Structure and Function of Hyperbranched Depsipeptides Formed Under Wet-Dry Cycling Conditions

A thesis submitted to

Indian Institute of Science of Education and Research Pune

in partial fulfilment of the requirements of BS-MS Dual Degree Program Programme by

Ajay Verma

Reg. No. 20151039

Fifth-year BS-MS Student

Department of Biology

Indian Institute of Science Education and Research Pune, India

Supervisor – Prof. Irena Mamajanov

Earth-Life Science Institute

Tokyo Institute of Technology, Tokyo, Japan

Expert – Prof. Sudha Rajamani

Department of Biology

Indian Institute of Science Education and Research Pune, India

EARTH-LIFE SCIENCE INSTITUTE
TOKYO INSTITUTE OF TECHNOLOGY

IISER PUNE

Certificate

This is to certify that this dissertation entitled "Structure and Function of Hyperbranched Depsipeptides Formed Under Wet-Dry Cycling Conditions" towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research (IISER), Pune represents study/work carried out by Ajay Verma at the Tokyo Institute of Technology, Tokyo, Japan under supervision of Prof. Irena Mamajanov, Earth-Life Science Institute (ELSI), during the academic year 2019-2020.

Ajay Verma	Prof. Irena Mamajanov	Prof. Sudha Rajamani
Student	Supervisor	Expert
IISER Pune, India	ELSI, Japan	IISER Pune, India

Declaration

I hereby declare that the matter embodied in the report entitled "Structure and Function of Hyperbranched Depsipeptides Formed Under Wet-Dry Cycling Conditions" are the results of the work carried out by me at the Tokyo Institute of Technology, Tokyo, Japan under the supervision of Prof. Irena Mamajanov, Earth-Life Science Institute and the same has not been submitted elsewhere for any other degree.

Ajay Verma	Prof. Irena Mamajanov	Prof. Sudha Rajamani
Student	Supervisor	Expert
IISER Pune, India	ELSI, Japan	IISER Pune, India

Abstract:

Proteins and peptides play a significant role in biology as structural components, enzymes, and transport carriers within cells. Peptides may have been important in prebiotic chemistry as well. In contemporary biology, peptide synthesis is achieved through highly sophisticated enzymatic machinery and energy harvesting mechanisms. However, peptide synthesis in the prebiotic world is difficult because the direct coupling of amino acids to form peptide is thermodynamically unfavourable. Also, peptides depend on their monomer sequences for folding patterns and functional properties, which is achieved by translation machinery. It is conceivable that in prebiotic chemistry, polymers other than peptides performed the necessary function. Here, we have studied polyamides formed upon condensation of polybasic carboxylic acids and amines, structures closely related to peptides, but formed under mild conditions. One of the polyamides was shown to have catalytic functions, suggesting that these structures are an attractive model for ancestral peptides. In modern biology, the peptides derive their function from sequence specificity, a property achieved through the highly evolved translation machinery. Here, we have studied the peptides and depsipeptides formed through the amide-exchange process, allowing for the insertion of amino acids into polyester matrixes. We have shown that the pattern of amino acid insertion is highly dependent on the topology of the polyesters and the chemical properties of amino acids. We, therefore, have concluded that the ester-amide exchange process could have facilitated a rudimentary "translation" in the peptide formation at the early stages of chemical evolution.

List of Figures:

List of Tables:

Acknowledgements:

1. INTRODUCTION:

Earth is estimated to be about 4.6 billion years old (Sleep et al., 1989). The currently accepted hypothesis to explain the formation of earth is the "planetary accretion hypothesis." The hypothesis states that all the planets in the solar system formed from the accretion of the flattened disk of gas and dust. The accretion resulted in 'planetesimals' which were mountain-size bodies. These planetesimals collided and interacted due to gravitational force among each other to form 'planetary embryos'. The collision of planetary embryos resulted in the planets present today (Chambers, 2004). At that time, the earth was molten, and extensively volcanic activities were happening on the surface caused by heavy bombardments from space (Abe, 1993). These impacts were significant enough to evaporate oceans. The atmosphere content was changing due to the vaporized water from oceans, delivery of extraterrestrial molecules, and volcanic outgassing (Sleep et al., 1989; Chyba and Sagan, 1992). In all the turmoil, life somehow emerged from the geological and chemical processes around at least 4 billion years ago (Mojzsis et al., 1996; Bell et al., 2015). One of the intriguing still unanswered questions is how these geological and chemical processes gave rise to biological systems.

My project aims to study a subpart of this big question, which is the prebiotic formation of functional biopolymers and their precursors. Many scientists have hypothesized mechanisms of the transition from chemistry to biology keeping in mind the possible prebiotic environments and the availability of chemical compounds of which we do not have a full record (Wächtershäuser, G. 2010; Greenwald sand Riek, 2012; Orgel, 2004; Ikehara, 2005). One of the famous hypotheses to explain the transition is the concept of the RNA World, which is based on the fact that RNA molecules are capable of self-replication and catalysis. The hypothesis assumes that when and if an RNA molecule arises spontaneously in the prebiotic world, the chemical evolution and eventually biological evolution process commences (Gilbert, 1986). Alternatively, life could have started from amyloids, aggregates of polypeptides characterized by fibrillar morphology (Sunde et al., 1997). Facile self-assembly of short peptides into an amyloid structure,

the ability of amyloids to auto catalyze their formation, and the emergent catalytic properties of the aggregates are attractive properties for the origins of life model system (Greenwald and Riek, 2012; Maury, 2018). The formation of biopolymers, such as RNA or peptides, is often central to the origins of life theories. In modern biology, biopolymers play an essential role in maintaining the structural and functional integrity of living systems. Polymers have different functions than their monomers. For instance, proteins have high tensile strength required for maintaining the cytoskeleton of cells (Wickstead and Gull, 2011). Also, stable folding of the proteins allows them to have catalytic and structural functions (Creighton, 1990). Polynucleotides such as DNA are chemically stable compared to their monomers; therefore, they can reliably store genetic information (DeVoe and Tinoco, 1962). In biology, as we know it, the above functions are inaccessible by small molecules or their assemblies. Therefore, understanding the chemical evolution of biopolymers is likely a crucial step in solving the origin of life.

The biosynthesis of biopolymers is made possible by highly evolved biological machinery, including energy harvesting mechanisms and enzymes. The abiotic synthesis of biopolymers is difficult due to the thermodynamic constraints and is considered a largely unsolved problem. Thermodynamics of peptide bond and phosphodiester bond formation is disfavored in aqueous solution and require activation and/or extreme conditions (Stojanoski and Zdravkovski, 1993; Dickson et al., 2000). The Gibbs free energy of the dimerization of two glycine molecules to form diglycine is +3.6kcal at 37.5°C suggesting that the reaction is non-spontaneous and only an insignificant amount of products forms at given conditions (Stojanoski and Zdravkovski, 1993). Similarly, at 25°C and pH 7, the free energy of the phosphodiester bond formation is 5.3kcal mol^{-1} showing the thermodynamic unfavorability of the reaction (Dickson et al., 2000).

To overcome the challenges of the abiotic polynucleotides synthesis, researchers proposed several mechanisms. In one of the approaches by Morávek, he showed that dry heating of nucleotides at 100°C yields 17% dinucleotides (Morávek, 1967). In another work, activated nucleotides in the presence of divalent metal ions like Pb^{2+} and

Co^{2+} at 20°C yield oligomers up to 16-mers (Sawai, 1976). In 2008, another group of researchers showed that the polymerization rate increases when nucleotides are encapsulated in lipid vesicles and subjected to wet-dry cycles (Rajamani et al., 2008).

Similarly, some prebiotic models describe the formation of peptides. The work by Fox and Harada shows that the mixture of amino acids containing an excess of dicarboxylic amino acids forms a polymer dubbed "proteinoid" when heated above 150°C. Proteinoids are abiotically synthesized polymers which contain biological amino acids and resemble proteins in their properties (Fox and Harada, 1958; Fox and Harada, 1959). In another work, researchers used metal ions mediated peptide formation under simulated tidal environments, which yielded tripeptides of alanine and glycine (Schwendinger and Rode, 1991). In 1998, Huber and Wächtershäuser heated a mixture of amino acids in aqueous solution at 100°C in the presence of (Fe, Ni)S precipitates under simulated hydrothermal environments and found dipeptide formation with 6.2% yield (Huber, 1998). Another report from Lahav and coworkers shows that glycine forms oligopeptides when heated at 94°C in the presence of clay (kaolinite). One solution to the polypeptide formation problem has been described in (Forsythe et al., 2015). The researchers invoked a thermodynamically favourable ester-amide process to transform polyester strands into mixed backbone ester-peptide polymers, depsipeptides. The synthesis of depsipeptides under wet-dry cycling conditions have ensured the enrichment of depsipeptides in peptide units and suggested a path to the facile formation of mostly pure peptides. The scheme for the ester amide exchange is shown in Figure 1.2. In the dry phase, the low water activity induces the production of ester linkages between monomers. Simultaneously, ester-amide exchange occurs between newly formed esters and amino acids. In wet-phase, after the addition of water, the products hydrolyze partially. Since ester bonds are more prone to hydrolysis than amide bonds, in the subsequent cycles, the polymers get enriched in amide bonds. The follow-up work by (Frenkel-Pinter et al., 2019) studied the reactivity of cationic amino acids towards depsipeptide formation and have shown that, at least in a small sampling, proteinaceous amino acids have incorporated into depsipeptides preferentially over non-proteinaceous acids. Another work by (Yu et al., 2017) shows that continuous

feeding of amino acids to the solution of depsipeptides formed through wet-dry cycling of hydroxy acids and amino acids results in an enrichment of peptide bond and chain growth of polymer.
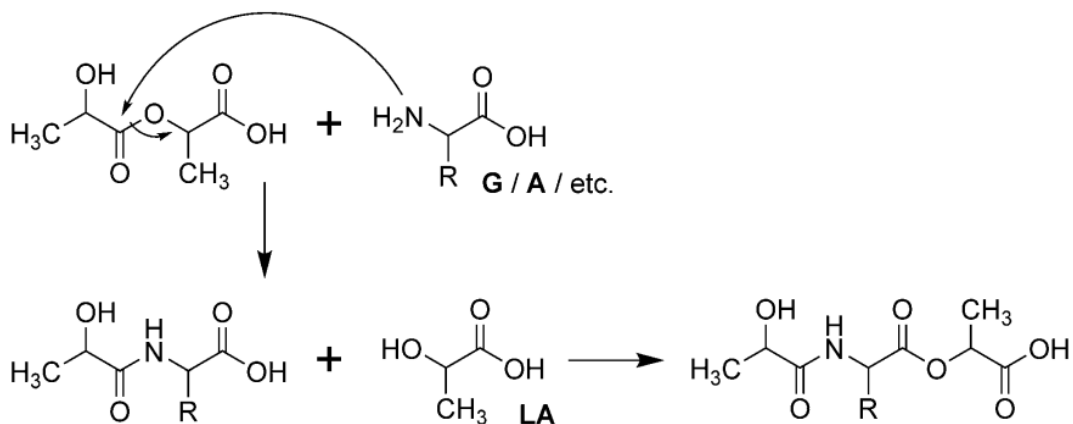
Figure 1.1: Schematic representation for the reaction between lactic acid (LA) and amino acids like glycine (G) or alanine (A) to form amide bond through ester amide-exchange (Forsythe et al., 2015).

The formation of peptide and phosphodiester bond in an abiotic scenario is challenging. Furthermore, very few abiotic synthesis models address the sequence specificity in the mixed peptides and nucleic acids. The sequence specificity is the means to store and manipulate the genetic information in nucleic acids and is responsible for the folding patterns in proteins that determine their functional properties. The flow of sequence specificity and genetic information in biological systems is explained by central dogma, which involves two steps, transcription and translation (Crick, 1970). In transcription, the enzyme RNA polymerase assisted by transcription factors binds promoter sequence on DNA and initiates unwinding of DNA (Buratowski et al., 1989). The RNA polymerase-transcription factors complex then binds to the sequence of DNA complementary to transcription start site sequence and starts synthesizing the new RNA strand (Hampsey, 1998). In the process of translation, protein biosynthesis happens by linking amino acids in a specific order determined by the messenger RNA (mRNA) sequence. The transfer RNA (tRNA) helps decode the mRNA sequence into a protein (Ramakrishnan, 2002).

The amino acid monomers are specified by three-nucleotide mRNA sequences (codons). tRNA molecules contain anticodon sequences capable of specific recognition of mRNA codons and are physically bound to the appropriate amino acids (Yusupov et al., 2001). When a tRNA recognizes a corresponding codon, the tRNA transfers the appropriate amino acid to the end of the growing amino acid chain (Gualerzi and Pon, 1990). This type of highly sophisticated biomolecular machinery would not have been present at the early stages of chemical evolution. So, controlling the sequence makeup of the precursors of these biopolymers in the prebiotic world might be difficult. So, my thesis aims to search and study a prebiotically plausible system where amino acid makeup is non-random.
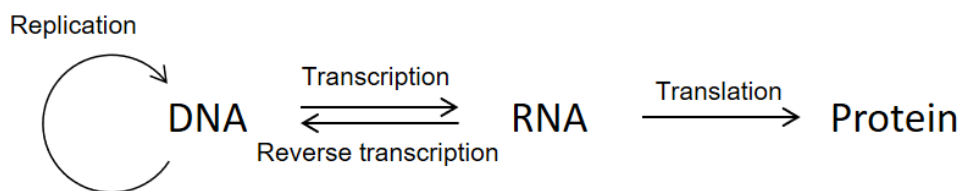
Figure 1.2: Schematic representation of the central dogma. Adapted from (Crick, 1970).

Multiple laboratory studies create systems that mimic early Earth conditions. In many reported experiments, biomolecules and biopolymers do not form in high yields. The major products usually consist of an intractable mixture of macromolecules, 'tar' (Miller, 1955; Moutou et al., 1995; Liebman et al., 1995; Butlerow, 1861). Tar polymers are messy chemical systems formed as a result of extensive polymerization of the substrate as well as formed biomolecules (Miller, 1955; Moutou et al., 1995; Liebman et al., 1995, Butlerow, 1861). For instance, the carbohydrates formed in the formose reaction polymerize further to form chemical messes (Decker et al., 1982). However, a subset of tar polymers can be potentially functional. Such as the proteinoid microsphere formed by thermal polymerization of α-amino acids have catalytic functions towards hydrolysis of p-Nitrophenyl acetate (Rohlfing and Fox, 1967). The work by Rohlfing and Fox suggested the idea of non-biological polymers having a function in the origin of life. Similar to tar, hyperbranched polymers (HBP) are formed through condensation

dehydration polymerization. Their one-step synthesis from prebiotically plausible molecules and the ability to mimic enzymatic function makes them an attractive candidate to study tar-like properties (Mamajanov et al., 2015; Kirkorian et al., 2012). The hyperbranched polymers are highly branched macromolecules with tree-like three-dimensional topology (Gao and Yan, 2004). The hyperbranched polymers are intrinsically globular capable of forming intramolecular cavities. Due to the high abundance of end groups, the hyperbranched polymers are efficient in binding small molecules. These properties of hyperbranched polymers are superficially similar to those proteins. The idea that hyperbranched polyesters (HBPEs) formed through a straightforward abiotic process could have preceded protein described in (Mamajanov et al., 2015). The HBPEs has been further shown to have catalytic properties (Mamajanov and Cody, 2017). How would the HBPE transition into functional peptides or proteins? Here, I investigate whether the ester-amide process could transform HBPEs into linear peptides while retaining any of HBPE properties.
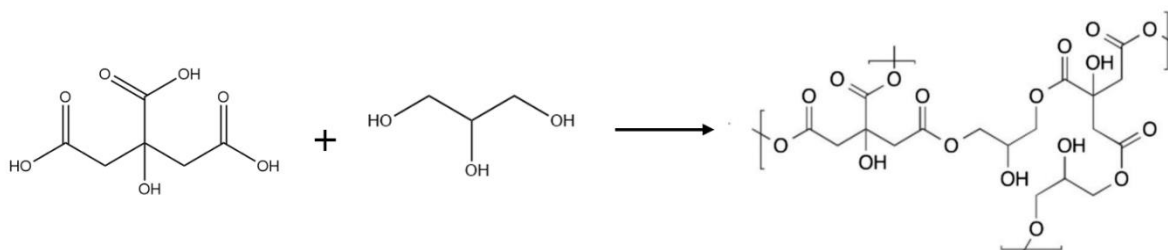
Figure. 1.3: Schematic representation of citric acid and glycerol polyesterification (Mamajanov et al., 2015).

My thesis project aims to introduce different amino acids to the hyperbranched polymers systems and study the effect of the HBPE matrix on the ester-amide exchange process. To mimic the prebiotic fluctuations of physicochemical parameters such as the temperature of the solution and concentration of substrates, we employed wet-dry cycling of the mixture. The intermittent drying of HBPEs with amino acids into linear depsipeptides would imprint some of the polyester properties into the newly formed polymers. The insertion of amino acids into the branched scaffold would not be random. In the system, the hydrophobic amino acids would incorporate into the less polar core of the hyperbranched polyesters, while polar amino acids would remain on the polymer's

periphery. Also, the smaller amino acids would be able to reach the more crowded core, whereas the longer amino acids will stay on the surface. To demonstrate this principle, I rely on HBPEs of various polarity and a minimal set of amino acids, glycine, alanine, aspartic acid, and valine, that covers a lot of peptides functionality as suggested in Kenji Ikehara's GADV theory (Ikehara, 2005). The resulting products are analyzed by multidimensional NMR spectroscopy, size-exclusion chromatography (SEC), and mass spectroscopy.
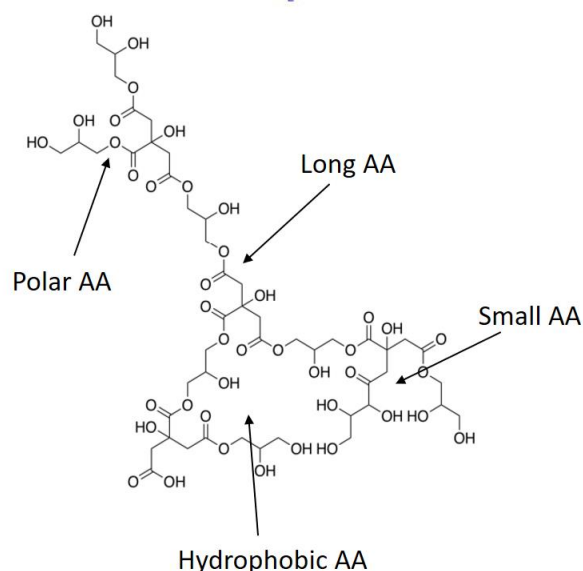
Figure 1.4: Hypothesis for the reaction between citric acid, glycerol polyesters and amino acids (AA). The amino acids will insert depending on their side groups.

It is usually assumed that peptides have been synthesized abiotically as is. However, the peptides might be the products of chemical evolution from more disordered backbone like polyamides. Polyamides are organic polymers formed upon reaction between carboxylic acids and amines; polyamides are called "peptides" when referred to the derivatives of alpha-amino acids. Polyamides are used in industrial applications such as insulators (Nylon6) and fibres (polyamide resin) (Reimschuessel, 2010), while peptides are mainly involved in the catalytic and structural functions of living systems. Polyamides are less controlled structures and devoid of sequence-based functionality. Unlike peptides, polyamide synthesis does not require highly sophisticated translation machinery, thus easier to make compared to peptides (Pattabiraman and Bode, 2011).

The current methods for industrial amide bonds require expensive coupling reagents and generate a large amount of waste. Therefore, the development of cheaper and greener methods of amide synthesis is an area of active research. (Pattabiraman and Bode, 2011). Several studies suggest that polybasic carboxylic acids and amines react to form amides at milder conditions. Vert and coworkers (Vert et al., 1991) have attempted to synthesize polyamides from citric acid and L-lysine in benzene. In previous works to polymerize citric acid, the polymeric end products were always cross-linked, thus insoluble in water. To prevent the cross-linking of functional groups, the multifunctional monomers were converted to bifunctional monomers by using protecting groups. The centre carboxylic group of citric acid was protected by condensation of the acid group with benzaldehyde and chloral to form oxolactonic derivative. Carboxylic acid group of L-lysine was protected by esterification with benzyl alcohol. After heating the mixture at 80°C with Pd/charcoal as a catalyst, the protecting groups in the polymerized products were removed, forming the required polyamides with a yield of 85%. In another report by Jikei's group (Jikei et al., 1999), condensing agents like triphenyl phosphite and pyridine assisted in the polymerization of aromatic amines and trimesic acid to form hyperbranched polyamides. Kroner and coworkers (Kroner et al., 1997) have described a method of amide-containing compound synthesis either under dry conditions or in aqueous solutions at temperatures ranging from 80°- 260°C The starting materials included polybasic acids such as citric acid, tartaric acid, aconitic acid, alone or mixed, and a variety of amines, including ammonia, amino acids, polyamines, alone or mixed. Kroner suggested that the procedures might yield amide products even at low temperatures. Can the simple procedure described by Kroner and coworkers applied to the synthesis of polyamides? Do the resulting polyamides possess protein-like properties? I intend to explore these questions in my thesis.

2. Materials and Method:

2.1 Materials

The monomers for polyester synthesis such as glycerol (G9012 Sigma), citric acid (791725 Sigma-Aldrich), DL-lactic Acid (W261114 Sigma-Aldrich), as well as commercial polyesters such as hyperbranched bis-MPA polyester–64–hydroxyl, generation 4 (686573 Aldrich) were used as received. The amino acids, including glycine (410225 Sigma-Aldrich), L-alanine (A7627 Sigma), L-aspartic acid (A9256 Sigma), L-glutamic acid (G1251 Sigma) and L-lysine (L5501 Aldrich) were used as received. HPLC solvents and additives such as 2-propanol (99.7%) and formic acid (98.0%) were purchased from Wako, Japan, and used as received. Barnstead™ Smart2Pure™ (Thermo Fisher Scientific, Waltham MA, USA) purified water was used in all experiments and HPLC analysis.

2.2.1 Hyperbranched Depsipeptide Synthesis

a) Polymerization: The procedure for polymer synthesis has been adapted from (Forsythe et al., 2015). In a typical polymerization reaction, 1mL of an aqueous solution containing polyesters and amino acids was allowed to evaporate at 85°C. The concentrations of substrates used in the reactions are mentioned in the Results and Discussion section. In some cases, rewetting protocols were implemented (see below). The mole ratio of reagents in the mixture used for polymerization is shown in Table 2.1.

b) Continuous and Intermittent Drying Cycles: For continuous drying, the aqueous samples were incubated, uncovered in 1.5mL Eppendorf tubes at 85°C on a heating block. Each 24-hour period constituted a single drying cycle. In the case of intermittent drying, the starting solutions were allowed to dry at 85°C for 24h, followed by rehydration and incubation at 85°C for 24h with the cap closed. The consecutive dehydration and rehydration of mixture for 24h each complete one intermittent drying cycle. Incubations of a maximum of eight cycles were performed for both continuous and intermittent drying cycles. Further, NMR, HPLC, and mass spectrometry analysis of the samples were carried out at the end of the 1st, 2nd, 4th, and 8th cycle.

Sr. No	Reagent A	Reagent B	Reagent C	Mole Ratio
1.	CA	Gly	-	1:2
2.	CA	Gly	A / G / A+G	1:2:1
3.	CA	-	A / G	1:1
4.	-	Gly	A / G	2:1
5.	CA	-	Lys	1:1
6.	CA	Gly	G, A, D, and V	1:2:1
7.	LA	-	G, A, D, and V	1:1
8.	Bis – MPA	-	V	1:1
9.	Bis – MPA	-	G, A, D, and V	1:1
10.	TA	-	Lys	1:1
11.	MS	-	Lys	1:1
12.	CA	-	S.TCl	1:1

Table 2.1: Reagents in different mole ratio for the polymerization series.

2.2.2 NMR Analysis

^{1}H and ^{13}C Nuclear Magnetic Resonance (NMR) 1D spectra were recorded on Bruker AVANCE III 400 spectrometer (Bruker Corporation, Billerica MA, USA) at 30°C. The acquisition and processing of the spectra were performed using the Bruker Topshim 4.0.7 software package. HSQC (Heteronuclear Single Quantum Coherence) and HMBC (Heteronuclear Multiple Bond Correlation) experiments were performed to confirm peak assignments. HSQC shows the correlation between carbon atoms and protons directly attached to them. HMBC correlates carbon and protons three or more bonds away. Prior to NMR analyses, all water-soluble samples subjected to deuterium exchange by repeated (3x) dissolution in 200µL D_2O and drying in the SAVANT SPD131DDA SpeedVac Concentrator (Thermo Fisher Scientific, Waltham MA, USA). The samples containing 2,2-bis (hydroxymethyl) propionic acid (bis-MPA) were dissolved in deuterated dimethyl sulfoxide (DMSO) without a prior deuterium exchange.

| Sr. No.	Command	Purpose
1.	lift	Remove the previous tube and insert a new sample
2.	edc	Create folder and name for storing sample spectra
3.	lock solvent	Lock ^{2}D signal to compensate for resonance frequency drift
4.	topshim	Automatic shimming command to correct magnetic field inhomogeneities
5.	atma	Automatically tuning of the nucleus
6.	rga	Automatically set the receiver gain value
7.	zg or multizg	Start acquisition (single or multiple experiments)
8.	efp or xfb	Process 1D or 2D data including Fourier transform, line broadening, linear prediction and weight functions
9.	apk	Automatic correct
10.	abs	Automatic baseline correct

Table 2.2: Steps of commands to perform an NMR experiment

Spectra were collected at a 30° inversion pulse at 10s acquisition time, and 1s recycle delay for ^{1}H. For ^{13}C-NMR, a 30° inversion pulse was applied to samples. HSQC and HMBC NMR were performed using a Bruker pulse program, *hsqcetgpsi2* and *hmbcgplpndqf* with a relaxation delay set to 1.5s. Table 2.2 shows the set of commands used in sequential order to perform an NMR analysis. The spectra were processed and visualized. The chemical shifts of the likely reaction products were predicted using the ^{1}H and ^{13}C NMR predictor package of the Spectrus Processor software (ACD Labs, Toronto ON, Canada).

2.2.3 Mass Spectrometry Analysis

Matrix-Assisted Laser Desorption Ionization (MALDI) is an ionization technique that uses a laser energy absorbing matrix to create ions from large molecules with minimal fragmentation. MALDI-MS spectra were collected on an ultrafleXtreme Bruker Daltonics MALDI-TOF-MS (Bruker Corporation, Billerica MA, USA) in a positive ion mode.

External mass calibration was conducted using standard peptide mixtures. Samples for the MALDI-MS analysis were prepared using either α-Cyano-4-hydroxycinnamic acid (CHCA) or SDHB matrices. Super-DHB consists of a 9:1(w/w) mixture of 2,5-Dihydroxybenzoic acid and 2-hydroxy-5-methoxybenzoic acid, respectively. Aqueous solutions of the polymers and matrix molecules at excess were deposited on a MALDI plate and allowed to dry. The samples containing bis-MPA were dissolved in hot water (65°C).
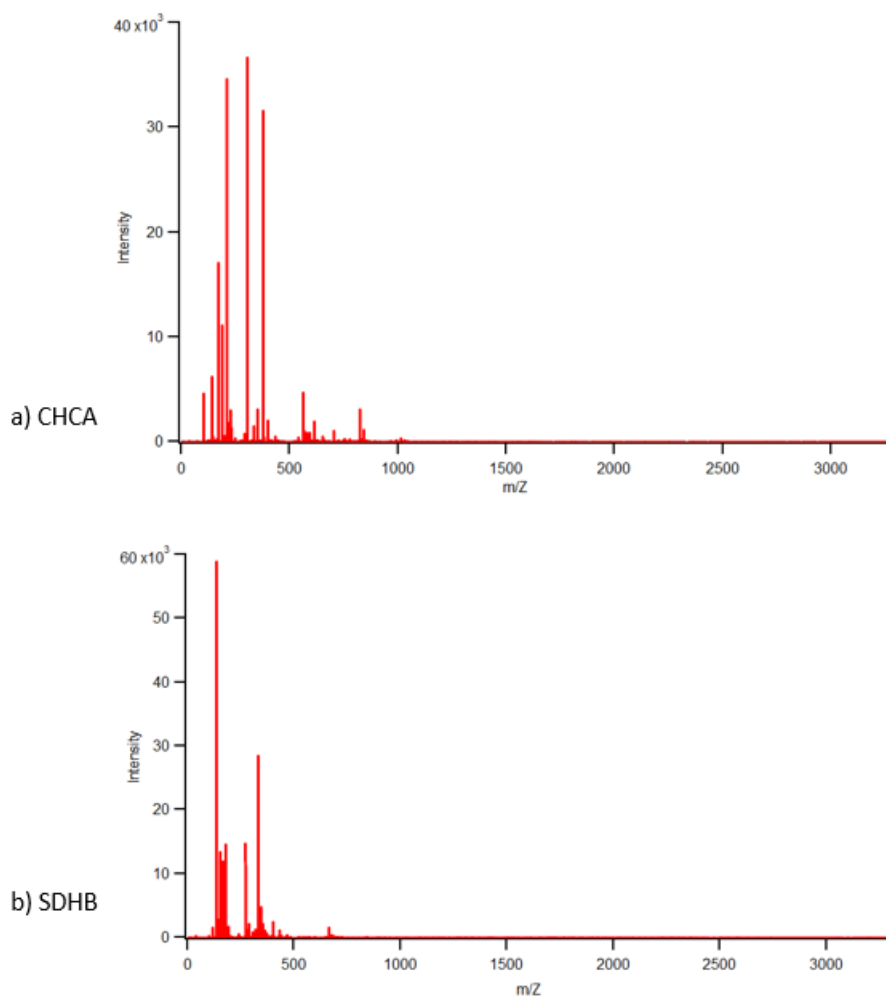
Fig 2.1: MALDI MS spectra of matrices used for analysis. a) CHCA and b) SDHB.

MALDI-MS data analysis for the depsipeptide products involved four steps:

1. *Creating a list of theoretically plausible products.* Ester-amide exchange involving four different amino acids, carboxylic acid, and alcohol monomers results in products, in which the monomers are bonded together in numerous combinations. To find the combinations of possible products, we wrote programming codes using python, as shown below.

// code to determine theoretical combinations of product for Glyc, CA, G, A, D, and V mixture

```
def truncate(n, decimals=0):
    multiplier = 10 ** decimals
    return int(n * multiplier) / multiplier

GLY = 92.09382
CA = 192.124
G = 75.07
A = 89.09
D = 133.11
V = 117.151

gly = 1
ca = 1
g = 0
a = 0
d = 0
v = 0

print ('CA ', 'Gly ', 'G ','A  ','D  ','V  ', 'M+ ')
 for ca in range(10):
  for gly in range(21):
   for g in range(21):
    for a in range(21):
     for d in range(21):
      for v in range(21):

        i = GLY*gly + CA*ca + G*g + A*a + D*d + V*v
        j = gly + ca + g + a + d + v - 1
        Molwt = i - 18.01528 * j
        if (gly + g + a + d + v) < (2*ca + 2) and gly > 0 and ca < (2*gly + 1) and i < 2600:
          print (ca,'  ',gly ,'  ',g,' ',a,'  ',d,'  ',v,'  ', truncate(Molwt , 2))
// end
```

2. *Mass spec data processing.* For this purpose, we used ACDLabs Spectrus Processor. Firstly, the mass spectra were centroided. Then, a threshold was applied to remove the baseline noise. The processed spectrum was exported into *.csv* format and inputted into the python code for analysis.

3. *Theoretical and experimental data comparison.* We wrote a python code that accesses files in *.csv* format. The code outputs values whose difference is less than or equal to a given threshold, as shown below.

// code to compare theoretical and experimental values

```
import csv
with open ('Exp.csv') as f1:
    csv_f1 = csv.reader(f1)
    a = list(f1)
with open('Theo M+23.csv') as f2:
    csv_f2 = csv.reader(f2)
    b = list(f2)
a = list(map(float, a))
b = list(map(float, b))
print('Theo.', 'Exp.')
for x in range(len(a)):
    for y in range(len(b)):
        if a[x] == b[y]:
            print (a[x], b[y])
        elif ((a[x] - b[y]) <0.5) and ((a[x] - b[y]) > 0):
            print (b[y], a[x])
```
//end

4. *Further comparative analyses.* We wrote additional programs to determine the total number of amino acids (T) found in each identified species and the calculated number of specific amino acids (N) present in all the detected mass peaks. Then, we used the median and its two adjacent values of 'T' for qualitative analysis of the composition of amino acids in the peaks. Also, the values of 'N' were used to analyze trends of different amino acids incorporation in polyesters.

//to perform qualitative analysis for determining the makeup of amino acids

```
import csv
import numpy as np
with open ('M+ stats.csv') as f1:
    csv_f1 = csv.reader(f1)
    z = list(csv_f1)
    z = np.array(z)
    z = z[:,:8]
    z = z.astype(np.float32)
row = len(z[:,1])
col = len(z[1,:])
for l in range(col):
    i=0
    for k in range(row):
        if z[k,l] == 0:
            i += 1
    if l == 3:
        print ('G0 =',i)
    if l == 4:
        print ('A0 =',i)
    if l == 5:
        print ('D0 =',i)
    if l == 6:
        print ('V0 =',i)
```
//end

2.2.4 Size – Exclusion Chromatography SEC Analysis

The analysis was performed on an Acquity Advanced Polymer Chromatography (APC) System (Waters Corporation, Milford MA, USA) equipped with an Acquity APC AQ column with the following specifications: pore size: 125Å, particle size: 2.5μm, inner diameter: 4.6 mm, length: 150mm. The temperature of the column compartment was set to 35°C; the refractive index (RI) detector was set to 40˚C. The RI detector operates by comparing the refractive index of the running sample and control (only solvent).

Before the analysis, the system was equilibrated by a mobile phase flush at a flow rate of 0.1mL/min overnight. The Photodiode array (PDA) detector lamp was turned on just before the experiment. PDA is a sensor that collects information about the entire spectrum in the UV-Vis wavelength range. In this case, it was set to measure spectra at 220nm and 310nm. The solutions (600mM, 250 μL) were diluted in a 1:1 volume ratio

with 0.3% formic acid solution to get the final volume of 500µL. The samples were transferred from Eppendorf tube to Centrifugal Filter Units tube (Millipore Sigma, Billerica MA, USA) equipped with a polyvinylidene fluoride (PVDF) membrane of 0.22µm. The samples were filtered using AS ONE (Osaka, Japan) mini-centrifuge at 5000 rpm. Finally, the analyte solutions were transferred to HPLC vials. An isocratic flow rate of 0.500mL/min of 0.15%formic acid in deionized water was maintained during the analysis. Injection volume and running time were set at 10µL and 10min, respectively. After every four runs, an injection of pure mobile phase was performed. The elution spectra were recorded by the RI and PDA detector.

2.2.5 Dialysis Protocol

Dialysis of the polymer was performed to remove unreacted monomers and low-molecular-weight products. Float-A-Lyzer G2d devices (Repligen, Waltham MA, USA) with molecular weight cut-offs (MWCO) of 100D-500D and 500–1000D were used. Before use, the device was soaked in deionized water for 30mins thrice to remove the glycerol introduced during the manufacturing process to reduce cracking during transportation and storage. Then, the sample was added to a dialysis tube, capped, and equipped with a Styrofoam floating device and place into a beaker containing 1L of deionized water and a magnetic stirrer. The water in the beaker was changed twice, after 1h and 12h. After the second change, the sample was dialyzed for additional 4 hours. The sample was then transferred to an Eppendorf tube and dried using SpeedVac. The Eppendorf tube was weighed before sample addition and following the drying to establish the weight of the resulting sample.

2.2.6 Kemp Elimination Assay

We prepared a 3mM stock solution of 5-nitro-1, 2-benzisoxazole in ethanol. The citric acid and spermidine trihydrochloride polymers were dialyzed, dried, and redissolved in water to get 5mM concentration. The reaction volume for the assay was 1mL. The aqueous solution containing 983uL 200mM Trizma buffer (pH 7.4) and 16.16uL 3mM kemp substrate was mixed and measured on Agilent Technology UV –Vis spectrometer (Agilent Cary 8454, Malaysia) as control. Since the presence of citric acid will affect the

pH of the solution after dialysis, the dried polymers were dissolved in 200mM Trizma buffer, and pH was adjusted to 7.4 with 1M NaOH. 983μL of polymer in buffer solution was mixed with 16.16μL of the kemp substrate in a quartz cuvette, at which time the time-lapse UV-vis measurement commenced. Spectra between 200nm to 1100nm were collected every 60 min for 20h. The maximum absorbance of the Kemp reaction product is observed to be at 380nm. The literature value for the molar extinction of p-nitrophenol $(18.1mM^{-1}cm^{-1})$ was used in the concentration determinations of the product, 2-cyano-4-nitrophenol (Mamajanov and Cody, 2017). The data was fitted as a pseudo-first-order reaction. The rate constant was derived from the slope of the function of ln (concentration) vs. time.

2.2.7 Catalyzed Hydrolysis Assays

a) Esterase Assay

The substrate used for the assay was p-Nitrophenyl acetate (p-NPA). We prepared a 50mM stock of p-NPA in ethanol. The citric acid and glycerol polymers were dialyzed and dissolved in 400mM HEPES buffer. HEPES stands for (4-(2-hydroxymethyl)-1-piperazine ethane sulfonic acids) and is one of the Good's buffer. The pH of the polymer solution was adjusted to 7.4 using 1M NaOH. The reaction volume for the assay was 3mL containing 2880μL of 400mM HEPES buffer, 60μL 2M polymers and 60μL 50mM. We ran blanked and then started time-based measurement for 16h. The spectra were collected from 200nm to 1100nm, and absorbance at 405nm was used for analysis.

b) Phosphatase Assay

We used p-Nitrophenyl phosphate (p-NPP) as a substrate for the assay. The reactions were run in 400mM HEPES buffer at room temperature, pH 7.4. The polymers were dialyzed, and the required concentration was prepared using the buffer. The reaction volume was 1mL containing 778μL of 514.13mM polymer solution and 222μL of 4.5mM p-NPP. The spectra were collected from 200nm to 1100nm, and absorbance at 405nm was used for analysis.

3. Results and Discussion:

### 3.1 Formation of hyperbranched polyesters in wet-dry cycling conditions:

Previous studies in the lab have shown the formation of hyperbranched polyesters at 85°C under both wet-dry and continuous drying conditions (Mamajanov et al., 2015). As a preliminary control experiment, we have repeated the synthesis using the above-developed protocol. Briefly, we subjected 0.33M citric acid (CA) and 0.66M glycerol (Glyc) to wet-dry conditions up to 8 cycles at 85°C. The resulting products were characterized by various analytical methods such as SEC, 1D and 2D NMR, and MALDI mass spectrometry. Figure 3.1 shows ^{1}H-NMR of 8th wet-dry cycles samples of CA and Glyc. The broadening of peaks and new peaks around 4.0 and 5.0 ppm shows the formation of polyester. Figure 3.2 depicts SEC data of polymeric samples of Glyc and CA (Glyc$_2$CA). Glyc monomer lacks a chromophore and therefore is not detected in the UV trace. Citric acid co-elutes with polymers presumably because of a hydration shell around it in aqueous solution. Due to this coelution of CA with the polymers, it is difficult to draw conclusions from the SEC analysis.
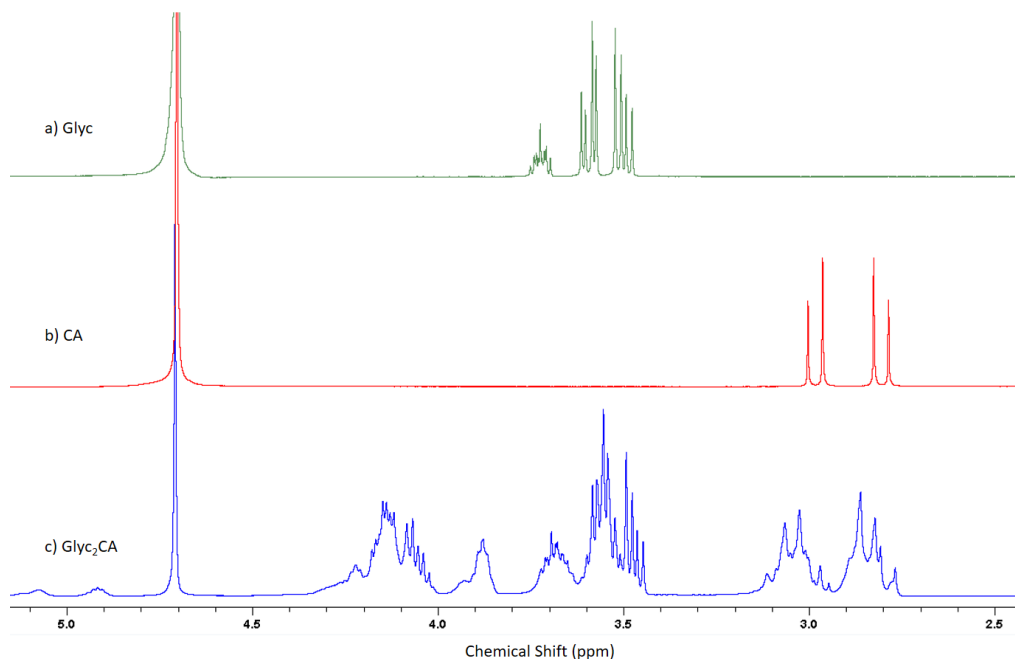
Figure 3.1: ^{1}H-NMR spectra to monitor the formation of hyperbranched polyesters from CA and Glyc mixture.
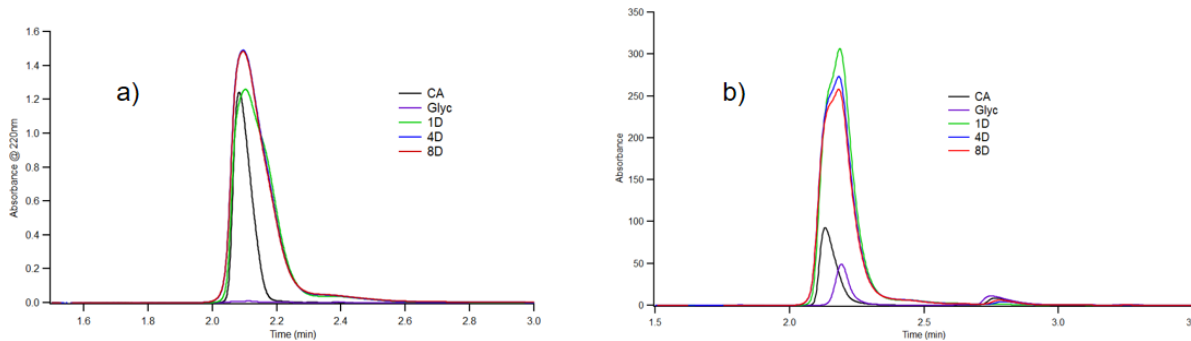
Figure 3.2: SEC data of CA and Glyc polymerization product. The chromatograms were detected by a) UV detector, 220nm trace, b) RI detector. The 1D, 4D and 8D mark the analyses of the samples that have undergone 1, 4 and 8 wet-dry cycles, respectively.
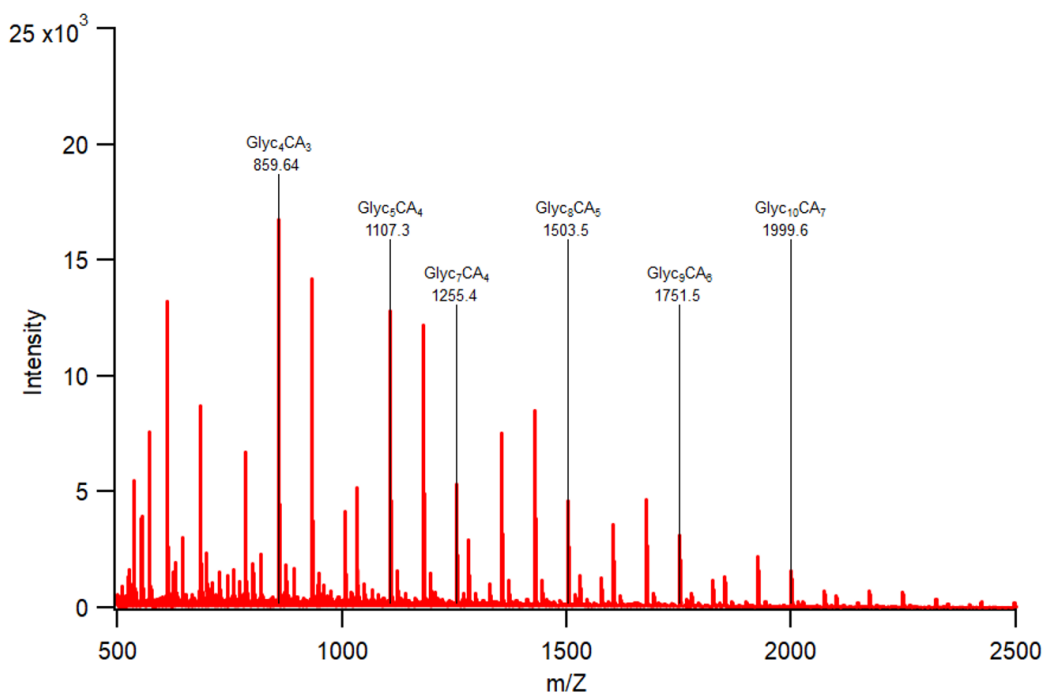

Figure. 3.3: MALDI MS data of 8 wet-dry cycles of CA and Glyc. The detected mass peaks correspond to polyesters formation between CA and Glyc.

MS data shows the formation of polymers up to 17-mers such as Gly_9CA_6, $Gly_{10}CA_7$, etc. (Figure 3.3). The polyester synthesis showed that we could use the same protocol and analytical methods for future experiments to characterize hyperbranched depsipeptides.

3.2 Hyperbranched depsipeptide formation between Glyc$_2$CA and amino acids under intermittent or continuous drying:

3.2.1 Thermal polymerization of CA, Glyc and alanine (A):

In this experiment, the goal was to introduce amino acids to hyperbranched polyesters and investigate its incorporation in the polymers to form hyperbranched depsipeptides. Towards this end, we used the citric CA, Glyc and A in 1:2:1 mole ratio (0.33M CA, 0.66M Glyc and 0.33M A). As control systems, we have also prepared solutions omitting either CA or Glyc. Then, we subjected the solutions to intermittent and continuous drying at 85°C for up to 4 cycles. The dried samples were analyzed using SEC and NMR (Figure 3.4 and Figure 3.5). In SEC data, the monomer CA co-elutes with the polymers. Due to this, we draw unequivocal conclusions from the SEC data.
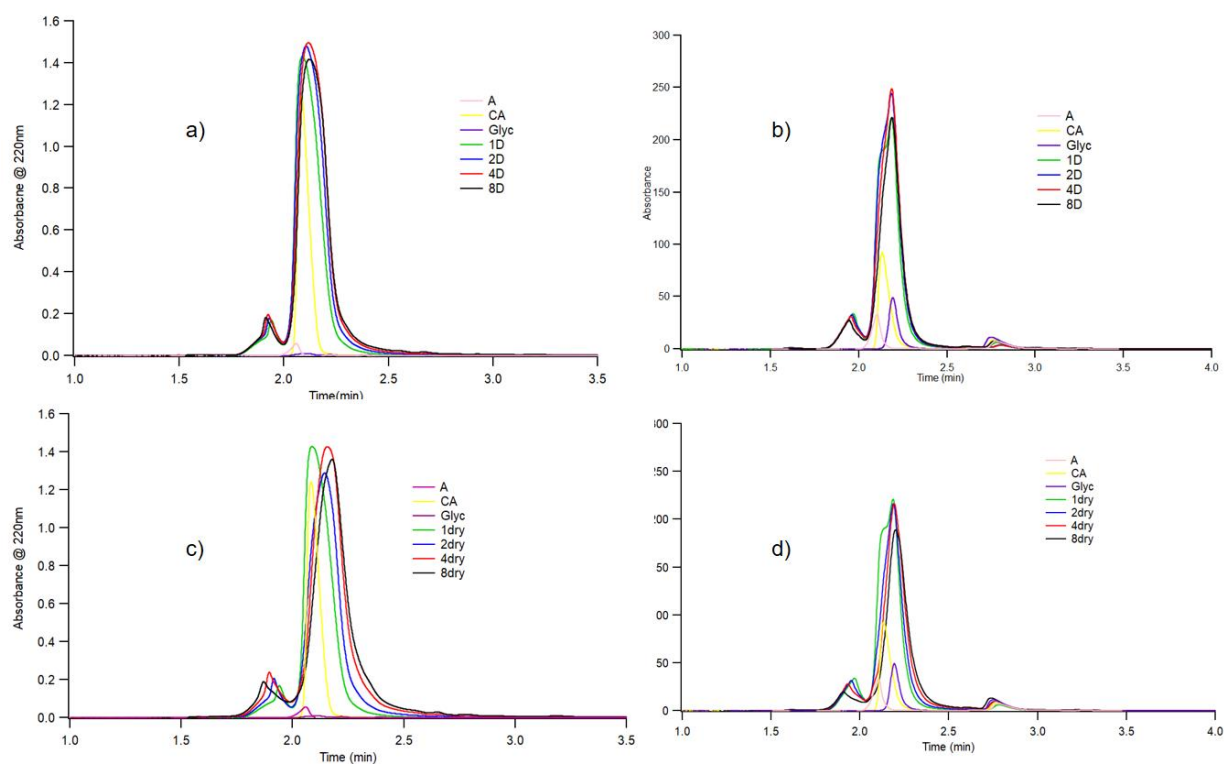

Figure 3.4: Size-exclusion chromatograms of the polymers formed upon reaction between CA, Glyc and A and the products of control experiments. a) UV and b) RI-detected chromatograms of the sample prepared under wet-dry conditions. c) UV and d) RI-detected chromatograms of the samples subjected to continuous drying. D and dry indicate intermittent and continuous drying, respectively.
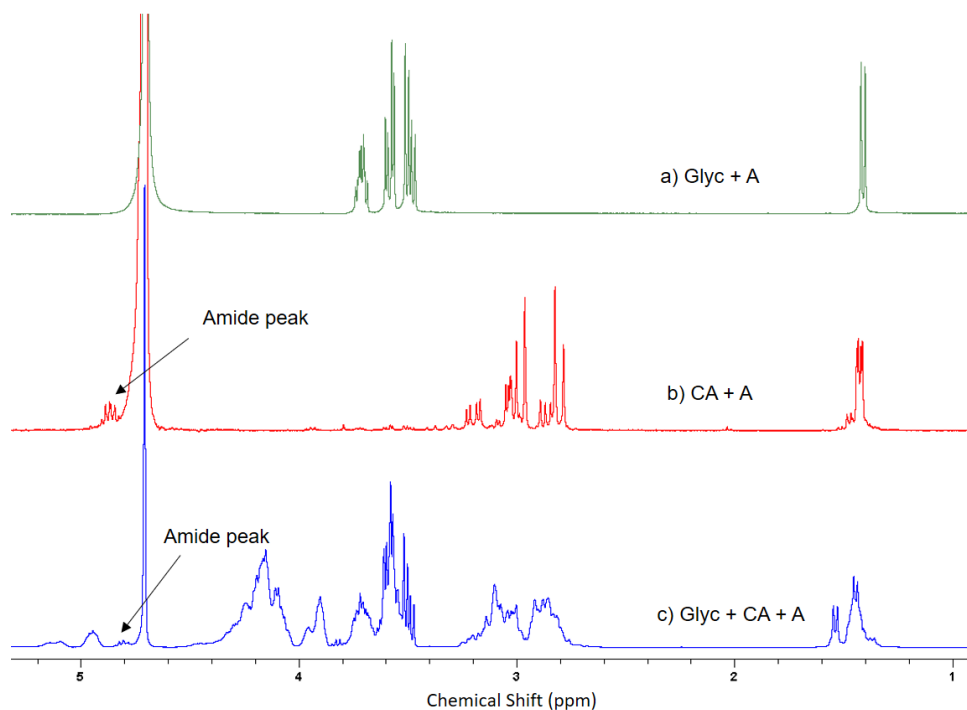
Figure 3.5: ^{1}H-NMR spectra of 4th wet-dry cycled sample of Glyc, CA and A.

The chemical shift values for the amide bond in the possible depsipeptide products were predicted using ACDLabs software mentioned in the Methods section. The chemical shift of α-H in amino acids is around ~3.8ppm. However, when the amino acid condenses to form the amide bond, the chemical shift of α-H changes to ~4.5ppm. The ^{1}H-NMR spectrum of the dried samples shows multiple peaks around 4.5ppm consistent with the formation of multiple amide bonds. The ^{1}H-NMR spectrum of Glyc and A sample suggests that no reaction has occurred. The ^{1}H-NMR spectrum of the CA and A sample, however, features a peak consistent with an amide proton. Further analysis is summarized in Results Section 2. To confirm the formation of hyperbranched depsipeptides, we analyzed the sample with MALDI MS.
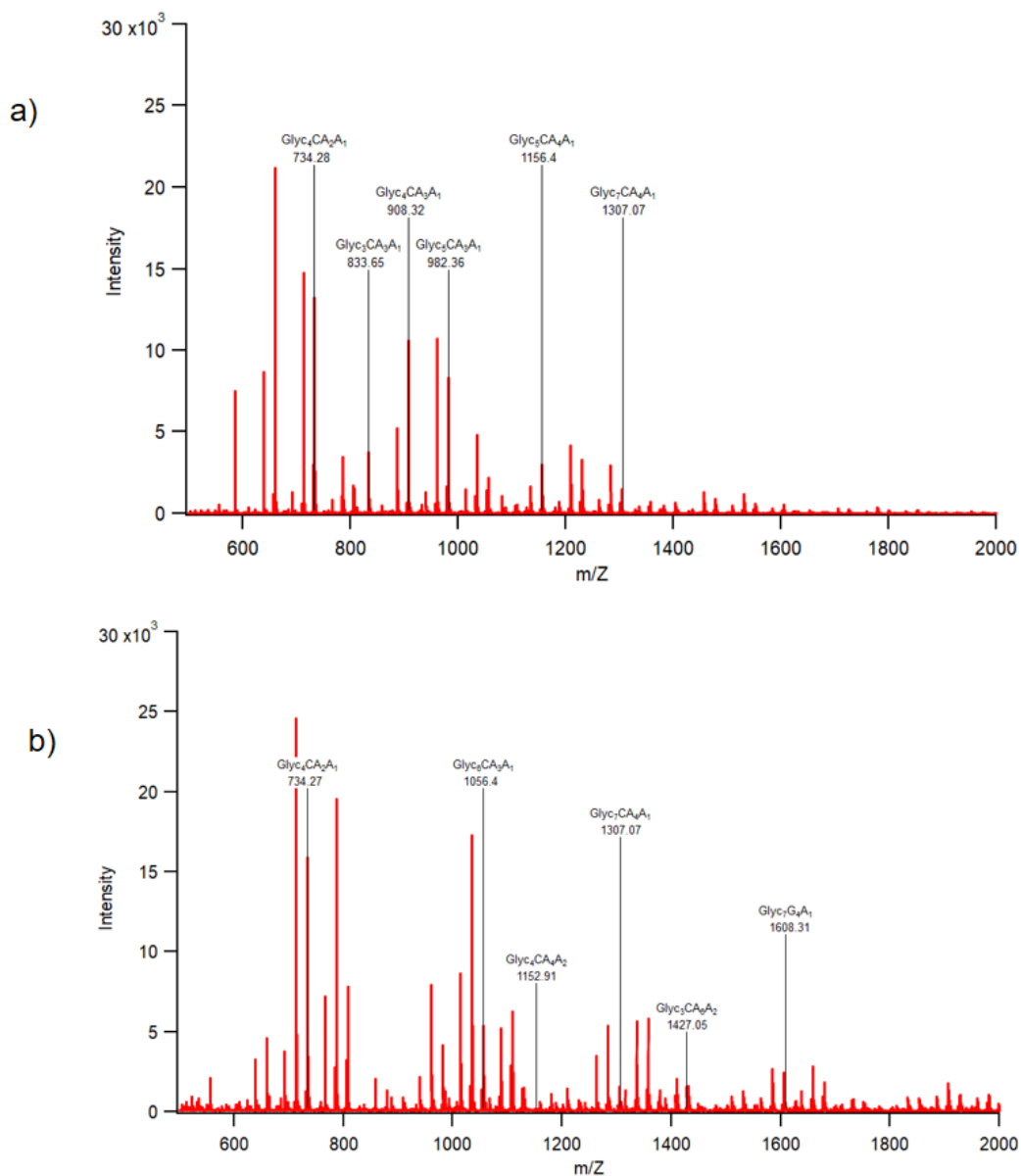
Figure 3.6: MALDI MS spectra of Glyc, CA and A. Both a) and b) represent intermittent and continuous drying conditions respectively for polymerizations.

The MS data indicated that both the intermittent and continuous drying methods produced polymers containing all three CA, Glyc, and A monomers. The intermittent drying samples show the formation of oligomers up to 11-mers, including species with a molecular formula of $Gly_4CA_5A_1$ and $Gly_5CA_6A_1$. In the case of continuous drying, the polymerization happens to a greater extent yielding heavier polymers due to less hydrolysis. The continuous drying produced oligomers up to 22-mers. Examples of

some of them include $Gly_7CA_9A_6$ and $Gly_6CA_9A_6$. The NMR and MS data confirmed that A could react with CA and Glyc under continuous or intermittent drying to form hyperbranched depsipeptides.

3.2.2 Thermal polymerization of CA, Glyc, A and G (glycine):

From the previous experiment, we found that A, when introduced to polyesters, it forms depsipeptides. To check whether the incorporation is not specific to A and to hone our analytical techniques toward multiple component polymers, we tested the depsipeptide formation protocols on a mixture of A and G. A solution of 0.25M G, 0.25M A, 0.25M CA, and 0.5M Glyc was subjected to intermittent drying. The products of the reaction were analyzed by NMR. The broad signals in the ^1H NMR spectrum (Figure 3.7) are consistent with polymer formation; however, the amide signals were difficult to find because of the strong water and methylene of CA peaks in the same region of the spectrum. To check the incorporation of both amino acids, we performed the MALDI MS. The MS spectrum shows multiple detected mass peaks of polymers up to 11-mers containing all four CA, Glyc, A and G, as shown in Figure 3.8. The results showed that the depsipeptide formation in our systems are not limited to A inclusion, and more than one amino acid could be incorporated in the same reaction setup.
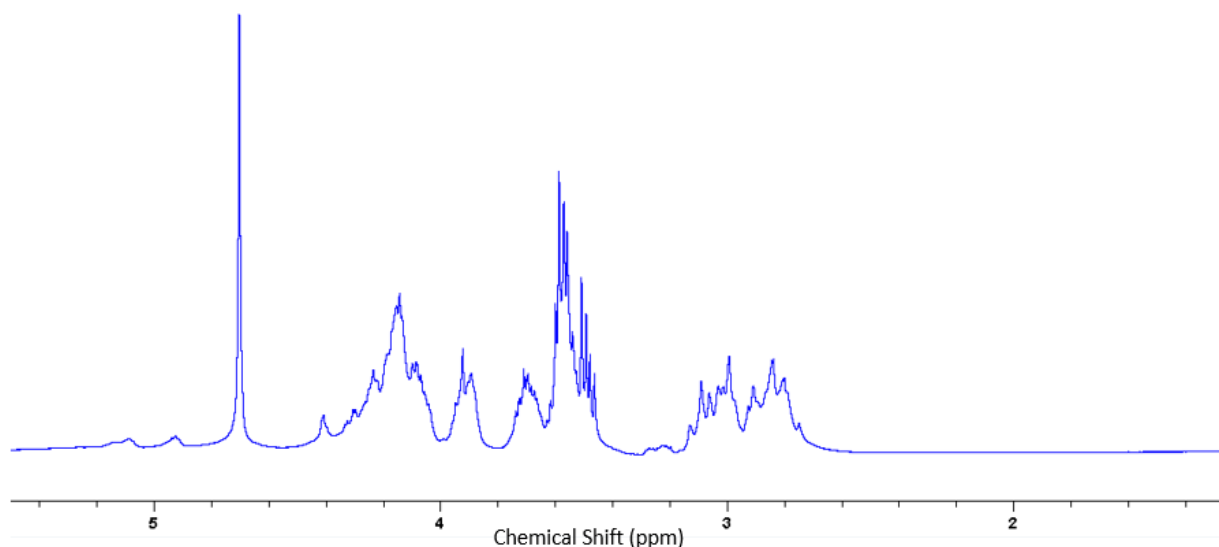
Figure 3.7: ^1H-NMR spectra of Glyc, CA, G and A containing depsipeptides formed after 4 intermittent drying cycles.
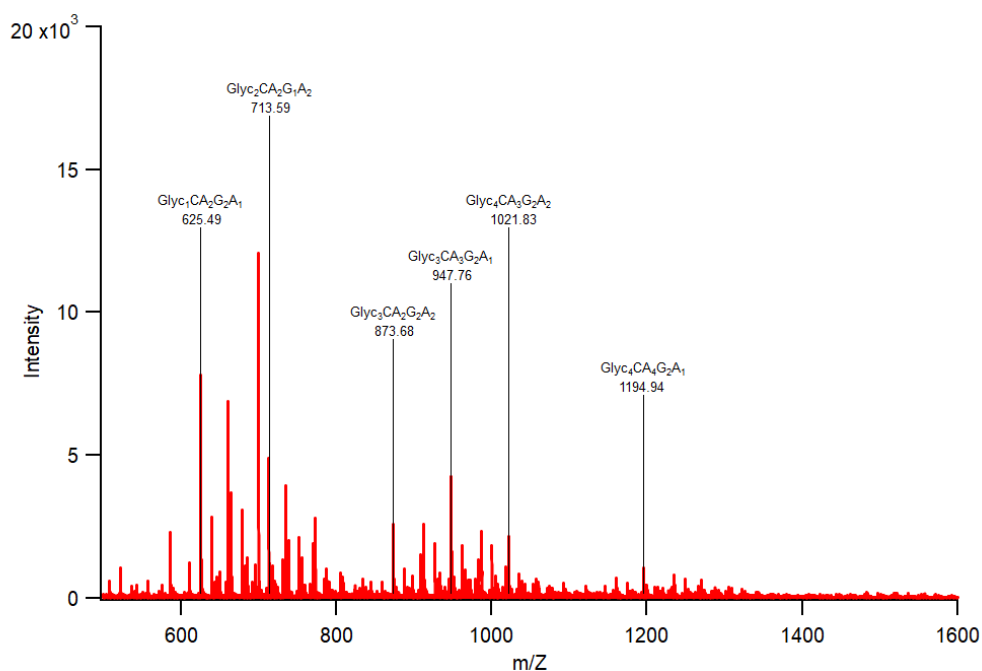
Figure 3.8: Mass spectrum of Glyc, CA, G and A depsipeptides formed after 4 intermittent drying cycles.

3.2.3 Depsipeptide formation upon thermal polymerization of CA, Glyc with G, A, aspartic acid (D), and valine (V):

Previous results show that CA and Glyc when subjected to intermittent or continuous drying with amino acids like A and G, form hyperbranched depsipeptides. The next aim of the project was to investigate whether the incorporation of amino acids into hyperbranched polyesters is dependent on the amino acid chemical properties, such as hydrophobicity, size, and identity of the side chain. To this end, we have selected to study the mixture of G, A, D, and V. The selection of these four amino acids was inspired by the hypothesis of K. Ikehara proposed that the ancestral proteins could have only used this minimal set of amino acids. Indeed, the G, A, D, and V set represents hydrophobic and charged amino acids of varying sizes (Ikehara, 2005). The mixtures containing 0.2M Glyc, 0.1M CA, 0.025M amino acids mixture (G, A, D and V each) were subjected to continuous and intermittent drying up to 8 cycles. The dried samples were characterized by ^{1}H-NMR, HSQC, and HMBC. ^{1}H-NMR analysis indicated the presence of multiple peaks around 4.5ppm on the ^{1}H-axis, which can be attributed to amide bond

formation between the different amino acids or CA or to the methylene group of CA (Figure 3.9). To assign the signals, HSQC NMR was performed. The chemical shift values of the amide moieties were compared to the predicted values using ACDLabs software, as shown in Table 3.1. In the control experiment, in which only the amino acids, G, A, D, and V were subjected to 8 intermittent drying cycles, no reaction occurs. In the presence of CA and Glyc, HSQC shows the insertion of all four amino acids into the reaction product (Figure 3.10).
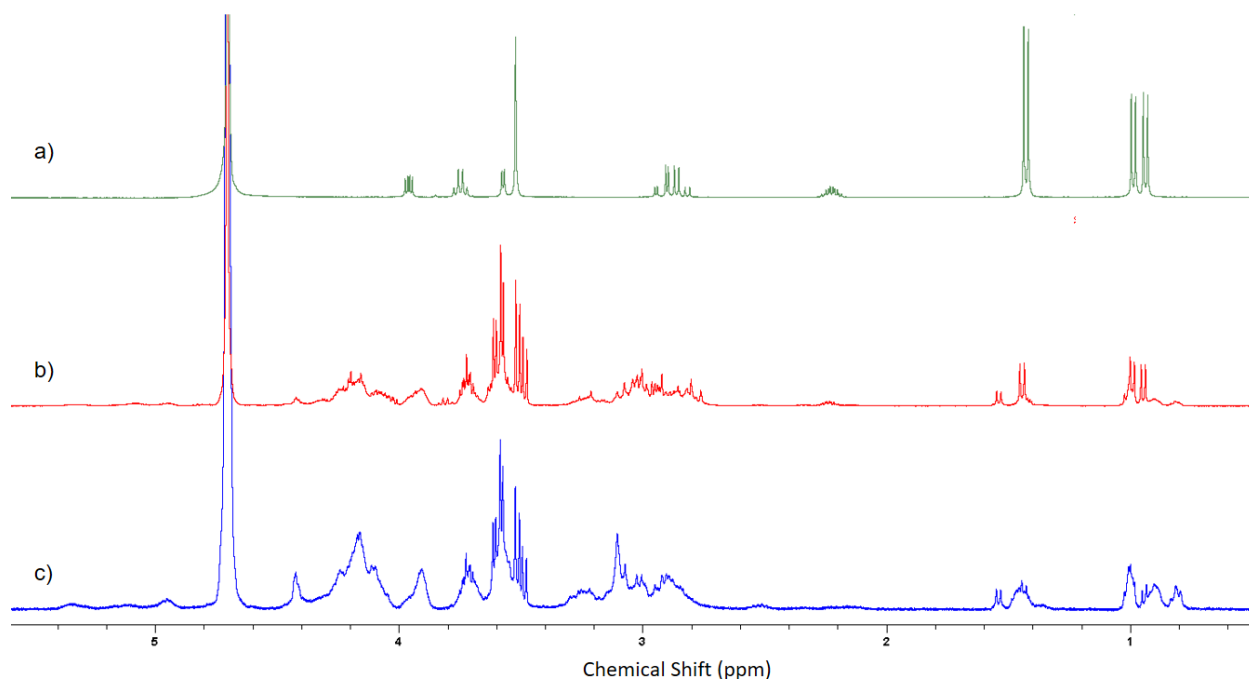

Figure 3.9: ^{1}H-NMR of the polymers (and controls) formed upon intermittent and continuous drying of Glyc, CA and the mixture of the amino acids (G, A, D, and V). a) G, A, D, and V mixture dried alone. b) and intermittent and c) continuous dried samples.

To gain a detailed view of the types of polymers produced, we performed the MALDI MS analysis (Figure 3.12). The MALDI MS spectra show a vast number of peaks indicating the formation of diverse products. A cursory analysis of most abundant masses revealed species of the following molecular formula, $CA_6Gly_4G_3A_1D_2V_1$, and $CA_7Gly_5G_1A_2D_1V_1$, indicating incorporation of all four amino acids into polymeric products.
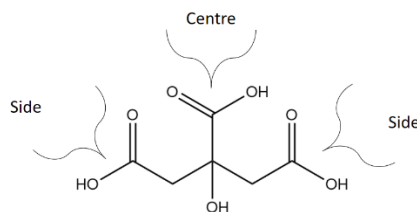
32
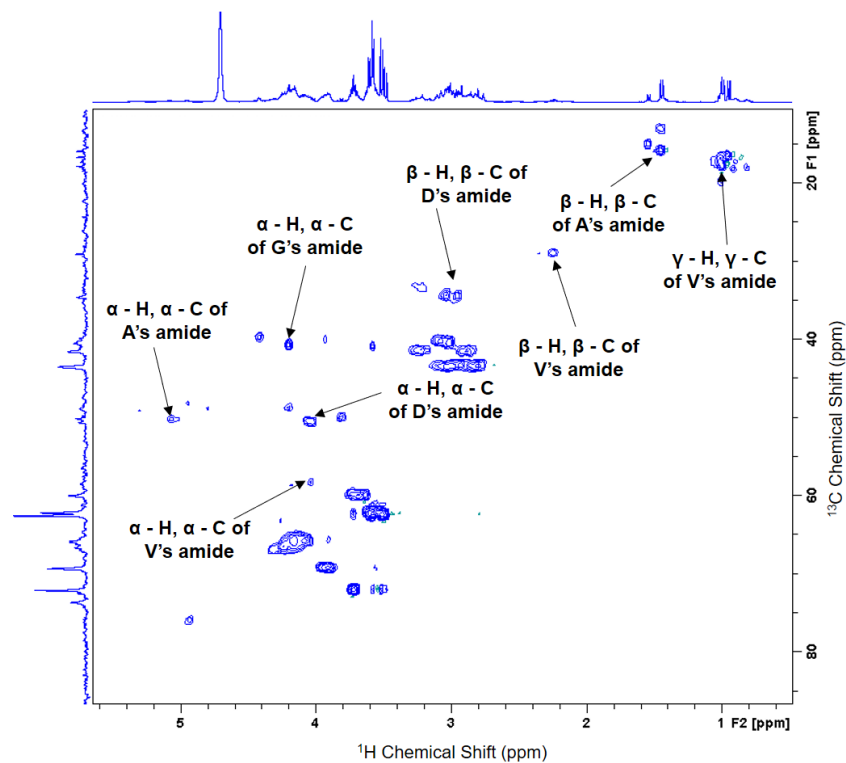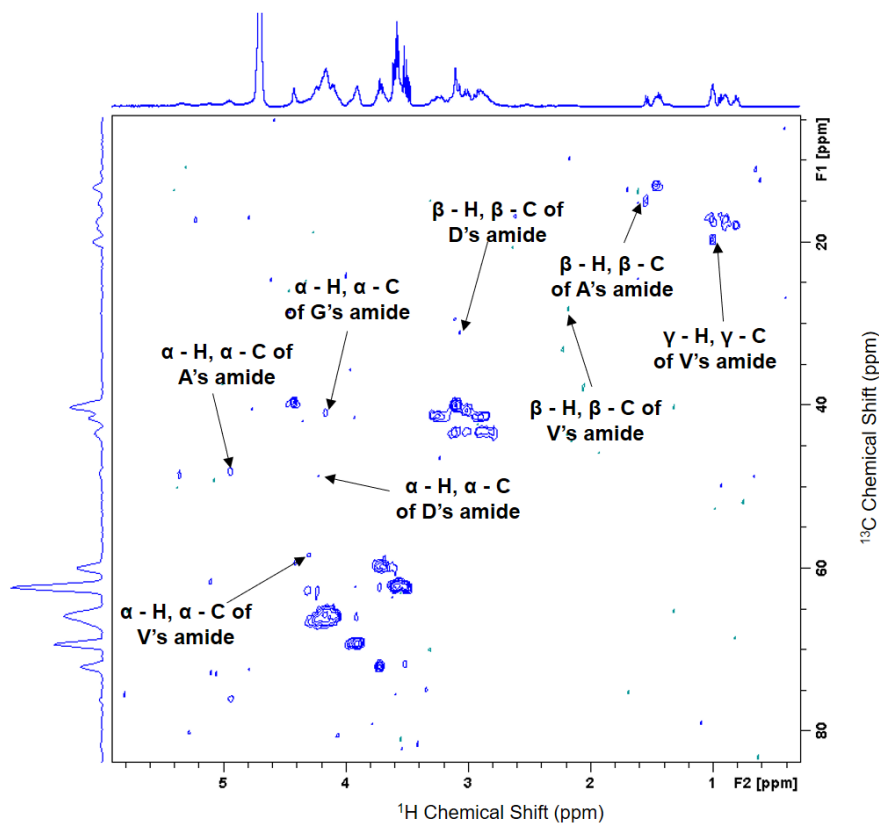
Figure 3.10: HSQC NMR of Glyc, CA, G, A, D, and V depsipeptides collected after 8 cycles of a) intermittent and b) continuous drying.

Figure 3.11: CA structure showing two different types of carboxylic acid: Side and Centre. Side represent the two equivalent carboxylic groups and Centre represents the 3rd carboxylic group of CA.

Amide moiety	Species	Theoretical shift (ppm)		Experimental shift (ppm)	
		Side	Centre	Wet-dry	Cont. dry
G	α - H	3.93	4.04	4.2	4.1
	α – C	44.08	42.05	40.68	41.0
A	α - H	4.75	4.72	5.06	4.94
	α - C	49.32	50.02	50.0	48.3
	β - H	1.41	1.43	1.44	1.54
	β - C	16.24	15.97	16.05	14.9
D	α - H	4.51	4.52	4.20	4.2
	α - C	51.25	51.95	48.99	49.1
	β - H	2.96	2.77	3.0	3.07
	β - C	35.51	35.90	34.5	31.6
V	α - H	4.40	4.55	4.09	4.4
	α - C	59.17	59.86	58.51	59.1
	β - H	2.17	2.28	2.23	2.18
	β - C	30.73	30.4	29.0	28.4
	γ - H	0.88	1.00	1.03	1.02
	γ - C	17.78	18.43	17.5	17.2

Table 3.1: Theoretical and experimental ^1H and ^{13}C chemical shifts of depsipeptides formed between G, A, D and V and CA. "Side" represent the two equivalent carboxylic acid groups of CA, while "centre" represents the third group, as shown in Figure 3.11.
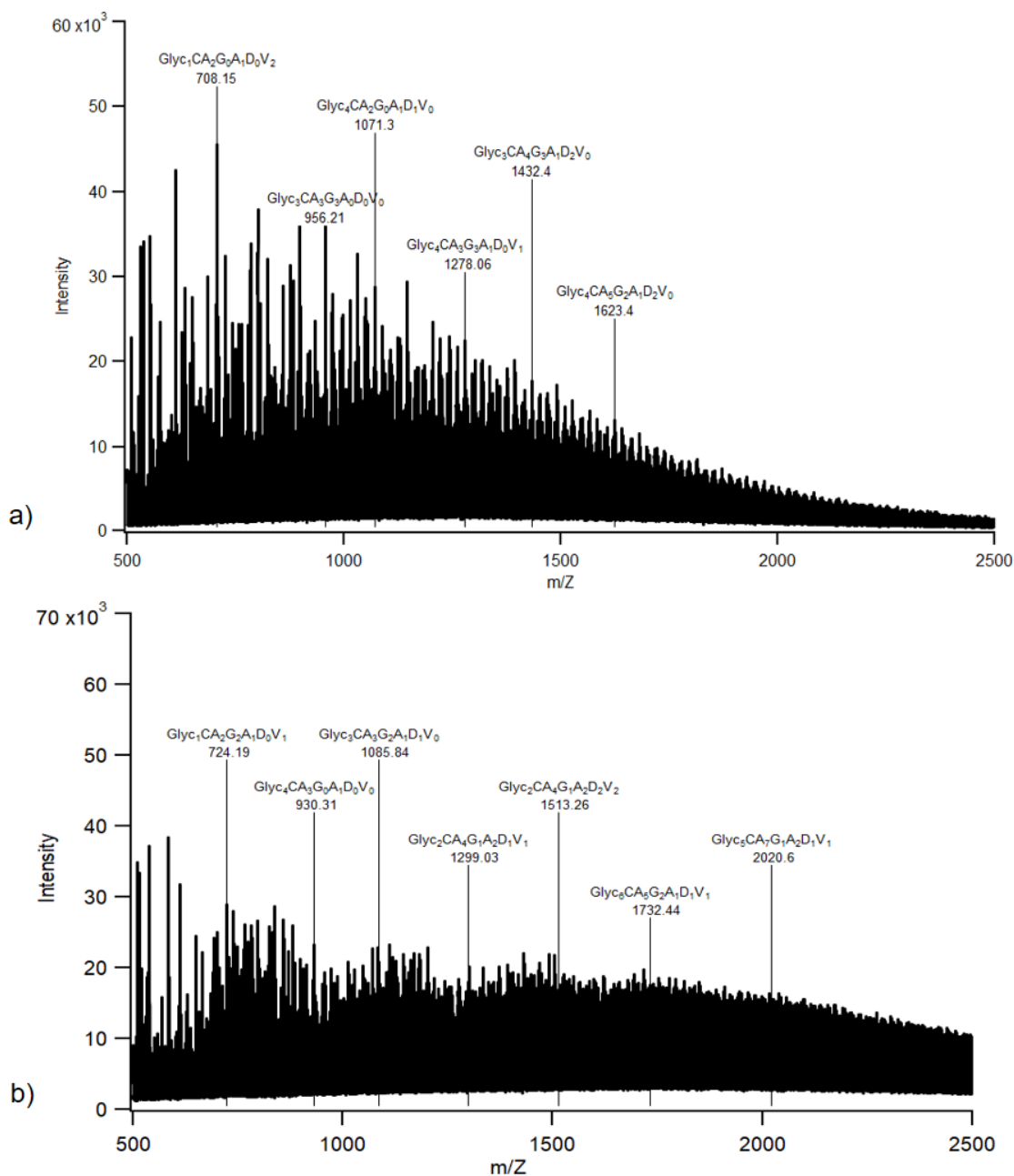
Figure 3.12: MALDI MS spectra of polymers formed upon a) intermittent and b) continuous drying of Glyc, CA, G, A, D, and V mixture for 8 cycles.
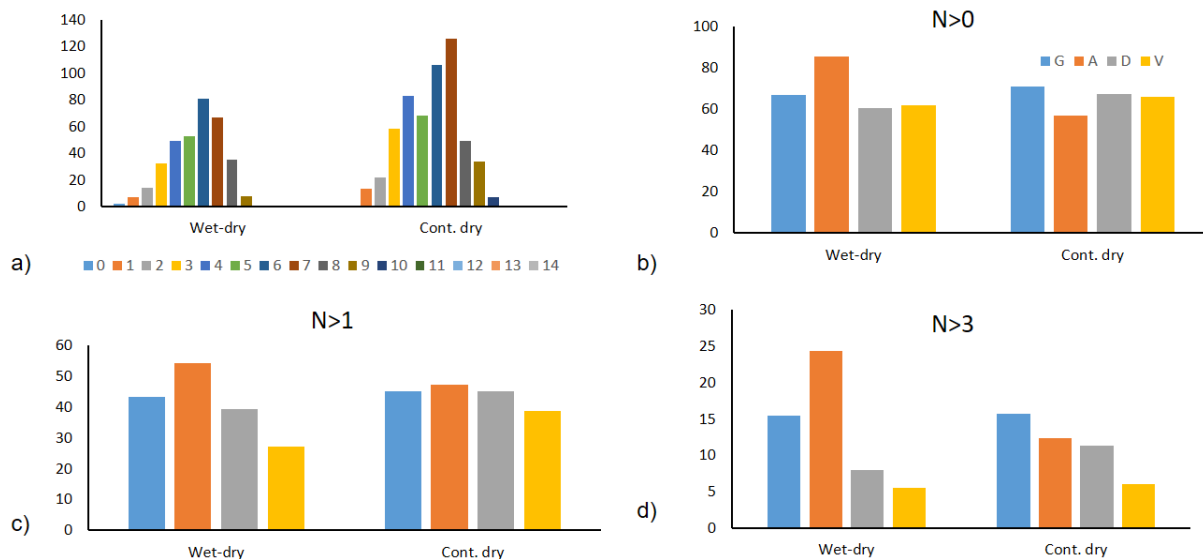
Figure 3.13: Qualitative product makeup analysis of CA, Glyc and G, A, D, and V depsipeptides. a) The number of mass peaks attributed to depsipeptides vs. the number of incorporated amino acids in each product. The number of mass peaks attributed to depsipeptides vs. the number of incorporated specific amino acids are shown in the next graphs. Incorporation of more than 0, 1 and 3 of the same acid are shown in b), c), and d), respectively.

We wanted to see whether G, A, D, and V can incorporate amino acids non-randomly upon amide-ester exchange. So, we performed a qualitative analysis of the MS data to check selective incorporation of different types of amino acids in polyesters. The protocol for the analysis is mentioned in the Materials and Methods Section. MS analysis shows that the total number of amino acids incorporated in the polymers varies from zero to ten (Figure 3.13a). The lower number for incorporation is probably due to the formation of the amide bond between CA and G, A, D, V mixture alone. The analysis of the median values for the number of same amino acids incorporation (N) indicates that for N>0 and N>1, all amino acids are incorporated relatively evenly (Figure 3.13b and 3.13c). We hypothesize that since the CA and Glyc polyesters are very reactive, the released monomers through ester-amide keep reacting to esters and engaging in ester-amide and scrambling the incorporation of the amino acid resulting in varying incorporations. Also, the MS analysis shows that as the N value increases, the

incorporation of V decreases compared to the other three amino acid monomers (Figure 3.13d). We hypothesize that the outer core of polyester is equally accessible to all four amino acids. However, the incorporation in the inner core is restricted due to the steric hindrance of amino acid side chains. Since V has the bulkiest side group among G, A, D, and V, it is difficult for V to reach the inner core of polyester, which results in its least incorporation. Similarly, G and A are more selected because of their smaller side-chain groups. The analysis also indicates that bulkier D is selected for less than G and A.

3.2.4 Thermal polymerization of Glyc$_2$CA polyesters with G, A, D, and V:

G, A, D, and V, when added in a mixture of Glyc and CA, forms hyperbranched depsipeptides upon intermittent and continuous drying. To check the extent of amino acid incorporation in already synthesized hyperbranched polyesters, we first prepared polyesters using 0.25M CA and 0.5M Glyc. The Glyc$_2$CA polyesters were synthesized by intermittent and continuous drying up to four cycles, and amino acids were added in afterwards (0.025M of G, A, D and V each). The mixture of polyesters and amino acids was subjected to 8 intermittent drying cycles and analyzed using NMR and MALDI. ^{1}H-NMR of cycled samples revealed similar results as the mixture of Glyc, CA, G, A, D and V.
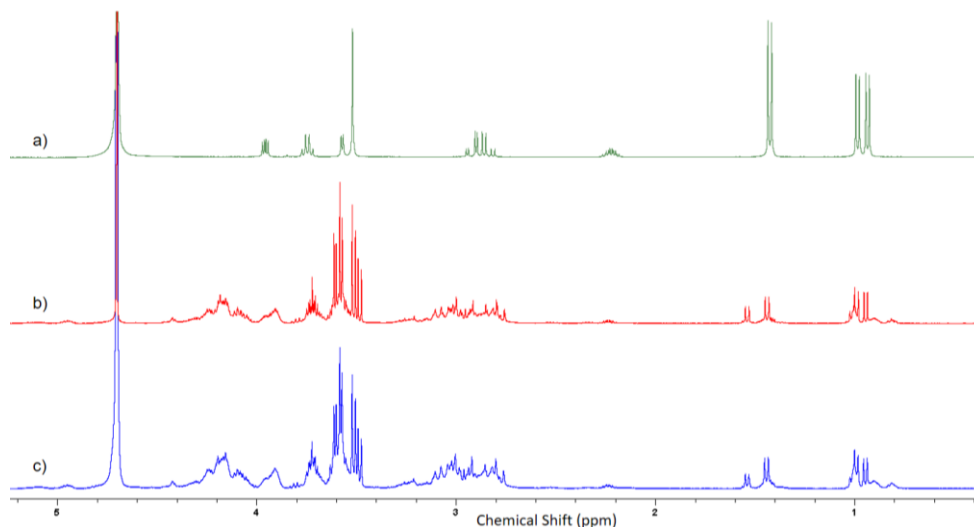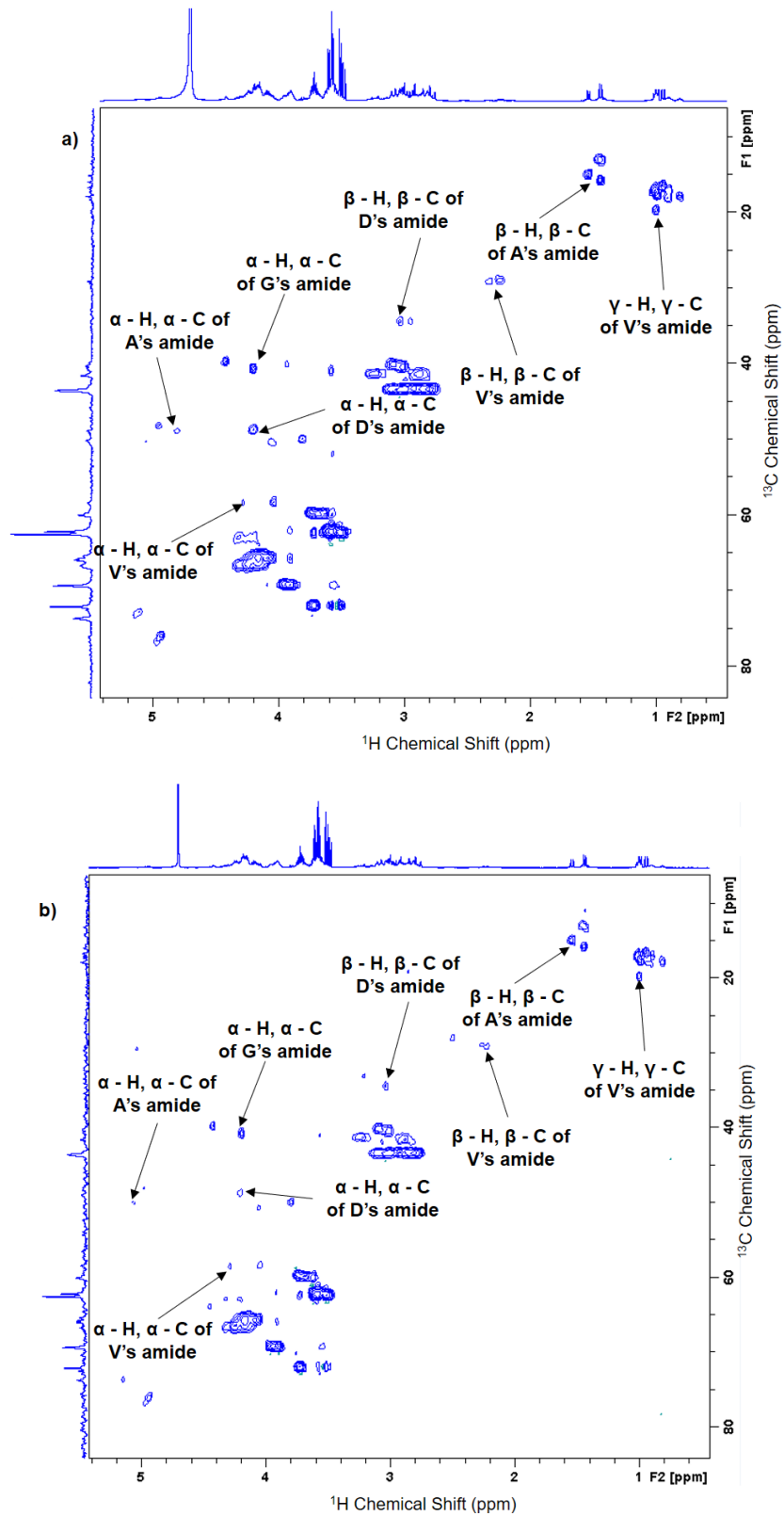

Figure 3.14: ^{1}H-NMR spectrum of Glyc$_2$CA polyesters, G, A, D, and V mixture. a) G, A, D, and V, b) and c) intermittently and continuously dried samples, respectively.

The HSQC NMR (Figure 3.15) along with ACD predictions (Table 3.2) helped us to assign peaks that correlate to amide bonds formed between $Glyc_2CA$ polyesters and all four amino acids suggesting that all four amino acids were incorporated into the depsipeptide. The mass spectra of the continuously and intermittently dried samples again show a vast diversity of depsipeptide products formed (Figure 3.16). Similar to Glyc, CA, G, A, D and V depsipeptides, the direct condensation of CA and amino acids to form amide bond is possible in this case as well and the reaction will not go through the ester-amide exchange. The detailed study of the CA and the amino acids reaction is shown in Results and Discussion section 3.6.

Amide moiety	Species	Theoretical shift (ppm)		Experimental shift (ppm)	
		Side	Centre	Wet-dry	Cont. dry
G	α - H	3.93	4.04	4.19	4.2
	α – C	44.08	42.05	40.7	41.0
A	α - H	4.75	4.72	4.95	4.9
	α - C	49.32	50.02	48.2	48.2
	β - H	1.41	1.43	1.43	1.45
	β - C	16.24	15.97	16.05	15.8
D	α - H	4.51	4.52	4.19	4.20
	α - C	51.25	51.95	48.7	48.6
	β - H	2.96	2.77	3.0	3.04
	β - C	35.51	35.90	34.7	34.5
V	α - H	4.40	4.55	4.28	4.3
	α - C	59.17	59.86	58.4	58.6
	β - H	2.17	2.28	2.23	2.26
	β - C	30.73	30.4	29.3	29.0
	γ - H	0.88	1.00	1.01	1.0
	γ - C	17.78	18.43	17.9	19.8

Table 3.2: Theoretical and experimental chemical shifts associated with the depsipeptides formed upon reaction between $Glyc_2CA$ and G, A, D, and V mixture. "Side" represent the two equivalent carboxylic acid groups of CA, while "centre" represents the third carboxylic group, as shown in Figure 3.11.
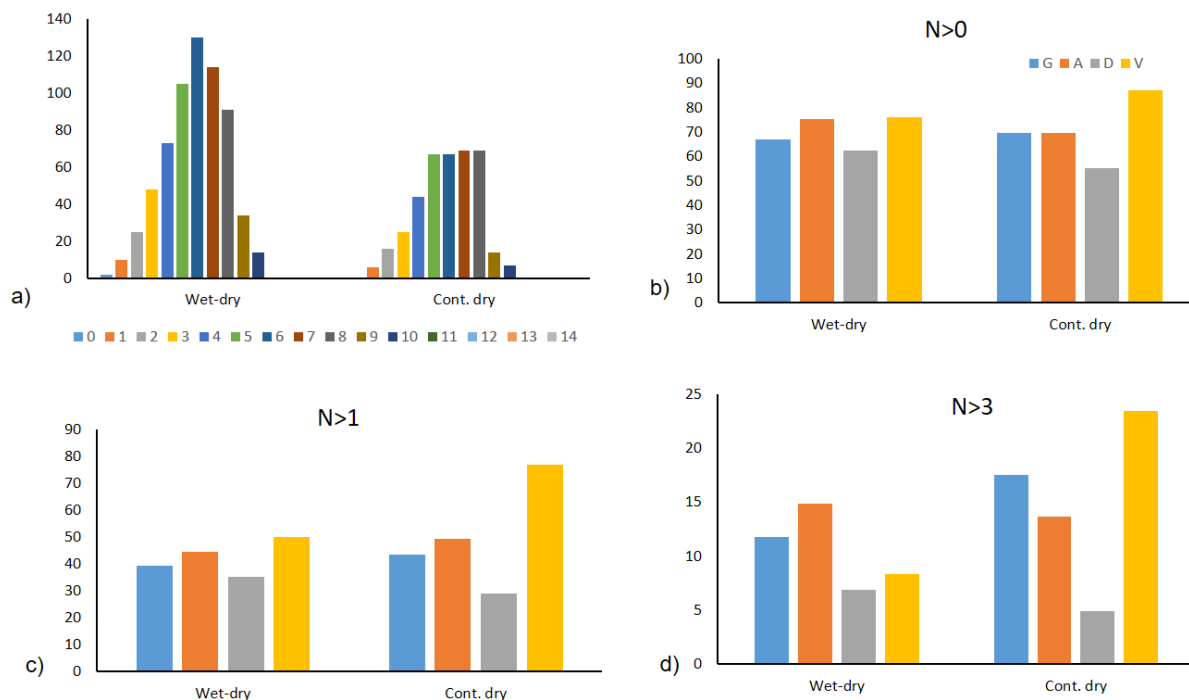
Figure 3.15: HSQC NMR spectra of the depsipeptides formed on a) intermittent and b) continuous drying of Glyc$_2$CA polyesters, G, A, D, and V subjected to for 8 cycles.
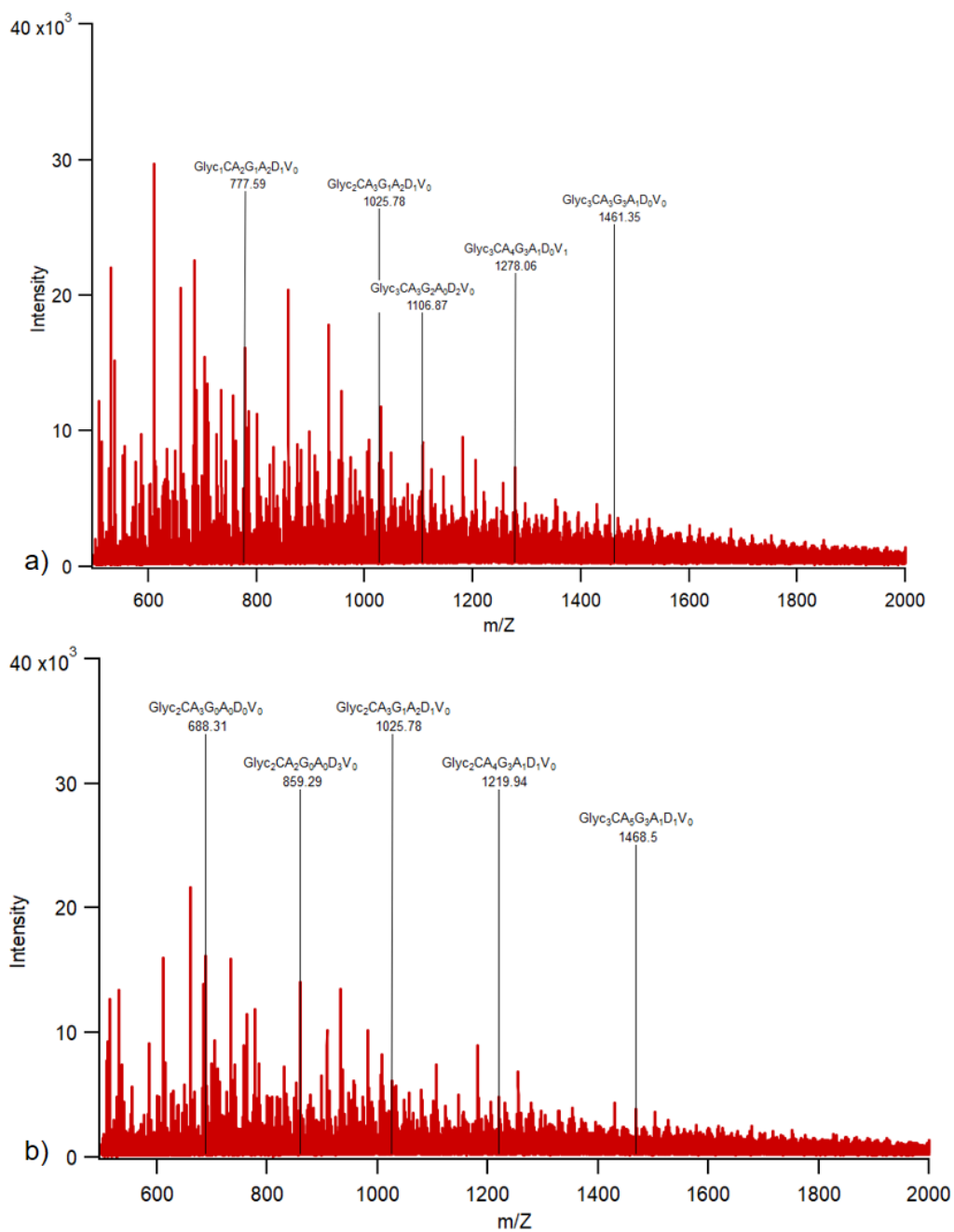
Figure 3.16: MALDI MS spectra of a) wet-dried and b) continuously dried sample of $Glyc_2CA$ polyesters, G, A, D, and V depsipeptides collected after 8 cycles.

Figure 3.17: Qualitative analysis of Glyc$_2$CA, G, A, D, and V depsipeptides. a) The number of mass peaks attributed to depsipeptides vs. the number of incorporated amino acids in each product. The number of mass peaks attributed to depsipeptides vs. the number of incorporated specific amino acids are shown in the next graphs. Incorporation of more than 0, 1 and 3 of the same acid are shown in b), c), and d), respectively.

The MS analysis (Figure 3.17) has detected depsipeptide species containing up to ten amino acids similar to the CA, Glyc, G, A, D, and V system. The analysis has also shown similar behavior for the repeated incorporation of the same amino acid into given depsipeptide species. The numbers for single and multiple incorporations are evenly distributed. The only outliers are the Glyc$_2$CA, G, A, D, and V depsipeptides prepared under continuous drying. These series of depsipeptides exhibit increased affinity for V. Further studies need to be conducted to understand this behavior.

3.3 Hyperbranched depsipeptide formation upon intermittent and continuous drying of bis-MPA polymers and amino acids:

The previous chapter discussed the formation and the makeup of hyperbranched depsipeptides upon reaction between CA and Glyc polyesters and amino acid mixture. $Glyc_2CA$ polyesters are branched and polar. As a next step, we wanted to compare the makeup of depsipeptides formed upon the reaction of the amino acid mixture with hydrophobic branched polyesters. So we used a polymer derived from 2,2-bis(hydroxymethyl) propionic acid (bis-MPA), a commercially available branched polymer that is more hydrophobic. Firstly, we used 20mM bis-MPA polymers and a mixture of G, A, D, and V (5mM each) in a 1:1 molar ratio and subjected it to 8 intermittent and continuous drying cycles. The resulting products were insoluble in room-temperature water, DMSO, acetonitrile and methanol. However, the dried sample was soluble in hot water at 60°C. So, we could not perform NMR analysis and analyzed the samples by MALDI MS alone. Figure 3.18 shows the MS spectra of the bis-MPA, G, A, D, and V mixture. The mass spectra signals, which correspond to depsipeptides formed between bis-MPA, G, A, D, and V shows the incorporation of single or more than one amino acids such as $bis\text{-}MPA_4A_3$ and $bis\text{-}MPA_6G_3A_6$. Also, some peaks represent the incorporation of all four amino acids, for example, $bis\text{-}MPA_4G_5A_1D_2V_1$ and $bis\text{-}MPA_6G_3A_3D_2V_1$.

The comparative MS analysis suggests that the majority of depsipeptides detected contain five or more amino acid moieties (Figure 3.19a). When considering multiple incorporations of specific amino acids, G is the most likely to incorporate more than once, while there are very few species containing more than 3 Vs (Figure 3.19). Also, the selection of the other three amino acids over V is more prominent in this case than that of CA and Glyc polyesters because bis-MPA polymers and monomers appear to be less reactive than CA and Glyc polyesters. The bis-MPA monomers, unlike the CA and Glyc polyesters, are hard to polymerize upon drying at 85°C and requires high temperature and catalyst (Magnusson et al., 2000). When a bis-MPA monomer is released in the process of ester-amide exchange, it is unlikely to form into new

polyesters that, in turn, can undergo further ester-amide exchange. Consequently, the selective incorporation of amino acids in this system is not prone to scrambling.
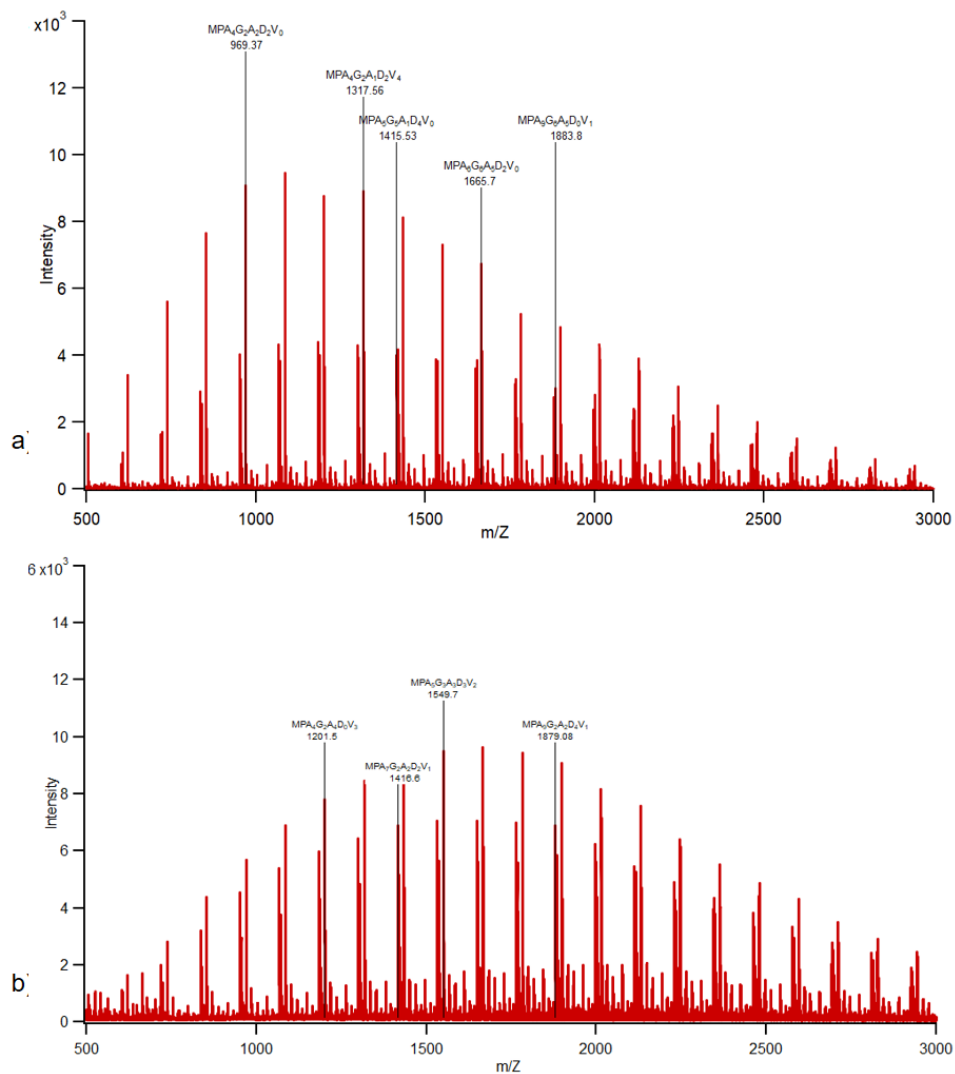
Figure 3.18: MALDI MS data of a) intermittently and b) continuously dried samples of bis-MPA polymer, G, A, D, and V depsipeptides collected after 8 cycles.
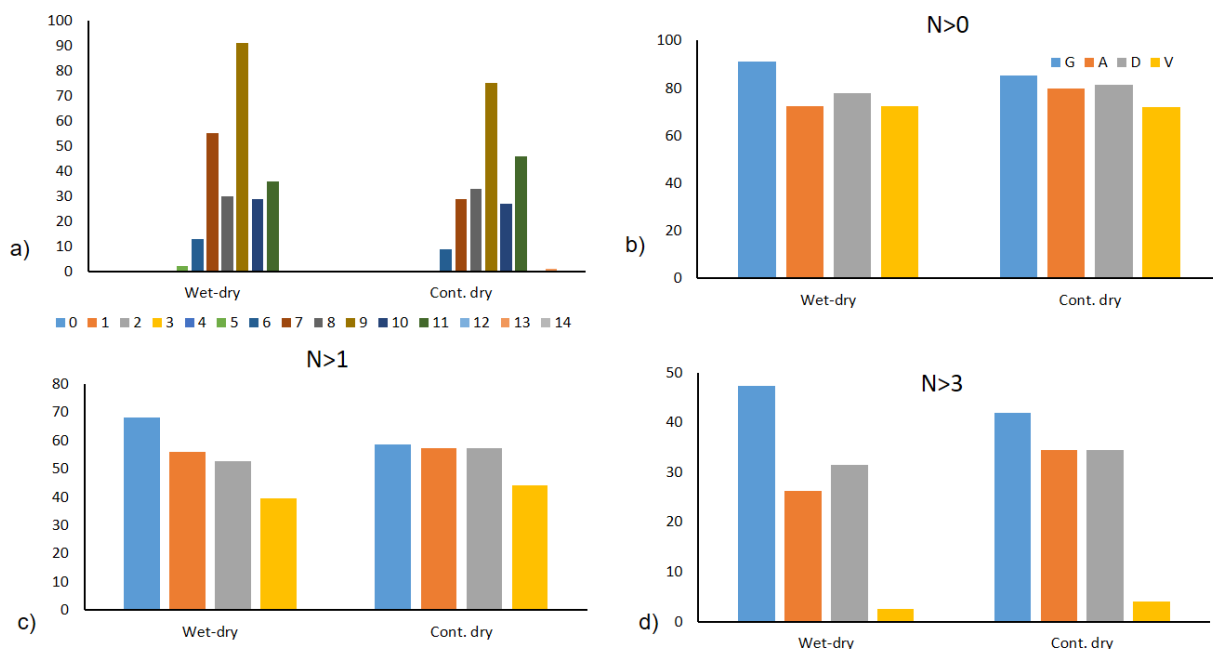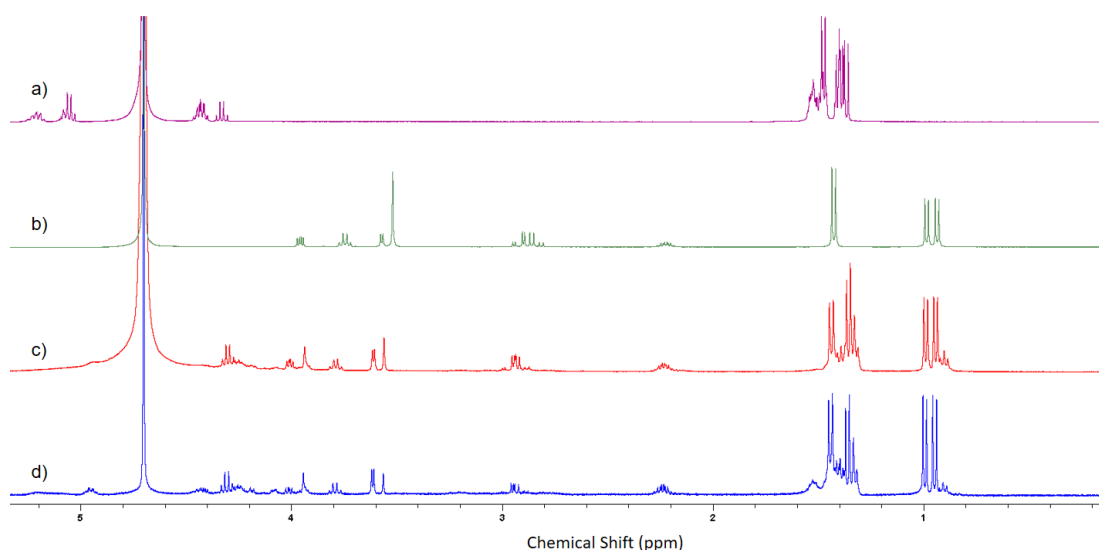
Figure 3.19: Qualitative analysis of bis-MPA, G, A, D, and V depsipeptides. a) The number of mass peaks attributed to depsipeptides vs. the number of incorporated amino acids in each product. The number of mass peaks attributed to depsipeptides vs. the number of incorporated specific amino acids are shown in the next graphs. Incorporation of more than 0, 1 and 3 of the same acid are shown in b), c), and d), respectively.

3.4 Thermal polymerization of lactic acid (LA), G, A, D, and V:

The earlier chapters show the generation of depsipeptides through the ester-amide exchange of branched polyesters. The next aim of the project was to look at the depsipeptides formed upon ester-amide exchange of linear polyesters and amino acids, and compare the makeup of the resulting depsipeptides to those formed using the branched matrices. For this purpose, we used LA, which polymerizes upon heating to yield polylactic acid, a linear polyester. We started the reaction with a mixture of 200mM LA and 50mM of each amino acids (G, A, D, and V). The mixture was subjected to 8 cycles of intermittent and continuous drying and the dried samples were analyzed using NMR and MALDI. In the ^{1}H-NMR spectrum, the chemical shift of α-H in lactic acid overlaps with that of the α-H in an amide bond, so, ^{1}H-NMR cannot be used for analysis

(Figure 3.20). Therefore, we performed the HSQC NMR of the dried samples. The HSQC NMR (Figure 3.21) along with ACD predictions (Table 3.3) helped us to assign peaks that correlate to amide bonds formed between LA and all four amino acids. The HSQC NMR data suggests that all four amino acids were incorporated into the depsipeptides. The NMR peaks shown in Figure 3.21 corresponds to amide bond formation between lactic acid and all four amino acids. The incorporation of all four amino acids in polyesters was also confirmed through MALDI MS. In Figure 3.22, the peaks correlate to depsipeptides formed between LA, G, A, D, and V. The largest detected mass in the spectra, which contains all four amino acids is $LA_4G_4A_1D_2V_2$.
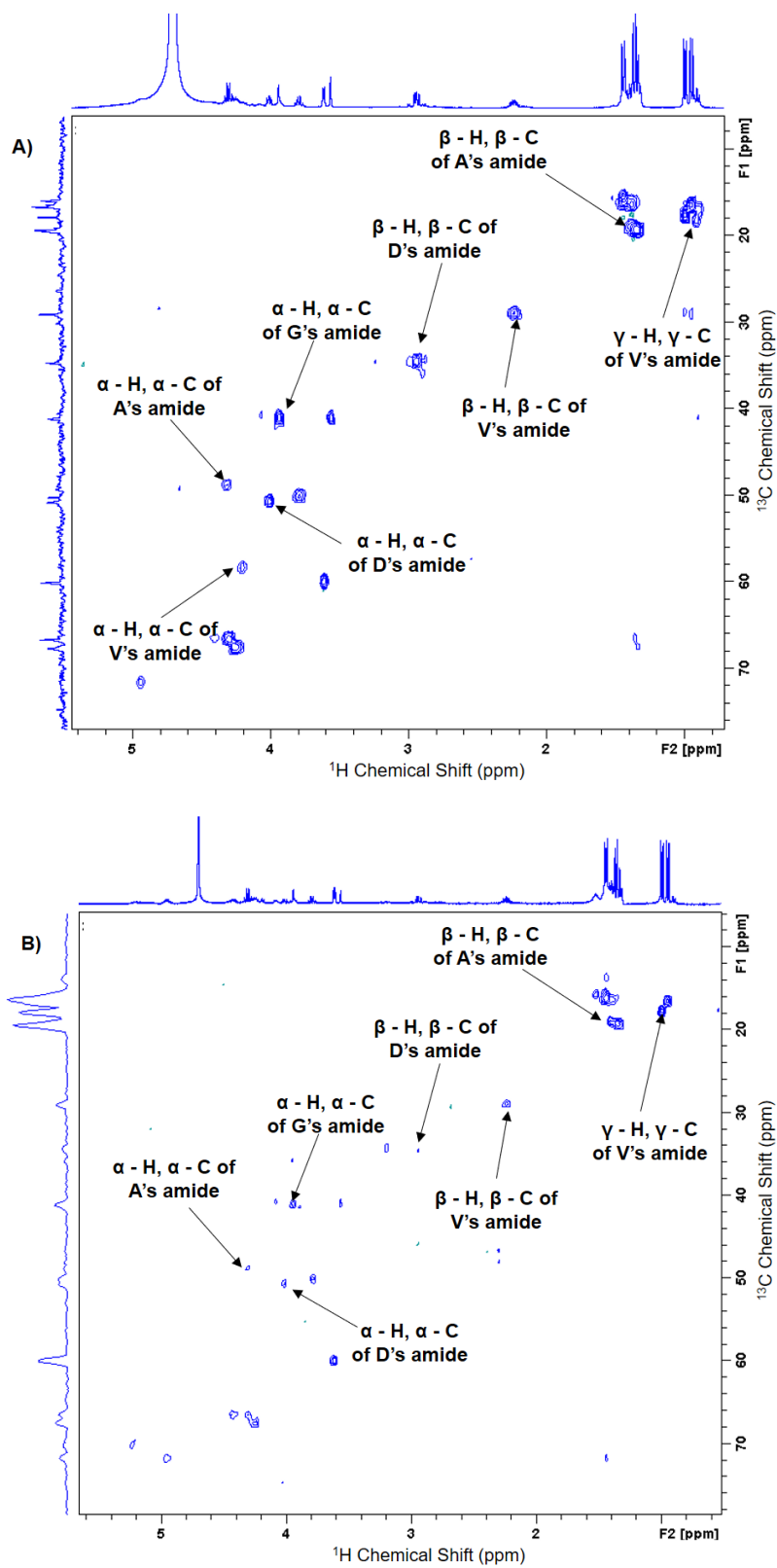

Figure 3.20: ^{1}H-NMR spectrum of LA, G, A, D, and V depsipeptides. a) LA, b) G, A, D, and V, c) and d) intermittently and continuously dried samples, respectively.

We also used qualitative analysis of detected mass peaks to study the differences in individual amino acid incorporation in linear depsipeptides and compare with branched depsipeptides (Figure 3.23). The qualitative analysis of mass peaks indicates that all the depsipeptides products contain a minimum of 5 amino acids. Also, the analysis of same amino acid incorporation (N) in depsipeptides shows that for N>3, D and V are less incorporated than that of A and G similar to the branched depsipeptides.

Amide moiety	Species	Theoretical shift (ppm)	Experimental shift (ppm)	
			Wet-dry	Cont. dry
G	α - H	3.95	3.94	3.95
	α – C	44.53	41.2	41.3
A	α - H	4.75	4.3	4.30
	α - C	50.07	48.9	48.9
	β - H	1.40	1.41	1.43
	β - C	16.21	16.2	16.0
D	α - H	4.60	4.00	4.02
	α - C	52.00	50.9	50.8
	β - H	2.8	2.95	2.93
	β - C	35.76	34.5	34.8
V	α - H	4.45	4.2	-
	α - C	59.91	58.2	-
	β - H	2.3	2.2	2.22
	β - C	30.73	28.9	29.1
	γ - H	0.9	0.99	1.07
	γ - C	17.86	17.2	17.7

Table 3.3: Theoretical and experimental chemical shifts associated with the depsipeptide amides formed upon reaction between LA and G, A, D, and V mixture. "Side" represent the two equivalent carboxylic acid groups of CA, while "centre" represents the third group, as shown in Figure 3.11.
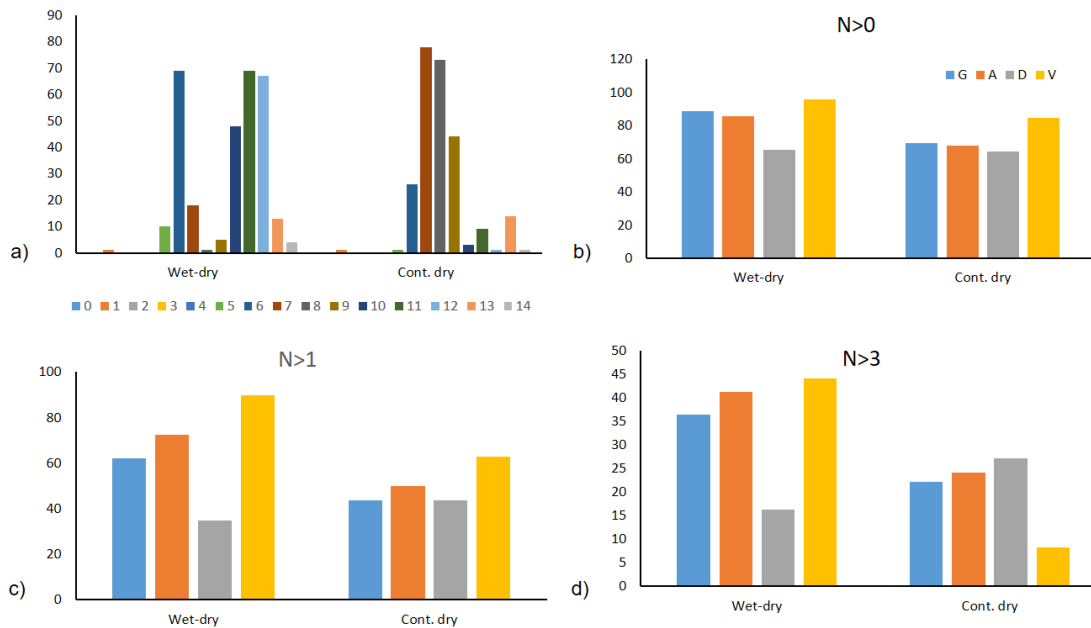
Figure 3.21: HSQC NMR spectra of the depsipeptides formed on a) intermittent and b) continuous drying of LA, G, A, D, and V subjected to for 8 cycles.
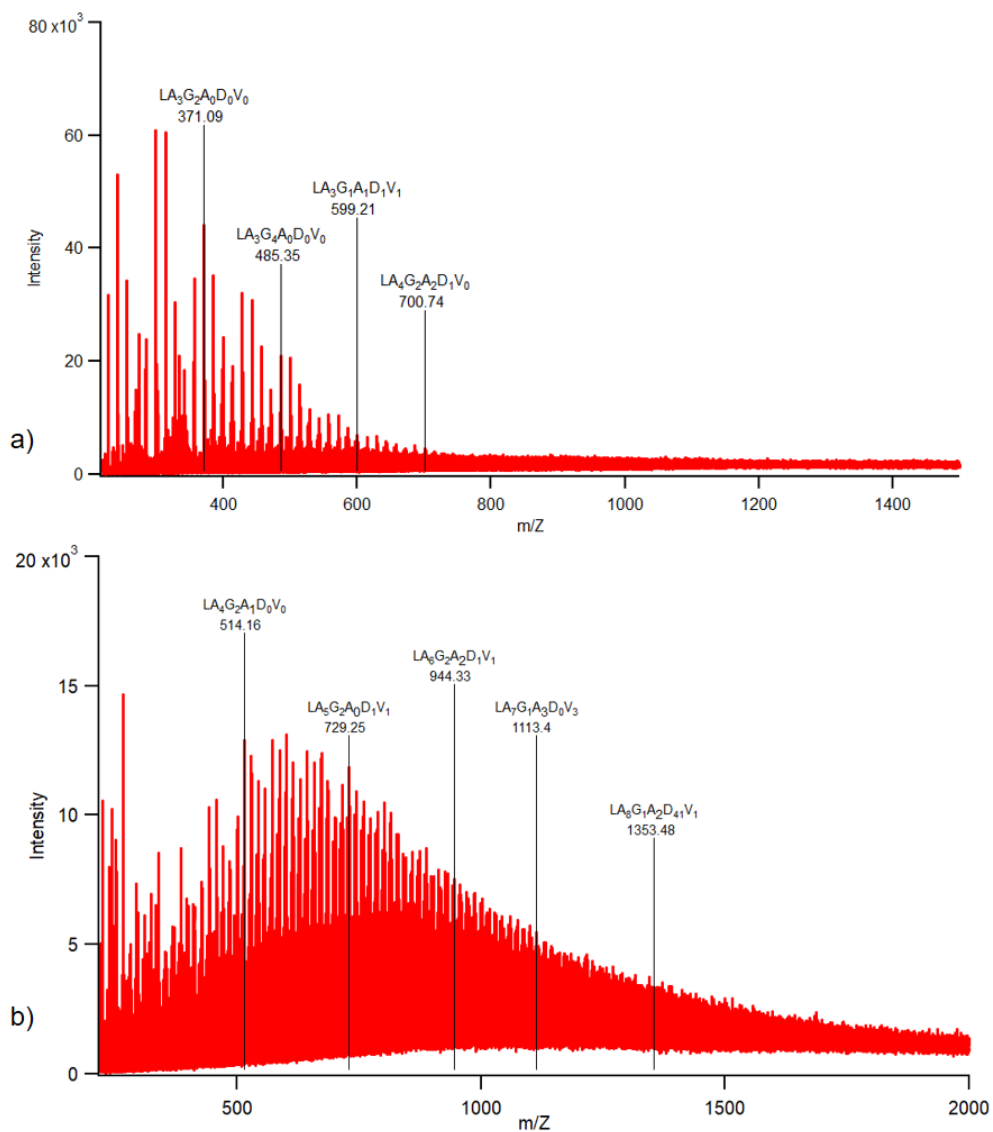
Figure 3.22: MALDI MS data of a) intermittently and b) continuously dried samples of LA, G, A, D, and V depsipeptides.

Figure 3.23: Qualitative analysis of LA, G, A, D, and V depsipeptides. a) The number of mass peaks attributed to depsipeptides vs. the number of incorporated amino acids in each product. The number of mass peaks attributed to depsipeptides vs. the number of incorporated specific amino acids are shown in the next graphs. Incorporation of more than 0, 1 and 3 of the same acid are shown in b), c), and d), respectively.

3.6 Amide bond formation upon a reaction between amino acids and polybasic carboxylic acids:

3.6.1 Amide bond formation between citric acid and alanine:

One of the controls for the experiments exploring the formation of depsipeptides between citric acid and glycerol polyesters (Glyc$_2$CA) and Glycine (G), Alanine (A), Aspartic acid (D) and Valine (V), we subjected 0.25M CA and 0.25M A solution to drying under the same conditions, namely intermittent or continuous drying at 85˚C for 8 cycles. ^{1}H-NMR of the resulting sample were compared to one of the starting materials. CA does not react alone, as depicted in Figure 3.24a. A also does not react on its own, as shown in Figure 3.9a. However, the dried sample of CA and A indicates the presence of amide peaks (Figure 3.24c). HSQC NMR and ACDLabs software were used to correlate the theoretical and experimental chemical shifts, which further confirmed amide bond formation between CA and A alone (Figure 3.25, Table 3.4).

49
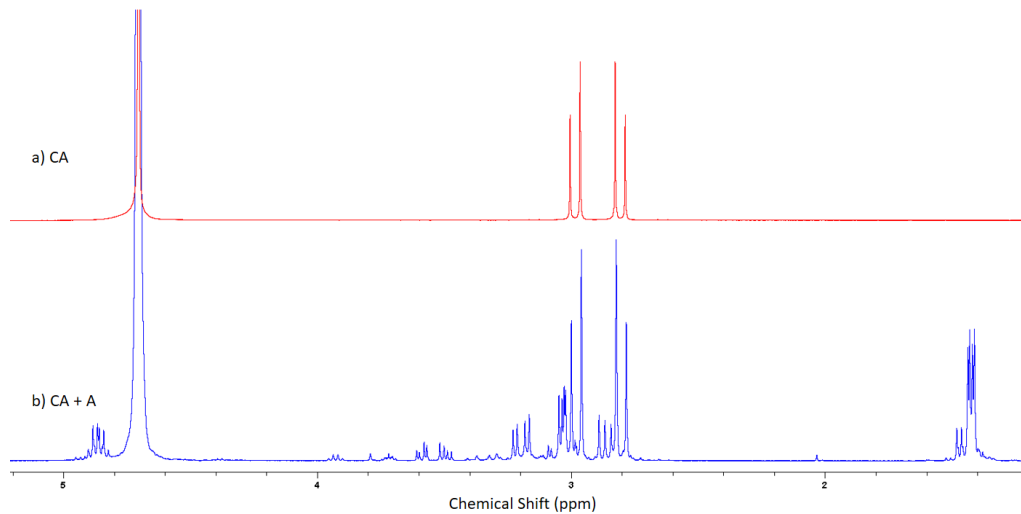
Figure 3.24: ^{1}H-NMR spectrum of CA and A amide a) CA and b) continuously dried samples of CA and A.
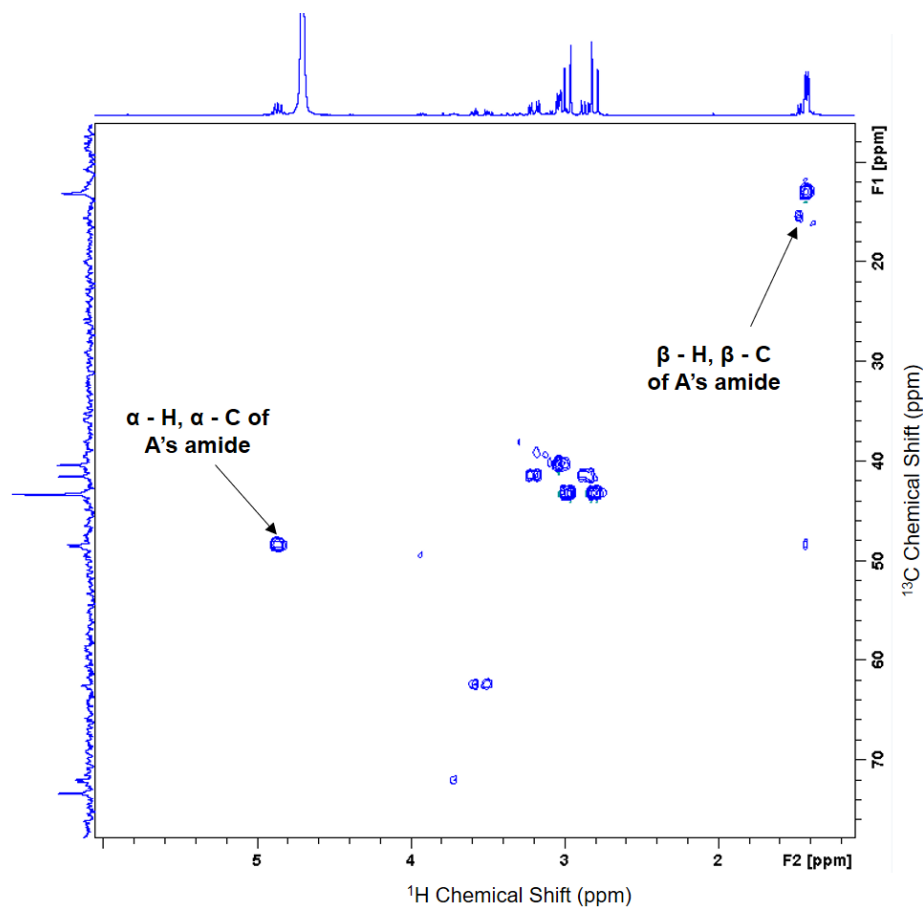
Figure 3.25: HSQC spectrum of CA and A amides. The peaks shown by the arrow corresponds to the amide bond between CA and A.

Amide moiety	Species	Theoretical shift (ppm)		Experimental shift (ppm)
		Side	Centre	
A	α - H	4.75	4.72	4.86
	α - C	49.32	50.02	48.4
	β - H	1.41	1.43	1.4
	β - C	16.24	15.97	15.6

Table 3.4: Theoretical and experimental chemical shift of amide moieties of CA and A.
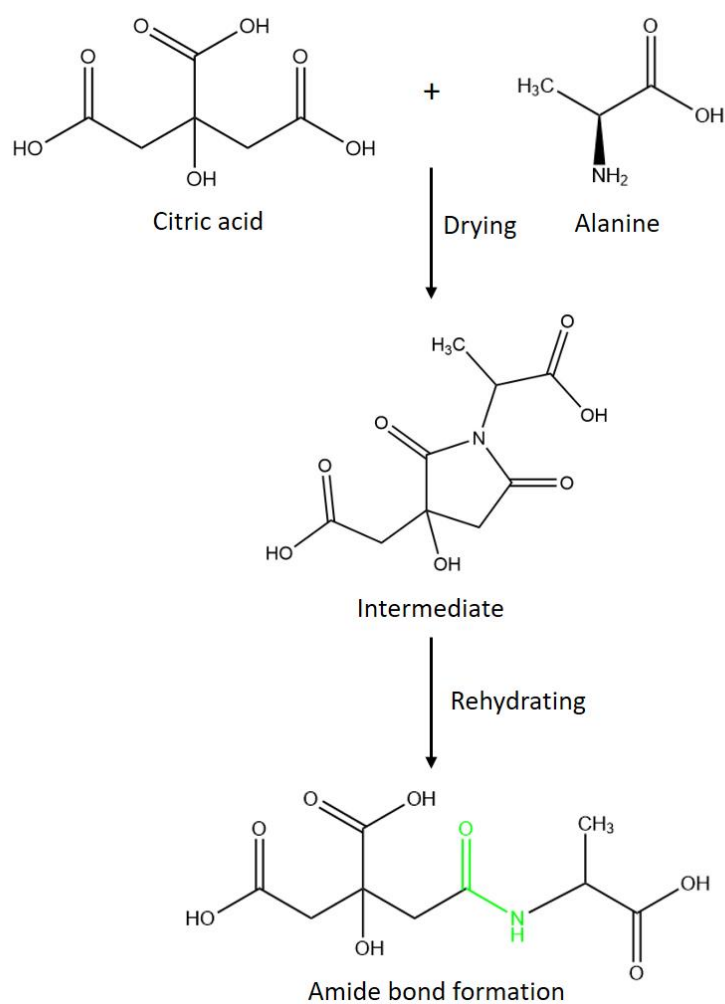
Figure 3.26: Schematic representation of a possible amide bond formation mechanism in a reaction between polybasic carboxylic acid and amine-containing molecule.

The direct coupling of carboxylic acid and amine to form an amide bond is thermodynamically unfavourable, and it happens at extreme conditions like temperature above 160°C or in the presence of activating and condensing agents as described in the Introduction section. However, in our case, we formed an amide bond using CA and A alone at 85°C. To explain these results, we hypothesize that the reaction goes through a cyclic succinamide-like intermediate formed between two carboxylic groups of the polybasic carboxylic acid and amine part of A upon condensation, as shown in Figure 3.26. In a similar system, Roweton and coworkers showed heat-driven condensation aspartic acid at 187.8°C forms polysuccinamide, which produces polyaspartic acid upon hydrolysis (Roweton et al., 1997).

3.6.2. Polyamide formation using CA and lysine (Lys):

The direct coupling of amino acids to form a peptide bond is thermodynamically unfavourable. In contemporary biology, this process happens through ester-amide exchange with the help of enzymes. The α-amino group of upcoming amino acids form an amide bond with the carboxylic group of ester linkage between tRNA and polypeptide (Jenni and Ban, 2003). Since we formed an amide bond at low temperature (85°C), we decided to test whether the above protocol can be used to form polyamides. A has only one amine-containing group, so we cannot form polymeric products using CA and A. To form polyamides, we subjected 0.5M CA and 0.5M Lys, a diamine, to continuous drying at 85°C up to 192h. The samples were analyzed using SEC, ^{1}H-NMR, HSQC and MALDI-MS. In SEC-chromatogram, the Lys monomer elutes around 2.2 min. However, in the UV spectra, Lys lacks chromophore that absorbs around 220nm so it is not visible in UV spectra. CA, co-elutes with the CA and Lys mixture (~2.0 min), as shown in Figure 3.27. So, we cannot infer anything from SEC data regarding polymer formation. The ^{1}H-NMR spectrum of the product formed upon reaction between CA and Lys shows peaks around 4.8ppm, consistent with an amide proton (Figure 3.28). To further assign the amide signals, we measured an HSQC and compared the results of chemical shift predictions. (Figure.3.29 and Table 3.5). The results show that both α and ε amine group of lysine form amides. To check the extent

of the polymerization, we analysed the sample by MALDI MS. The MS data revealed the formation of oligomers up to 8-mers (Figure. 3.30).
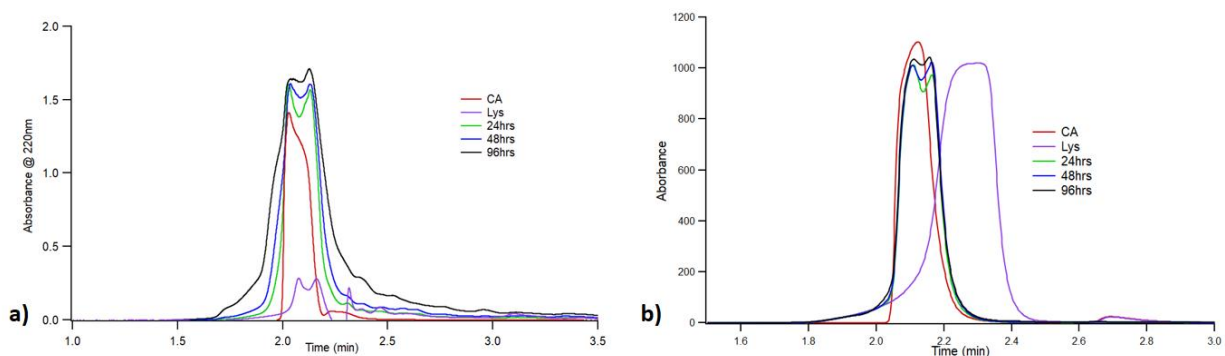
Figure 3.27: SEC data of a continuously dried sample of CA and Lys heated for 96h. The chromatogram was recorded using a) UV absorbance at 220nm and b) changes in the refractive index.
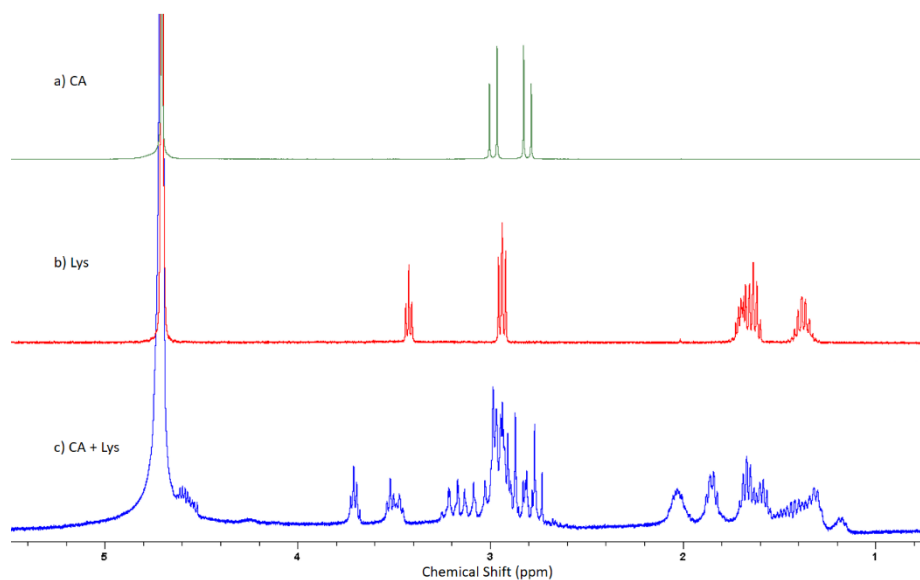
Figure 3.28: ^{1}H-NMR of CA and Lys samples. a) CA, b) Lys and c) continuously dried CA and Lys mixture for 96h.
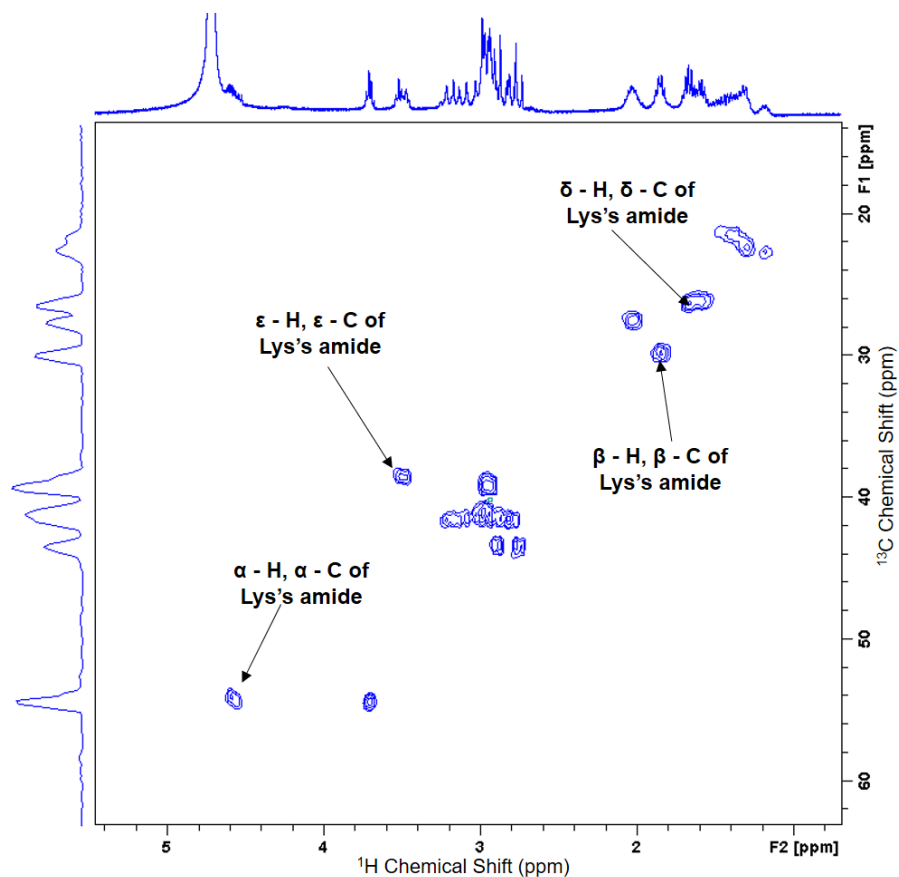
Figure 3.29: HSQC NMR of continuous dried CA and Lys for 192h showing two different amide peaks between CA and Lys.

Amide moiety	Species	Theoretical shift (ppm)		Experimental shift (ppm)
		Side	Centre	
Lys	α - H	4.35	4.35	4.57
	α - C	54.11	54.36	54.2
	β - H	1.87	1.78	1.83
	β - C	30.01	29.76	29.6
	ε - H	3.16	3.21	3.48
	ε - C	39.66	38.21	38.5
	δ - H	1.50	1.53	1.61
	δ - C	28.16	27.91	26.1

Table 3.5: Theoretical and experimental chemical shift of amide moieties of CA and Lys.
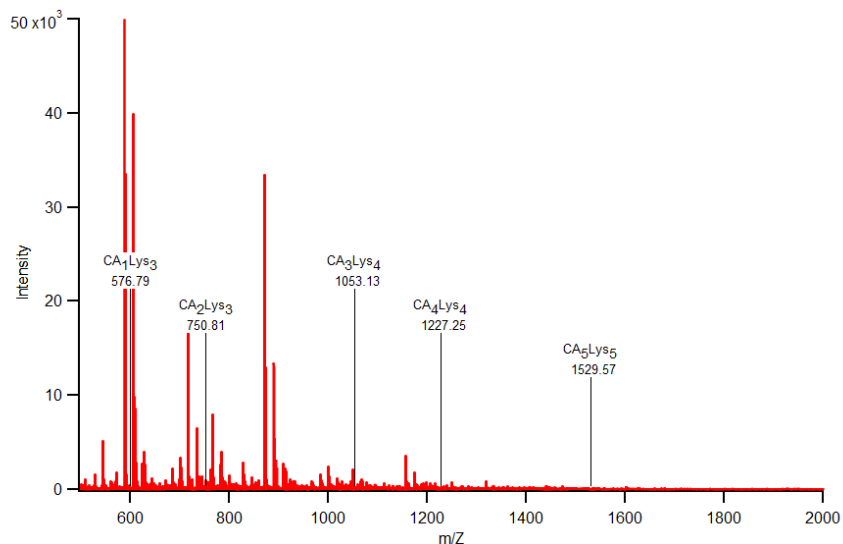
Figure 3.30: MS data of CA and Lys mixture collected after continuous drying for 192h.

3.6.3 Hydrolysis assays using CA and Lys polyamide:

Peptides are known for their catalytic property. So, we wanted to check whether the CA and Lys polyamides can perform some catalytic reactions or not. To check the catalytic properties of the CA-Lys polyamide, we performed esterase and phosphatase assay whose protocols are mentioned in the Materials and Method. In the esterase assay, the chosen substrate has an ester bond, which is prone to hydrolysis. The resulting product after the cleavage absorbs at 405nm, which is used to determine the extent of reaction. Similarly, in the case of phosphatase assay, the hydrolysis of phosphate group results in a chromophore. For determining the catalytic activity of the polymer, we subjected the substrate to hydrolysis in the presence and absence of the CA and Lys polyamides. The reaction was monitored by time-lapse UV-vis measurements. The presence of polyamides did not increase the rate of the hydrolysis reaction, as shown in Figure 3.28. It shows that the CA and Lys polyamides have no catalytic activity for the assays. Besides, the polyamides presence in the solution seems to inhibit the rate of hydrolysis by providing some sort of protection to the substrate and limiting its access to water.

55
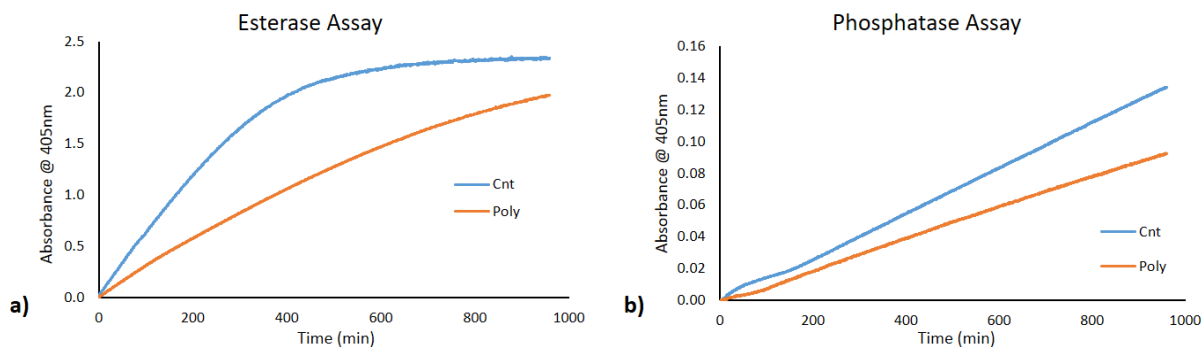
Figure 3.31: Graph of absorbance vs time for a) esterase and b) phosphatase assay

3.6.4: Formation of polyamides at different pH

To test the effect of pH on polyamide formation, we tried to synthesize CA and Lys at pH 2, 3.8 (native), 7, and 9. Then, they were subjected to continuous drying for 192h. The dried samples were analyzed using ^{1}H-NMR. Since the pH of the polymer solution is different, the values of the chemical shifts were not consistent. The NMR analysis shows no polymer formation at pH 7.0 and 9.0 (Figure: 3.32). Also, the polyamides form only in the case of pH 2.0 and 3.8). The amide peak intensity is higher than in the case of pH 3.8 than that of pH 2, indicating a higher polymer yield.
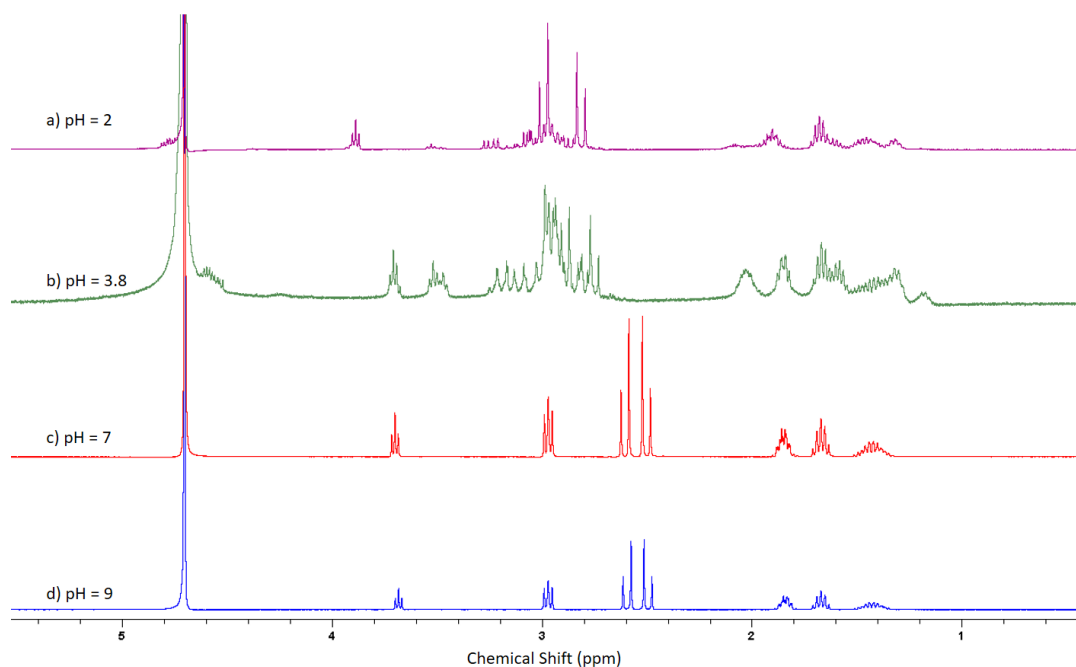

Figure 3.32: ^{1}H-NMR spectra of polyamides formed in the reaction between CA and Lys at a) pH=2.0, b) pH=3.8 c) pH=7.0, and d) pH=9.0.

56

3.6.5: Stability of CA and Lys polyamides against hydrolysis

To check the stability of polyamides against hydrolysis, we heated the polymer solution at 85°C for 24h, 48h, 96h and 192h. The samples were analyzed using ^{1}H-NMR (Figure 3.33). The NMR spectra of polyamides are almost unchanged during the 192h. The spectra show a slight increase in sharpness of CA peak as the duration of heating increases indicated by the doublets around 2.9 ppm and 3.0 ppm consistent with unreacted CA. We cannot quantify the CA recovery through the NMR peak integration because the CA and the product peaks are not well-resolved. But, the overall effect on the spectra is small as shown by the amide peak around 4.8ppm, which not disappear even after heating for 192h. Thus, the ^{1}H-NMR analysis indicates that the rate of polyamide hydrolysis is slow.
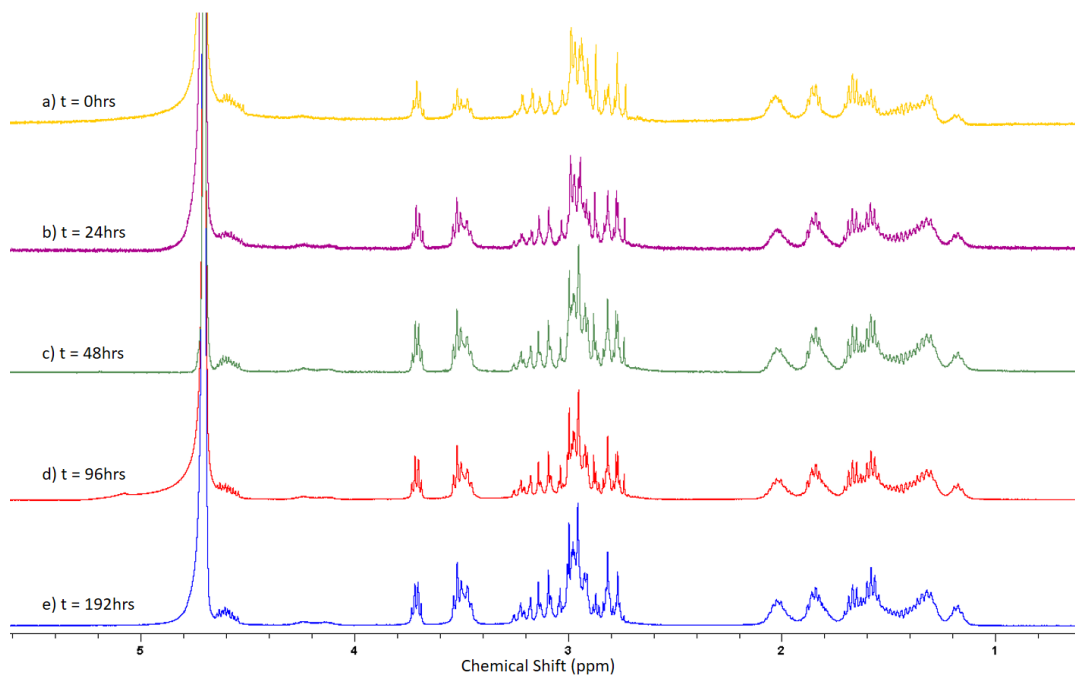

Figure 3.33: ^{1}H-NMR experiment of hydrolysis of CA and Lys polyamides by heating it in water for time (t). a) t = 0h, b) t =24h, c) t = 48h, d) t = 96h and e) t = 192h.

3.6.6: Formation of polyamides between various polybasic carboxylic acids and amines.

Our previous results show that polyamides of up to 8-mers can be synthesized in milder conditions such as 85°C using CA and Lys. Our next goal was to explore these polyamides systems using the same protocol and form different polyamides with different starting materials.

a) Tartaric acid (TA) and Lys: To synthesize polyamides, we used a natural polybasic carboxylic acid, TA and Lys. We subjected a mixture of 0.5M (TA) and 0.5M Lys to heating at 85°C for 24h, 48h, 96h, and 192h. There was the formation of bubbles during the process of heating, which indicates decarboxylation of TA. The dried sample was analyzed using ^{1}H-NMR. The ^{1}H-NMR spectrum did not contain any peaks for amide bond or any polymeric products, as shown in Figure 3.34.
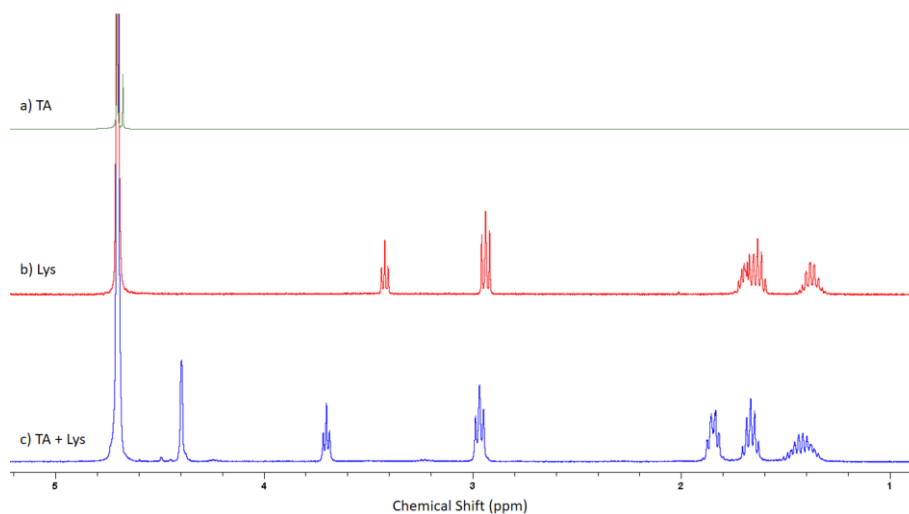
Figure 3.34: ^{1}H-NMR spectrum of TA and Lys mixture. a) TA, b) Lys and c) CA + Lys. The shifts in Lys signals in the spectra b) and c) are due to pH change.

b) Methyl succinic acid (MSA) and Lys: The TA and Lys mixture did not show any reaction among themselves. We tried another hydroxy acid, MSA with Lys for thermal polymerization. The mixture of 0.5M MSA and 0.5M Lys was heated for 24h, 48h, 96h, and 192h at 85°C. The dried sample was analyzed using ^{1}H-NMR. The spectrum shows the presence of amide peaks around 4.5ppm as shown in Figure 3.35. However, the

yield of polymers is less than that of CA and Lys polyamides. We hypothesize that in the pH of CA and Lys mixture is buffered by unreacted CA. Whereas, the MSA is less acidic than CA the extent of buffering capacity is lesser in the case of MSA and Lys mixture. Previous results have shown that polyamides synthesis is favoured in acidic conditions. So, the MSA and Lys mixture will have a lesser yield than that of CA and Lys.
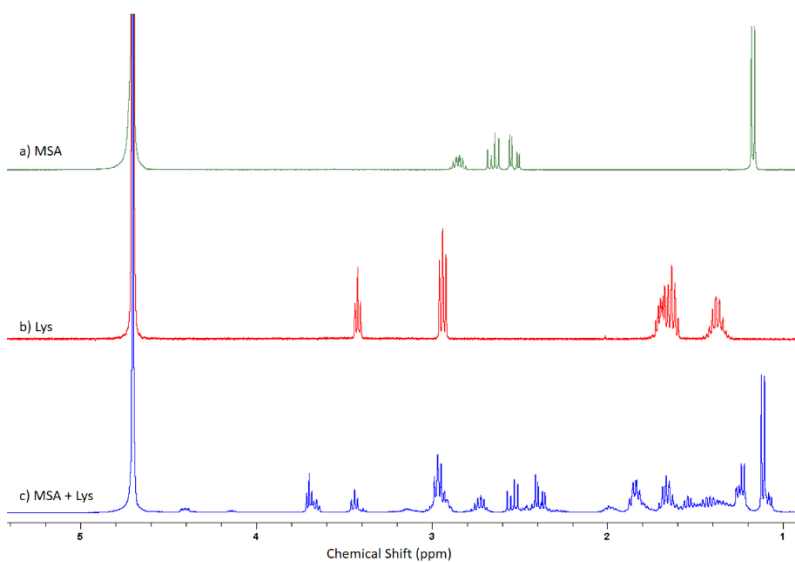
Figure 3.35: ^{1}H-NMR spectrum of MSA and Lys heated samples. a) MSA, b) Lys and c) MSA + Lys.

c) CA and spermidine trihydrochloride (STCl): In another experiment, we used CA and STCl in an equal molar ratio as starting materials for polyamide synthesis. The mixture of 0.125M CA and 0.125M STCl was heated for 24h, 48h, 96h, and 192h at 85°C. STCl with a chemical formula of $NH_2(CH_2)_3NH(CH_2)_4(NH_2).3HCl$ is a three amine group-containing molecule with two aliphatic moieties, which makes it more hydrophobic. Due to the presence of three amine groups, we expect it to form branched polyamides with CA. ^{1}H-NMR of cycles sample shows an amide peak around 3.8 ppm (Figure 3.36). HSQC analysis of the sample heated for 192h shows the formation of an amide bond between CA and all three amines of spermidine trihydrochloride (Figure 3.37). We performed SEC analysis of the samples that were run for different duration, to check the extent of polymerization across an increase in time (Figure 3.38). We cannot infer the formation of the polymers by SEC data due to the overlap of monomer CA

peak with the oligomeric or polymeric products (Figure 3.38). The MALDI mass spec analysis of samples confirmed the formation of polyamides up to 6-mers (Figure 3.39).
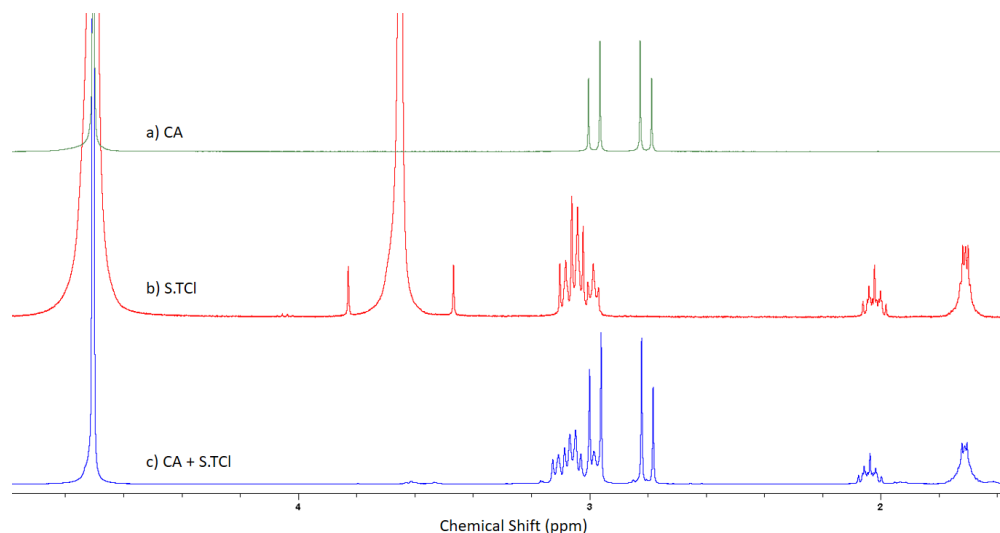
Figure 3.36: ^{1}H-NMR of CA and STCl mixture subjected to continuous drying for 8 cycles.
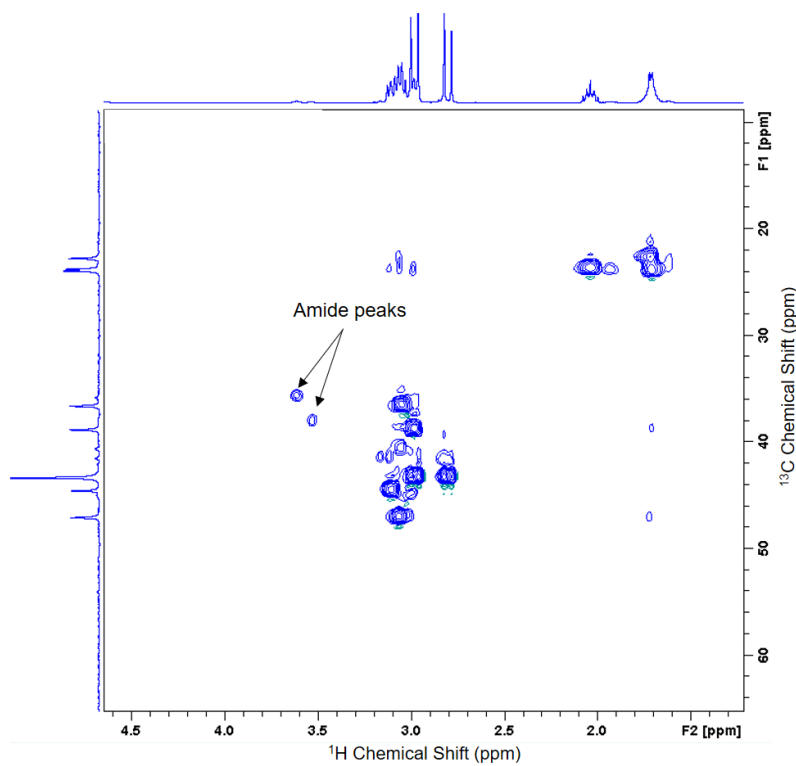
Figure 3.37: HSQC NMR of CA and STCl mixture shows the presence of two amide peaks.
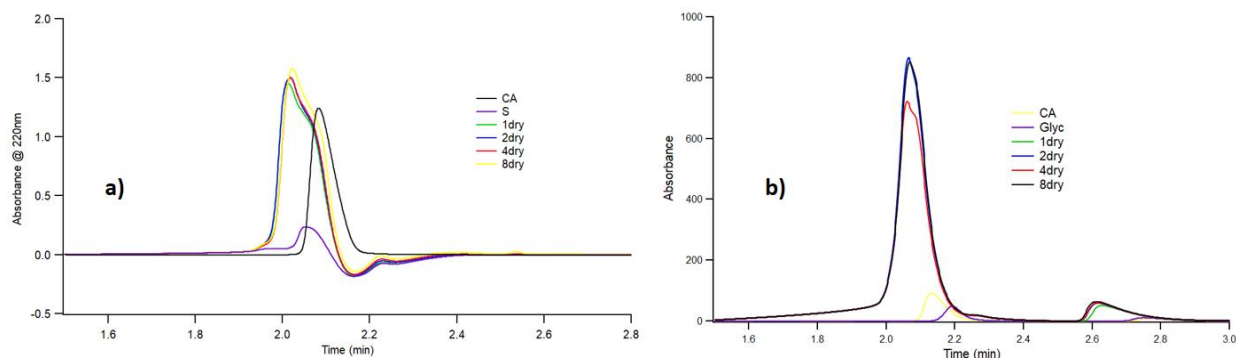
Figure 3.38: SEC data of CA and STCI across multiple continuous drying cycles. The chromatogram was taken using a) UV and b) RI detector.
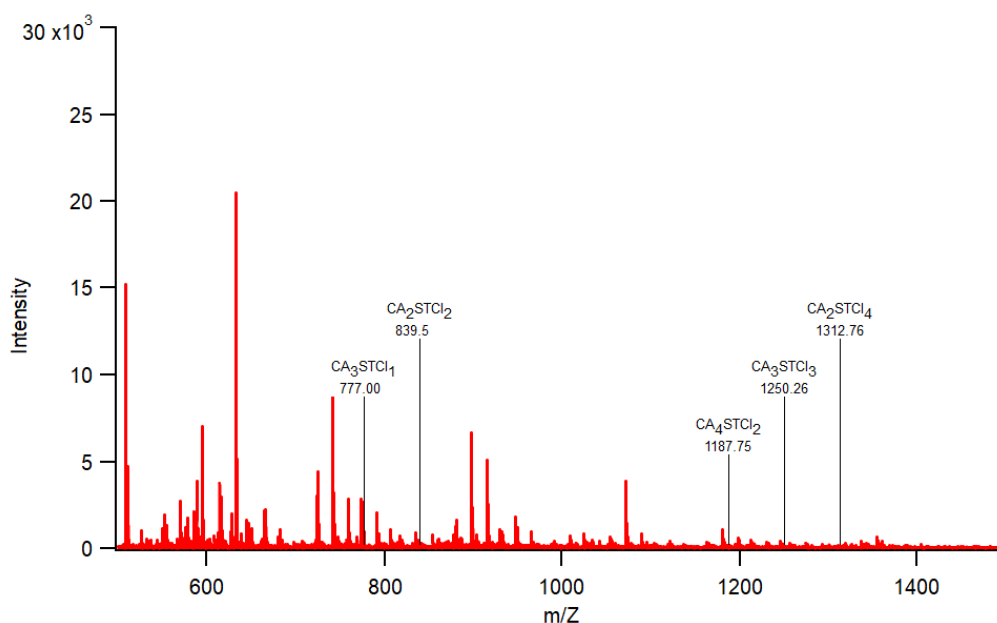
Figure 3.39: MALDI MS data of CA and STCL collected after 8 continuous drying cycles.

3.6.7 Kemp elimination assay using CA and STCI:

STCI is a multifunctional molecule with hydrophobic moieties. So, the polymers formed using STCI will be hyperbranched having hydrophobic pockets because of the aliphatic moieties of STCI. These types of hyperbranched polymers containing hyperbranched moieties have been shown to catalyze the Kemp elimination reaction (Mamajanov and Cody, 2017). To check the catalytic function of the CA and STCI polymer, we employ Kemp-elimination assay using 5-nitro-1, 2-benzioxazole as substrate. The reaction

61

starts when a catalytic base extracts a proton from benzioxazole ring, which results in the ring-opening by cleavage of a NO bond to form 2-cyano-4-nitrophenolate. The elimination proceeds by a single transition state, which is stabilized by dispersion forces. The UV-Vis spectra show that the rate of increase in absorbance is more when polyamides are present in the solution (Figure 3.40a). The rate constant for the solution containing polyamides and substrate is twice as of control, which indicates an increase in the catalytic activity of the reaction (Figure 3.40b). The reaction of CA and STCl polyamides with the substrate was performed three times. The error bars for the experiment is shown in Figure 3.40b.

a)

b)
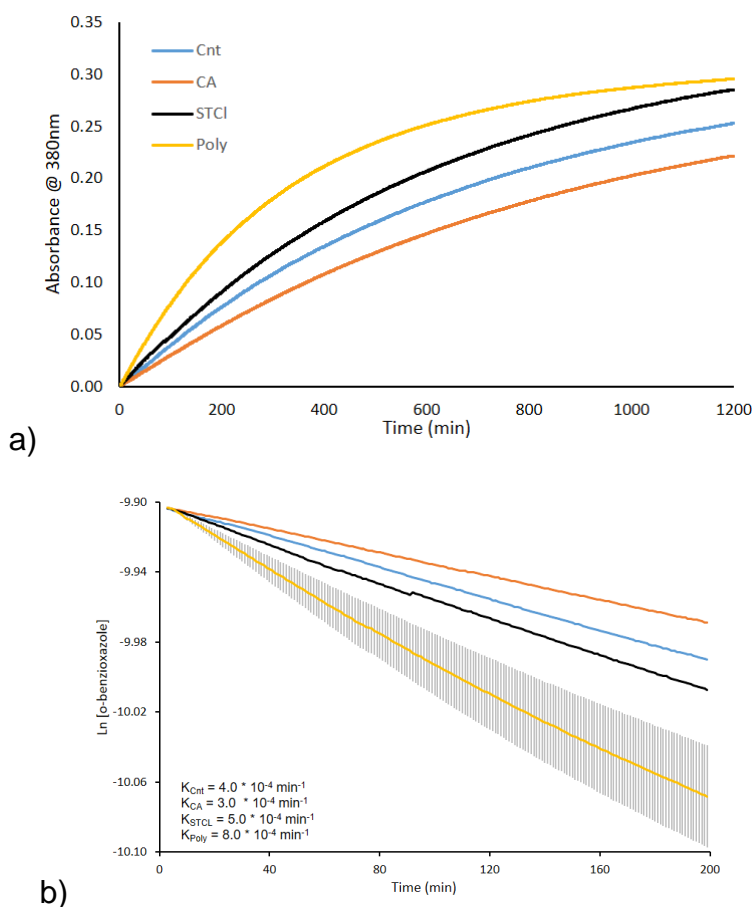
Figure 3.40: Kemp Elimination Assay. a) UV-Vis spectra of time-based measurement at 380nm. The reactions were performed with substrate alone (Cnt), CA and substrate (CA), STCl and substrate (STCL), Polyamides and substrate (Poly). b) To calculate the rate constant, we plotted ln *[o-benzioxazole]* vs time assuming pseudo-first-order kinetics.

4. Conclusion:

Our results show that both linear polyesters synthesized from lactic acid (LA) and hyperbranched polyesters synthesized from citric acid and glycerol (Glyc$_2$CA), and 2,2-bis(hydroxymethyl)propionic acid (bis-MPA) form depsipeptides when subjected to intermittent and continuous drying in the presence of amino acid mixture containing glycine (G), alanine (A), aspartic acid (D), and valine (V). The results indicate that the polymers formed in the intermittent and continuously dried samples are huge combinatorial libraries of depsipeptide molecules. To study the makeup of the formed sequences, we performed an analysis of the products. We have created databases of all possible molecular formulae of the depsipeptides in our system and compared them with the experimental mass spec data. Within the experimental data, we evaluated the likelihood of single and repeated incorporation of each amino acid into depsipeptides. Overall, all amino acids get incorporated relatively evenly distributed in both linear and branched depsipeptides. In most cases, there are fewer V and D moieties in the depsipeptides, probably owing to their larger size compared to G and A. The analysis of repeated incorporations of the same amino acid into the same depsipeptide varies. The analysis shows that in all polyesters, there are relatively fewer sequences that contain more than three incorporations of V. This feature is most prominent in the case of the bis-MPA polyester matrix because of its hyperbranched structure. LA forms linear polyesters, so all amino acids are equally accessible to the polyesters. Glyc$_2$CA, despite being a hyperbranched polymer degrades into a mixture of linear and hyperbranched polyesters upon ester-amide exchange. Also, citric acid monomers of Glyc$_2$CA can directly react with amino acids, which causes scrambling of amino acids incorporation. Unlike the Glyc$_2$CA and LA polyesters, the released bis-MPA monomers in the process of ester-amide exchange do not react further to form new polyesters, so the scrambling of amino acids does not happen. These results show different patterns of incorporation of amino acids into linear and branched polyester matrixes. Therefore, the mechanism of depsipeptides synthesis can act as rudimentary "translation machinery" for controlling the makeup of amino acids in depsipeptides and, eventually, peptides.

Abiotic peptide synthesis is a thermodynamically unfavourable process and requires a high temperature that is destructive to both monomers and products. This notion leads us to believe that maybe peptides were not synthesized de novo from the prebiotic soup. It is conceivable that easier to synthesize polyamides performed some of the peptide functions at the early stage of evolution, and peptides replaced them once some of the biological machinery came into existence. Here, we studied the abiotic synthesis of amide bond formation in milder conditions, which require prolonged heating of a mixture of a polybasic carboxylic acid and an amine at 85°C. Using this method, we formed different polyamides from different starting materials such as citric acid (CA) and lysine (Lys), CA and spermidine trihydrochloride (STCl), and methyl succinic acid (MSA) and Lys. Some of these polyamides like CA and Lys polyamides have a chain length of up to 8-mers, and they are slow to hydrolyze. We have also tried to search for catalytic functions in these polyamides. The CA and Lys polyamides do not show any activity towards esterase and phosphatase assay. However, The CA and STCl polyamides increase the rate of Kemp elimination reaction by twice the controls. The catalytic activity of the CA and STCl polyamides suggests the possibility that in prebiotic chemistry if we stepped away from making modern peptides, we can straightforwardly synthesize similarly, but different polyamides and those can be functional. The polyamides could have served as a viable peptide alternative in the prebiotic world.

5. References:

Abe, Y. (1993). Physical state of the very early Earth. Lithos *30*, 223–235.

Bell, E.A., Boehnke, P., Harrison, T.M., and Mao, W.L. (2015). Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. Proc. Natl. Acad. Sci. U. S. A. *112*, 14518–14521.

Boustta, M., Huguet, J., & Vert, M. (1991, July). New functional polyamides derived from citric acid and L-lysine: Synthesis and characterization. In Makromolekulare Chemie. Macromolecular Symposia Vol. 47, No. 1, pp. 345-355.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell *56*, 549–561.

Butlerow, A. (1861). Formation synthétique d'une substance sucrée. CR Acad. Sci, 53, 145-147.

Chambers, J.E. (2004). Planetary accretion in the inner Solar System. Earth Planet. Sci. Lett. *223*, 241–252.

Chyba, C., and Sagan, C. (1992). Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: An inventory for the origins of life. Nature *355*, 125–132.

Creighton, T. E. (1990). Protein folding. Biochemical journal, 270(1), 1.

Crick, F. (1970). Central dogma of molecular biology. Nature, 227(5258), 561-563.

Decker, P., Schweer, H., and Pohlamnn, R. (1982). Bioids. X. Identification of formose sugars, presumable prebiotic metabolites, using capillary gas chromatography/gas chromatography-mas spectrometry of n-butoxime trifluoroacetates on OV-225. J. Chromatogr. A *244*, 281–291.

DeVoe, H., and Tinoco, I. (1962). The stability of helical polynucleotides: Base contributions. J. Mol. Biol. *4*, 500–517.

Dickson, K.S., Burns, C.M., and Richardson, J.P. (2000). Determination of the free-

energy change for repair of a DNA phosphodiester bond. J. Biol. Chem. *275*, 15828–15831.

Forsythe, J.G., Yu, S.S., Mamajanov, I., Grover, M.A., Krishnamurthy, R., Fernández, F.M., and Hud, N. V. (2015). Ester-Mediated Amide Bond Formation Driven by Wet-Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. Angew. Chemie - Int. Ed. *54*, 9871–9875.

Fox, S.W., and Harada, K. (1958). Thermal copolymerization of amino acids to a product resembling protein. Science (80-.). *128*, 1214.

Fox, S.W., and Harada, K. (1960). The Thermal Copolymerization of Amino Acids Common to Protein. J. Am. Chem. Soc. *82*, 3745–3751.

Frenkel-Pinter, M., Haynes, J.W., Martin, C., Petrov, A.S., Burcar, B.T., Krishnamurthy, R., Hud, N. V., Leman, L.J., and Williams, L.D. (2019). Selective incorporation of proteinaceous over nonproteinaceous cationic amino acids in model prebiotic oligomerization reactions. Proc. Natl. Acad. Sci. U. S. A. *116*, 16338–16346.

Gao, C., and Yan, D. (2004). Hyperbranched polymers: From synthesis to applications. Prog. Polym. Sci. *29*, 183–275.

Gilbert, W. (1986). The RNA world Superlattices point ahead. Nature *319*, 618.

Greenwald, J., and Riek, R. (2012). On the possible amyloid origin of protein folds. J. Mol. Biol. *421*, 417–426.

Gualerzi, C.O., and Pon, C.L. (1990). Initiation of mRNA translation in prokaryotes. Biochemistry *29*, 5881–5889.

Hampsey, M. (1998). Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery. Microbiol. Mol. Biol. Rev. *62*, 465–503.

Huber, C. (1998). Peptides by Activation of Amino Acids with CO on (Ni,Fe)S Surfaces: Implications for the Origin of Life. Science (80-.). *281*, 670–672.

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

Jenni, S., and Ban, N. (2003). The chemistry of protein synthesis and voyage through the ribosomal tunnel. Curr. Opin. Struct. Biol. *13*, 212–219.

Jikei, M., Chon, S.H., Kakimoto, M.A., Kawauchi, S., Imase, T., and Watanebe, J. (1999). Synthesis of hyperbranched aromatic polyamide from aromatic diamines and trimesic acid. Macromolecules *32*, 2061–2064.

Kirkorian, K., Ellis, A., and Twyman, L.J. (2012). Catalytic hyperbranched polymers as enzyme mimics; Exploiting the principles of encapsulation and supramolecular chemistry. Chem. Soc. Rev. *41*, 6138–6159.

Kroner, M., Hartmann, H., Boeckh, D., Baur, R., Kud, A., & Schwendemann, V. (1997). U.S. Patent No. 5,639,723. Washington, DC: U.S. Patent and Trademark Office.

Lahav, N., White, D., and Chang, S. (1978). Peptide formation in the prebiotic era: Thermal condensation of glycine in fluctuating clay environments. Science (80-.). *201*, 67–69.

Liebman, S.A., Pesce-Rodriguez, R.A., and Matthews, C.N. (1995). Organic analysis of hydrogen cyanide polymers: Prebiotic and extraterrestrial chemistry. Adv. Sp. Res. *15*, 71–80.

Magnusson, H., Malmström, E., and Hult, A. (2000). Structure buildup in hyperbranched polymers from 2,2-bis(hydroxymethyl)propionic acid. Macromolecules *33*, 3099–3104.

Mamajanov, I., and Cody, G.D. (2017). Protoenzymes: The case of hyperbranched polyesters. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. *375*.

Mamajanov, I., Callahan, M.P., Dworkin, J.P., and Cody, G.D. (2015). Prebiotic Alternatives to Proteins: Structure and Function of Hyperbranched Polyesters. Orig. Life Evol. Biosph. *45*, 123–137.

Maury, C.P.J. (2018). Amyloid and the origin of life: self-replicating catalytic amyloids as prebiotic informational and protometabolic entities. Cell. Mol. Life Sci. *75*, 1499–1507.

Miller, S.L. (1955). Production of Some Organic Compounds under Possible Primitive Earth Conditions. J. Am. Chem. Soc. *77*, 2351–2361.

Mojzsis, S.J., Arrhenius, G., McKeegan, K.D., Harrison, T.M., Nutman, A.P., and Friend, C.R.L. (1996). Evidence for life on Earth before 3,800 million years ago. Nature *384*, 55–59.

Morávek, J. (1967). Formation of oligonucleotides during heating of a mixture of uridine 2′(3′) -phosphate and uridine. Tetrahedron Lett. *8*, 1707–1710.

Moutou, G., Taillades, J., and Mansani, R. (1995). Formaldehyde , Hydrogen Cyanide and Ammonia in Aqueous Solution : Industrial and Prebiotic. *8*, 721–730.

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

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

Rajamani, S., Vlassov, A., Benner, S., Coombs, A., Olasagasti, F., and Deamer, D. (2008). Lipid-assisted synthesis of RNA-like polymers from mononucleotides. Orig. Life Evol. Biosph. *38*, 57–74.

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

Reimschuessel, H.K. (2010). Polyamide Learn more about Polyamide Polyamides Chemistry , Formulation , and Proper- Polyamides (Nylons) Specialty Polymers & Polymer Process- ing.

Rohlfing, D.L., and Fox, S.W. (1967). The catalytic activity of thermal polyanhydro-α-amino acids for the hydrolysis of p-nitrophenyl acetate. Catalysis by thermal polyamino acids. Arch. Biochem. Biophys. *118*, 122–126.

Roweton, S., H, S.J., and Swift, G. (1997). Poly (Aspartic acid): Synthesis , Biodegradation , and Current Applications t j ,. Polymer (Guildf). *5*.

Sawai, H. (1976). Catalysis of Internucleotide Bond Formation by Divalent Metal Ions. J. Am. Chem. Soc. *98*, 7037–7039.

Schwendinger, G., and Rode, M. (1991). Salt-induced conditions. *186*, 247–251.

Sleep, N.H., Zahnle, K.J., Kasting, J.F., and Morowitz, H.J. (1989). Annihilation of ecosystems by large asteroid impacts on the early Earth. Nature *342*, 139–142.

Stojanoski, K., and Zdravkovski, Z. (1993). Textbook forum: On the formation of peptide bonds. J. Chem. Educ. *70*, 134–135.

Sunde, M., Serpell, L.C., Bartlam, M., Fraser, P.E., Pepys, M.B., and Blake, C.C.F. (1997). Common core structure of amyloid fibrils by synchrotron X-ray diffraction. J. Mol. Biol. *273*, 729–739.

Wächtershäuser, G. (2010). Chemoautotrophic origin of life: The iron–sulfur world hypothesis. In Geomicrobiology: Molecular and environmental perspective (pp. 1-35). Springer, Dordrecht.

Wickstead, B., and Gull, K. (2011). The evolution of the cytoskeleton. J. Cell Biol. *194*, 513–525.

Yu, S.S., Solano, M.D., Blanchard, M.K., Soper-Hopper, M.T., Krishnamurthy, R., Fernández, F.M., Hud, N. V., Schork, F.J., and Grover, M.A. (2017). Elongation of Model Prebiotic Proto-Peptides by Continuous Monomer Feeding. Macromolecules *50*, 9286–9294.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D., and Noller, H.F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. Science (80). *292*, 883–896.