 2.63 mmol) was added. A hydrogen balloon was added to the top of the flask and stirred for 1.5 hours. The mixture was then filtered through a pad of celite and the pad washed with copious amounts of methanol. The filtrate was concentrated *in vacuo* to give **Gelator B** as a white powder in a 99 % yield (0.94 g, 2.37 mmol). **Mp** 76-79 °C; **IR** (ATR): ν_{max} 3285, 2918, 2851, 1637, 1558, 1540, 1231 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -67.4 (*c* = 1.0, chloroform); **^{1}H NMR** (400 MHz, $CDCl_3$); δ 7.84 (1H, br d, J= 9.2, N-H), 6.07 (1H, br t, J= 6.5 Hz, N-H), 4.07 (1H, dd, J= 9.2, 7.8 Hz, H-7), 3.29-3.16 (2H, m, H-11), 3.12-3.09 (1H, m, H-5), 2.87 (1H, ddd, J= 10.1, 5.0, 5.0 Hz, H-2),

2.35 (3H, s, H-1), 2.27-2.17 (1H, m, H-2), 2.20-2.10 (1H, dspt, J= 7.8, 6.9 Hz, H-8), 1.86-1.70 (4H, m, H-3, H-4), 1.50-1.42 (2H, m, H-12), 1.27-1.21 (18H, m, H-13 – H-21), 0.93 (3H, d, J= 6.9 Hz, H-9), 0.90 (3H, d, J= 6.9 Hz, H-9), 0.86 (3H, t, J= 6.9 Hz, H-22); **^{13}C NMR** (400 MHz, $CDCl_3$) δ 175.1 (CO), 171.1(CO), 69.0 (C-1), 58.4 (CH), 56.7 (CH), 41.8 (CH_2), 39.6 (CH_2), 32.0 (CH_2), 31.4 (CH_2), 30.7 (C-8), 29.8 (CH_2), 29.8 (CH_2), 29.7 (CH_2), 26.7 (CH_2), 29.6 (CH_2), 29.5 (CH_2), 29.4 (CH_2), 27.0 (CH_2), 24.5 (CH_2), 22.8 (CH_2), 19.7 (CH_3), 18.3 (CH_3), 14.2 (CH_3); **HRMS** (ESI): $[M+H]^+$ HRMS found 396.3586, $C_{23}H_{46}N_3O_2$ required 396.3585. $[M+Na]^+$ HRMS found 418.3417, $C_{23}H_{45}N_3NaNO_2$ required 418.3404.

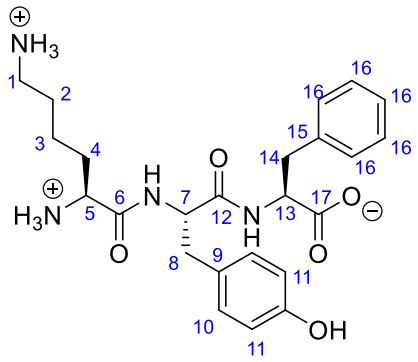
L-Proline-L-phenylalanine dodecylamide (Gelator C)



TFA (1.73 mL, 22.7 mmol) was added to a stirred solution of *N*-Boc-L-proline-L-phenylalanine-dodecylamide **222** (600 mg, 1.13 mmol) in DCM (10 mL). After 17 hours the solution was neutralised with a solution of saturated sodium bicarbonate (10 mL) and extracted three times with DCM (3 x 10 mL). The combined organic extracts were dried over magnesium sulfate, filtered and concentrated *in vacuo* to give the title compound, **Gelator C**, as a white powder in a 92 % yield (445 mg, 1.04 mmol). **Mp** 155-157 °C; **IR** (ATR): ν_{max} 3287, 2918, 2850, 1638, 1558, 1537, 1226 cm^{-1}; **$[\alpha]_D^{25}$** (deg cm^3 g^{-1} dm^{-1}) -46.2 (c= 1.0, chloroform); **^{1}H NMR** (400 MHz, $CDCl_3$); δ 8.11 (1H, d, J = 8.2 Hz, N-H amide), 7.27-7.16 (5H, m, H-9), 6.25 (1H br t, J= 6.5 Hz, H-N-H amide) 4.53 (1H, ddd, J= 8.2, 8.2, 8.2 Hz, H-6), 3.69 (1H, dd, J= 9.2, 5.0 Hz, H-4), 3.20 (2H, td, J= 7.0, 6.5 Hz, H-11), 3.14

(1H, dd, J= 13.7, 8.2 Hz, H-7), 2.99 (1H, dd, J= 13.7, 8.2 Hz, H-7), 2.91 (1H, dt, J= 10.2, 6.9 Hz, H-1), 2.70 (1H, dt, J= 10.2, 6.1 Hz, H-1), 2.35 (1H, br s, N-H), 2.00 (1H, dddd, J= 12.7, 9.2, 6.9, 6.1 Hz, H-3), 1.65 (1H, dddd, J= 12.7, 6.9, 6.1, 5.0 Hz, H-3), 1.57 (1H, dtt, J= 12.7, 6.9 Hz, 6.1 Hz, H-2), 1.41 (1H, dtt, J= 12.7, 6.9 Hz, 6.1 Hz, H-2), 1.39-1.34 (2H, m, H-12),1.28-1.15 (18H, m, H-13 - H21), 0.86 (3H, t, J= 6.9 Hz, H-22); **^{13}C NMR** (400 MHz, CD_3OD) δ 175.5 (CO amide), 170.9 (CO amide), 137.2 (C-8), 129.3 (C-9), 128.6 (C-9), 126.9 (C-9), 60.3 (C-4), 54.2 (C-6), 47.2 (C-1), 39.6 (C-11), 37.9 (C-3), 39.6 (CH_2), 37.9 (CH_2), 32.0 (CH_2), 30.7 (CH_2), 29.7 (CH_2), 29.4 (CH_2), 29.3 (CH_2), 26.9 (CH_2), 25.9 (CH_2), 22.8 (CH_2), 14.2 (C-22); **HRMS** (ESI): $[M+H]^+$ HRMS found 30.3407, $C_{26}H_{44}N_3O_2$ required 430.3428.

L-Lysine-L-tyrosine-L-phenylalanine (K-Y-F)

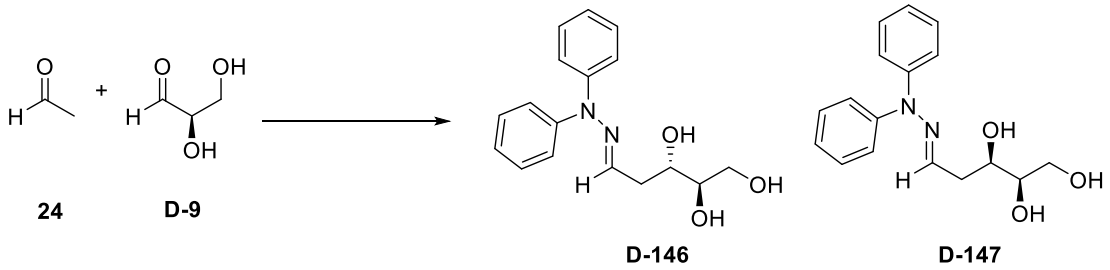


L-Lysine-L-tyrosine-L-phenylalanine **(K-Y-F)** was prepared using the general procedure for one pot peptide synthesis using cleave method A (**Section 10.3.3**) to remove the peptide from resin. The tripeptide was acquired as a white powder (as the TFA salt). **Mp** 216-219 °C; **IR** (ATR): ν_{max} 3064.9, 1647, 1515, 1444, 1177, 1137 cm^{-1}; **^{1}H NMR** (400 MHz, DMSO d^6) δ 9.28 (1H, br s, O-H), 8.53 (1H, d, J = 9.0 Hz, NH amide), 8.52 (1H, d, J = 8.5 Hz, N-H amide), 8.02 (3H, br s, NH_3), 7.74 (3H, br s, NH_3), 7.25-7.15 (5H, m, H-16), 7.02 (2H, d, J = 8.2 Hz, H-10), 6.60 (2H, d, J = 8.2 Hz, H-11), 4.47 (1H, ddd, J = 9.0, 9.0, 3.2 Hz, H-7), 4.41

(1H, ddd, J = 8.5, 8.5, 4.6 Hz, H-13), 3.66-3.62 (1H, m, H-5), 3.03 (1H, dd, J = 14.0, 4.6 Hz, H-14), 2.87 (1H, dd, J= 14.0, 3.2 Hz, H-8), 2.86 (1H, dd, J= 14.0, 8.5 Hz, H-14), 2.68-2.62 (2H, m, H-1), 2.60 (1H, dd, J = 14.0, 9.0 Hz, H-8), 1.64-1.58 (2H, m, H-4), 1.48-1.41 (2H, m, H-2), 1.26-1.17 (2H, m, H-3); **^{13}C NMR** (400 MHz, DMSO d^6) δ 172.8 (C-17), 171.1 (CO amide), 168.4 (CO amide), 158.0, 156.0, 137.5, 130.2, 129.2, 128.2, 127.5, 126.5, 115.0, 54.4 (C-7), 53.5 (C-13), 51.8 (C-5), 38.0 (C-1), 36.7 (C-14), 30.9 (C-4), 26.5 (C-2), 20.9 (C-3); **HRMS** (ESI) $[M+H]^+$ HRMS found 457.2479, $C_{24}H_{33}N_4O_5$ required 457.2445. $[M+Na]^+$ HRMS found 479.2283, $C_{24}H_{32}N_4NaNO_5$ required 479.2265.

6.3. General Procedures

6.3.1. General procedure for 2- deoxy-D-ribose forming reaction



Acetaldehyde **24** (44 mg, 1 mmol) and D-glyceraldehyde **D-9** (90 mg, 1 mmol) were added to a flask containing catalyst (20 mol %) and aqueous medium (3 mL) and stirred for 24 hours at room temperature. The solvent was removed *in vacuo* and the residue was redissolved in methanol (5 mL). *N*-*N*-Diphenyl hydrazine **144** (550 mg, 3 mmol) and acetic acid (2 drops) were added and the reaction stirred for 1 hour before being the concentrated *in vacuo* to give a red/brown oil. The products were isolated by flash column chromatography (methanol:DCM 3:97 to 10:90). Further purification *via* preparative thin layer chromatography (90:10 ethyl acetate:hexane) afforded the product as a mixture of diastereomers **D-146** and **D-147**.

The diastereomeric ratio was determined *via* ^{1}H NMR spectroscopy using the azomethine peaks as a reference. There are two examples shown below. **Figure 6.1.** is a ^{1}H NMR spectrum of the two sugar standards mixed together and the azomethine peak used for diastereomeric ratio determination. **Figure 6.2.** is a ^{1}H NMR spectrum of the assay after purification.

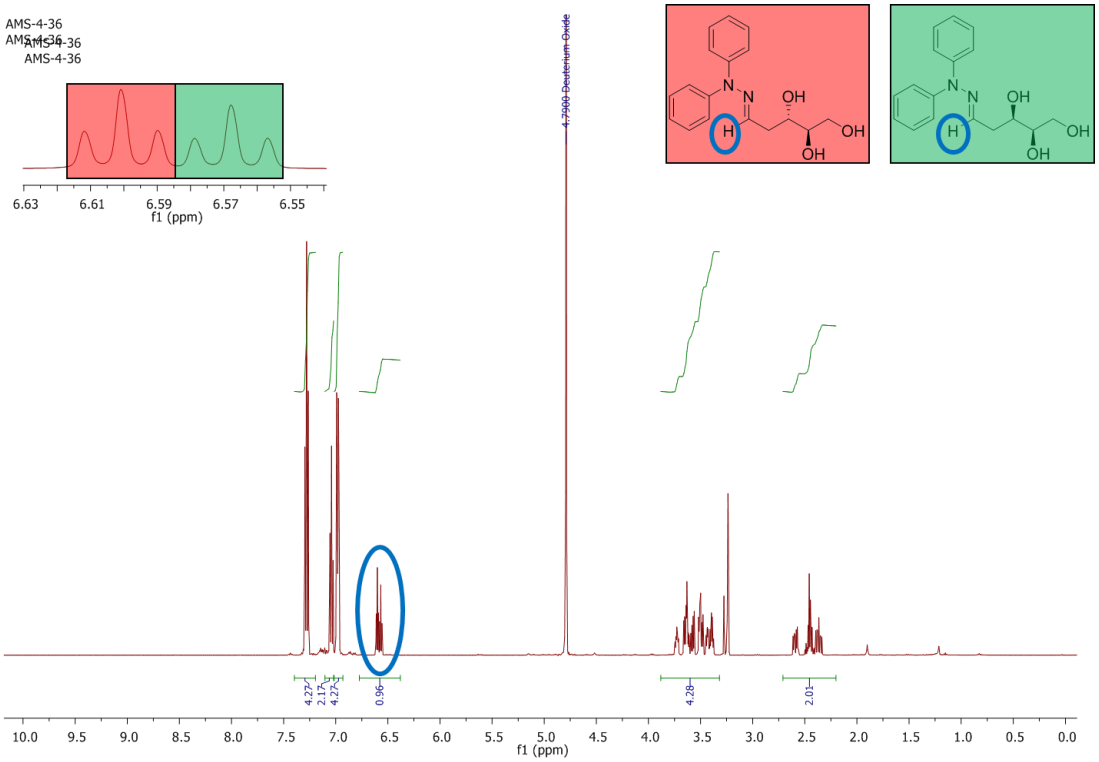


Figure 6.1. ^{1}H NMR spectrum of the two sugar standards.

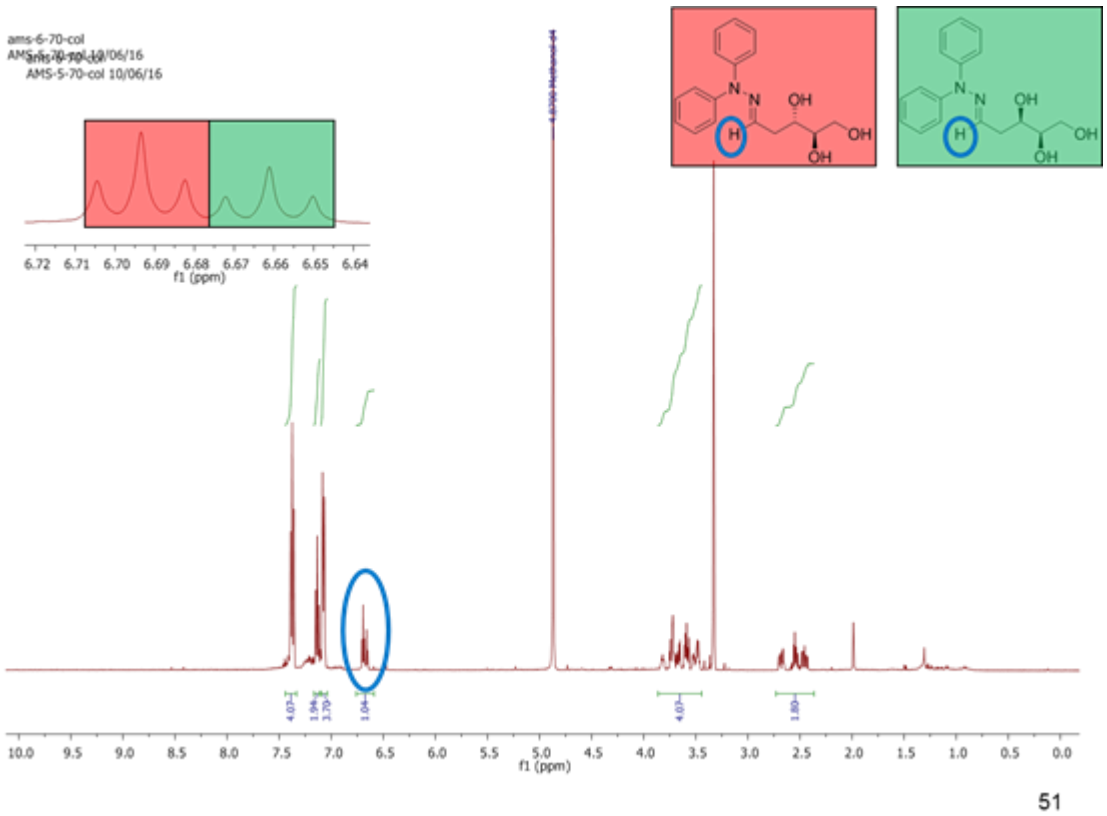
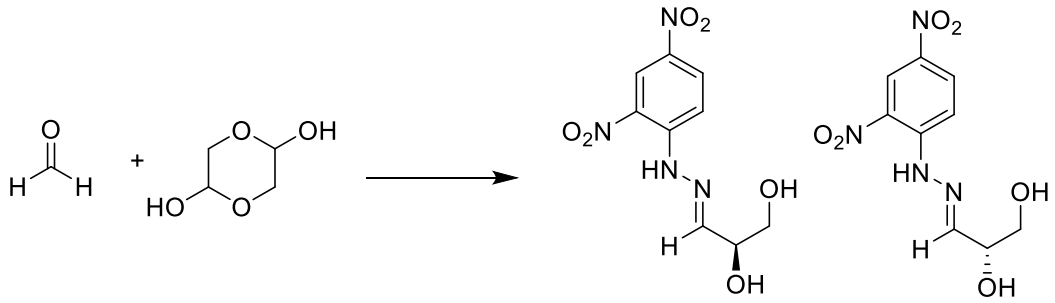


Figure 6.2. ^{1}H NMR spectrum of the deoxyribose forming assay after purification.

6.3.2. General procedure for glyceraldehyde forming reaction



L-Valine nitrile **L-157** (20 mg, 0.20 mmol) and glycolaldehyde dimer (60 mg, 0.5 mmol) were dissolved in pH 7 phosphate buffer (1.7 mL). Paraformaldehyde (750 mg) was dissolved in water (25 mL) and heated to 60 °C for 2 hours. Upon cooling, an aliquot of the solution (1.3 mL) was added to the reaction mixture and stirred at room temperature for 24 hours. Dinitrophenyl hydrazine (510 mg, 2.6 mmol) was added to the reaction mixture and stirred for a further 24 hours. After this time the reaction was concentrated *in vacuo* to afford the crude mixture of hydrazones as an orange solid. The crude mixture was purified by flash column chromatography (5:95 methanol:DCM) followed by preparative thin layer chromatography (2:98 methanol:DCM) to yield the glyceraldehyde-trapped hydrazone product **115** as a yellow solid (3 mg, 0.01 mmol). The enantiomeric excess of hydrazone product was analysed *via* HPLC using a chiralpak AD column (15:85 isopropanol:hexane) at a flow rate of 1.0 mL/min.

Figure 6.3 and **Figure 6.4** below show the chromatograms of (A) the reaction mixture and (B) the reaction mixture doped with authentic D-glyceraldehyde hydrazone standard.

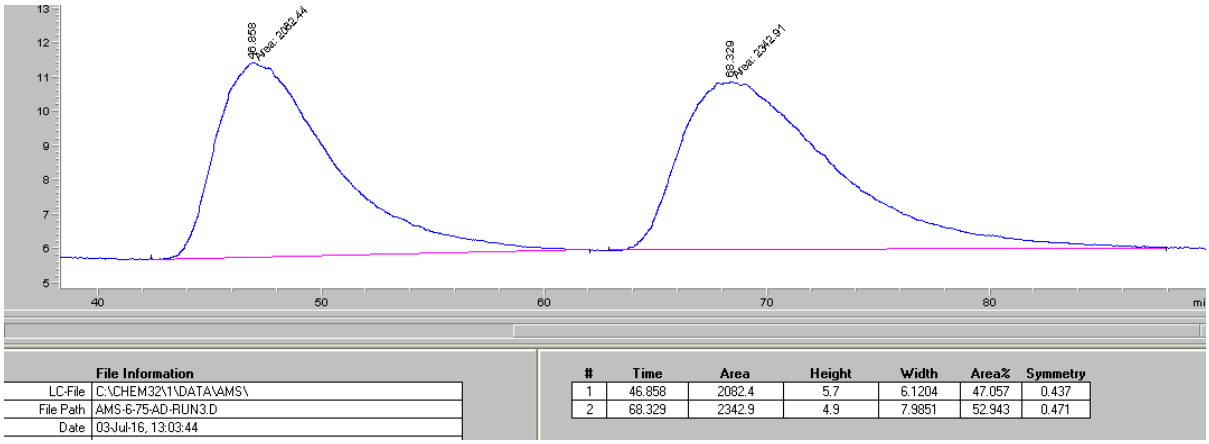


Figure 6.3. There is a 6% ee in favour of D-glyceraldehyde hydrazone **D-115**. L-glyceraldehyde **L-115** elutes at ~46 minutes and D-glyceraldehyde**D-115** at ~68 minutes.

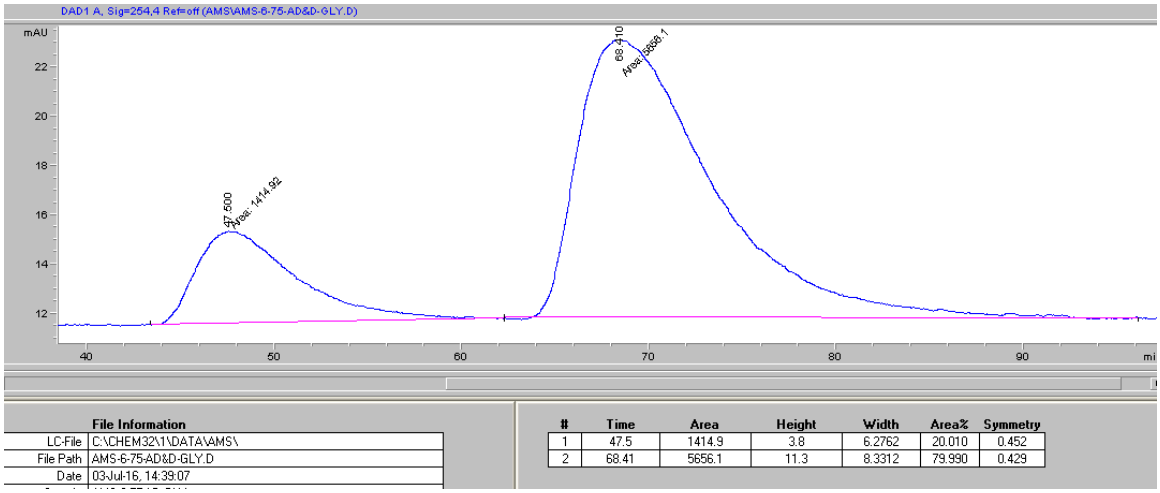


Figure 6.4. HPLC run of glyceraldehyde hydrazone products from reaction doped with authentic D-glyceraldehyde hydrazone **D-115**.

6.4. General procedure for peptide synthesis on 2-chlorotrityl resin

Resin loading: Chlorotrityl resin preloaded with the first amino acid in the sequence was swollen in DMF (3 mL) and shook for 30 min. The DMF was then filtered off

Coupling reaction: A solution of Fmoc-protected amino acid (5 equiv), HCTU (5 equiv) and DIPEA (5 equiv) in DMF were added to the resin. The reaction was shaken for 1 h. The solvent was filtered off and shaken with DMF (3mL) for 2 minutes and the DMF then filtered off. This was repeated twice more with DMF.

Fmoc deprotection: A solution of 20% piperidine in DMF was added to the resin and the reaction shaken for 2 minutes and the solvent filtered off. This process was repeated 5 times. The resin was then washed with DMF and shaken for 2 minutes before filtering. This was repeated 3 times.

Final amino acid coupling: The process was repeated the same as above but the amino acid used was protected with a Boc group.

Cleavage and Isolation: The resin was washed with DCM (3 × 2 min with rotation) followed by methanol (3 × 2 min with rotation). The resin was then dried on a high vacuum line overnight.

- Cleavage method A – Cleavage and global deprotection

A 3 mL solution of TFA:H_2O:TIPS (95:2.5:2.5) (3 mL) was added to the resin and shaken for one hour. The blood red solution was filtered into ice cold ether to produce a white precipitate. The white powder was pressed into a pellet through the use of a centrifuge and the ether decanted and fresh ether added (This procedure was repeated 3 times). The resultant white pellet was dissolved in water (5 mL) and freeze dried for 48 hours.

- Cleavage method B – Cleavage from resin with protecting groups intact

A solution of 20 % hexafluoroisopropanol (HFIP) in DCM was added to the resin and shaken for 1 hour. The rest of the procedure was the same as cleavage method A.

6.5. HPLC traces for 2-deoxyribose reaction

Below are HPLC chromatograms from various HPLC columns used to identify hydrazone-trapped products as starting materials of the various sugar-forming reactions.

A) HPLC traces using a chiralpak OD-H column at 1.0 mL / min in a 5:95 isopropanol/hexane solvent system.

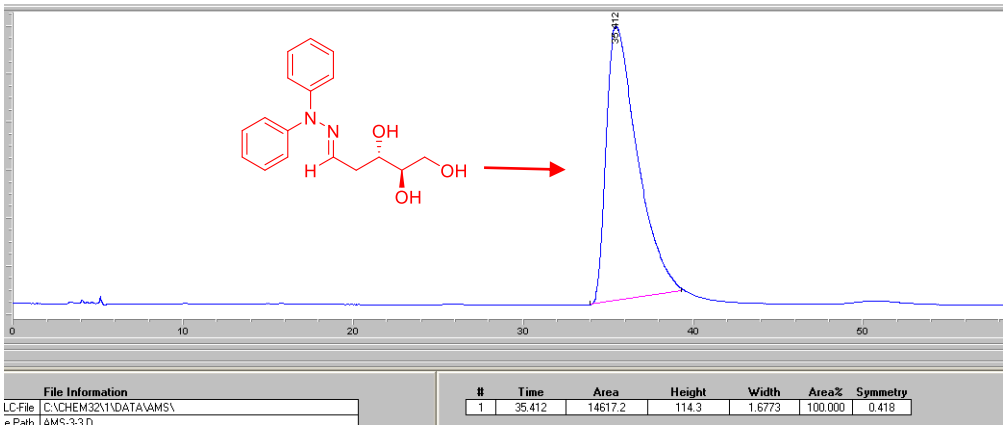


Figure 6.5. HPLC trace of 2-deoxy-D-ribose hydrazone standard **D-146**. The compound elutes at approximately 35 minutes.

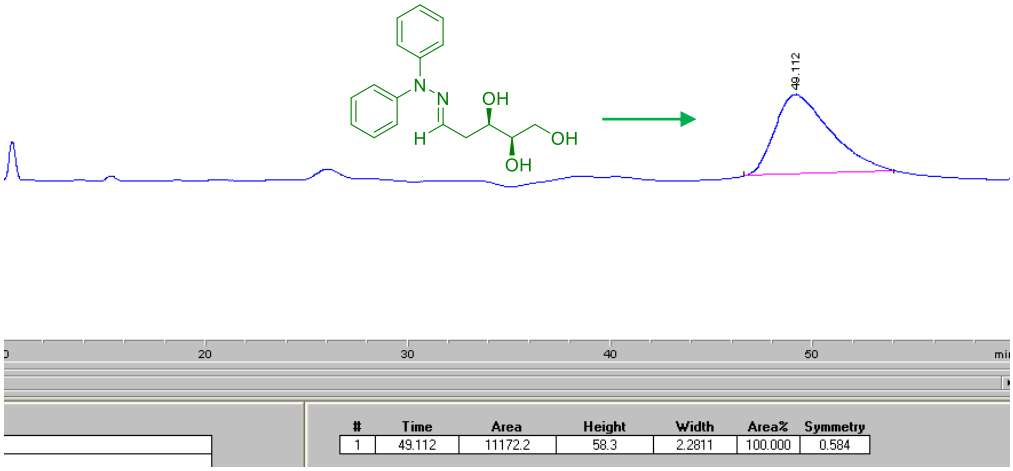


Figure 6.6. HPLC trace of 2-deoxy-D-threopentose hydrazone standard **D-147**. The compound elutes at approximately 49 minutes.

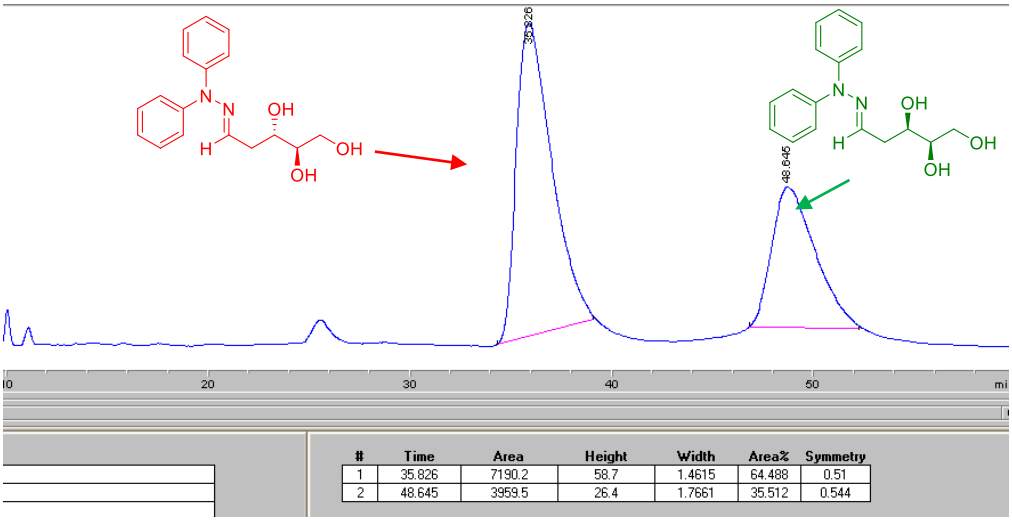
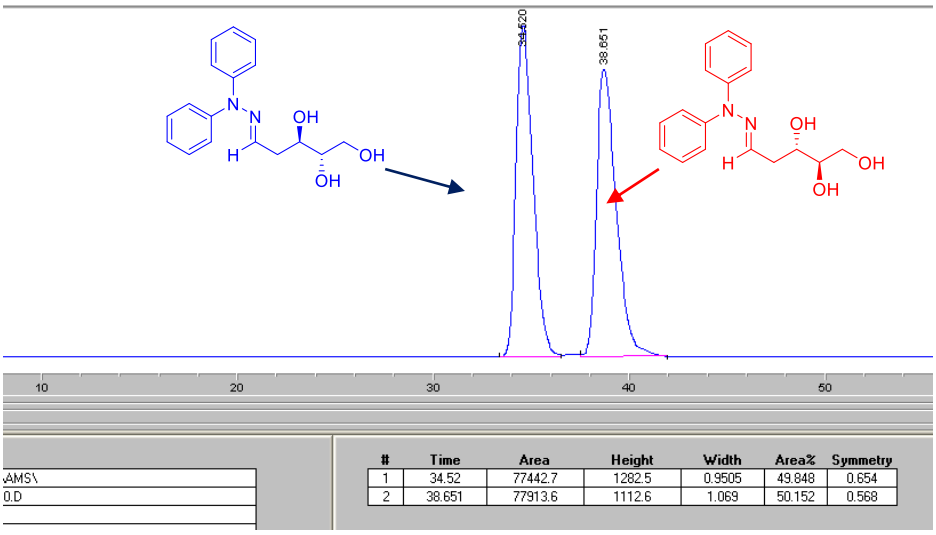


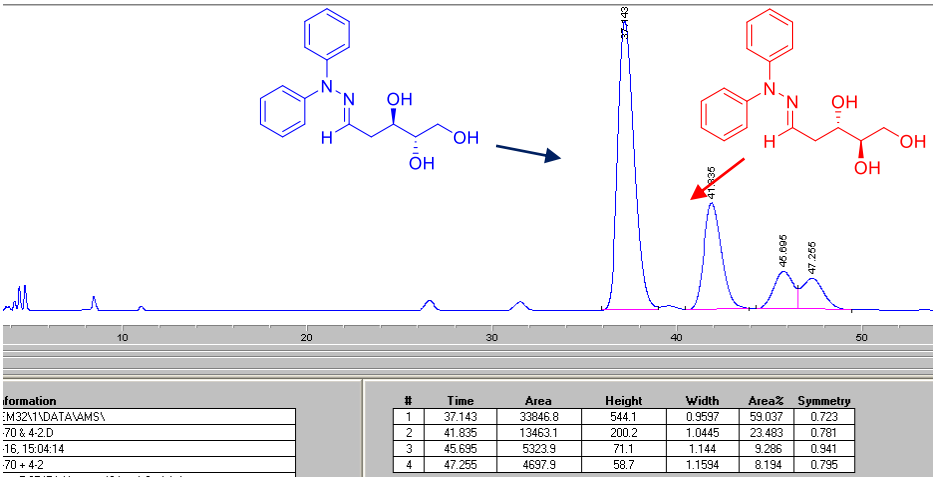
Figure 6.7. Genuine run of isolated products from the deoxyribose-forming reaction. 2-deoxy-D-ribose **D-146** elutes at 35.8 mins. 2-deoxy-D-threopentose **D-147** elutes at 48.7 mins.

B) For the investigation of the deoxyribose reaction using racemic glyceraldehyde **rac-9** a chiralpak IC column at 1.0 mL / min in a 5:95 isopropanol/hexane solvent system was used.



#	Time	Area	Height	Width	Area%	Symmetry
1	34.52	77442.7	1282.5	0.9505	49.848	0.654
2	38.651	77913.6	1112.6	1.069	50.152	0.568

Figure 6.8. HPLC trace of 2-deoxy-D-ribose hydrazone **D-146** standard and 2-deoxy-L-ribose hydrazone **D-147** standard.



#	Time	Area	Height	Width	Area%	Symmetry
1	37.143	33846.8	544.1	0.9597	59.037	0.723
2	41.835	13463.1	200.2	1.0445	23.483	0.781
3	45.695	5323.9	71.1	1.144	9.286	0.941
4	47.255	4697.9	58.7	1.1594	8.194	0.795

Figure 6.9. Products of the reaction of racemic glyceraldehyde **rac-9** and acetaldehyde **24** spiked with authentic 2-deoxy-L-ribose hydrazone **L-146**. 2-deoxy-L-ribose hydrazone **L-146** elutes at 37 mins and 2-deoxy-D-ribose hydrazone **L-147** elutes at 42 minutes.

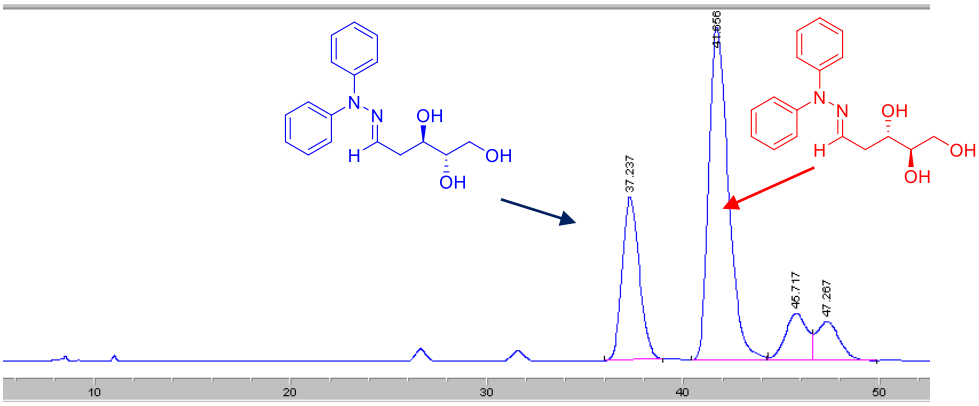


Figure 6.10. Products of the reaction of racemic glyceraldehyde **rac-9** and acetaldehyde **24** spiked with authentic 2-deoxy-D-ribose hydrazone **D-146**. 2-deoxy-L-ribose hydrazone **L-146** elutes at 37 mins and 2-deoxy-D-ribose hydrazone **D-146** elutes at 42 minutes.

6.6. General assay procedures

6.6.1. Dynamic Light Scattering (DLS) experiments of micelles

Dynamic Light Scattering (DLS) experiments to determine the size of supramolecular aggreagtes were carried out using a Zetasizer Nano (Malvern Instruments Ltd., Worcestershire, UK), based on the principle of measurement of the backscattered light fluctuations at an angle of 173° and the calculation of an autocorrelation function. Data were recorded from 5-10 runs per single measurement, each of which was carried out at 25°C using folded capillary cells (DTS 1060). The monomer solutions were prepared by dissolving the monomer (1 mg) in deionised water (1 mL). All samples were agitated and incubated at 25°C for 5 minutes prior to measurement. The data reported in Chapter 9 for compounds **210** and **219** are based on both volume distribution and intensity distribution.

6.7. General procedure for Nile Red (211) assay

The monomer (at a concentration of 250 μM for L-proline dodecylamide **210** and 75 μM for P-E-F-dodecylamide) was dissolved in phosphate buffered saline (PBS, 0.01 M, endowed with NaCl (138 mM) and KCl (2.7 μM)). In a cuvette, an aliquot of this solution was diluted by addition of PBS to a total volume of 1 mL before Nile red **211** (1 μL, 2.5 mM in ethanol) was added. Following inversion to ensure mixing, fluorescence intensity at 635 nm was recorded using a 550 nm excitation wavelength. The fluorescence intensity was measured three times at each concentration and an average of the three results taken. The logarithm of concentration vs. absorbance was plotted and the point of inflection determined as the critical micelle concentration.

6.8. Transition Electron Microscopy (TEM) imaging

In the case of the hydrogels, the monomer (20 mg) was gelled in deionised water (3 mL) through the cycle of heating, agitation and cooling. The gel was allowed to rest for 24 hours before imaging. In the case of the micellar aggregates, the monomer (10 mg) was dissolved in deionised water (1 mL) (well above the CMC limit). The samples were agitated at 25 °C before imaging.

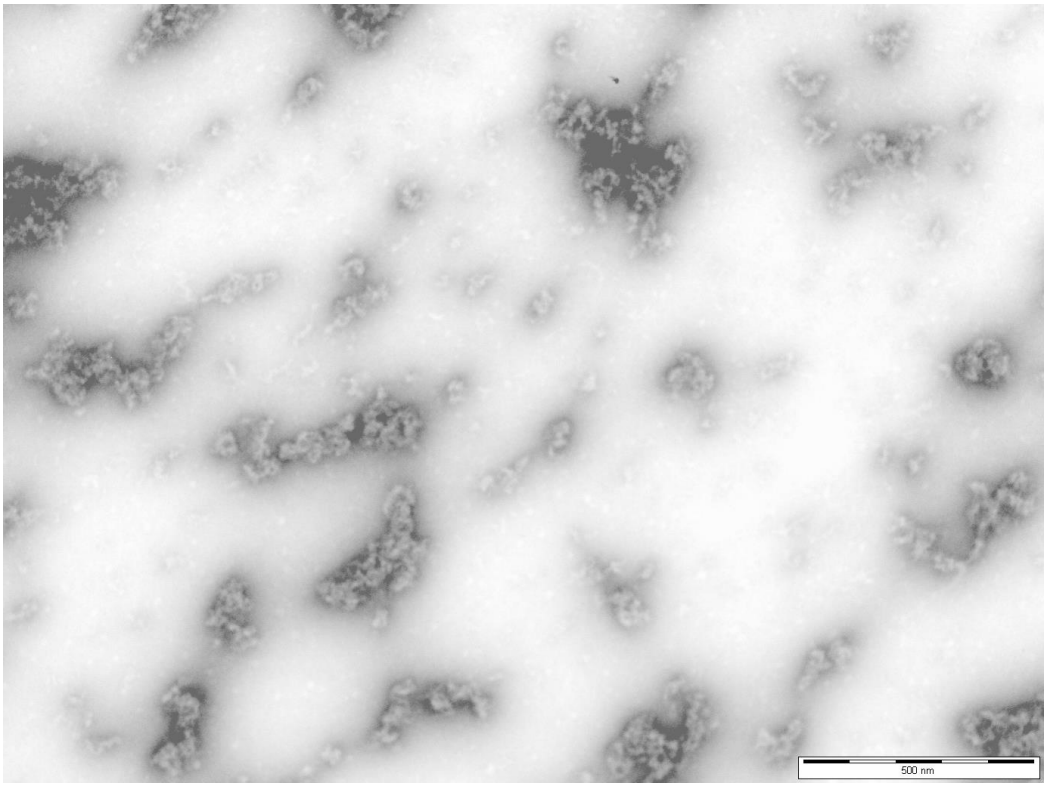


Figure 6.11. P-E-F-dodecyl amide aggregates. Scale bar = 500 μm.

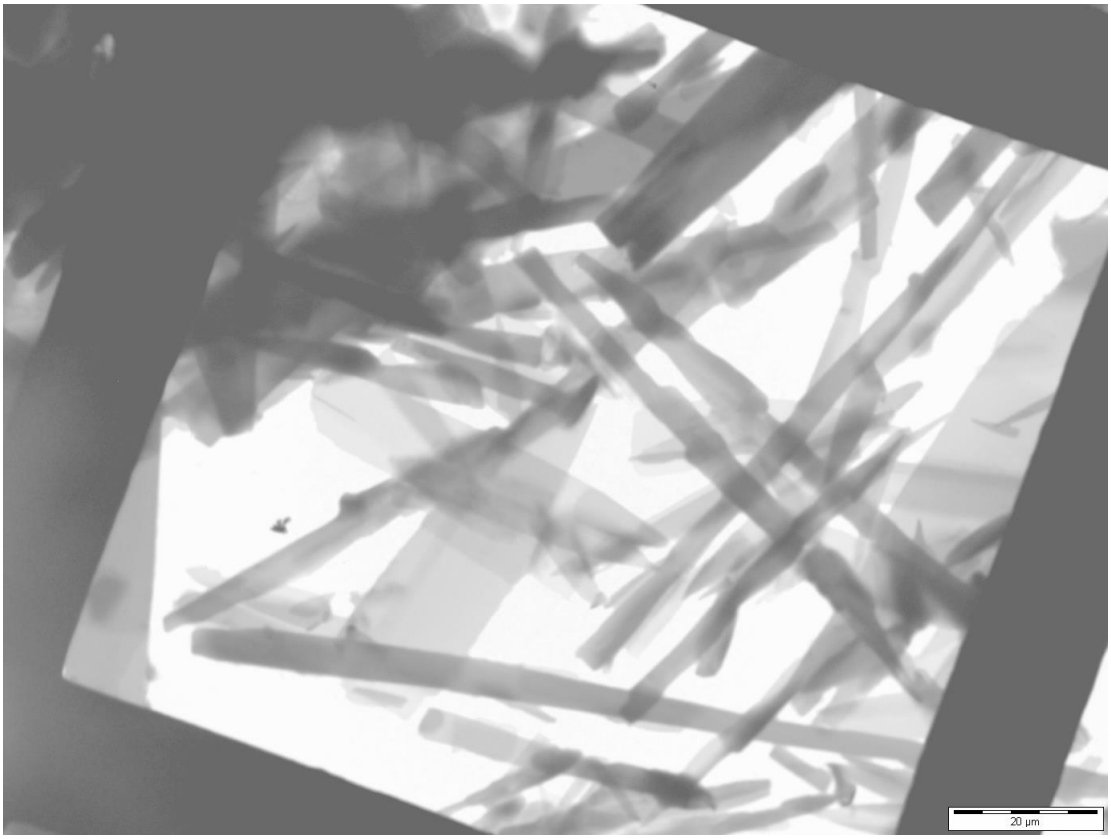


Figure 6.12. L-Phenylalanine-dodecylamide hydrogel. Scale bar = 20 μm.

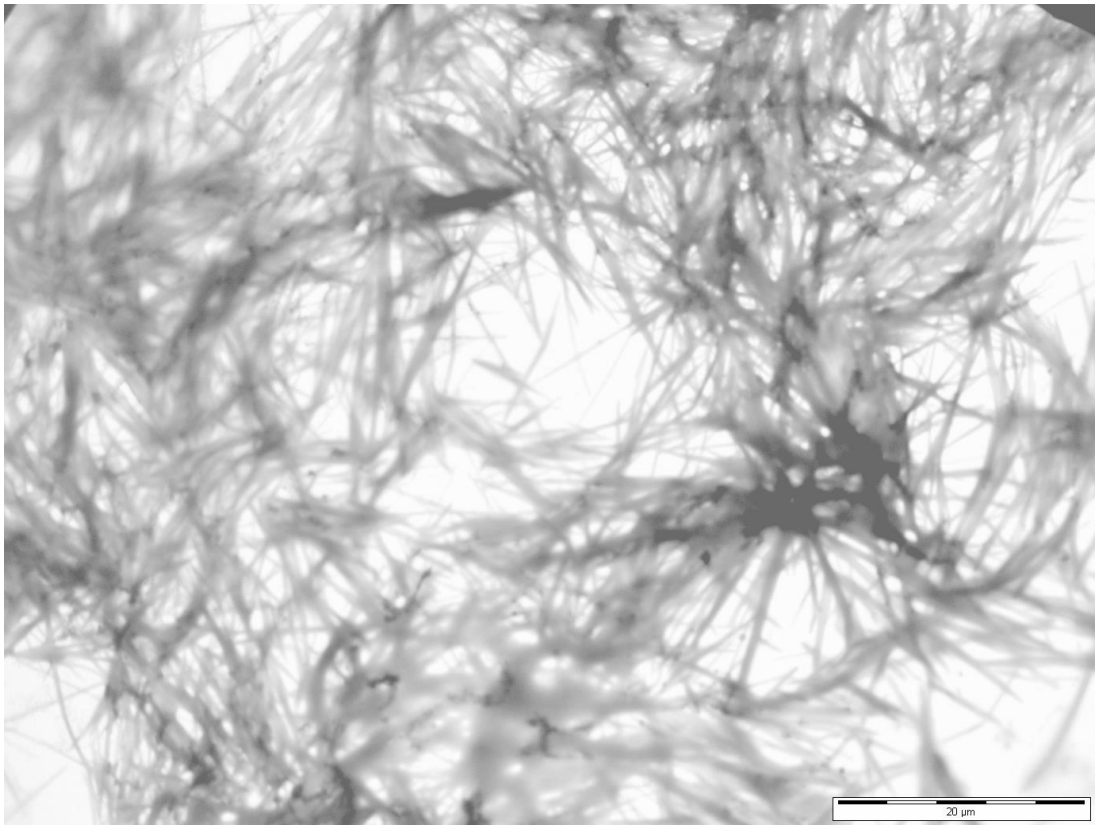


Figure 6.13. L-Proline-L-valine-dodecylamide hydrogel. Scale bar = 20 µm.

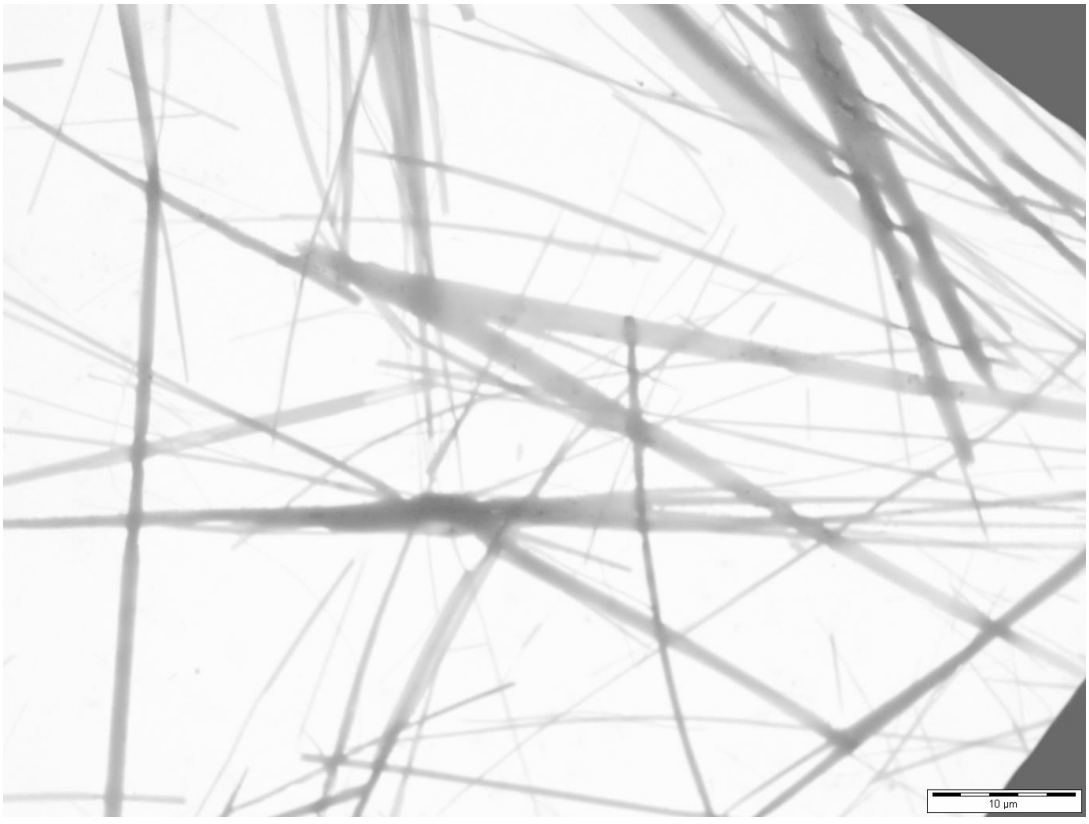


Figure 6.14. L-Proline-L-phenylalanine-dodecylamide hydrogel. Scale bar = 10 µm.

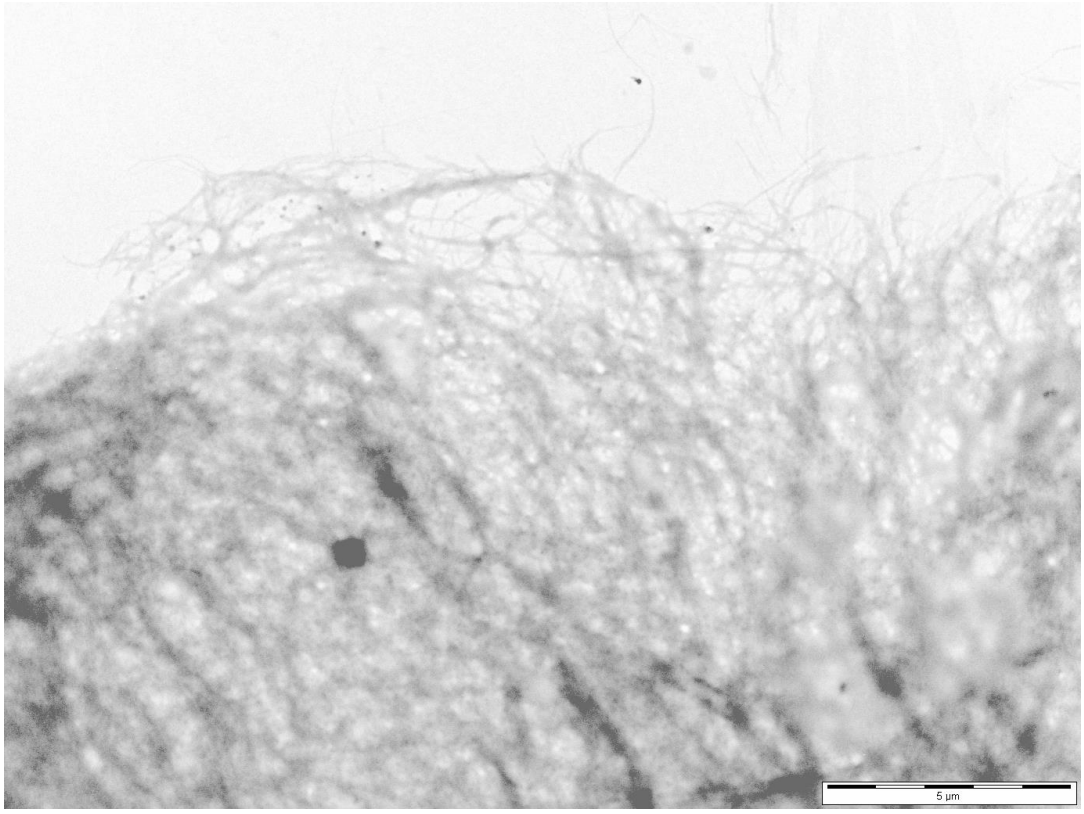


Figure 6.15. *N*-Methyl-L-proline-L-valine-dodecylamide hydrogel. Scale bar = 5 µm.

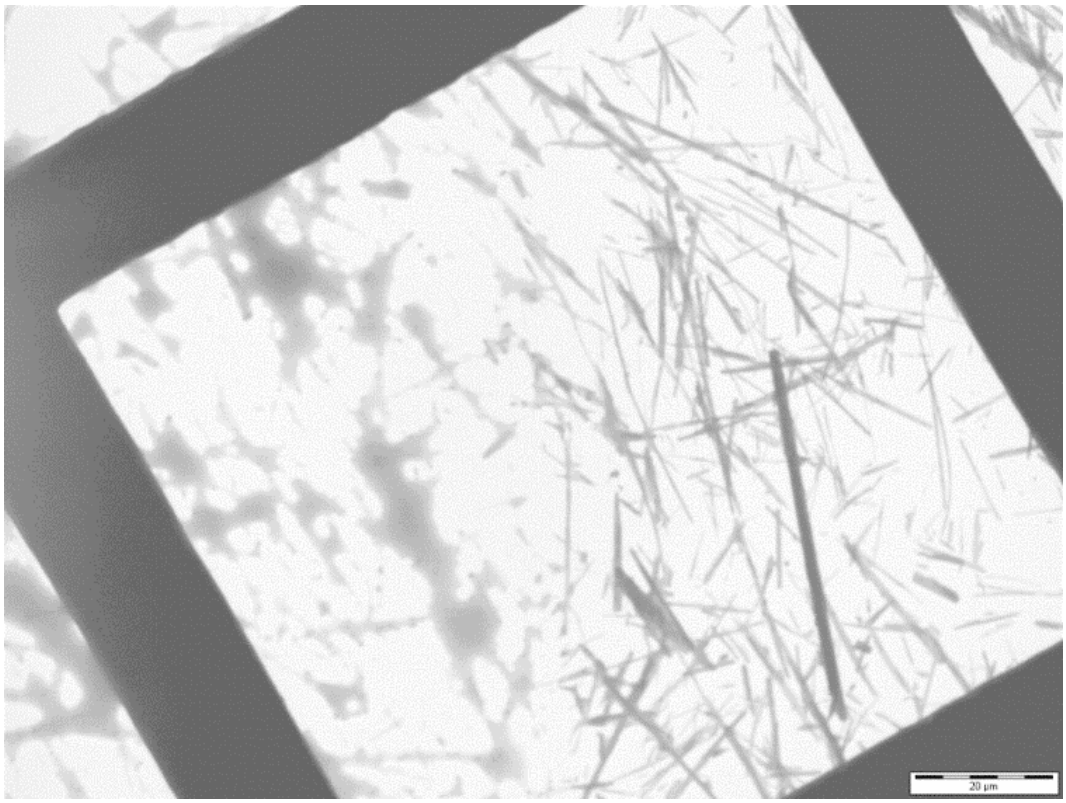


Figure 6.16. L-Valine-dodecylamide hydrogel. Scale bar = 20 µm.

Abbreviations

Ac	acetyl
AIBN	2,2’-Azobis(2-methylpropionitrile)
Ala	alanine
Asp	asparagine
aq	aqueous
Asp	aspartic acid
Bn	benzyl
Boc	*tert*-butyl oxycarbonyl
Cbz	Carboxybenzyl
DAHP	dihydroxyacetone phosphate
DBU	1,8-Diazabicylo[5.4.0]undec-7-ene
DCM	dichloromethane
DLS	Dynamic Light Scattering
DMAP	4-dimethylaminpyridine
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DNBSA	Dinitrobenzyly sulfuric acid
d.r.	diasteriomeric ratio
EDC	1-ethyl-3(3’-diemthylaminopropyl)carbodiimide
e.e.	enantiomeric excess
ESI	electrospray ionisation
Eq	equivalents
Fmoc	Fluorenylmethyloxycarbonyl
GC	gas chromatography
Gln	glutamine
Glu	glutamic acid
Gly	glycine

h	hour(s)
HBA	hydrobromic acid
HFIP	hexafluoroisopropanol
His	Histidine
HMDS	hexamethyldisilazane
HPLC	High Performance Liquid Chromatography
HRMS	High Resolution Mass Spectrometry
Hz	Hertz
IR	Infra-red
Ile	Isoleucine
Isoval	Isovaline
IPA	Isopropanol
J	coupling constant
MHz	megahertz
min	minute(s)
mp	melting point
MS	mass spectrometry
NBA	nitrobenzoic acid
NMR	Nuclear Magnetic Resonance
PBS	Phosphate Buffered Saline
Ph	phenyl
Phe	phenylalanine
Pro	proline
RNA	ribonucleic acid
rt	room temperature
Ser	serine
Suc	succinimide
TBDPS	*tert*-butyldiphenyl silyl
TEM	Transmission Electron Microscopy
TFA	Trifluoroacetic acid

TFAA	Trifluoroacetic anhydride
T_{gel}	Temperature of gelation
THF	tetrahydrofuran
Thr	threonine
TIPS	tri*iso*propylsilyl
TLC	thin-layer chromatography
Val	valine

References

[1] Lazcano, Miller, S. L., *Cell*, **1996**, *85*, 793

[2] Orgel, L. E., *Orig. Life Evol. Biosph.*, **2003**, *33*, 211

[3] Crick, F., *Nature*, **1970**, *227*, 561

[4] Accessed from: https://en.wikipedia.org/wiki/Messenger_RNA on 12th November 2017

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

[6] Woese, C., *The genetic code*, **1967**, Harper & Row, New York, p179-195

[7] Crick, F. H. C., *J. Mol. Biol.,* **1968**, *38*, 367

[8] Orgel, L. E., IJ. Mol. Biol.I, **1968**, *38*, 381

[9] Joyce. G. F., Orgel, L. E., *The RNA World*, **1993**, Cold Spring Harbor Laboratory Press 0-87969-380-0

[10] Robertson, P. M., Miller, S. R., *Nature*, **1995**, *375*, 772

[11] Orgel, L. E., *Crit. Rev. Biochem. Mo. Biol.*, **2004**, *39*, 99

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

[13] Bean, H. D., Sheng, Y., Collins, J. P., Anet, F. A. L., Leszczynski, J., Hud, N. V., *J. Am. Chem. Soc.*, **2007**, *129*, 9556

[14] Powner, M. W., Sutherland, J. D., *Angew. Chem. Int. Ed.*, **2010**, *49*, 4641

[15] Boutlerow, A., *Comptes Rendus,* **1861**, *53*, 145

[16] Decker, P., Schweer, H., Pohlmann, R., *J. Chrom.,* **1982**, *244*, 281

[17] Siminov, A., Pestunova, O. P., Matvienko, L. G., Parmon, V. N., *Kinet. Catal.,* **2007**, *48*, 245

[18] Socha, R. F., Weiss, A. H., Sakharov, M. M., *J. Catal.,* **1981**, *67*, 207

[19] Kopetzki, D., Anotnietti, M., *New J. Chem.* **2011**, *35*, 1787

[20] Breslow, R., *Tetrahedron Lett.,* **1959**, *1*, 22

[21] Appayee, C., Breslow, R., *J. Am. Chem. Soc.*, **2013**, *136*, 3720

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

[23] Kim, H.-J., Ricardo, A., Illangkoon, H. I., Kim, M. J., Carrigan, M. A., Frye, F., Benner, S. A., *J. Am. Chem. Soc.*, **2011**, *133*, 9457

[24] Lambert, J. B., Gurusamy-Thangavelu, S. A., Ma, K., *Science*, **2010**, *327*, 984

[25] Lambert, J. B., Lu, G., Singer, S. R., Kolb, V. M., *J. Am. Chem. Soc.*, **2004**, *126*, 9611

[26] Miller, S. L., *Science*, **1953**, *117*, 528

[27] Miller, S. L., *J. Am. Chem. Soc.*, **1955**, *77*, 2351

[28] Copper, G. J. T., Surnan, A. J., McIver, J., Colón-Santos, S. M., Gromski, P. S., Buchwald, S., Marina, I. S., Cronin, S., *Angew. Chem. Int. Ed.,* **2017**, *56*, 8079

[29] Abelson, P. H., *Proc. Natl. Acad. Sci. USA.,* **1966**, *55*, 1365

[30] Sutherland, J. D., Whitfield J. N., *Tetrahedron*, **1997**, *53*, 11493

[31] Miller, S. L., *The endogenous synthesis of organic compounds*. In: Brack A (ed) The molecular origins of life: assembling pieces of the puzzle. **1998,** Cambridge University Press, Cambridge, 59

[32] Cleaves, J. H., Chalmers, J. H., Lazcano, A., Miller, S. L., Bada, J. L., *Orig. Life Evol. Biosph.*, **2008**, *38*, 105

[33] Shock, E. L., *Orig. Life Evol. Biosph.*, **1990**, *20*, 331

[34] Hennet, R. J.-C., *Naturewissenschaften*, **1992**, *79*, 361

[35] Meierhenrick, U. J., *Chem. Soc. Rev.*, **2012**, *41*, 5447

[36] Takano, Y., Sato, R., Kaneko, T., Kobayashi, K., Marumo, K., *Org. Geochem.*, **2003**, *34*, 1491

[37] Takano, Y., Kobayashi, K., Yamanaka, T., Marumo, K., Urabe, T., *Earth Plan. Sci. Lett.,* **2004**, *219*, 147

[38] Yoshino, K., Anders, E., *Geochim. Cosmochim. Acta,* **1971**, *35*, 927

[39] Ritson, D. J., Sutherland, J. D., *Angew. Chem. Int. Ed.,* **2013**, *52*, 5845

[40] Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D., Sutherland J. D., *Nature Chem.,* **2015**, *7*, 301

[41] Martins, Z., Modica, P., Zanda, B., D'Hencecourt, L. L. S., *Meteor. Plan. Sci.,* **2015**, *50*, 926

[42] Kvenvolden, K., Lawless, J., Perine, K., Peterson, E., Flores, J., Ponnamperuma, C., *Nature*, **1970**, *228*, 923

[43] Bada, J. L., Cronin, J. R., Ho, M.-S., Kvenvolden, K. A., Lawless, J. G., Miller, S. L., Steinberg, S., *Nature*, **1983**, *301*, 494

[44] Engel, M. H., Macko S. A., *Nature*, **1997**, *389*, 265

[45] Pizarello, S., *Acc. Chem. Res.,* **2006**, *39*, 231

[46] Pizarello, S., Feng, X., Epstein, S., Cronin, J. R., *Geochim. Cosmochim. Acta,* **1994**, *58*, 5579

[47] Strecker, A., *Ann. Chem. Pharm.,* **1850***, 1*, 27

[48] Strecker, A., *Ann. Chem. Pharm.*, **1854**, *3*, 349

[49] Stern, S. A., Colwell, J. E., *Astrophys. J.,* **1997**, *490*, 879

[50] Bailey, J., *Orig. Life Evol. Biosph.*, **2001**, *31*, 167

[51] Bailey, J., Chrysostomou, A., HoughJ., Gledhill, T., McCall, A., Clark, S., Menard, F., Tamaura, M., *Science*, **1998**, *281*, 672

[52] Buschermohle, M., Whittet, D., Chrysostomou, A, Hough, J., Adamson, A., Whiteny, B., Wolff, M., *Astrophys. J.*, **2005**, *624*, 821

[53] Flores, J., Bonner, W., *J. Am. Chem. Soc.*, **1977**, *99*, 3622

[54] Breslow, R., Levine, M., Cheng, Z.-L., *Orig. Life Evol. Biosph.,* **2010**, *40*, 11

[55] Levine, M., Kenesky, S., Mazori, D., Breslow, R., *Org. Lett.*, **10**, *10*, 2433

[56] List, B., *Chem. Rev.*, **2007**, *107*, 5413

[57] Einhorn, A., Hollandt, F., *Liebigs Ann. Chem.*, **1898**, *301*, 95

[58] Hajos, Z. G., Parrish, D. R., *J. Org. Chem.*, **1974**, *39*, 1615

[59] Eder, U., Sauer, G., Wiechert, R., *Angew. Chem. Int. Ed.*, **1971**, *10*, 496

[60] List, B., Lerner, R. A., Barbas III, C. F., *J. Am. Chem. Soc.,* **2000**, *122*, 2395

[61] Zotova, N., Franzke, A., Armstrong, A. M., Blackmond, D. G., *J. Am. Chem. Soc.,* **2007**, *129*, 15100

[62] Mlynarski, J., Gut, B., *Chem. Soc. Rev.,* **2012**, *41,* 587

[63] Dabre, T., Machuquiero, M., *Chem. Commun.,* **2003**, 1090

[64] Andreau, C., Varea, T., Gregio, A., *Tetrahedron*, **2011**, *67*, 8705

[65] Zhu, X., Tanaka, F., Hu, Y., Heine, A., Fuller, R., Zhing, G., Olson, A. J., Lerner, R. A., Barbas, C. F., III,, Wilson, I. A., *J. Mol. Biol.,* **2004**, 1509

[66] Mase, N., Nakai, Y., Ohara, N., Yoda, H., Takabe, K., Tanaka, F., Barbas, C. F., III, *J. Am. Chem. Soc.,* **2006**, *128*, 734

[67] Amdejkouh, M., *Tetrahedron : Asymmetry*, **2007**, *18*, 390

[68] Ma, J. C., Dougherty, D. A., *Chem. Rev.,* **1997**, *97*, 1303

[69] Bisai, V., and Sing, V. K., *Synlett,* **2011**, 481

[70] Tsutsui, A., Takeda, H., Kimura, M., Fujimoto, T., Machinami, T., *Tetrahedron Lett.,* **2007***, 48*, 5213

[71] Pedatella, S., De Nisco, M., Mastroianni, D.M., Naviglio, D., Nucci, A., Caputo, R., *Adv. Synth. Catal,* **2011***, 353*, 1443

[72] Lipshutz, B., Ghoria, S., *Org. Lett.,* **2011**, *14*, 422

[73] De Nisco, M., Pedatella, S., Ullah, H., Zaidi, J. H., Naviglio, D., Ozdamar, O., Caputo, R., *J. Org. Chem.,* **2009**, *74*, 9562

[74] Triandafillidi, I., Aikaterini, B., Voutyritsa, E., Galiatsatou, G., Kokotos, C. G., *Tetrahedron,* **2015**, *71*, 932

[75] Córdova, A., Zou, W., Dziedzic, P., Ibrahem, I, Reyes, E., Xu, Y., *Chem. Eur. J.,* **2006**, *12*, 5383

[76] Northrup, A. B., MacMillan, D. W. C., *Science,* **2004**, *305*, 1752

[77] Pizzarello, S., Weber, A. L., *Science,* **2004**, *303*, 1151

[78] Weber, A. L., Pizarello, S., *Proc. Nat. Acad. Sci.* **2006**, *103*, 12713

[79] Dickerson T. J., Janda, K. D., *J. Am. Chem. Soc.*, **2002**, *124*, 3220

[80] Rogers, C. J., Dickerson, T. J., Brogan, A. P., Janda, K. D., *J. Org. Chem.*, **2005**, *70*, 3705

[81] Burroughs, L., Clarke, P. A., Forintos, H., Gilks, J. A., Hayes, C. J., Vale, M. E., Wade, W., Zvytniewski, M., *Org. Biomol. Chem.*, **2012**, *10*, 1565

[82] Snyder, L. E., Buhl, D., Zuckerman, B., Palmer, P., *Phys. Rev. Lett.,* **1969**, *22*, 679

[83] Hollis, J. M., Lovas, F. J., Jewell, P. R., *Astrophys. J.*, **2000**, *540*, L107

[84] Jørgensen, J. K., Favre, C., Bisschop, S. E., Bourke, T. L., van Dishoeck, E. F., Schmalzl, M., *Astrophys. J. Lett.*, **2012**, *757*, L4

[85] Breslow, R., Cheng, Z.-L., *Proc. Nat. Acad. Soc. USA,* **2010**, *107*, 5723

[86] Breslow, R., Ramalingam, V., Appayee, C., *Orig. Life Evol. Biosph.,* **2013**, *43*, 323

[87] Hein, J. E., Blackmond, D. G., *Acc. Chem. Res.*, **2012**, *45*, 2045

[88] Blackmond, D. G., Moran, A., Hughes, M., Armstrong, A., *J. Am. Chem. Soc.*, **2010**, *132*, 7598

[89] Seebach, D., Beck, K. A., Badine, D. M., Limbach, M., Eschenmoser, A., Treasurywala, A. M., Hobi, R., *Helvetica Chimica Acta,* **2007**, *90*, 425

[90] Klussmann, M., Iwamura, H., Mathew, S. P., Wells, D. H., Pandya, U., Armstrong, A., Blackmond, D. G., *Nature,* **2006**, *441*, 621

[91] Anastasi, C., Crowe, M. A., Powner, M. W., Sutherland, J. D., *Angew. Chem. Int. Ed.*, **2006**, *45*, 6176

[92] Powner, M. W., Gerland, B., Sutherland, D., *Nature*, **2009**, *459*, 239

[93] Cooper, G., Rios, A. C., *Proc. Nat. Acad. Sci.,* **2016**, *133,* E3322

[94] Meinert, C., Myrgorodska, I, de Marcellus, P., Buhse, T., Nahon, L., Hoffmann, S. V., le Sergeant d'Hendecourt, L., Meierhenrich, U. J., *Science,* **2016**, *352*, 208

[95] Wagner, A. J., Zubarev, D. Y., Aspuru-Guzik, A., Blackmond, D. G., *ACS Cent. Sci.,* **2017**, *3*, 322

[96] Frank, F. C., *Biochem. Biophys. Acta.,* **1953**, *11*, 459

[97] Soai, K., Shibata, T., Morioka, H., Choji, K., *Nature*, **1995**, *378,* 767

[98] Blackmond. D. G., *Proc. Nat. Acad. Soc.* **2004**, *101*, 5732

[99] Pizarello, S., Weber, A. L., *Orig. Life Evol. Biosph.*, **2010**, *40*, 3

[100] Breslow, R., Appayee, C., *Proc. Nat. Acad. Sci.,* **2013**, *110*, 4187

[101] Oro, J., Cox, A. C., *Fed. Proc.*, **1962**, *21*, 80

[102] Ritson, D. J., Sutherland, J. D., *J. Mol. Evol.* **2014**, *78*, 245

[103] Image of DNA labelled for noncommerical reuse and modification. Taken from: upload.wikimedia.org/Wikipedia/commons/thumb/8/82/Eukaryote_DNA.svg/1280px-Eukaryote_DNA.svg.png

[104] Bock, K., Lundt, I., *Carbohydr. Res.*, **1981**, *90*, 17

[105] Falentin, C., Beaupere, D., Demailly, G., Stasik, I., *Tetrahedron*, **2008**, *64*, 9989

[106] Kawasaki, T., Takamatsu, N., Aiba, S., Tokunaga, Y., *Chem. Commun.*, **2015**, *51*, 14377

[107] Barton, D. H. R., Kirby, G. W., *J. Chem. Soc.*, **1962**, 806

[108] Pincock, R. E., Perkins, R. R., Ma, A. S., Wilson, K. R., *Science*, **1971**, *174*, 1018

[109] Noorduin, W. L. *et al.*, *Angew. Chem. Int. Ed.,* **2008**, *47*, 6445

[110] Viedma, C., *Phys. Rev. Lett.*, **2005**, *94*, 065504-1

[111] McBride, J. M., Tully, J. C., *Nature*, **2008**, *452*, 161

[112] Miyagawa, S., Yoshimura, K., Yamazaki, Y., Takamatsu, N., Kuraishi, T., Aiba, S., Tokunaga, Y., Kawasaki, T., *Angew. Chem. Int. Ed.,* **2017**, *56,* 1055

[113] Hedstrom, E., *Chem. Rev.*, **2002**, *102*, 4501

[114] Hoang, C. T., Alezra, V., Guillot, R., Kouklovsky, C., *Org. Lett.*, **2007**, *9*, 2521

[115] Alberhart, L., *J. Am. Chem. Soc.*, **1973**, *95*, 7859

[116] Chitale, S., Derasp, J. S., Hussain, B., Tanveer, K., Beauchemin, A. M., *Chem. Comm.*, **2016**, *52*, 13147

[117] Coggins, A. J., Powner, M. W., *Nat. Chem.,* **2017**, *9*, 310

[118] Menten, K. M., Wyrowski, F., *Interstellar Molecules, Spring*er *tracts in modern physics*, **2011**, *2411*, 27

[119] Friestad, G. K., Massari, S. E., *J. Org. Chem.*, **2004**, *69*, 863

[120] Schrum, J. P., Zhu, T. F., Szostak, J. W., *Cold Spring Harb. Perspect. Biol.,* **2010**, *2*, a00212

[121] Ruiz-Mirazo, K., Briones, C., Escosura, A., *Chem. Rev.,* **2014**, *114*, 285

[122] Kee, P. T., Monnard, P.-A., *Elements*, **2016**, *12*, 419

[123] Deamer, D. W., *Nature*, **1985**, *317*, 792

[124] Deamer, D. W., Pashley, R. M., *Orig. Life Evol. Biosph.*, **1989**, *19*, 21

[125] Cistola, D. P., Atkinson, D., Hamilton, J. A., Small, D. M., *Biochemistry,* **1986**, *25*, 2804

[126] Haines, T., *Proc. Natl. Acad. Sci. USA,* **1983**, *80*, 160

[127] Dejanoic, B., Noethig-Laslo, V., Sentjurc, M., Walde, P., *Chem. Phys. Lipids*, **2011**, *164*, 83

[128] Monnard, P. A., Apel, C. L., Kanavarioti, A., Deamer, D. W., *Astrobiology*, **2002**, *2*, 139

[129] Apel, C. L., Deamer, D. W., Mautner, M. N., *Biochim. Biophys. Acta*, **2002**, *1559*, 1

[130] Maurer, S. E., Deamer, D. W., Boncella, J. M., Monnard, P. A., *Astrobiology*, **2009**, *9*, 979

[131] Namani, T., Deamer, D. W., *Orig. Life Evol. Biosph.,* **2008**, *38*, 329

[132] Adamala, K. P., Engelhart, A.E., Szostak, J. W., *Nat. Commun.*, **2016**, *7*, 11041

[133] Chen, I. A., Salehi-Ashtiani, K., Szostak, J. W., *J. Am. Chem. Soc.*, **2005**, *127*, 13213

[134] Adamala, K., Szostak, J. W., *Nature Chem.*, **2013**, *5*, 495

[135] Estroof, L. A., Hamilton, D. A., *Chem. Rev.*, **2004**, *104*, 1201

[136] Du, X., Zhou, J., Xu, B., *Chem. Rev.*, **2015**, *115*, 13165

[137] Singh, N., Kumar, M., Miravet, J. F., Uljin, R. V., Escuder, B., *Chem. Eur. J.*, **2017**, *23*, 981

[138] Tanford, C., *Science*, **1978***, 200,* 1012

[139] Rodríguez-Llansola, F., Miravet, J. F., Escuder, B., *Chem. Commun.,* **2009**, 7303

[140] Xie, Z. G., Zhang, A. Y., Ye, L., Feng, Z. G., *Soft Matter*, **2009**, *5*, 1474

[141] Bayoumi, M., Bayley, H., Maglia, G., Sapra, J. T., *Sci. Rep.* **2017***, 7*, 45167

[142] Pollack, G. H., *Cells, Gels and the Engines of Life,* Ebner and Sons publishers, **2001**, 305

[143] Trevors, J. T., Pollack, G. H., *Prog. Biophys. Mol. Biol.*, **2005**, *89*, 1

[144] Tako, M., Nakamura, S., *Carbohydr. Res.* **1988**, *180*, 277

[145] Xiong, J.-Y., Narayanan, J., Liu, X.-Y., Chong, T. K., Chen, S. B., Chung, T.-S., *J. Phys. Chem. B*, **2005**, *109*, 5638

[146] Letherby, M. R., Young, D. A., *J. Chem. Soc. Faraday Trans. 1*, **1981**, *77*, 1953

[147] Escuder, B., Roderiguez-Llansola, Miravet, J. F., *New J. Chem.,* **2010**, *34*, 1044

[148] Diaz-Oltra, S., Berdugo, C., Miravet, J. F., Escuder, B., *New. J. Chem.,* **2015**, *39*, 3785

[149] Kyte, J., Doolittle, R. F., *J. Mol. Biol.*, **1982**, *157*, 105

[150] Frederix, P. W. J. M., Scott, G. G., Abul-Haija, Y. M., Kalafatovic, D., Pappas, C. G., Javid, N., Hunt, N. T., Uljin, R. V., Tuttle, T., *Nature Chem.,* **2015**, *7*, 30

[151] Jia, T. Z., Pappas, C. G., Ulijn, R. V., Szostak, J. W., *XVIII Int. Conf. on Origin of Life,* **2017**, *No. 1967*, 4020

[152] Tena-Solsona, M.,Nanda, J., Díaz-Oltra, S., Chotera, A., Ashkenasy, G., Escuder, B., *Chem. Eur. J.*, **2016**, *22*, 6687

[153] Stuart, M. C. A., van der Pas, J. C., Engberts, J., *J. Phys. Org. Chem.*, **2005**, *18*, 929

[154] Lim, Y.-B., Lee, E., Lee, M., *Angew. Chem. Int. Ed.*, **2007**, *46*, 9011

[155] Hirst, A. R., Coates, I. A., Boucheteau, T. R., Miravet, J. F., Escuder, B., Castelletto, V., Hamley, I. W., Smith, D. K., *Angew. Chem. Int. Ed.,* **2008**, *130*, 9113

[156] Burroughs, L., PhD Thesis, University of York, Chemistry Department, **2011**

[157] Gasparini, F., Vogel, P., *J. Org. Chem.*, **1990**, *55*, 2451

[158] Friestad, G. K., Massari, S. E., *Org. Lett.*, **2000**, *2*, 4237

[159] Rossi, J.-C., Marull, M., Bioteua, L., Taillades, J., *Eur. J. Org. Chem.*, **2007**, 662

[160] Mangette, J. E., Johnson, M. R., Le, V.-D., Shenoy, R. A., Roark, H., Stier, M., Belliotti, T., Capiris, T., Guzzo, P. R., *Tetrahedron*, **2009**, *65*, 9536

[161] An, X.-D., Yu, S., *Org. Lett.,* **2015**, *17*, 5064

[162] Lawandi, J.,Toumieux, S., Seyer, V., Campbell, P., Thielges, S., Juillerat-Jeanneret, L., Moitessier, N., *J. Med. Chem.*, **2009**, *52*, 6672

[163] Caputo, C. A., Carneiro, F. D. S., Jennings, M. C., Jones, N. D., *Can. J. Chem.*, **2007**, *85*, 85

[164] Bagley, M. C., Buck, R. T., Hind, S. L., Moody, C. J., *J. Chem. Soc. Perkin Trans. 1*, **1998**, 591

[165] Hoang, C. T., Alezra, V., Guillot, R., Kouklovsky, C., *Org. Lett.*, **2007**, *9*, 2521

[166] Kawaski, K., Kawasaki, C., Tsuda, Y., Yagu, M., Okada, Y., Yamaji, K., Takagi, T., Tanizawa, O., *Chem. Pharm. Bull.*, **1980**, *28*, 2699

[167] Moresch, G. W., Rebstock, M. C., Wittle, E. L., Tinney, F. J., Nicolaides, E. D., Hutt, M. P., Mich, T. F., Vandenbelt, J. M., Edgren, R. E., *J. Med. Chem.*, **1979**, *22*, 935

[168] Lu, J.-Y., Riedrich, M., Wojtas, K. P., Arndt, H.-D., *Synthesis*, **2013**, *45*, 1300

[169] Il'ina, A. V., Davidovich, Y. A., Rogozhin, S. V., *J. Gen. Chem. USSR (Engl. Transl.)*, **1984**, *54*, 1904

[170] Dendrinos, K. G., Kalivretenos, A. G., *Tetrahedron Lett.*, **1998**, *39*, 1321