

**ESTER-MEDIATED AMIDE BOND FORMATION: A POSSIBLE
PATH TO PROTO-PEPTIDES ON THE PREBIOTIC EARTH**

A Dissertation
Presented to
The Academic Faculty

by

Sheng-Sheng Yu

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemical & Biomolecular Engineering

Georgia Institute of Technology
May 2017

COPYRIGHT © 2017 BY SHENG-SHENG YU

**ESTER-MEDIATED AMIDE BOND FORMATION: A POSSIBLE
PATH TO PROTO-PEPTIDES ON THE PREBIOTIC EARTH**

Approved by:

Dr. Martha A. Grover, Advisor
School of Chemical & Biomolecular
Engineering
Georgia Institute of Technology

Dr. William J. Koros
School of Chemical & Biomolecular
Engineering
Georgia Institute of Technology

Dr. F. Joseph Schork, co-Advisor
School of Chemical & Biomolecular
Engineering
Georgia Institute of Technology

Dr. Athanasios Nenes
School of Chemical & Biomolecular
Engineering
Georgia Institute of Technology

Dr. Charles L. Liotta
School of Chemical & Biomolecular
Engineering/ School of Chemistry
Georgia Institute of Technology

Date Approved: March 29, 2017

ACKNOWLEDGEMENTS

I owe a great debt to my two advisors, Dr. Martha Grover and Dr. F. Joseph Schork. They brought me into a topic that I would never have a chance to do in my home country. I am grateful for their efforts on educating me to be an independent researcher and for their tireless commitment to research.

I also had the opportunity to work with three other advisors during my time in the Center for Chemical Evolution (CCE): Dr. Nicholas Hud, Dr. Facundo Fernández and Dr. Ramanarayanan Krishnamurthy. Their peerless expertise and wisdom make my work better than my imagination.

I would like to thank my thesis committees: Dr. William Koros, Dr. Charles Liotta and Dr. Athanasios Nenes. They provided me valuable feedbacks and kept me on the right track.

Special thanks to Dr. Jay Forsythe, now a professor at College of Charleston. He is not only a good friend, but also a mentor to me. I still remember the exciting moment when we first detected depsipeptides from our beloved Synapt G2 mass spectrometer. I hope that he will enjoy his time in Charleston and educate more young people about science.

I would like to thank all people I met in the CCE, especially the unrivaled proto-peptide team: Aaron Mckee, Dr. Eric Parker, Chelsea Walker, Dr. Molly Hopper, Martin Solano, Dr. Anton Petrov, Dr. Moran Pinter and Dr. Irena Mamajanov. It is the creative and enthusiastic environment in the center that helps me overcome obstacles in my graduate study.

In addition, I would like to acknowledge the great help from CCE's administrative team: Dr. Christine Conwell and Shantel Floyd. Things would descend into chaos without their vigilance.

I would also like to express my gratitude to all my Taiwanese friends in graduate school: Dr. Ping-Hsun Chu, Yu-Han Chu, Dr. Wei-Ming Yeh, Dr. Li-Chen Lee, Dr. Li-Wei Chou, Dr. Sheng-Chaing Yang, Chi-Ta Lee and Ya-Dong Chiang. Studying in a place so far away from our island is never easy. Together, we faced challenges in this foreign continent and helped each other walk further than alone.

I greatly acknowledge the friendship and support from the Grover group: Dr. Andres Hernandez Moreno, Dr. Huayu Li, Dr. Daniel Griffin, Dr. Xun Tang, Dr. Christine He, Ming-Chien Hsieh, Nils Persson, Tristan Kernick, Michael McBride, Matthew McDonald, Yi Sun and Chiamaka OBianyor.

I would like to thank three undergraduate researchers: Catherine Psarakis, Matthew Blanchard and Xiaosi Gao. They offered exceptional assistance to my research. I hope that they are on the right track to their future career.

I am thankful to my family: my mothers, my sister and my two brothers. They supported my decision to study in the United States without doubts. I only wish my father could see me earn my degree before he left us.

Finally, I would like to sincerely thank all amazing people that I do not have a chance to mention in this abbreviated list. It is an honor to work with you.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS AND NOMENCLATURE	xviii
SUMMARY	xxi
CHAPTER 1. Introduction	1
1.1 A brief review of prebiotic chemistry	1
1.1.1 Hypothesis for origin of life	2
1.1.2 Building blocks of biopolymers	4
1.2 Challenges of the peptide problem	5
1.2.1 How peptides are generated nowadays	5
1.2.2 Challenge of the polycondensation of amino acids	6
1.2.3 Previous works for the prebiotic synthesis of peptides	7
1.3 Proto-biopolymers	8
1.3.1 Hydroxy acid-based polyester	9
1.3.2 Depsipeptide formation by the ester-amide exchange reaction	11
1.4 Objectives	12
CHAPTER 2. Synthesis of proto-peptides from ester-amide exchange reaction	14
2.1 Introduction	14
2.2 Experimental method	15
2.2.1 Materials	15
2.2.2 Reaction condition: Day-night cycling	15
2.2.3 Mass spectrometry analysis	16
2.2.4 FT-IR analysis	18
2.2.5 NMR analysis	19
2.3 Results & discussion	19
2.3.1 Proof of principle experiment	19
2.3.2 Evolution of LA/A depsipeptides	28
2.3.3 The effect of diketopiperazines	33
2.3.4 Sequence analysis	36
2.3.5 Analysis of reaction mechanism	40
2.3.6 Reaction at neutral pH	44
2.3.7 Two-dimensional NMR analysis and quantitative NMR analysis of LA/G oligomers	47
2.3.8 Exploring the chemical space	57
2.4 Conclusion	68

CHAPTER 3. Kinetics of prebiotic depsipeptide formation from the ester-amide exchange reaction	69
3.1 Introduction	69
3.2 Experimental method	70
3.2.1 Materials	70
3.2.2 Reaction procedure	70
3.2.3 Polymer characterization	71
3.3 Model development	72
3.3.1 Modeling of depsipeptide copolymerization	72
3.3.2 Calculation of heats and entropies of activation	81
3.4 Results and discussion	82
3.4.1 Characterization of depsipeptides by HPLC method	82
3.4.2 Estimation of rate constants	87
3.4.3 Flux analysis	98
3.4.4 Activation parameters	100
3.5 Conclusion	102
CHAPTER 4. Growth of proto-peptides by continuous feeding of monomers	104
4.1 Introduction	104
4.2 Experimental method	105
4.2.1 Materials	105
4.2.2 Day-night machine	105
4.2.3 Oligomer characterization	106
4.2.4 Synthesis of standard compounds	107
4.3 Results & discussion	108
4.3.1 Quantitative analysis of the GA/G depsipeptides	108
4.3.2 Oligomer distribution under different feeding policies	117
4.3.3 The effect of hydrolysis treatment on amide bonds	127
4.4 Conclusion	133
CHAPTER 5. Conclusion and recommendations	134
5.1 Conclusion	134
5.2 Recommendations	136
5.2.1 Secondary structures and assemblies of depsipeptides	136
5.2.2 Catalytic activity of depsipeptides	138
5.2.3 The rise of homochirality	139
5.2.4 Some other prebiotic conditions	140
5.2.5 Application of the ester-mediated amide bond formation:	141
APPENDIX A. Illustrative examples of the ester-amide exchange model	142
A.1 Simple example	142
A.1.1 Ejection of $P_{0,0}^O$ from parent oligomers $P_{1,1}^O$	142
A.1.2 Growth of $P_{0,2}^O$ with new amide bond from $P_{1,1}^O$	143
A.2 Advanced example	144
A.2.1 Ejection of $P_{0,1}^O$ from parent oligomers $P_{2,2}^O$	144

A.2.2 Growth of $P_{1,3}^o$ with new amide bond from $P_{2,2}^o$ 146

REFERENCES 148

LIST OF TABLES

Table 2.1	– Chemical shifts of assigned resonances observed in HSQC and HMBC spectra of the glycine/lactic acid mixture after 16 cycles. Initial amount of lactic acid (0.02 mmol) was added at 8 th cycle to reactivate the reaction.	52
Table 2.2	– List of peaks used for quantification and their corresponding integrals used in the example shown above.	55
Table 3.1	– The notation of species and the list of polymerization events in the model.	73
Table 3.2	– UV response factors at 210 nm for LA/V oligomers.	87
Table 4.1	– UV response factors at 210 nm of GA/G oligomers.	116

LIST OF FIGURES

Figure 1.1	– The polymerization of amino acids and the assembled structure. Reprinted from Reference [13].	3
Figure 1.2	– Reaction pathway to form amino acids.	4
Figure 1.3	– The typical pathway to polypeptides.	6
Figure 1.4	– The cyanohydrin pathway leads to alpha-hydroxy acid.	9
Figure 1.5	– Comparison between amino acids and hydroxy acids.	10
Figure 1.6	– The simple mechanism of ester-amide exchange reaction.	12
Figure 2.1	– Demonstration of depsipeptide formation from lactic acid (LA) and glycine (G) through ester-amide exchange. (a) Negative electrospray mass spectrum of LA after 4 wet-dry cycles. Polyesters are observed with $n \geq 10$. (b) Mass spectrum of a 1:1 mol LA/G mixture after 4 cycles. The most intense signals correspond to mixed depsipeptides. LA-G depsipeptides are observed up to $n = 10$ (inset).	20
Figure 2.2	– Mass spectrum of the glycine control sample (pH 3.0; 4 cycles).	21
Figure 2.3	– Depsipeptide formation at a lower dry-state temperature of 65°C. Both the ester-linked and amide-linked depsipeptide are observed. CID collision energy = 10 eV (lab-frame).	22
Figure 2.4	– LA/G depsipeptides formation with 10 times dilution after 4 cycles 10 mM of LA and G monomers were used, respectively. MS analysis was performed on Agilent 6130 single quadrupole mass spectrometer. Samples were diluted in ultrapure deionized water to 0.0825 mg mL ⁻¹ and directly injected into the mass spectrometer by electrospray ionization in negative-ion mode.	23
Figure 2.5	– Tandem MS of the LA/G dimer with 100 times dilution after 1 cycle. 1 mM of LA and G monomers were used, respectively. Tandem MS sequencing of 1LA+1G depsipeptide. CID collision energy = 5 eV (lab-frame).	23
Figure 2.6	– FT-IR spectra of (top) glycine, (middle) diglycine, and (bottom) triglycine standards.	25

Figure 2.7	– FT-IR spectra of (top) lactic acid monomer and (bottom) polylactic acid standards.	26
Figure 2.8	– Emergence of amide bonds as demonstrated by IR spectroscopy. Spectra are of (orange) LA + glycine (G) monomers before cycling, (yellow) LA/G mixture after 1 cycle, (light green) LA/G mixture after 8 cycles, and (green) LA/G mixture after 16 cycles, where LA was re-added after the 8 th cycle.	27
Figure 2.9	– Mass spectra of lactic acid and alanine oligomers. Initial ratio of between the two monomers is 1:1. Only the dominant peak in each group is labeled.	28
Figure 2.10	– Amino acid/peptide bond enrichment in depsipeptides following environmental cycling. Mass spectral region shows LA-A pentamers after 1, 4, 8, and 12 cycles.	30
Figure 2.11	– Relative amounts of A incorporated in each oligomer length after different environmental cycles.	30
Figure 2.12	– Mass spectra of LA/A mixtures after 8 cycles with varying mole ratios of starting monomer. Depsipeptide formation and alanine incorporation is optimal when a higher ratio of lactic acid is present. Sodiated alanine monomer clusters, marked with asterisks, are observed in the 90% alanine sample and not the others (e.g., m/z 199.0676 corresponds to $[2(A^-)+Na^+]^-$; theo. m/z = 199.0696), reflecting a higher amount of unreacted alanine monomer.	32
Figure 2.13	The largest oligomers in the LA/A depsipeptides observed using electrospray IM-MS (n = 12–14). These depsipeptides are made using 30% alanine and 70% lactic acid after 8 environmental cycles. Ion mobility traveling wave velocity = 900 m s ⁻¹ . Traveling wave height = 40 V.	33
Figure 2.14	– Investigation on the formation of diketopiperazines (DKPs) in the LA/A system. Experiment was done by positive-ion mode MS (2.50 kV capillary voltage, 40 V sampling cone, 4 V extraction cone). In the above mass spectrum, 25 μM alanine DKP is spiked in the sample.	34
Figure 2.15	– Depsipeptide formation with GG DKP. Tandem MS sequencing of 1LA+1G depsipeptide found in LA + G and LA + G-G DKP samples. CID collision energy = 8 eV (lab-frame).	35

Figure 2.16	– Tandem MS sequencing of LA/A heterodimer (parent ion $[M-H]^- = 160.06$ Da), over 1, 4, 8, and 12 cycles. CID collision energy = 15 eV (lab-frame).	37
Figure 2.17	– Tandem MS sequencing of LA/G depsipeptides after 1 cycle. (a) 1LA+2G and (b) 3LA+1G. CID collision energy = 15 eV (lab-frame). Primary sequences are LA-G-G and LA-LA-G-LA, respectively.	38
Figure 2.18	– Sequencing of depsipeptides by tandem MS. (a) Sequencing of 2LA+4G, theoretical precursor ion $[M-H]^- = 389.13$ Da, formed after 8 cycles. CID collision energy = 20 eV. The primary sequence is LA-G-G-G-G-LA. (b) Sequencing of 2LA+3A, theoretical precursor ion $[M-H]^- = 374.16$ Da, formed after 12 cycles. CID collision energy = 15 eV. The primary sequence is LA-A-A-A-LA. (c) Sequencing of 2LA+2G+1A, theoretical precursor ion $[M-H]^- = 346.13$ Da. Formed after 4 cycles. CID collision energy = 20 eV. LA-A-G-G-LA, LA-G-A-G-LA, and LA-G-G-A-LA sequences are observed.	39
Figure 2.19	– Proposed scheme of the ester-mediated amide bond formation.	42
Figure 2.20	– Proposed scheme of the ester-mediated amide bond formation (continued).	43
Figure 2.21	– Depsipeptide formation at pH 7. pH was adjusted with various bases. For parts (a)-(c), sodiated cluster signals are labeled with asterisks. (a) Mass spectrum of NaOH-adjusted sample after one cycle. (b) Mass spectrum of NH_3 -adjusted sample after one cycle. (c) Mass spectrum of triethylamine-adjusted sample after one cycle. Depsipeptides were readily formed. (d) Tandem MS sequencing of 4LA+1A from pH 7 triethylamine sample, theoretical parent ion $[M-H]^- = 376.12$ Da. Formed after 1 cycle. CID collision energy = 10 eV.	45
Figure 2.22	– LA/A mixture with triethylamine at pH 7. (top) Full mass spectrum after 4 cycles, adjusted to pH 7 with triethylamine. (middle-bottom) Enrichment of alanine-containing depsipeptides from 1 to 4 cycles. The rate of enrichment may be lower than that of unbuffered solutions. Signals with m/z 378.11 correspond to ^{13}C -containing 5LA oligoesters.	46
Figure 2.23	– 1H - ^{13}C HSQC (green) and HMBC (red) spectra of the 16-cycle glycine/lactic acid sample with replenishment of lactic acid at the 8 th cycle. (a) Full spectra. (b-e) Enlarged spectral regions of interest. Red lines indicate the correlation of methylene or carbonyl carbon to the methyl protons of lactic acid. Blue lines	49

show the correlation of methylene protons to two different carbonyl carbons.

- Figure 2.24 – Confirmation of C-terminal amide bond. (a) ^1H - ^{13}C HMBC spectrum of the triglycine standard. Five peaks were assigned based on the correlation between methylene protons and carbonyl carbons. (b) HMBC spectra overlap of the 16 cycles sample (green) and the triglycine standard (red). C-terminus correlations (Peaks 4 and 5) were also observed in the 16-cycle mixture suggested the presence of C-terminal amide bond. 50
- Figure 2.25 – ^1H - ^{13}C HSQC (green) and HMBC (red) spectra of the 1 cycle glycine/lactic acid mixture. 51
- Figure 2.26 – ^1H - ^{13}C HSQC (green) and HMBC (red) spectra of the 8 cycle glycine/lactic acid mixture. 51
- Figure 2.27 – ^1H NMR spectrum of glycine/lactic acid mixture after 16 cycles. Lactic acid was replenished at the 8th cycle. 55
- Figure 2.28 – Quantitative analysis of the LA/G reaction across different cycles. (a) The amounts of glycine monomer, total lactic acid (monomer + residues in products). (b) Overall yield of oligomers. 56
- Figure 2.29 – Mass spectra for the 1:1 mol leucine and lactic acid system after 1 cycle, 4 cycles, and 8 cycles. 58
- Figure 2.30 – Tandem MS sequencing of leucine-lactic acid depsipeptides. (top) Leucine/lactic acid heterodimer sequence after 8 cycles is LA-L. Collision energy = 15 eV. (middle) Two 2LA+2L sequences are observed after 8 cycles, LA-L-L-LA and LA-L-LA-L. Collision energy = 15 eV. (bottom) Two 2LA+3L sequences are observed after 8 cycles, LA-L-L-L-LA and LA-L-L-LA-L. Collision energy = 18 eV. 59
- Figure 2.31 – Mass spectra of lactic acid, glycine, alanine, and leucine (3:1:1:1 mol ratio) after 1, 4, and 8 cycles. Depsipeptides containing various mixtures of the three amino acids are observed, and depsipeptides have higher amino acid content as cycling increases. 60
- Figure 2.32 – Tandem MS sequencing of isobaric 2LA+1G+1A+1L depsipeptides (theo. parent ion $[\text{M-H}]^- = 402.1877$ Da). Formed after 8 cycles. Collision energy = 15 eV. One lactic acid residue is almost exclusively at the N-terminus, and the other lactic acid residue is either at the C-terminus or the residue adjacent to the C-terminus. 12 isobaric sequences could be present: (1) LA-A-G-L-LA, (2) LA-A-G-LA-L, (3) LA-A-L-G-LA, (4) LA-A-L-LA-G, (5) LA-G-A-L-LA, (6) LA-G-A-LA-L, (7) LA-G-L-A-LA, (8) 61

LA-G-L-LA-A, (9) LA-L-A-G-LA, (10) LA-L-A-LA-G, (11) LA-L-G-A-LA, and (12) LA-L-G-LA-A. All potential combinations are shown.

- Figure 2.33 – Mass spectra of serine and lactic acid system after 1 and 8 cycles. # corresponds to loss of CH₂O from serine side-chain in serine-lactic acid dimer. 62
- Figure 2.34 – Tandem MS of serine/lactic acid depsipeptides after 1 cycle. The primary sequences observed are S-LA/LA-S and S-LA-S. Collision energy = 20 eV. 63
- Figure 2.35 – Mass spectra of 1:1 mol GA/G depsipeptides after 1 cycle and after 8 cycles. After 1 cycle, the majority of depsipeptide signals correspond to water-loss species, suggesting they may be cyclic instead of linear. However, after 8 cycles, it appears the depsipeptides preferentially adopt a linear structure over the water-loss species, similar to those formed using lactic acid. 64
- Figure 2.36 – Tandem MS sequencing of GA/G heterodimer after 1 and 8 cycles. Collision energy = 8 eV. After 1 cycle, both the ester and amide-linked heterodimer are present. After 8 cycles, only the amide-linked heterodimer remains. 65
- Figure 2.37 – Mass spectra of LA/A oligomers with different chiral configuration. (top) Zoomed-in mass spectrum of L-alanine and L-lactic acid pentamers after 1 cycle; (middle) zoomed-in mass spectrum of D-alanine and L-lactic acid pentamers after 1 cycle; (bottom) zoomed-in mass spectrum of D-alanine and L-lactic acid pentamers after 4 cycles. 67
- Figure 3.1 – The closed reactor. (a) The oven used to perform the reaction. (b) The blue print of the closed tube reactor. 71
- Figure 3.2 – The formation of $P_{i,j}^O$ from the exchange reaction (Reaction D in Table 3.1). The parent oligomer, P_{i_0,j_0}^O , is attacked by the nucleophile ($P_{r,s}^N$). $P_{i,j}^O$ with a specified number of ester and amide bonds forms either from (1) ejection or (2) growth. The numbers of possible polymer sequences were calculated using binomial coefficients. 76
- Figure 3.3 – The formation of valine/lactic acid depsipeptides in the bent-tube closed reactor. R is the methyl group for lactic acid and R₁ is the isopropyl group for valine. The solution was sealed in a bent glass tube and inserted into an oven through an insulating wall. During the drying process, evaporated water and lactic acid were 83

collected on the cold side of the tube outside the oven. The copolymerization between valine and lactic acid proceeded in the drying mixture (hot side) via esterification and exchange reactions.

Figure 3.4	– HPLC-UV/MS analysis of the LA/V depsipeptide mixture. The sample was prepared from 33.34 mM valine and 100 mM lactic acid, 300 μ l solution. The solution was dried in the closed reactor for 24 hours.	84
Figure 3.5	– Mass spectra of LA/V depsipeptides after HPLC separation.	85
Figure 3.6	– Changes of UV peak area for the two hetero-dimers during hydrolysis. (a) The mixed dimer eluted at 1.5 min. (b) The mixed dimer eluted at 6.5 min. Sample were prepared by drying the monomer mixture at 85°C in the closed reactor for 24 hours.	86
Figure 3.7	– Monomer and oligomer distributions versus time profile in the hydrolysis stage. The reaction temperatures are denoted as different colors: purple (95°C), red (85°C), green (75°C) and blue (65°C). Solid lines are the model predictions. The experimental data points are represented by \square , \diamond , ∇ and \times symbols, respectively.	88
Figure 3.8	– The amount of water left in the drying side of the reaction at different temperatures. Purple (95°C), red (85°C), green (75°C) and blue (65°C). Solid lines are the model predictions. The experimental data points are represented by \square , \diamond , ∇ and \times symbols, respectively.	89
Figure 3.9	– Arrhenius plot of the hydrolysis rate constants evaluated from the lactic acid hydrolysis experiments.	89
Figure 3.10	– Arrhenius plot of the water evaporation rate constants.	90
Figure 3.11	– Oligomer distributions versus time profile in the drying stage. The reaction temperatures are denoted as different colors: purple (95°C), red (85°C), green (75°C) and blue (65°C). Solid lines are the model predictions. The experimental data points are represented by \square , \diamond , ∇ and \times symbols, respectively. Molar ratio (LA/V) is 100/10.	91
Figure 3.12	– The Arrhenius plot of the four rate constants evaluated from the drying experiments. The estimated rate constants at each temperature are plotted along with their 95% confidence intervals.	92
Figure 3.13	– Experimental results and model prediction for low valine loading. The results for different LA/V ratios are denoted as different colors: blue (100/5), green (100/7.5) and red (100/10).	94

Solid lines are the model predictions. The experimental data points are represented by the same colors and by the \times , ∇ and \diamond symbols, respectively. All simulations were calculated by using the rate constants obtained from the 100/10 experiment.

- Figure 3.14 – Experimental results and model prediction for the 100/0 and 100/5 experiment. The results for different LA/V ratios are denoted as different colors: black (100/0) and blue (100/5). Solid lines are the model predictions. The experimental data points are represented by the same colors and by the \diamond and \times symbols, respectively. Optimized rate constants from the data was used for the model prediction of the 100/0 experiment. The simulation for 100/5 ratio experiment was calculated by using the rate constants obtained from the 100/10 experiment. 95
- Figure 3.15 – Experimental results and model prediction for high valine loading. The results for different LA/V ratios are denoted as different colors: red (100/10), purple (100/20) and orange (100/50). Solid lines are the model predictions. The experimental data points are represented by the same colors and by the \diamond , ∇ and \square symbols, respectively. All simulations were calculated by using the rate constants obtained from the 100/10 experiment. 96
- Figure 3.16 – Flux analysis of the studied reactions. The fluxes were calculated from the simulation of 24 hours reaction of 100/10 LA/V mixture at 85°C. The magnitudes of fluxes ($\mu\text{mole/hr}$) are the average rates over 24 hours. The width of each arrow indicates the relative magnitudes of each flux. Only the fluxes of forward reactions are included because the fluxes of hydrolysis and reverse reactions are negligible in the dry state. 99
- Figure 3.17 – Flux analysis for the reactions in 6 hours. The average fluxes ($\mu\text{mole/hr}$) were calculated from the simulation of 6 hr reaction of 100/10 LA/V mixture at 85°C. The width of each arrow indicates the relative magnitudes of each flux. Only fluxes of forward reactions are included because the fluxes of hydrolysis and reverse reactions are negligible in the dry state. 100
- Figure 3.18 – Activation parameters of the ester-amide exchange pathway and the pathway for direct peptide bond formation. All parameter values were estimated for 85°C. 102
- Figure 4.1 – The day-night machine used for the feeding experiment. 106
- Figure 4.2 – Using hydrolysis to simplify the oligomer mixture. 109

Figure 4.3	– HPLC-MS analysis of the GA/G oligomer mixture before the hydrolysis treatment. (a) The full ion chromatogram. (b) Mass spectrum of Peak 3. (c) Mass spectrum of Peak 5. (d) Mass spectrum of Peak 11. (e) Mass spectrum of Peak 12. The sample was prepared by drying a mixture of glycolic acid (100 mM) and glycine (100 mM) at 95°C for 16 days. 20 mM solution of glycolic acid and glycine was added to the dry mixture every 24 hours.	110
Figure 4.4	– HPLC-MS analysis of the GA/G oligomer mixture after the hydrolysis treatment. (a) The full ion chromatogram. (b) Mass spectrum of Peak 3. (c) Mass spectrum of Peak 5. (d) Mass spectrum of Peak 11. (e) Mass spectrum of Peak 12. The sample was prepared by drying a mixture of glycolic acid (100 mM) and glycine (100 mM) at 95°C for 16 days. 20 mM solution of glycolic acid and glycine was added to the dry mixture every 24 hours. The sample was hydrolyzed at 95°C for 48 hours prior to the LC-MS experiment.	111
Figure 4.5	– Log-ratio analysis of the LC-MS data before and after the hydrolysis treatment. The sample was prepared by drying a mixture of glycolic acid (100 mM) and glycine (100 mM) at 95°C for 16 days. 20 mM solution of glycolic acid and glycine was added to the dry mixture every 24 hours.	112
Figure 4.6	– Confirmation of glycine DKP. (a) UV traces. (b) MS positive ion trace. (c) MS signal of DKP from reaction mixture. (d) MS signal of DKP from standard.	114
Figure 4.7	– UV responses curves of 1GA-1G, 1GA-2G and DKP.	115
Figure 4.8	– UV responses curves of 1G, 2G and 3G.	116
Figure 4.9	– Tandem MS sequencing analysis of the hydrolyzed GA/G mixture. Sample was prepared by drying a mixture of glycolic acid (100 mM) and glycine (100 mM) at 95°C for 16 days. 20 mM solution of glycine was added to the dry mixture every 24 hours.	117
Figure 4.10	– Instantaneous yield and conversion in different feeding scenarios. (a) and (b): yields of 1GA-nG, nG (cyan) and DKP (dark blue) over time. (c) and (d): Instantaneous conversion of unreacted glycine monomer (red) and the total yield of products (blue), the sum of 1GA-nG, nG and DKP.	119
Figure 4.11	– Yields of glycine peptide (nG) and DKP in a cycling experiment without glycolic acid. The initial solution was acidified by phosphoric acid.	120

Figure 4.12	– Chain length distribution of 1GA-nG under different feeding policies. The y-axis is the number of glycine units in each oligomer.	122
Figure 4.13	– The number of oligomer chains in 1GA-nG.	123
Figure 4.14	– Evaporation rate of lactic acid and glycolic acid at 85°C. The evaporation rates were determined by using the closed reactor as described in Chapter 3. The concentrations of lactic acid and glycolic acid were measured by ¹ H NMR.	123
Figure 4.15	– The largest oligomers observed. The sample was prepared by drying a mixture of glycolic acid (100 mM) and glycine (100 mM) at 95°C for 16 days. 20 mM solution of glycine was added to the dry mixture every 24 hours.	126
Figure 4.16	– MALDI-MS analysis. The sample was prepared by drying a mixture of glycolic acid (100 mM) and glycine (100 mM) at 95°C for 24 days. 20 mM solution of glycine was added to the dry mixture every 24 hours.	126
Figure 4.17	– Hydrolysis rate of 1GA-2G at 95°C. Solid lines are the model predictions. The experimental data are represented by the ■ symbols. Experimental data for 1GA is unavailable due to its low UV response. Only its model prediction is shown and it is not included in the fitting process.	129
Figure 4.18	– Estimated oligomer distribution before the hydrolysis treatment.	131
Figure 4.19	– Average number of amide bonds in each oligomer distribution. The results of the two feeding policies are represented by red (1:1) and blue (0:1). Lines with light and strong colors represent the experimental and estimated distribution, respectively.	132
Figure A.1	– The ejection formation of $P_{0,0}^O$ from $P_{1,1}^O$ by a $P_{0,0}^N$.	143
Figure A.2	– The growth of $P_{0,2}^O$ from $P_{1,1}^O$ by a $P_{0,0}^N$.	144
Figure A.3	– The ejection formation of $P_{0,1}^O$ from $P_{2,2}^O$ by a $P_{0,1}^N$.	146
Figure A.4	– The growth of $P_{1,3}^O$ from $P_{2,2}^O$ by a $P_{0,1}^N$.	147

LIST OF ABBREVIATIONS AND NOMENCLATURE

Abbreviations

A, Ala	Alanine
ACN	Acetonitrile
ATR-FT-IR	Attenuated total reflection Fourier transform infrared spectroscopy
C18	Octadecyl carbon chain bound silica
DKP	Diketopiperazine
ESI	Electrospray ionization
GA	Glycolic acid
G, Gly	Glycine
HILIC	Hydrophilic interaction chromatography
HPLC	High-performance liquid chromatography
KHP	Potassium hydrogen phthalate
L, Leu	Leucine
LA	Lactic acid
LC-MS	Liquid chromatography-mass spectrometry
MALDI	Matrix-assisted laser desorption/ionization
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
S	Serine
V	Valine

Nomenclature

k_{11}	Rate constant for the esterification between two hydroxy-terminated oligomers ($\text{L}\cdot\text{mol}^{-1}\cdot\text{h}^{-1}$)
k_{12}	Rate constant for the esterification between hydroxy-terminated and amine-terminated oligomers ($\text{L}\cdot\text{mol}^{-1}\cdot\text{h}^{-1}$)
k_c	Cyclization rate constant to form DKP
k_{h1}	Rate constant for the hydrolysis of ester ($\text{L}\cdot\text{mol}^{-1}\cdot\text{h}^{-1}$)
k_{ha}	Amide bond hydrolysis rate constant ($\text{L}\cdot\text{mol}^{-1}\cdot\text{h}^{-1}$)
k_e	Rate constant for the exchange reaction ($\text{L}\cdot\text{mol}^{-1}\cdot\text{h}^{-1}$)
K_{pLA}	Rate constant for the evaporation of lactic acid ($\mu\text{mole}\cdot\text{cm}^{-1}\cdot\text{h}^{-1}$)
K_{pw}	Rate constant for the evaporation of water ($\mu\text{mole}\cdot\text{cm}^{-1}\cdot\text{h}^{-1}$)
$N_{w,z}$	Flux of water ($\mu\text{mole}\cdot\text{h}^{-1}\cdot\text{cm}^{-2}$)
$P_{i,j}^O$	Hydroxy-terminated oligomers with i ester bonds and j amide bonds (μmole)
$P_{i,j}^N$	Amine-terminated oligomers with i ester bonds and j amide bonds (μmole)
P_i^O	Hydroxy-terminated oligomers with i ester bonds (μmole)
P_i^N	Amine-terminated oligomers with i ester bonds (μmole)
P_{air}	Partial pressure of air (kPa)
P_w^*	Vapor pressure of water (kPa)
$R_{v,w}$	Evaporation of water ($\mu\text{mole}\cdot\text{h}^{-1}$)
$R_{v,LA}$	Evaporation rate of lactic acid ($\mu\text{mole}\cdot\text{h}^{-1}$)
S_A	Cross-section area (cm^2)
T_H	Drying temperature ($^{\circ}\text{C}$)
V	System volume (μL)
W	Water (μmole)

- x_w Liquid mole fraction of water
- x_{LA} Mole fraction of lactic acid in liquid
- $y_{w,1}, y_{w,2}$ Gas mole fraction water at (1) liquid-gas interface and (2) the top of high temperature region
- z_1 Liquid height (cm)
- z_2 Length of tube inside oven (cm)

SUMMARY

The formation of polypeptides on the early Earth has been a long-standing problem in the field of prebiotic chemistry. Although it is generally accepted that amino acids were present on the prebiotic Earth, the plausible mechanism to form long chain polypeptides is still unclear. Because of the high activation energies and the formation of side-products, direct peptide bond formation is slow unless high temperature or activating agents are used.

This thesis describes a simple system composed of hydroxy acids and amino acids that is capable of forming peptide bonds under mild conditions. Hydroxy acids form metastable oligoesters in the oscillating (hot dry/cool wet) environment and transform into mixed copolymers via the ester-amide exchange reaction. This pathway enables the amide bond formation in lower temperature and leads to a library of oligomers with random sequences when multiple amino acids mix together.

To further understand the kinetic behavior of this copolymerization, the growth of initial species from a valine/ lactic acid mixture in a closed system reactor was tracked. A mathematical model was developed to simulate the reactions and evaluate the rate constants at different temperatures. These reactions can be described by the empirical Arrhenius equation even when the reaction occurred in the solid (dry) state. Further calculations for activation parameters showed that the ester-mediated pathway facilitates amide bond formation by lowering activation entropies.

In the final part of this thesis, fresh monomers were added to the oligomer mixture periodically in the wet-dry cycling. The feeding composition was found to affect the

growth rate of oligomer chain length. The combination of the esterification and ester-amide exchange is similar to the behavior of the living polymerization. Adding more hydroxy acids creates more active chains, but does not necessary elongate the oligomer chains. The average chain length grows more rapidly when the number of active chains is limited.

The results from this thesis demonstrate a simple system exhibiting key features for the emergence of peptides at mild conditions and provide a theoretical framework that illustrates why the ester-mediated pathway would have been more favorable on the early Earth, compared to peptide bond formation without the aid of hydroxy acids.

CHAPTER 1. INTRODUCTION

1.1 A brief review of prebiotic chemistry

The origin of life on our Earth is a long-standing mystery. Even with the advances in technology, how non-living matter transformed into living organisms is still a mystery to us. Back to 1920s, Oparin proposed that simple molecules on the early Earth might have undergone a series of chemical reactions and become complex biomolecules we have today [1]. Oparin himself did not have the opportunity to test this idea. Until sixty years ago, Miller and Urey started the field of prebiotic chemistry by a simple experiment [2]. They applied electric discharge to a mixture of simple inorganics like water, methane and ammonia. After a few days of reaction, various organic compounds related to our life emerged. Their simulation of early Earth, along with Watson and Crick's discovery of the DNA double helix, initiated scientists to find a possible route leading to the formation of simple life forms based on chemistry [3].

However, even in the simplest life form today, biopolymers with different functions such as DNA, RNA, sugar, proteins and lipids all work together to keep the machine of life going. One example is the replication of living organisms. In modern biology, the genetic information stored in DNA is first transcribed to RNA. With the help of the ribosome, the information produces proteins that act as enzymes. Those enzymes made by proteins are necessary to make each step happen. It leads to a simple chicken and egg problem: if every component of life depends on each other, how did they first appear on our planet [4]?

1.1.1 Hypothesis for origin of life

The concept of primordial soup proposed by Oparin is a major theory dominating the field of prebiotic chemistry [5]. Simple compounds formed from the early atmosphere might have accumulated in various locations like a small pond and eventually form some of the early biopolymers. So far, the RNA world hypothesis is no doubt the most famous scenario [6, 7]. RNA can potentially serve two roles: storing information and performing catalysis [4, 8]. The non-enzymatic polymerization of nucleotides would lead to RNA with random sequences. Some of the early RNA with the right sequences might catalyze the copying and reproduction of RNA itself. Through this natural selection process, functional RNA could sustain and eventually lead to the DNA/RNA/protein world today.

While the RNA world hypothesis greatly simplifies the mission to solve the origin of life mystery, the non-enzymatic production of RNA is still challenging. Each nucleotide of RNA is made by a ribose sugar, a base and a phosphate group. Those building blocks are connected by phosphate bonds in a specific order. While the prebiotic production of each component has been proposed [9, 10], polymerization of nucleotides still remains as a challenge.

On the other hand, the non-enzymatic synthesis of polypeptides seems to be an easier case. In polypeptides, small building blocks known as amino acids connect to each other by peptide bonds. Ideally, the polycondensation of amino acids should produce polypeptides directly. When sufficient amounts of amino acids are arranged in a specific order, polypeptides can form higher order structures such as alpha-helices and beta-sheets by hydrogen-bonding as summarized in Figure 1.1. Those special structures allow

polypeptides to coordinate reactants in certain orders and to act as catalysts for many reactions [11]. The peptide bond is also highly stable. Its half-life was estimated to over 200 years at pH 7 [12]. Therefore, it seems to be more plausible to accumulate polypeptides on the early Earth.

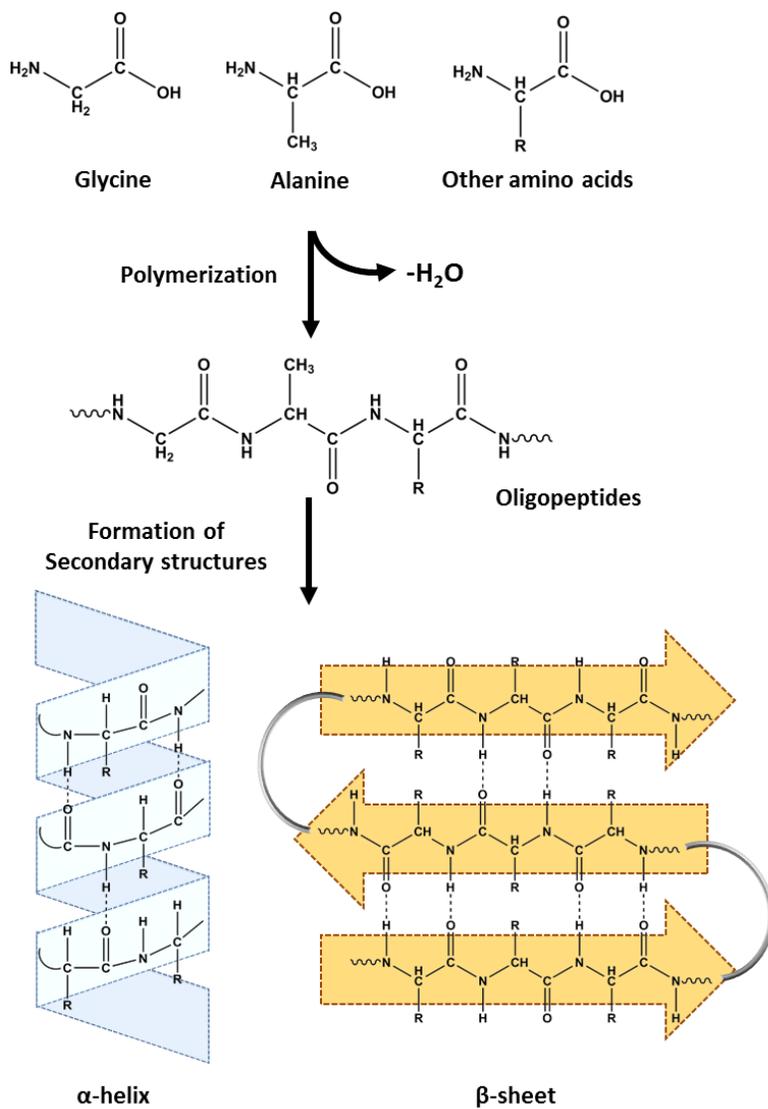


Figure 1.1 – The polymerization of amino acids and the assembled structure. Reprinted from Reference [13].

1.1.2 Building blocks of biopolymers

Today, twenty different amino acids are used to make proteins in most of life on Earth. The major difference between each amino acid is the side-group on the alpha-carbon. The original Miller-Urey experiment has identified simple amino acids, such as glycine, alanine and aspartic acid [2]. Generally, the amino acids in the Miller-Urey experiment were synthesized through the Strecker reaction pathway as shown in Figure 1.2 [14].

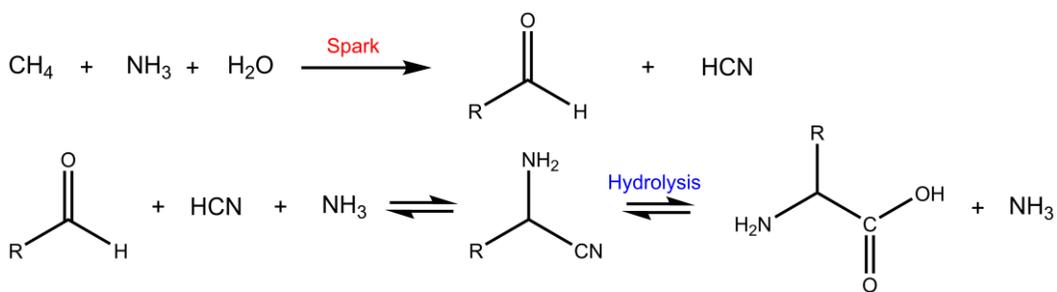


Figure 1.2 – Reaction pathway to form amino acids.

In a more recent research, the samples from Miller-Urey experiment were analyzed again by state of the art instruments. The results show the mixture contains more species than what had been identified sixty years ago. More complicated amino acids like serine and leucine have also been found [15, 16]. Besides the Miller-Urey mixture, a mixture of amino acids was also found in meteorites such as the one that fell near Murchison, Australia. Some of the building blocks could have been delivered to the early Earth by extraterrestrial sources [17, 18]. The side-groups of amino acids dictate the potential secondary structures. For example, some amino acids, like methionine, alanine, leucine, glutamate, and lysine, have higher tendency to form alpha-helices [19]. When more amino acids are allowed to reacted together, some functional polymers might emerge from the random sequences library.

1.2 Challenges of the peptide problem

1.2.1 *How peptides are generated nowadays*

In living organisms, polypeptides are synthesized by ribosomes combined with mRNA [20]. Each amino acid is attached its corresponding tRNA using the energy from ATP. When the mRNA is read by the ribosome, tRNA brings an amino acid into the ribosome and links each amino acids by a peptide bond. By such machinery, ribosome assembles amino acids into proteins with a correct order.

In synthetic biology, in order to synthesize artificial polypeptides with the same accuracy as ribosomes, solid-phase peptide synthesis was developed [21]. The basic principle is the repeated cycling of deprotection-wash-coupling-wash. The amine groups of the amino acids are first protected. Condensing agents such as carbodiimides are used to form peptide bonds between the amino acid carboxylate group and the free amine group on the solid support. The protected amine group is then deprotected and used to connect another amino acid in the next cycle. This method allows the synthesis of artificial peptides with high sequence accuracy, but it requires significant amounts of organic solvents and activating agents. The product yield is also generally low when longer polypeptide is desired [22].

The complex mechanism used by the ribosomes and the solid-phase synthesis are unlikely to have been present on the early Earth. The primary goal in this field is to seek other simple chemistries that can form peptides in a prebiotically reasonable condition.

1.2.2 Challenge of the polycondensation of amino acids

The simplest path to form polypeptides is the polycondensation of amino acids, but this reaction is difficult to happen in general. Figure 1.3 summarizes the general pathways for the polycondensation of amino acids. This reaction is thermodynamically unfavorable in aqueous solution [23]. Fox and Harada first used a simple drying experiment to produce copolymers of amino acids, but it requires high reaction temperature, generally ranging from 150°C to 200°C [24, 25]. In addition, diketopiperazine (DKP) can form easily from the cyclization of dipeptides and become a major side-product that prevents further polymerization [26, 27].

Also, the amino acids can be unstable at high temperature. Even the simplest amino acid, glycine, can be decarboxylated into methylamine [28]. Under such condition, amino acids with more complex side-groups are less likely to survive. Recently, Rodriguez-Garcia *et al.* re-investigated the polymerization of amino acids under drying-rehydration condition [29]. Their study shows the polymerization of glycine could occur at temperatures as low as 90°C under highly acidic or basic conditions, but this condition might still be challenging if one considers the early condition of Earth and the stability of monomers.

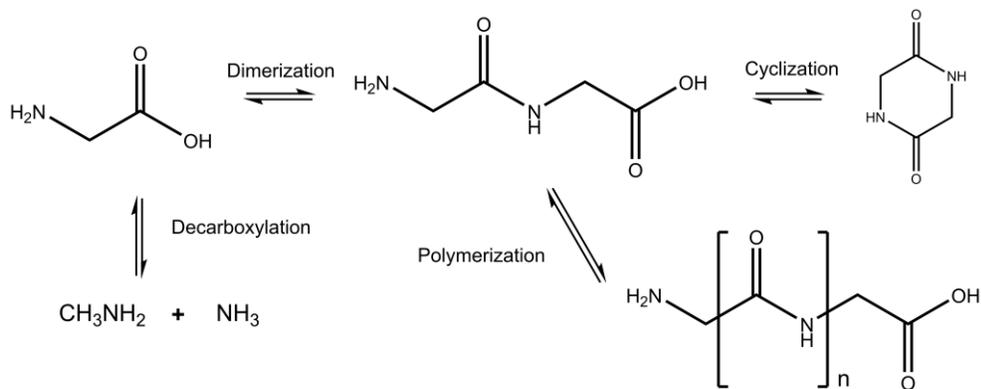


Figure 1.3 – The typical pathway to polypeptides.

1.2.3 Previous works for the prebiotic synthesis of peptides

Because the direct polycondensation of amino acids by drying is difficult under mild conditions, various scenarios catalysts were considered. Minerals have been considered as heterogeneous catalysts for peptide polymerization. Several types of mineral including kaolinite [26], silica and alumina [30, 31] have been investigated based on their abundance on Earth. The catalytic effect of minerals have been proposed through the adsorption and alignment of amino acids on the mineral surface [32]. Besides minerals, Rode *et al.* also used NaCl and Cu (II) salt to catalyze peptide formation [33, 34]. The catalytic activity is achieved by the dehydrating effect of NaCl and the complex of Cu (II) and amino acids. They also demonstrated the combination of minerals and salt catalysts leads to longer peptides up to 6-mers [35].

Alternative environments other than drying-rehydration have also been investigated. Submarine hydrothermal vents have been recognized as a potential candidate because of their high-pressure and high-temperature environment. Imai *et al.* used a flow reactor to simulate the submarine hydrothermal vents where the thermal synthesis of oligomers was achieved in a hot and pressurized region [36]. However, even under such harsh condition, the yield is still low and a significant amount of DKP forms. Also, oligomers longer than trimers are hard to synthesize because of hydrolysis. Actually, considering the typical time that reactants are exposed in the hydrothermal region, most of the peptides would depolymerize by hydrolysis and decomposition [37]. A more recent study used the air-water interface as an alternative hydrophobic environment to promote peptide bond formation [38]. However, they started with amino acid esters instead of amino acids. The

generation of amino acid esters is difficult by itself in aqueous solution because it requires the removal of water.

Another route to synthesize peptides is through the chemical activation of amino acids. The activation can be achieved by using condensing agents such as carbonyl imidazole [39], carbonyl sulfide [40] and nitrogen oxide [41, 42]. While the activation by carbonyl imidazole is efficient, its availability the early Earth is questionable considering its high reactivity. Carbonyl sulfide facilitates the peptide synthesis via the formation of N-carboxyanhydride intermediates. They are simple molecules, but their abundance on the early Earth depends on the surrounding environment. Sufficient amounts of carbonyl sulfide might be available if the reaction took place near volcanic sites. Nitrogen oxides were proposed to activate amino acids in repeated wet/dry cycles. This route only happens if sufficient amounts of nitrogen oxides and oxygen could present in the early atmosphere. In summary, the activation of amino acids depends on the sporadic bursts of condensing agents. Whether or not sufficient amounts of these active reactants could have been formed is still problematic.

1.3 Proto-biopolymers

The challenge of directly forming biopolymers might imply some simpler systems preceded initially. For RNA, alternative structures like glyoxylate [43] and peptide linkages [44] have been proposed due to the difficulty to connect nucleotides by phosphate bonds directly. Polyesters were suggested as one possible precursor for polypeptides [45]. The previous report from our group has shown that polyesters can form from alpha-hydroxy acids under repeat wet-dry cycles [46]. Although the role of esters in modern biochemistry

is not as important as peptides, the possibility to form longer peptides by esters deserves more detailed investigation.

1.3.1 Hydroxy acid-based polyester

Hydroxy acids are another building blocks in the Miller-Urey mixture [47, 48]. They have similar chemical structures as the amino acids, except for the hydroxy group. Similar to amino acids, hydroxy acids can be synthesized from ketone or aldehyde precursors, but via the cyanohydrin pathway in the electric discharge experiment (Figure 1.4) [14]. The relative abundance between amino acids and their hydroxy acid counterparts is similar, although it could depend on the pH and the concentration of NH_3 in the reaction medium [14, 49]. Hydroxy acids were also found in the Murchison meteorite and in simulated volcanic condition with similar abundance to amino acids [50-53].

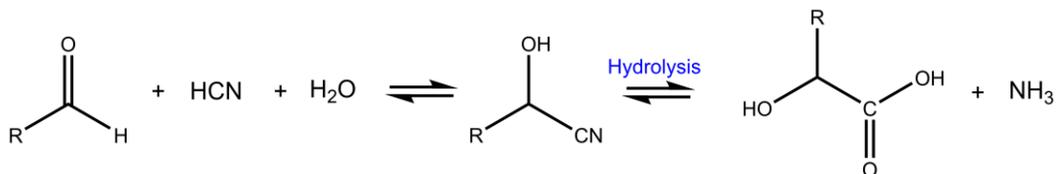


Figure 1.4 – The cyanohydrin pathway leads to alpha-hydroxy acid.

In the Miller-Urey mixtures and meteorite samples, the reaction seems to favor the formation of simple amino acids like glycine and alanine. Their counterparts, glycolic acid and lactic acid, were also the most abundant hydroxy acids. Although more complex monomers with other side-groups were also detected, their abundance is typically much lower than those simple building blocks.

While the chemical structures of amino acids and hydroxy acids are similar, the reactivity for polymerization is very different as summarized in Figure 1.5. Esterification

typically has lower bond formation energy compared to amide bond formation [54]. Hydroxy acids also form cyclic dimers, but unlike DKP, cyclic dimers of hydroxy acids like lactide are less stable toward hydrolysis. Therefore, they can be recycled by gentle hydrolysis.

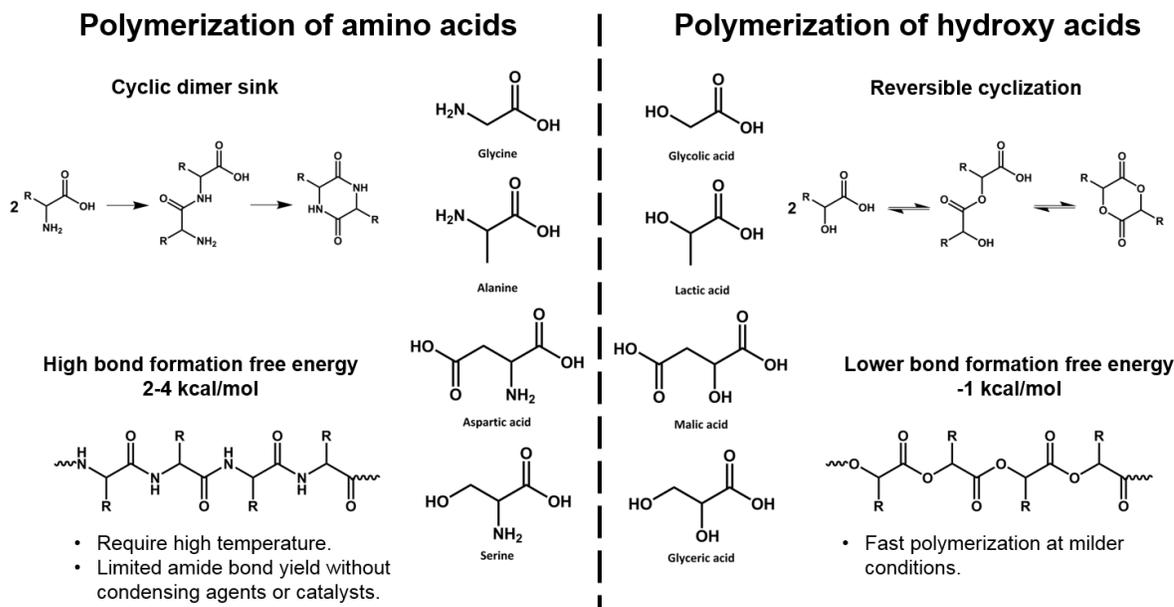


Figure 1.5 – Comparison between amino acids and hydroxy acids.

In the modern industry, simple hydroxy acids like glycolic acid and lactic acid have been widely used to synthesize polyesters. Hydroxy acids can form polyesters through polycondensation or ring-opening polymerization very easily [55]. Hydroxy acid-based polyesters like high molecular weight polylactic acid are well-behaved thermoplastics that have similar properties to polystyrene. It can be slowly degraded by ester bond hydrolysis. Therefore, polyesters have been used as biodegradable polymers to replace petroleum-based plastics [56]. Polylactic acids also received attention in the field of medical application because of its biocompatibility. It is one of the major materials for tissue scaffolds [57, 58]. Nanoparticles made of poly(glycol-lactic) acids have also been tested as

drug delivery vehicles [59-61]. Compare to polypeptides, synthesis of polyester from hydroxy acids received much more attention and the pathways leading to long chain polymer are more heavily studied.

1.3.2 Depsipeptide formation by the ester-amide exchange reaction

Along with the development of polyester-based materials, other chemistries like amidation have been used to modify the properties of polymers [62]. In the amidation or ester-amide exchange reaction, the ester groups are replaced by free amine groups into more stable amide bonds. This simple reaction has been used to improve the wettability, dyeability and other surface properties of polyesters fibers [63, 64].

As mentioned before, the formation of polyester from hydroxy acids can occur easily through a dehydration process. It is possible when amino acids and hydroxy acids are mixed together as they are in the Miller-Urey mixture, hydroxy acids form metastable oligoesters first and transform into mixed copolymers via the ester-amide exchange reaction (Figure 1.6). This reaction can produce copolymers of hydroxy acids and amino acids, also known as depsipeptides.

Wet-dry cycling is a physical environment that has been used before to promote amino acid polymerization via the dehydration process. The rehydration phase of the wet-dry cycle, corresponding to the rain and dew on the early Earth, would re-suspend dry materials in each cycle. This will not only re-mix the products, but also provide a route for gentle hydrolysis and selection. While the polycondensation of amino acid requires a high-temperature environment as mentioned before, the formation of polyester can occur at moderate temperatures from 65°C to 95°C [46]. Therefore, it is worth knowing if the

hydroxy acid can promote peptide bond formation in a mild condition such as day-night cycling.

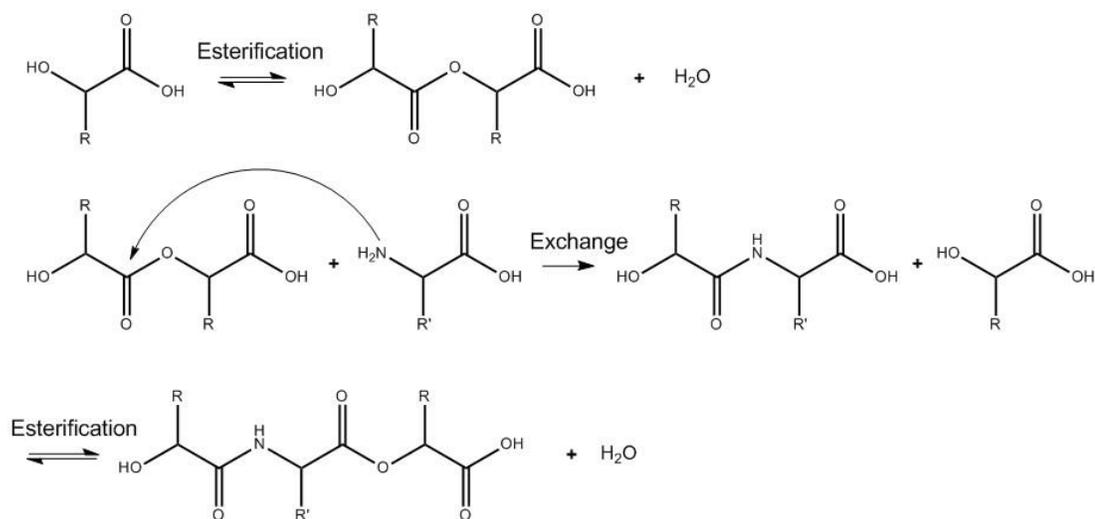


Figure 1.6 – The simple mechanism of ester-amide exchange reaction.

Outside the field of prebiotic chemistry, depsipeptides have been studied for their unique properties and application. For example, cyclic depsipeptides such as Kahalalide F that originates from the marine mollusk exhibits anti-tumor activity [65, 66]. Depsipeptides have also been used as alternatives to current polyester based materials and drug delivery vehicles [67]. By introducing amino acids with reactive side-chains into polyesters, the stability, solubility and thermal transition behavior of polyesters can be adjusted [68]. The reactive side-chain further allows drugs to be covalently linked in the depsipeptide-based nanoparticles [69, 70].

1.4 Objectives

In this thesis, I will address the challenge of non-enzymatic peptide synthesis under mild conditions. From a mixture of hydroxy acids/amino acids, the potential precursor of modern polypeptides can form through esterification/ester-amide exchange reaction.

In Chapter 2, I will describe the procedure we used to synthesize depsipeptides through repeated day-night cycling. Detailed analysis of the complex mixture of depsipeptides is accomplished by mass spectrometry, infrared spectroscopy and NMR spectroscopy.

In Chapter 3, the focus is how to develop a model to describe the copolymerization process. The effects of temperature and reaction time were studied to evaluate the major pathway of the copolymerization. The results provide a theoretical framework to explain why the ester-mediated pathway could have been more favorable on the early Earth.

In Chapter 4, I will explore the possibility of forming long chain depsipeptides by feeding of monomers continuously. It describes a procedure to simplify and quantitate the complex reaction mixtures. The effect of feeding composition on the growth of chain length is evaluated.

Chapter 5 contains the summary of my work and my recommendation to further explore the prebiotic chemistry of depsipeptides.

CHAPTER 2. SYNTHESIS OF PROTO-PEPTIDES FROM ESTER-AMIDE EXCHANGE REACTION

2.1 Introduction

As discussed in the previous chapter, the non-enzymatic synthesis of peptides in the context of prebiotic chemistry remains a challenge. We hypothesize that a mixture of hydroxy acids and amino acids would produce oligomers containing both ester and amide linkages through esterification and ester-amide exchange reactions. The high tendency of hydroxy acids to form polyesters should allow the formation of depsipeptides under mild conditions such as the day-night environmental cycles on the early Earth.

In this chapter, we provide our initial survey on the depsipeptide synthesis from the ester-amide exchange reaction in repeated wet-dry cycles. This chapter discusses a proof-of-principle experiment to demonstrate the efficiency of the ester-mediated pathway. The copolymerization between hydroxy acids and amino acids usually forms a complex and large library of products, which are difficult to be analyzed by conventional methods. Mass spectrometry (MS) provides detailed analysis on the overall oligomer and sequence distribution. Infrared spectroscopy and two-dimensional NMR spectroscopy also confirm the presence of amide bonds in the products. The ability of the ester-mediated pathway to form peptide bonds under different conditions, including temperature, concentration, side-groups and chirality of amino acids, is also explored.

Procedure and results in this chapter are reprinted with permission from [71].

2.2 Experimental method

2.2.1 Materials

Glycine, L-lactic acid, glycolic acid, L-alanine, D-alanine, L-leucine, L-serine, L-phenylalanine, triglycine, hydrochloric acid, sodium hydroxide, ammonium hydroxide, triethylamine, and deuterium oxide (99.9 mol %) were obtained from Sigma-Aldrich. Potassium hydrogen phthalate was from Fisher Chemical. Ultra-pure water (18.2 M Ω /cm) was from a Barnstead Nanopure Diamond system (Van Nuys, CA).

2.2.2 Reaction condition: Day-night cycling

Typical reactions were started with 200 μ L aqueous solution of L-lactic acid (100 mM) and of amino acid (100 mM). In the warm/day stage, the solution was allowed to dry at 85°C for 18 hours. In the following cool/night stage, the dried samples were rehydrated with 200 μ L of deionized water for 30 minutes. at room temperature and capped at 65°C for 5.5 hours, resulting in one 24-hour cycle. This process was repeated for the reported number of cycles. Before MS and/or NMR analysis, dried samples obtained after the warm/dry stage were rehydrated and diluted to the desired concentrations in either ultrapure water (MS) or D₂O (NMR), except for the lactic acid control sample which was dissolved in 90% acetonitrile and 10% water. No filtration was used for any sample.

For the glycine-only control experiment, solution pH was acidified using 0.5 M HCl in order to mimic the acidic environment of lactic acid. For pH-adjusted experiments, the pH of 100 mM L-lactic acid and 100 mM L-alanine aqueous solution was raised using 1 M sodium hydroxide, 1 M ammonium hydroxide, or 0.2 M triethylamine to pH 5, 7, or 9.

After the drying process, reaction mixtures were rehydrated and the pH of the solution was adjusted to the initial value.

2.2.3 *Mass spectrometry analysis*

MS analysis was performed on Waters Synapt G2 HDMS traveling-wave ion mobility and Waters Xevo G2 mass spectrometers, both equipped with a quadrupole/time-of-flight (qTOF) mass analyzer (Resolution mode; $m/\Delta m = 20,000$ FWHM). Samples were diluted in ultrapure deionized water to concentrations ranging from 0.013 to 0.044 mg mL⁻¹ (of starting monomers, concentrations of individual oligomers are lower) and directly infused into the mass spectrometer by electrospray ionization (10 μ L min⁻¹ for 2 min) in negative-ion mode. For all samples, the capillary was set at 2.00 kV, the sampling cone voltage was 30, the extraction cone voltage was 3.0, the source temperature was 90°C, the cone gas flow was 20 L hr⁻¹, the desolvation gas flow was 650 L hr⁻¹, and the desolvation temperature was 250°C. The TOF was calibrated daily using sodium formate clusters from m/z 50 to 1200 with a threshold of 1 ppm. For accurate mass data (<10 ppm), peaks were centered based upon area and LockMass correction was applied using leucine enkephalin ($[M-H]^- = m/z$ 554.2615). Tandem MS was performed using collision-induced dissociation (CID) using argon gas in the transfer collision cell (Synapt G2). All specified collision energies are given in electron volts (eV) and are lab-frame. When ion mobility separation was used, the wave velocity ranged from 550–900 m s⁻¹, the wave amplitude was set at 40 V, and the mobility gas flow was 90 mL min⁻¹ N₂ (corresponding to a pressure of 3.67 mbar in the cell). Ion mobility data extraction was performed using Driftscope v2.1 software. Spectra were plotted in OriginPro 9.0 software.

Unless otherwise noted, all depsipeptide peak assignments correspond to $[M-H]^-$ ions. In traditional proteomic analysis, proteolytic peptides are commonly analyzed using positive-ion mode. However, because the depsipeptides studied here contain an -OH instead of an -NH₂ on the “N” terminus, we found the sensitivity to be lower in positive mode. Additionally, tandem MS sequencing was more effective in negative-ion mode as essentially all fragment ions contained the C-terminus. This is consistent with literature on negative-ion mode peptide fragmentation (“alpha-cleavages”) [72]. Several DKP-like cyclic by-products could only be detected in positive-ion mode and not in negative-ion mode.

For the data given in Figure 2.11, ¹³C isotopic deconvolution was performed manually in order to determine the relative amounts of alanine and lactic acid within each oligomer. The mass difference between lactic acid and alanine residues is 0.9840 Da, and thus X+1 isotopic signal from one combination of lactic acid and alanine (e.g., 1LA+2A) often overlapped with subsequent combinations of lactic acid and alanine (e.g., 2LA+1A). In order to quantitatively correct for this effect, signals were integrated and peak areas were obtained for each combination of lactic acid and alanine. Based upon the number of carbons, the relative contribution of ¹³C in the mass spectrum was determined (1.109% per carbon) and subtracted out, resulting in the net peak area from each unique combination of lactic acid and alanine. These areas were summed for each oligomer (e.g., $n = 3$) and the relative amount of each combination was calculated. The average amounts of alanine and lactic acid were determined by multiplying each percentage by the number of residues within the depsipeptide. This process does assume that differences in ionization efficiency between combinations of lactic acid and alanine of the same size (e.g., all $n = 3$ LA+A depsipeptides)

are negligible. If positive-ion mode had been used, this assumption may not be valid, as backbone nitrogen atoms can be protonated. However, in negative-ion mode, we find it unlikely for deprotonation to persist anywhere other than the C-terminus [73] (that is, without collisional activation), especially considering the side-chains of lactic acid, alanine, and glycine. An example isotopic correction calculation is below:

Determining relative amounts of LA+A trimer ($n = 3$) for 12 cycles.

m/z #1 (corresponding to 1LA+2A; $C_9H_{15}N_2O_5^-$) = 231.10; peak area = 1.37E6. ^{13}C contribution = 0; net area = 1.37E6.

m/z #2 (corresponding to 2LA+1A; $C_9H_{14}NO_6^-$) = 232.10; peak area = 1.46E5. ^{13}C contribution = (# carbon atoms)*(^{13}C relative isotopic abundance)*(area) = (9)*(0.01109)*(1.37E6) = 1.37E5; net area = 1.46E5 - 1.37E5 = 9.26E3.

Sum area (all $n = 3$ oligomers) = 1.37E6 + 1.37E5 + 9.26E3 = 1.52E6.

$$\mathbf{1LA+2A} = (1.37E6 + 1.37E5)/1.52E6 = \mathbf{99.4\%}$$

$$\mathbf{2LA+1A} = 9.26E3/1.52E6 = \mathbf{0.6\%}$$

$$\mathbf{\%A} = 0.994*(2/3 \text{ residues}) + 0.006*(1/3 \text{ residues}) = \mathbf{66\%}$$

$$\mathbf{\%LA} = 0.994*(1/3 \text{ residues}) + 0.006*(2/3 \text{ residues}) = \mathbf{34\%}$$

Therefore, for $n = 3$ depsipeptides containing alanine and lactic acid, the relative abundance of alanine is 66% and the relative abundance of lactic acid is 34%.

2.2.4 FT-IR analysis

IR data was obtained on a Thermo Nicolet 4700 FT-IR Spectrometer. Dry samples after the final cycles were collected from the reactor, and then were pressed directly on the Attenuated Total Reflectance (ATR) crystal. Spectra were background-subtracted from

400 to 4000 cm^{-1} and signal-averaged (16 scans per spectrum). Data processing (normalization and subtraction) was performed using OriginPro 9.0 software.

2.2.5 NMR analysis

The NMR spectra were recorded by a Bruker Avance II-500 spectrometer. Dried samples obtained after the last warm/dry period were rehydrated in 600 μL D_2O . For quantitative ^1H NMR analysis, a 1 mm capillary with potassium hydrogen phthalate (KHP) D_2O solution was used as external standard in the 5 mm diameter NMR tube, and data were collected using 30 degree pulse program with 30 seconds relaxation delay. HSQC and HMBC data were obtained by standard pulse program, hsqcetgpsi2 and hmbcgp1pndqf, with 1.5 seconds relaxation delay [74]. Spectra width of ^1H is 3501.401 Hz. For ^{13}C , the spectra width is 25153.822 Hz. All spectra were processed and plotted by MestReNova 9.1.

2.3 Results & discussion

2.3.1 Proof of principle experiment

For our initial experiments, we examined the capability of our ester-mediated mechanism by studying the copolymerization between lactic acid and glycine. When L-lactic acid (LA) alone is subjected to 4 wet-dry cycles (dry phase 85°C , 18 hr, wet phase 65°C , 5.5 hr) linear polyesters were formed (Figure 2.1(a)). When glycine (G) is mixed with LA in a 1:1 mole ratio and subjected to the same wet-dry cycles, essentially all mass spectral signals correspond to depsipeptides (Figure 2.1(b)). G-containing depsipeptides of up to 10 residues in length are observed, containing varying amounts of amide and ester linkages (Figure 2.1(b), inset).

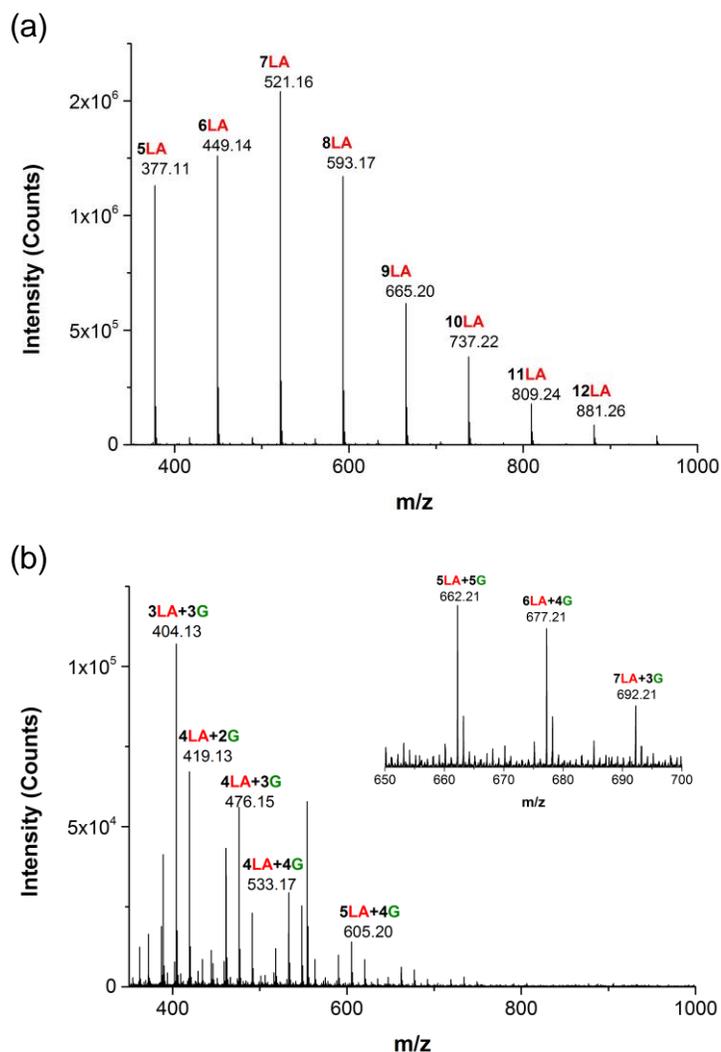


Figure 2.1 – Demonstration of desipeptide formation from lactic acid (LA) and glycine (G) through ester-amide exchange. (a) Negative electrospray mass spectrum of LA after 4 wet-dry cycles. Polyesters are observed with $n \geq 10$. (b) Mass spectrum of a 1:1 mol LA/G mixture after 4 cycles. The most intense signals correspond to mixed desipeptides. LA-G desipeptides are observed up to $n = 10$ (inset).

In contrast, when glycine alone is subjected to 4 wet-dry cycles (pH = 3.0 by HCl), no peptides are observed in the mass spectrum (Figure 2.2). Glycine monomer is present but no glycine peptides are detected at m/z 131.0458 (di-Gly), 188.0673 (tri-Gly), or 245.0888 (tetra-Gly). The majority of signals in this m/z range correspond to trace contaminants

found in the solvent blank, including m/z 157.1241, 171.1389, 185.1572, 194.0496, 199.1718, 213.1869, 227.2039, 255.2346, 283.2603, and 297.1497. For example, we tentatively identify m/z 255.2346 as palmitic acid, a common ESI contaminant with a theoretical $[M-H]^-$ of 255.2330 Da (~ 6 ppm error). However, it is also a possible side-product in low abundance from environmental cycling.

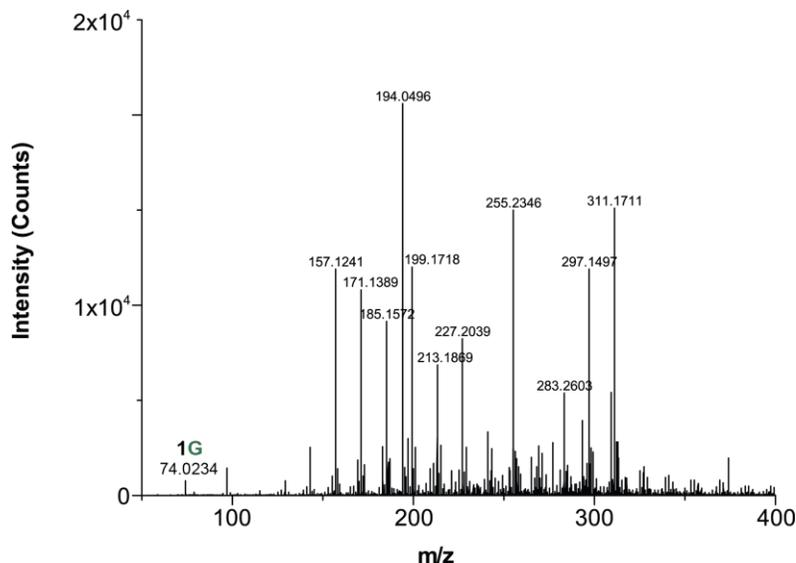


Figure 2.2 – Mass spectrum of the glycine control sample (pH 3.0; 4 cycles).

To test the temperature limitation for this reaction, experiments under lower drying temperature was conducted. For samples containing mixtures of LA and G, the formation of amide bonds was observed at 65°C (Figure 2.3). Based on this data, we confirm that amide bonds are produced under milder conditions than the protocols of Fox and Harada, which required temperatures above 150°C [24], and with longer oligomer lengths than produced by salt-catalysis [34]. This property makes ester-mediated amide bond formation more compatible with some models put forth for the origins of life.

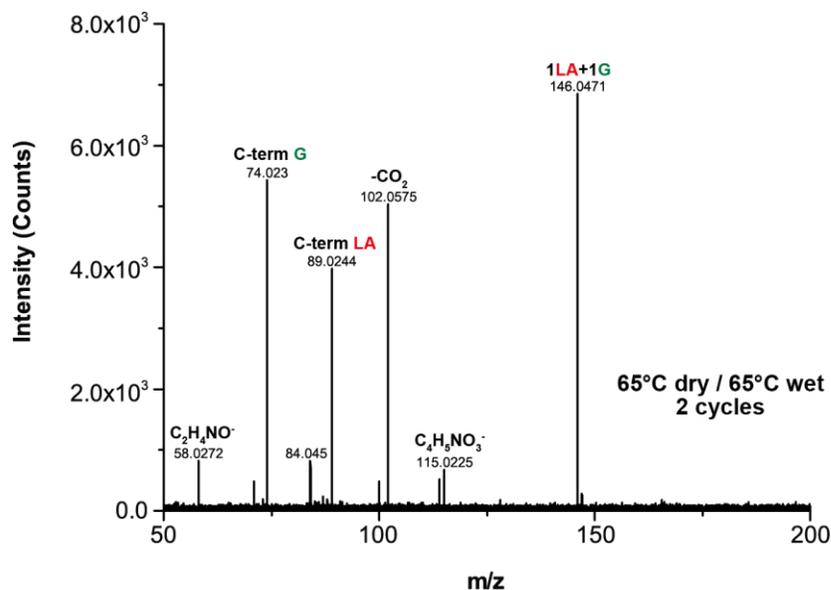


Figure 2.3 – Dipeptide formation at a lower dry-state temperature of 65°C. Both the ester-linked and amide-linked dipeptide are observed. CID collision energy = 10 eV (lab-frame).

We have also verified the effect of initial concentration on the copolymerization. Our typical protocol used 100 mM of each monomer as initial concentration. The concentration was chosen to ensure enough materials for analysis. Such high concentration is considered high in the context of the prebiotic scenario. To show the effectiveness of this reaction, experiments were carried under lower concentration. Figure 2.4 shows the product distribution of LA/G oligomers when 10 times dilution was used. Under such condition, LA/G dipeptides still form. Even when 100 times dilution was applied, the tandem MS still shows both the ester-linked and amide-linked dipeptide dimers, suggesting the reaction is not hindered by the low concentration (Figure 2.5). These results confirm that dipeptides can form even in dilute concentration. This robustness might arise from the drying procedure we used in day-night cycling. The drying step not only removes water generated by condensation, but also concentrates the reaction mixture to circumvent the dilution problem in prebiotic scenarios.

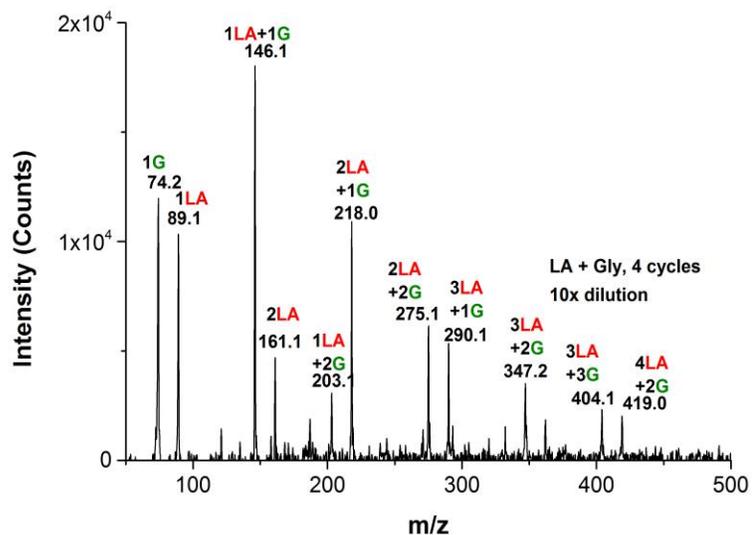


Figure 2.4 – LA/G depsipeptides formation with 10 times dilution after 4 cycles 10 mM of LA and G monomers were used, respectively. MS analysis was performed on Agilent 6130 single quadrupole mass spectrometer. Samples were diluted in ultrapure deionized water to $0.0825 \text{ mg mL}^{-1}$ and directly injected into the mass spectrometer by electrospray ionization in negative-ion mode.

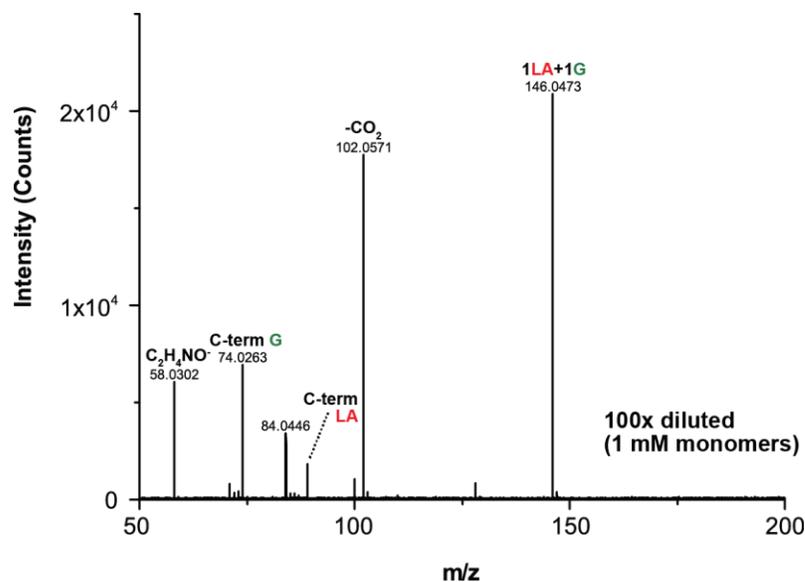


Figure 2.5 – Tandem MS of the LA/G dimer with 100 times dilution after 1 cycle. 1 mM of LA and G monomers were used, respectively. Tandem MS sequencing of 1LA+1G depsipeptide. CID collision energy = 5 eV (lab-frame).

Besides mass spectrometry, FT-IR spectroscopy also confirms the emergence of peptide bonds. In Figure 2.6, we first examined the IR spectrum of glycine oligomers. The Amide I and II bands are clearly observed in the tripeptide ($\sim 1640\text{ cm}^{-1}$ and $\sim 1520\text{ cm}^{-1}$, respectively). For the case of lactic acid oligoesters (Figure 2.7), the signals of interest are the C=O band shifts from 1714 cm^{-1} (free acid) to 1757 cm^{-1} (ester).

Figure 2.8 shows the change of IR spectra for the LA/G oligomers during environmental cycling. The summed IR spectrum of G and LA monomers contains a strong band at $\sim 1710\text{ cm}^{-1}$, corresponding to carboxylic acids. After subjecting a 1:1 mixture of LA and G to 1 wet-dry cycle, the intensity of the carboxylic acid band is reduced and a shoulder appears at $\sim 1730\text{ cm}^{-1}$, indicative of ester bonds. With continued cycling, we observe growth of Amide I and Amide II bands ($\sim 1640\text{ cm}^{-1}$ and $\sim 1520\text{ cm}^{-1}$, respectively), and the spectrum of the sample subjected to 16 cycles begins to resemble that of polyglycine [75]. These observations are fully supportive of our proposed mechanism of peptide bond enrichment through repeated ester-amide exchange.

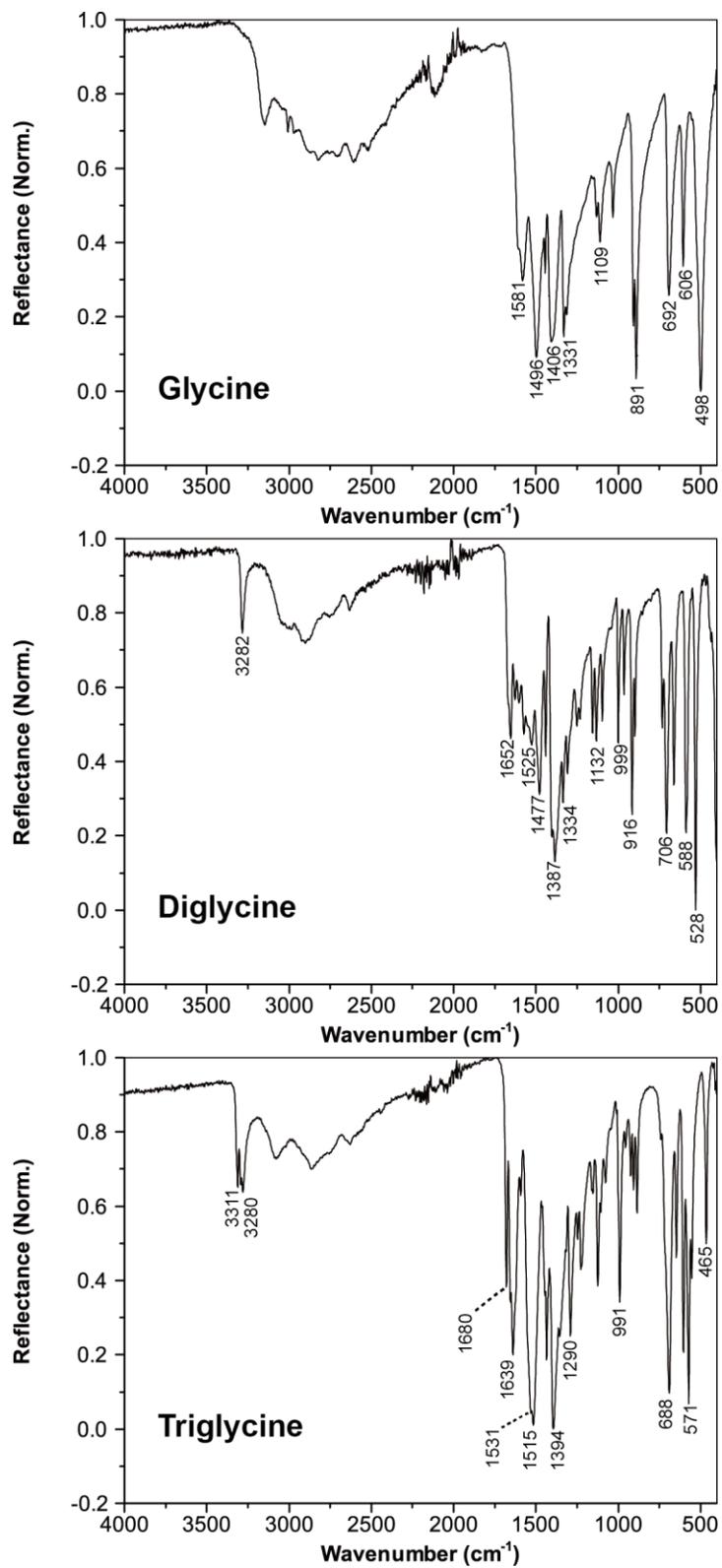


Figure 2.6 – FT-IR spectra of (top) glycine, (middle) diglycine, and (bottom) triglycine standards.

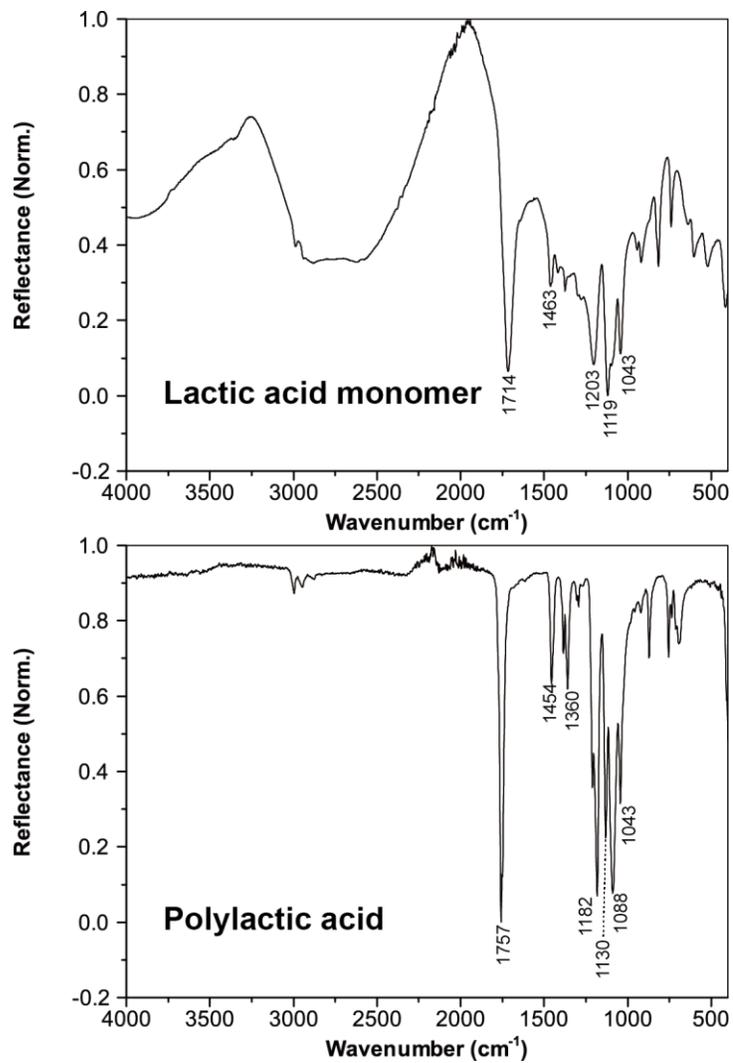


Figure 2.7 – FT-IR spectra of (top) lactic acid monomer and (bottom) polylactic acid standards.

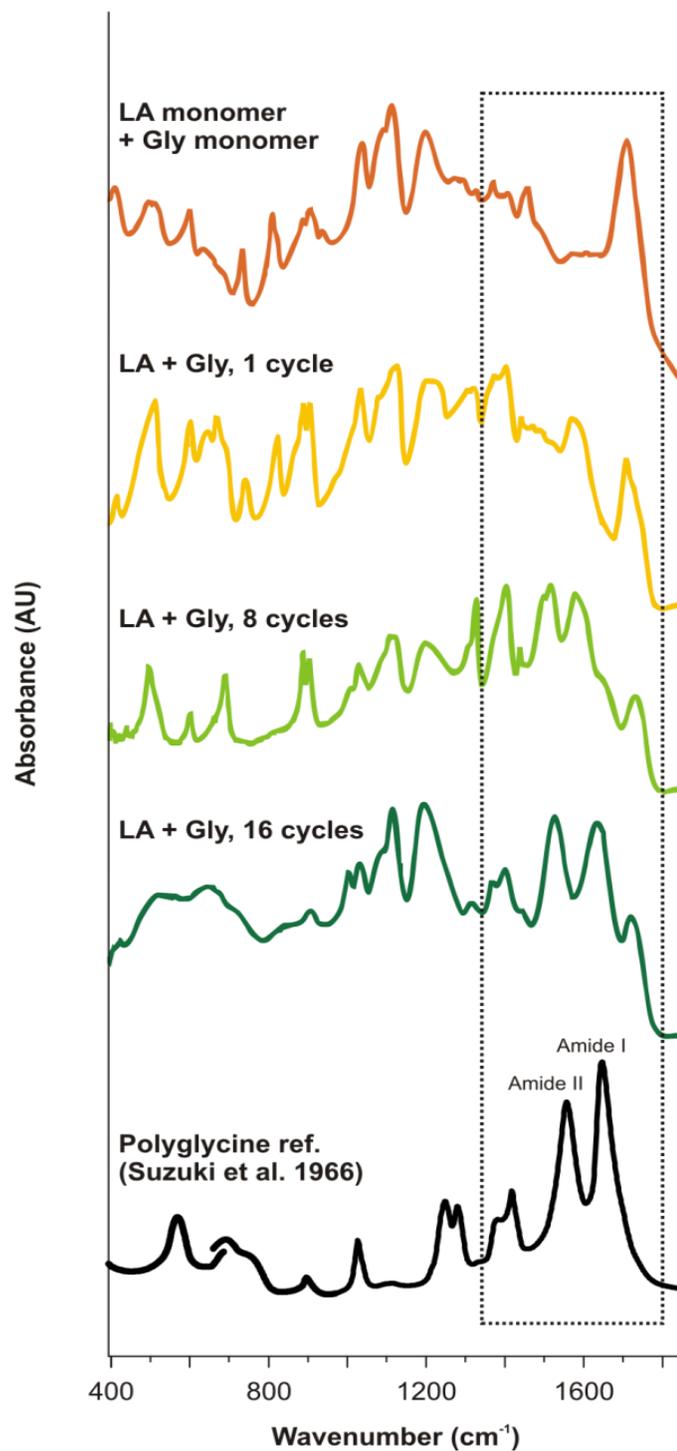


Figure 2.8 – Emergence of amide bonds as demonstrated by IR spectroscopy. Spectra are of (orange) LA + glycine (G) monomers before cycling, (yellow) LA/G mixture after 1 cycle, (light green) LA/G mixture after 8 cycles, and (green) LA/G mixture after 16 cycles, where LA was re-added after the 8th cycle.

2.3.2 Evolution of LA/A depsipeptides

Our proposed mechanism suggests that the peptide linkages should gradually increase at the expense of the less stable ester bond. To demonstrate this phenomenon, we studied the copolymerization from a 1:1 mixture of lactic acid (LA) and alanine (A).

Figure 2.9 shows the mass spectra of such a mixture after 1, 4, 8 and 12 environmental cycles. Because the LA and A neutral residue masses differ by approximately 0.984 Da, only the overall chain length distribution can be seen on the full range mass spectra plots. Notably, however, the overall length of the depsipeptides does not appear to change significantly.

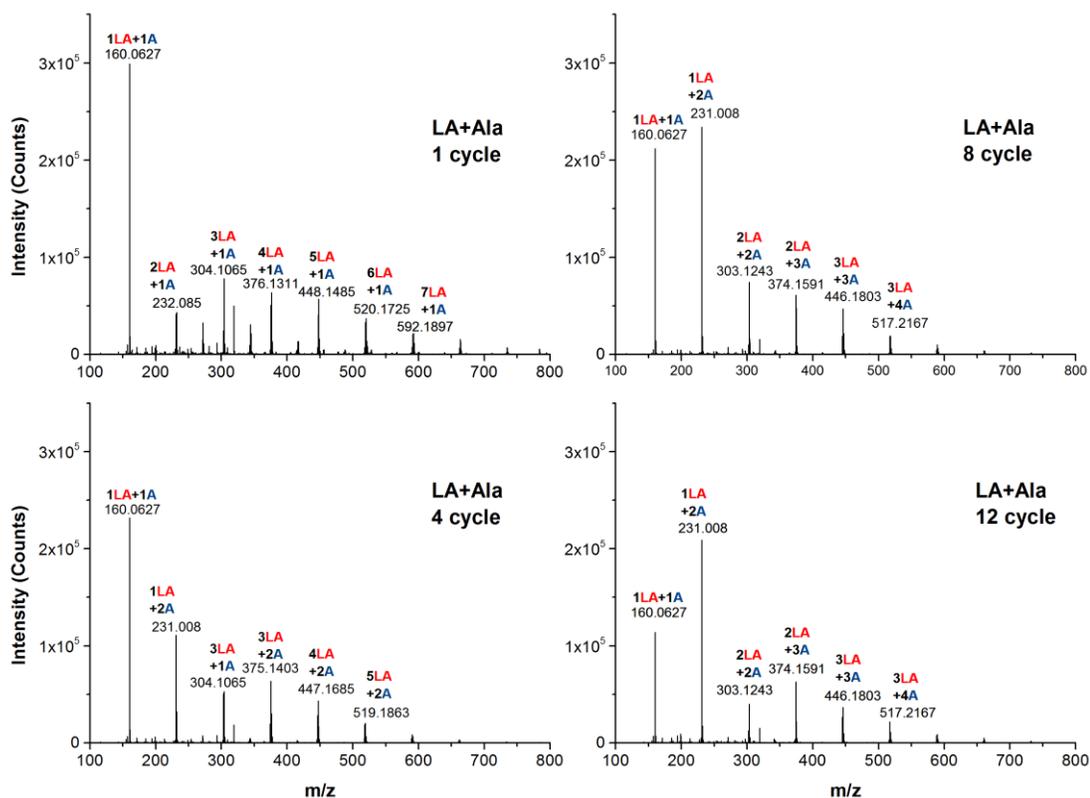


Figure 2.9 – Mass spectra of lactic acid and alanine oligomers. Initial ratio of between the two monomers is 1:1. Only the dominant peak in each group is labeled.

For a LA-A depsipeptide of a given residue length, a shift toward lower molecular weight corresponds to A enrichment. Copolymerization of LA and A could produce a mixture of depsipeptides with all possible combinations. For example, the pentamer series ($n = 5$) could contain the 5LA polyester, along with the 4LA+1A, 3LA+2A, 2LA+3A, and 1LA+4A depsipeptides, each separated by 1 Da. After 1 cycle we observe pentamers with each of these possible LA/A combinations, except for the 1LA+4A depsipeptide (Figure 2.10, top). As the number of cycles increases, the distribution of these species shifts to lower mass-to-charge values, corresponding to increasing incorporation of A (Figure 2.10, subsequent spectra). After 12 environmental cycles, the pentamer series contains 1LA+4A, 2LA+3A, and 3LA+2A depsipeptides. The 4LA+1A depsipeptide and 5LA polyester are no longer present.

To provide a semi-quantitative demonstration of the composition change, isotopically-corrected MS data for $n = 2$ through $n = 8$ oligomers were used to show the relative amounts of A incorporated as a function of environmental cycle (Figure 2.11). The relative abundance of alanine in oligomers gradually increases over continuous environmental cycling. Following 20 cycles, the average composition of the depsipeptides is approximately 65% A and 35% LA. The theoretical limit of A abundance for each oligomer (shown in black) is provided for comparison, based on the assumption that one LA residue remains attached to the *N*-terminus of the peptide. The relative abundance of short oligomers, like dimer and trimer, approached the theoretical limit in the first few cycles. We expect amino acid abundance would continue to approach this theoretical limit with additional cycling, as long as catalytic hydroxy acids continue to be present.

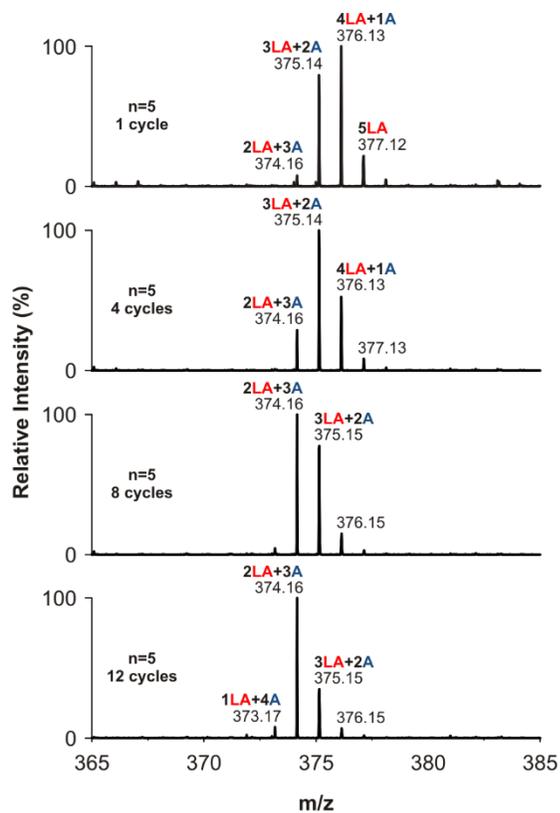


Figure 2.10 – Amino acid/peptide bond enrichment in depsipeptides following environmental cycling. Mass spectral region shows LA-A pentamers after 1, 4, 8, and 12 cycles.

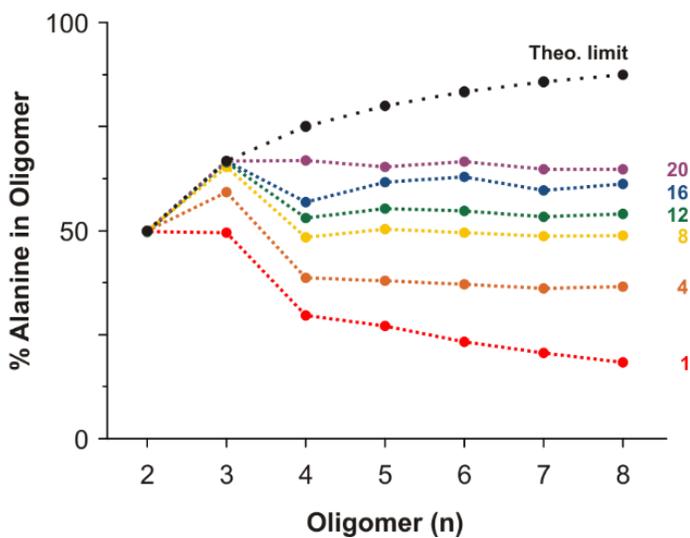


Figure 2.11 – Relative amounts of A incorporated in each oligomer length after different environmental cycles.

Our typical experiment protocol used a 1:1 mixture of hydroxy acid and amino acid. Here, we further examined the change of MS signals when the reaction started from 9:1, 3:7 and 1:9 mixtures of A and LA. As shown in Figure 2.12, when 10 mol % of LA is present initially, only dimer and trimer with relatively low abundance are detected. The reaction seems to progress better when LA is in excess, as longer oligomers were detected.

Previous results from Cronin and Moore [76] and Peltzer and Bada [50] show the amounts of alanine and lactic acid in the Murchison meteorite to be $3.5 \mu\text{g g}^{-1}$ and $5.9 \mu\text{g g}^{-1}$, respectively. This mass ratio corresponds to a relative mole ratio of approximately 40% alanine and 60% lactic acid starting monomers. Notably, we observe depsipeptides form readily from similar starting ratios (e.g., 50% alanine and 50% lactic acid, as shown previously, and 30% alanine and 70% lactic acid, as shown in Figure 2.12). For the 3:7 mixture, we further used ion-mobility (IM) to spread the signal into a second dimension. This decreases peak overlap and improves S/N for larger oligomers (Figure 2.13). Although the actual abundance might be low, we confirmed oligomers up 14-mer form after 8 environmental cycles.

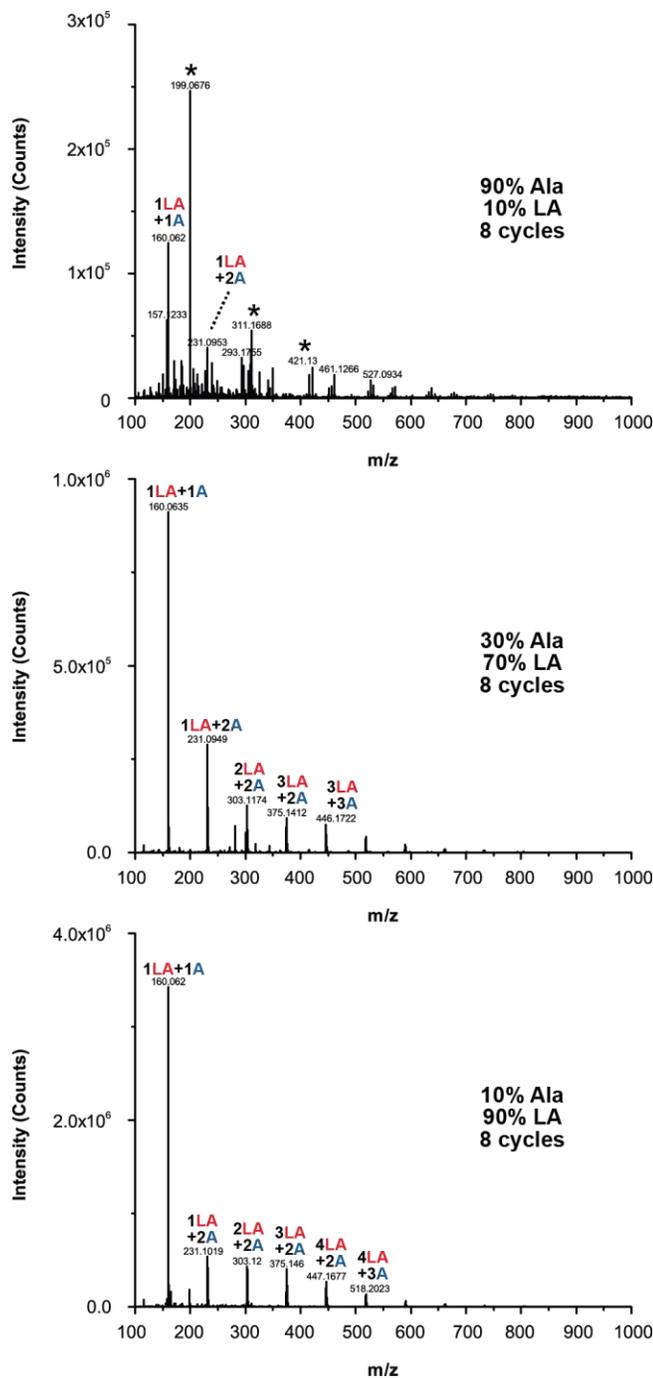


Figure 2.12 – Mass spectra of LA/A mixtures after 8 cycles with varying mole ratios of starting monomer. Depsipeptide formation and alanine incorporation is optimal when a higher ratio of lactic acid is present. Sodiated alanine monomer clusters, marked with asterisks, are observed in the 90% alanine sample and not the others (e.g., m/z 199.0676 corresponds to $[2(A^-)+Na^+]^-$; theo. m/z = 199.0696), reflecting a higher amount of unreacted alanine monomer.

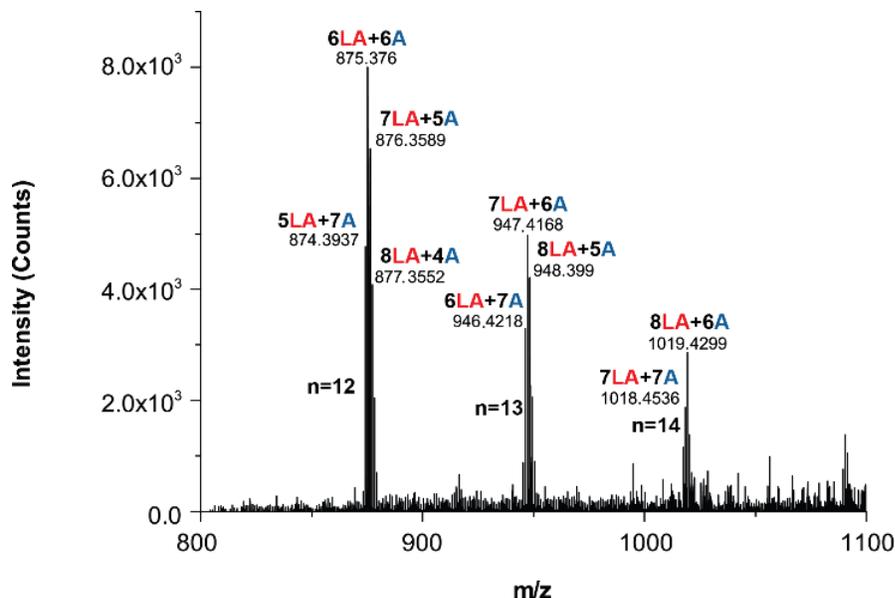


Figure 2.13 – The largest oligomers in the LA/A depsipeptides observed using electrospray IM-MS ($n = 12$ – 14). These depsipeptides are made using 30% alanine and 70% lactic acid after 8 environmental cycles. Ion mobility traveling wave velocity = 900 m s^{-1} . Traveling wave height = 40 V.

2.3.3 The effect of diketopiperazines

A number of studies on prebiotic amide bond formation have noted the generation of diketopiperazines (DKPs). [34, 77-80] We tentatively identified cyclic LA (cyclic 2LA) and cyclic mixed LA-A (cyclic 1LA+1A) species in our MS data, but have not observed the 2A DKP (Figure 2.14) The signal intensities for 1LA+1A and 2LA cyclic dimers are too low to for tandem MS fragmentation. To see whether a proper signal of 2A DKP can be obtained by MS, we spiked the sample with 2A DKP, which is commercially available. After the spiking, the signal for 2A DKP with expected mass was observed, indicating 2A DKP was not present in the unspiked sample. We hypothesize the 2LA and 1LA+1A cyclic species can reversibly interchange with their respective linear forms. Although both linear and cyclic 2A were not detected, we expect the rate of interchange of the 2A DKP with its linear form to be slower.

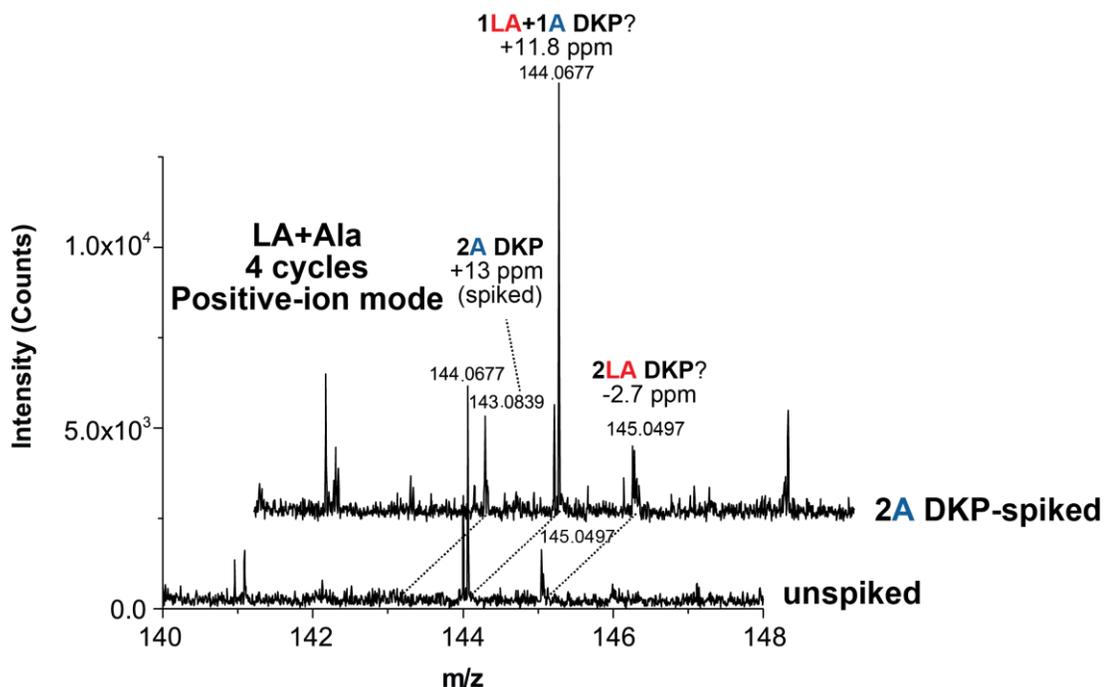


Figure 2.14 – Investigation on the formation of diketopiperazines (DKPs) in the LA/A system. Experiment was done by positive-ion mode MS (2.50 kV capillary voltage, 40 V sampling cone, 4 V extraction cone). In the above mass spectrum, 25 μ M alanine DKP is spiked in the sample.

To examine the reactivity of DKP for the ester-mediated amide bond formation, we carried out an experiment starting from LA and G-G DKP (Figure 2.15). Comparing to the experiment from LA and G, only a small amount of 1LA-1G is detected, suggesting DKP is not as reactive as the free monomer to produce depsipeptides.

Whether or not DKPs are involved in depsipeptide formation remains unclear at this time; however, it is clear DKP formation is not an obstacle to forming linear depsipeptides and, eventually, peptide sequences.

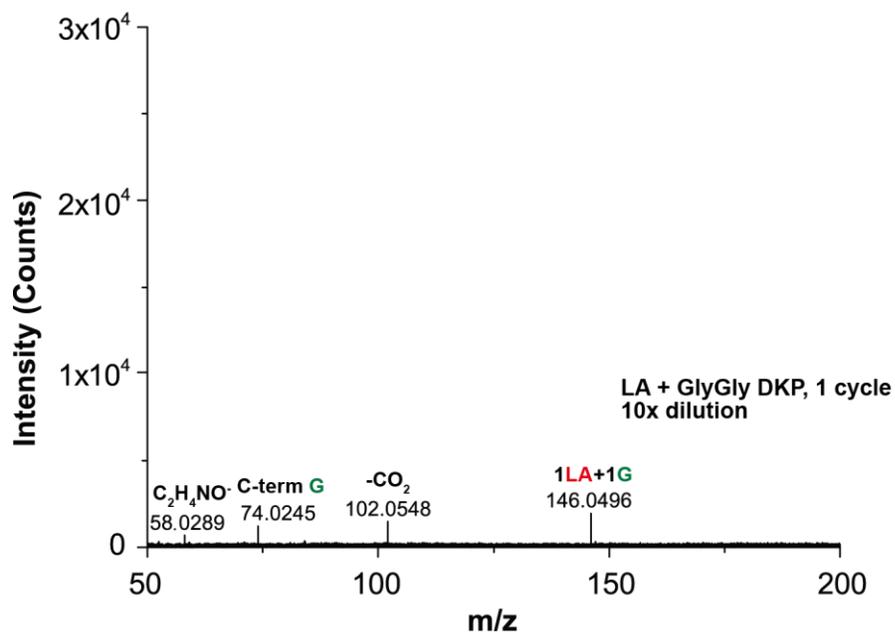
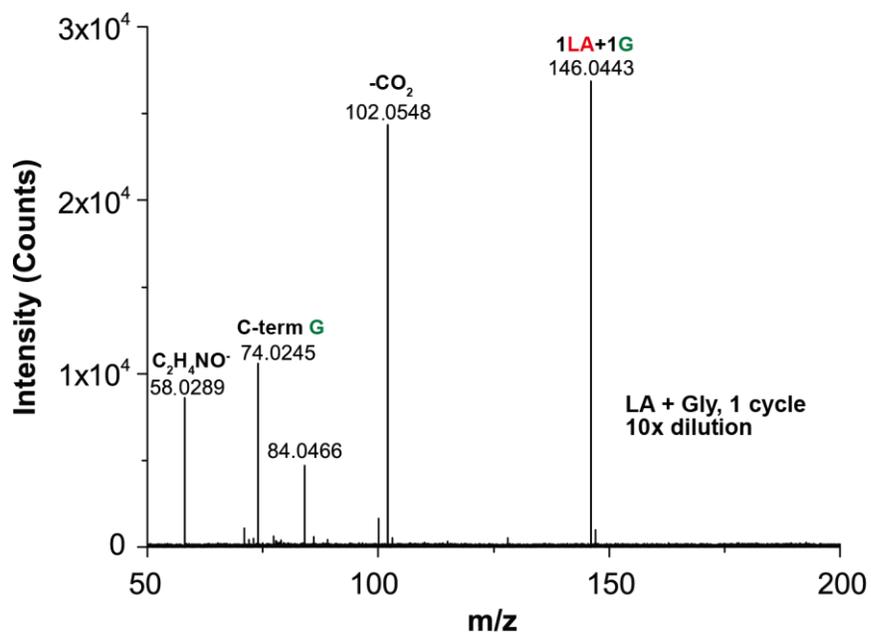


Figure 2.15 – Depsipeptide formation with GG DKP. Tandem MS sequencing of 1LA+1G depsipeptide found in LA + G and LA + G-G DKP samples. CID collision energy = 8 eV (lab-frame).

2.3.4 Sequence analysis

In the previous section, it is clear that the ester-mediated copolymerization produces oligomers with various compositions. In addition, each oligomer may also have different isomers depends on the potential arrangement of the building blocks. Even the heterodimer has two different sequence isomers. Depsipeptide residue sequences were determined using tandem MS. Figure 2.16 shows the sequence analysis of the hetero-dimer from LA/A reaction mixture. Two sequence isomers, LA-A (amide linkage) and A-LA (ester linkage), were observed after the first cycle. As the number of cycles increases, the relative abundance of ester bond (A-LA) decreases and the relative abundance of amide or peptide bond (LA-A) increases. After 8 cycles, essentially all heterodimer is LA-A.

For the LA/G reaction, G-G amide bond sequence was found immediately after 1 cycle as shown in Figure 2.17(a). LA-G amide bond was found in longer oligomers like 3LA-1G (Figure 2.17(b)). After continuous environmental cycling, we observed internal peptide sequences. Specifically, the G-G-G-G internal peptide sequence is observed after 8 cycles in the LA/G reaction system (Figure 2.18(a)); the three-amino acid residue A-A-A sequence is observed after 12 cycles in the LA/A reaction system (Figure 2.18(b)); and the three possible 2G+1A internal sequences are detected after only 4 cycles of the glycine-LA/G/A reaction system (Figure 2.18(c)). We note that sequences of the peptide core in depsipeptides appear to be random. The exact abundance of each sequence isomer is unavailable to us. Consistent with the proposed reaction mechanism, LA residues are progressively displaced from the core of depsipeptides, leading to continuous peptide sequences predominantly in the middle of the oligomers sequence. From the large library of depsipeptides with many possible sequences, some of them might be functional and be

able to survive through a natural selection process. Therefore, the randomness of sequences could be an advantage to the next step of evolution.

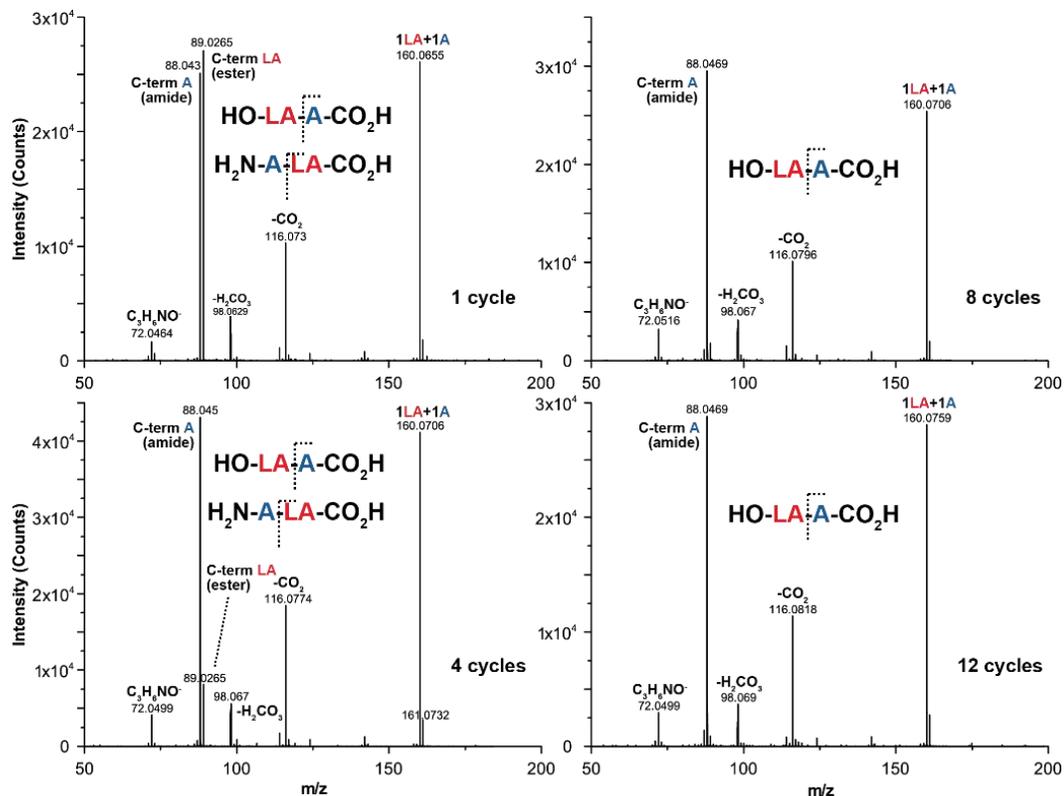


Figure 2.16 – Tandem MS sequencing of LA/A heterodimer (parent ion $[M-H]^- = 160.06$ Da), over 1, 4, 8, and 12 cycles. CID collision energy = 15 eV (lab-frame).

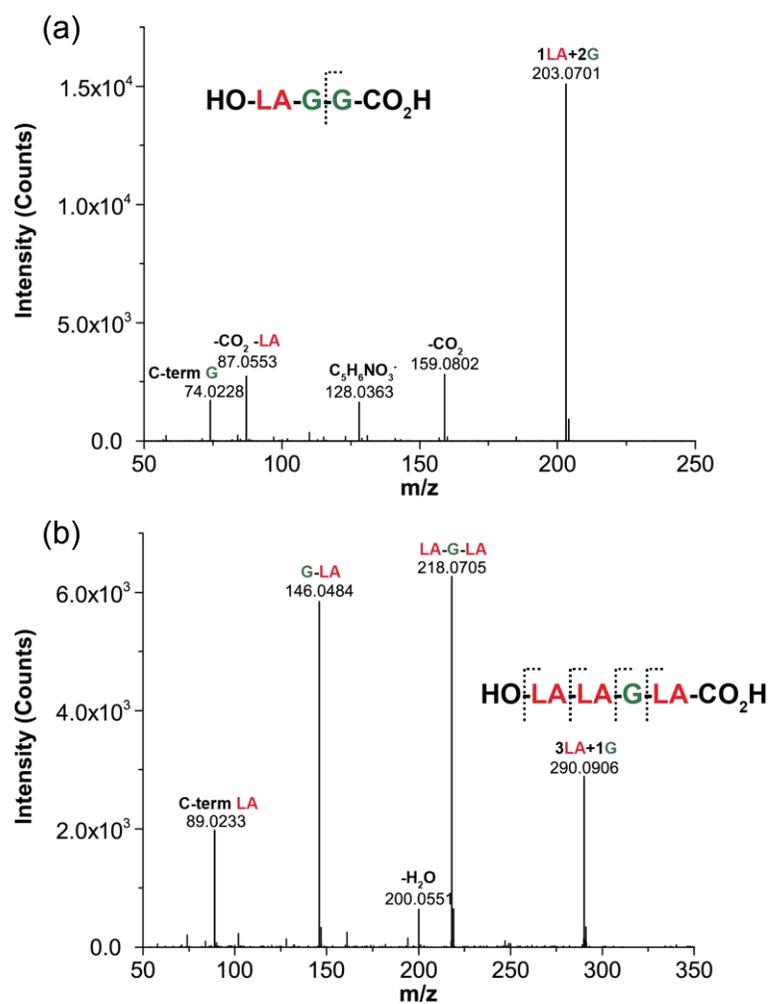


Figure 2.17 – Tandem MS sequencing of LA/G depsipeptides after 1 cycle. (a) 1LA+2G and (b) 3LA+1G. CID collision energy = 15 eV (lab-frame). Primary sequences are LA-G-G and LA-LA-G-LA, respectively.

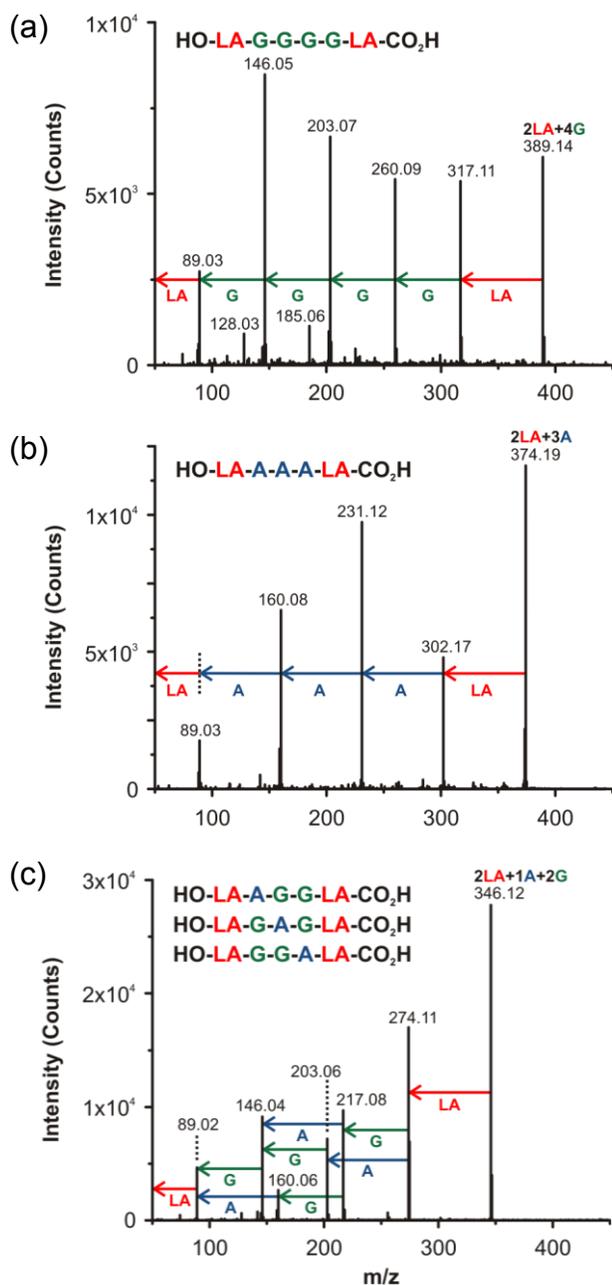


Figure 2.18 – Sequencing of depsipeptides by tandem MS. (a) Sequencing of 2LA+4G, theoretical precursor ion [M-H]⁻ = 389.13 Da, formed after 8 cycles. CID collision energy = 20 eV. The primary sequence is LA-G-G-G-G-LA. (b) Sequencing of 2LA+3A, theoretical precursor ion [M-H]⁻ = 374.16 Da, formed after 12 cycles. CID collision energy = 15 eV. The primary sequence is LA-A-A-A-LA. (c) Sequencing of 2LA+2G+1A, theoretical precursor ion [M-H]⁻ = 346.13 Da. Formed after 4 cycles. CID collision energy = 20 eV. LA-A-G-G-LA, LA-G-A-G-LA, and LA-G-G-A-LA sequences are observed.

2.3.5 Analysis of reaction mechanism

Based on our experimental data, we are able to provide a more detailed analysis of the reaction mechanism (Figure 2.19 and Figure 2.20). We propose that two major pathways, labeled Pathway A and Pathway B, through which depsipeptide formation and amino acid enrichment may occur. Lactic acid (LA, **1**) and amino acid (AA, **2**) are initially mixed together and dried down. As water evaporates, pathway **A** favors the formation of ester bonds between lactic acid residues to form a lactic acid dimer (**3**). This linear dimer is likely in equilibrium with the cyclic lactide, the ester equivalent of a diketopiperazine or DKP. The lactide dimer is less stable than an amide-linked DKP, however, and is expected to reversibly interchange with the linear form. The amine of an amino acid monomer attacks the ester carbonyl carbon in **3** to produce a heterodimer (**4**) containing one lactic acid residue and one amino acid residue linked with an amide bond. This species may cyclize as well. In the previous section, we do tentatively detect the 2LA and 1LA+1AA cyclic dimers in our samples. It is currently unclear whether the linear or cyclic forms are the active species; however, we note that depsipeptide formation does appear to be much less efficient starting from DKP than from the glycine monomer. Upon repeated cycling, the amide-linked heterodimer **4** continues to grow to **5** and, eventually, to **12**. Spontaneous amide-ester exchange occurs on **12** to form **6**, a linear depsipeptide with a lactic acid residue at the “N”-terminus and amino acid residues elsewhere. Further esterification leads to lactic acid residues being found at both termini and amino acid sequences being enriched internally. It is likely that not only monomers but also other short depsipeptides (e.g., the heterodimer **4**) may add onto **6**, resulting in the introduction of an internal lactic acid residue (**13**). Indeed, we do observe depsipeptides with a lactic acid residue between

internal amino acids and C-terminal amino acid. However, with increasing cycles we expect these species to convert further to pure peptide internal sequences (**14**). Pathway **A** appears to be most consistent with our data; therefore, we propose this is the main pathway for depsipeptide formation and the resulting transition towards peptide sequences.

Another route, Pathway **B**, is also possible. In Pathway **B**, a reversible ester-linked heterodimer (**7**) is formed as water is either removed or added. This heterodimer is observed in mass spectrum shown in Figure 2.16; however, it disappears as cycling increases. Subsequent attack on the ester carbonyl carbon by the amino acid leads to irreversible formation of dipeptide (**8**). Upon cycling, lactic acid adds to the C-terminus of the peptide to form **9**, followed by amide-ester exchange to form the tripeptide **10**. A likely side-product in Pathway **B** is a DKP (**11**). Although we tentatively identify mixed ester-amide DKP-like cyclic species, pure amino acid DKPs have not been detected by MS. Moreover, we do not observe pure peptides in our samples without containing at least one lactic acid residue, and depsipeptides with lactic acid residues solely on C-termini are extremely rare. Therefore, we propose that Pathway **B** provides only a minor contribution.

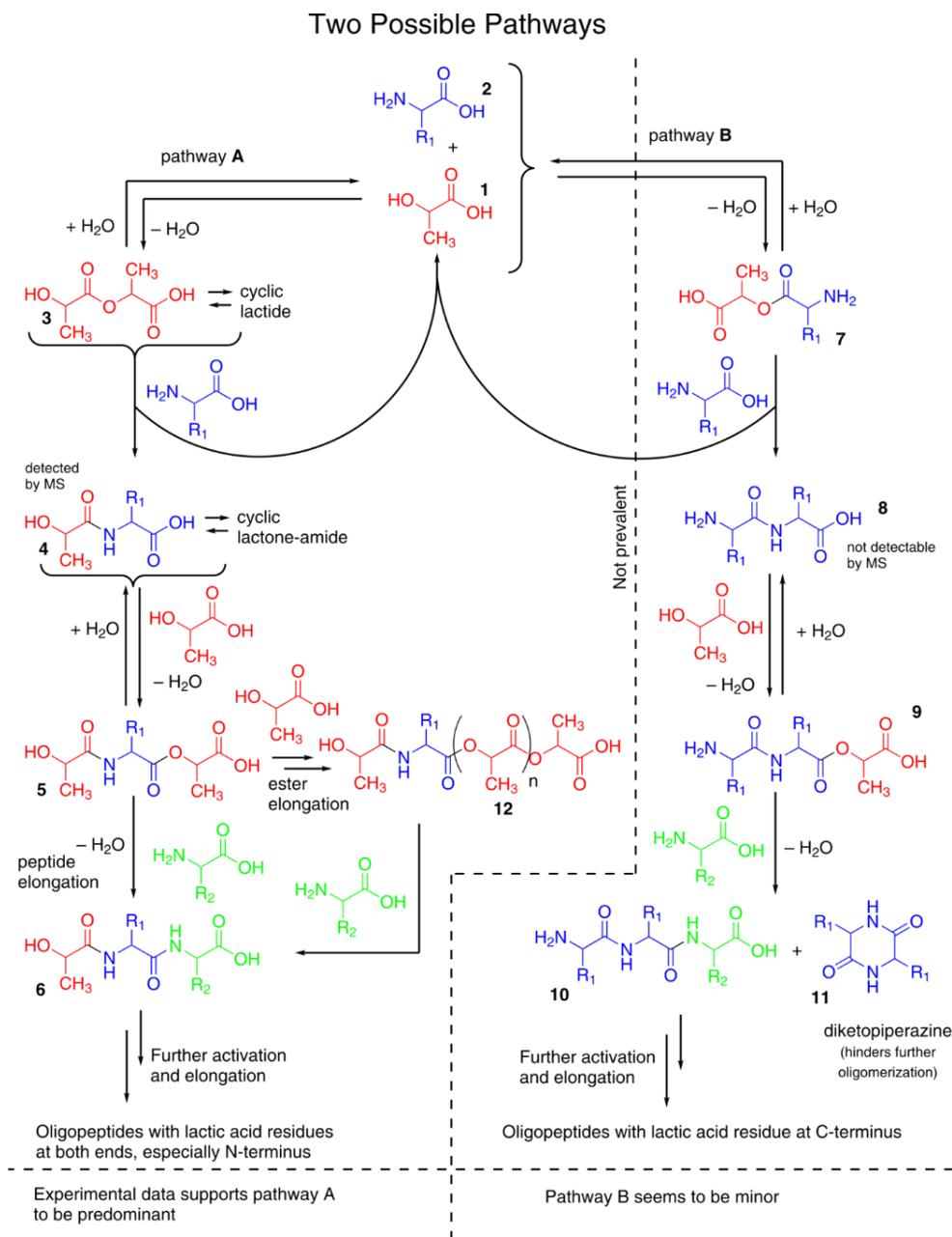


Figure 2.19 – Proposed scheme of the ester-mediated amide bond formation.

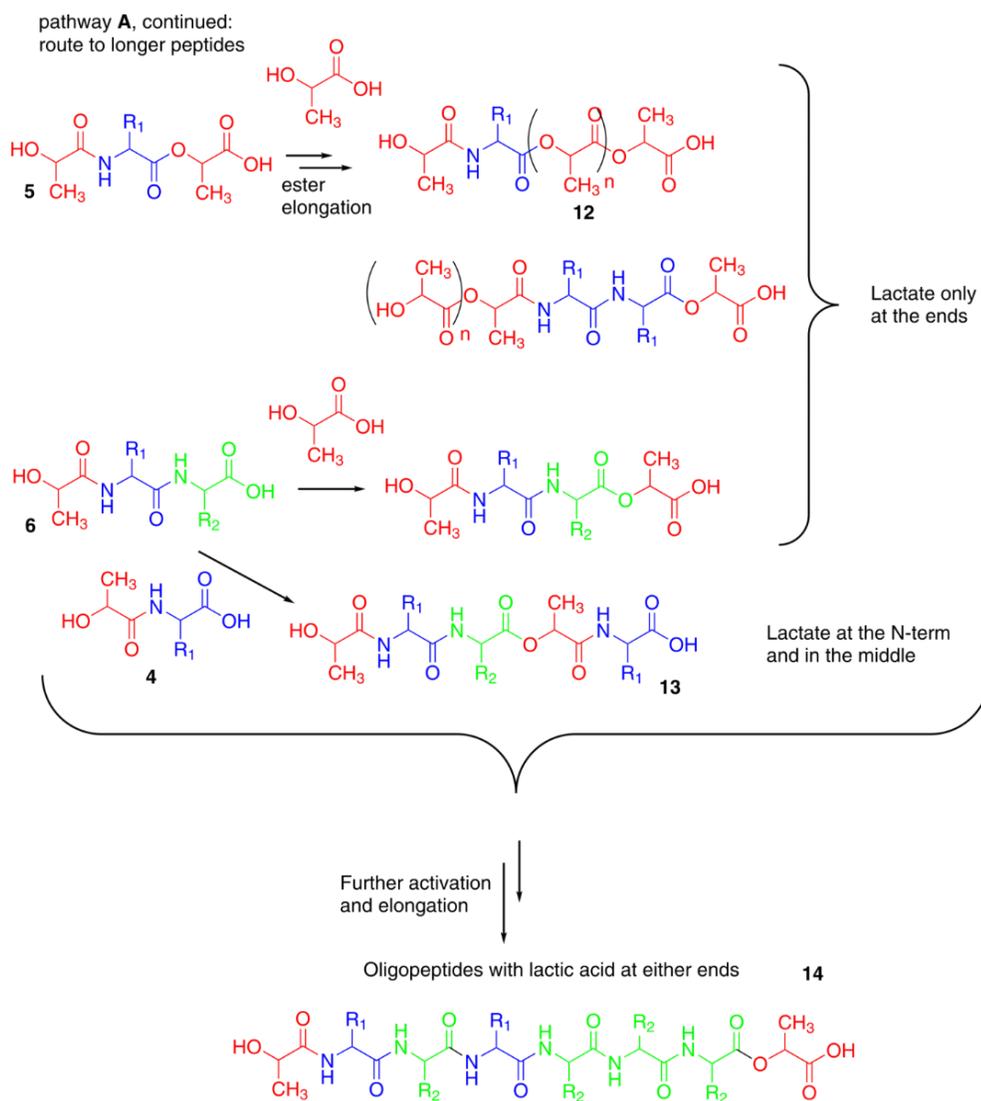


Figure 2.20 – Proposed scheme of the ester-mediated amide bond formation (continued).

2.3.6 *Reaction at neutral pH*

In the experiments described previously, reaction mixtures were initially around pH 3 due to unbuffered carboxylic acid moieties. While acidic pH can facilitate ester bond formation, protonation of the amine also reduces its ability to act as a nucleophile in the ester-amide exchange reaction. Moreover, from an origin of life perspective, prebiotic environments of pH 3 would have likely been rare [81]. To test the impact of pH on amide bond formation, sodium hydroxide, ammonium hydroxide, and triethylamine were used to adjust the pH to 7 for a 1:1 mixture of LA and A. Depsipeptides is formed with all three bases; however, depsipeptide formation is more efficient with ammonium hydroxide or triethylamine than sodium hydroxide (Figure 2.21). When sodium hydroxide is used to adjust the pH, depsipeptide formation is minimal. Peaks labeled with asterisks correspond to sodiated monomer cluster signals and not oligomers (Figure 2.21(a)). For example, m/z 201.04 corresponds to $[2(\text{LA}^-)+\text{Na}^+]$. The lactic acid and depsipeptide dimer are observed only with low signal intensity. In the cases of using ammonium hydroxide and triethylamine, more depsipeptides are observed and the tandem MS shows alanine is incorporated in the backbone by amide-linkage. We hypothesize this is due to the fact that sodium hydroxide cannot provide a proton to make the carboxylate a good leaving group in the ester-amide exchange reaction. Other factors include that ammonia and trimethylamine are more volatile, and the ionization efficiency in MS analysis is likely reduced in NaOH due to sodium-induced ion suppression.

In addition to the ability of the ammonium ion and protonated triethylamine to act as general acids to provide protons, the volatility of ammonia and triethylamine during evaporative cycles is also expected to facilitate ester bond formation. Upon rehydration,

the samples neutralized with ammonia or triethylamine exhibited a decrease in pH from 7 to approximately 3.5. To account for this effect, the base was re-added in each rehydration step to return the solution to pH 7. After 1 triethylamine-adjusted cycle, both LA oligoesters and mixed depsipeptides are observed (Figure 2.22). Following 8 cycles, enrichment of A-containing depsipeptides is observed, although at a slower rate than in pH-unadjusted samples.

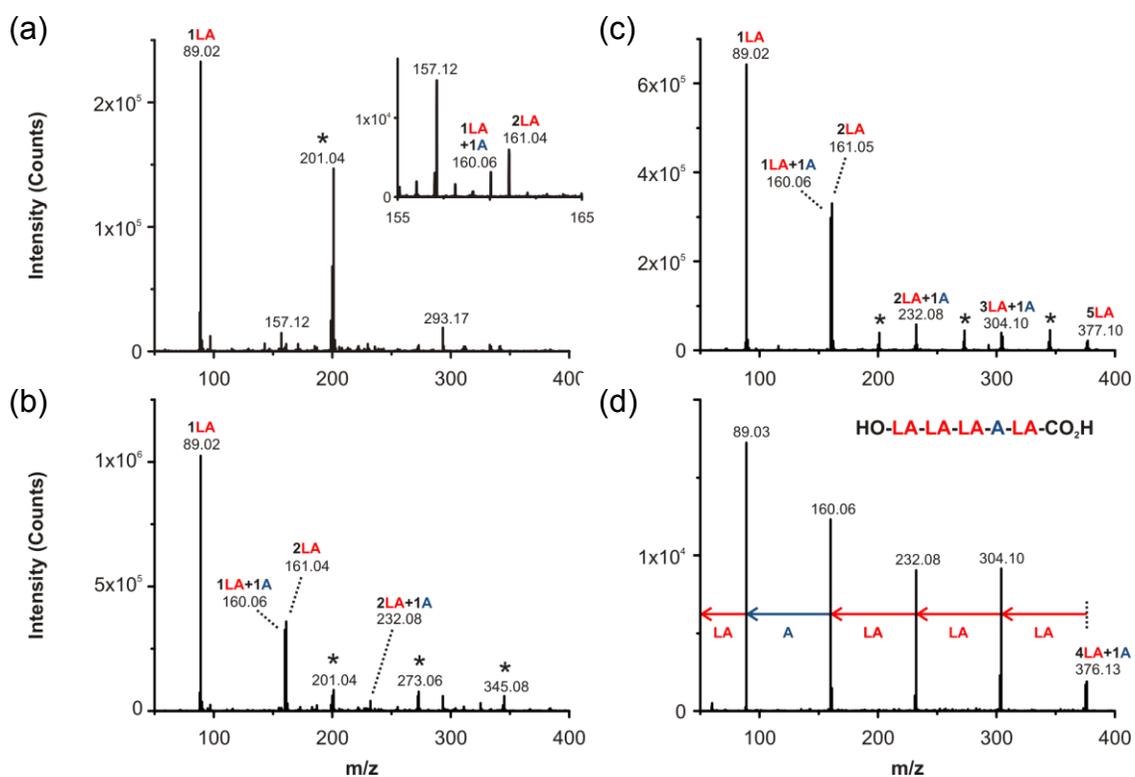


Figure 2.21 – Depsipeptide formation at pH 7. pH was adjusted with various bases. For parts (a)-(c), sodiated cluster signals are labeled with asterisks. (a) Mass spectrum of NaOH-adjusted sample after one cycle. (b) Mass spectrum of NH₃-adjusted sample after one cycle. (c) Mass spectrum of triethylamine-adjusted sample after one cycle. Depsipeptides were readily formed. (d) Tandem MS sequencing of 4LA+1A from pH 7 triethylamine sample, theoretical parent ion $[M-H]^- = 376.12$ Da. Formed after 1 cycle. CID collision energy = 10 eV.

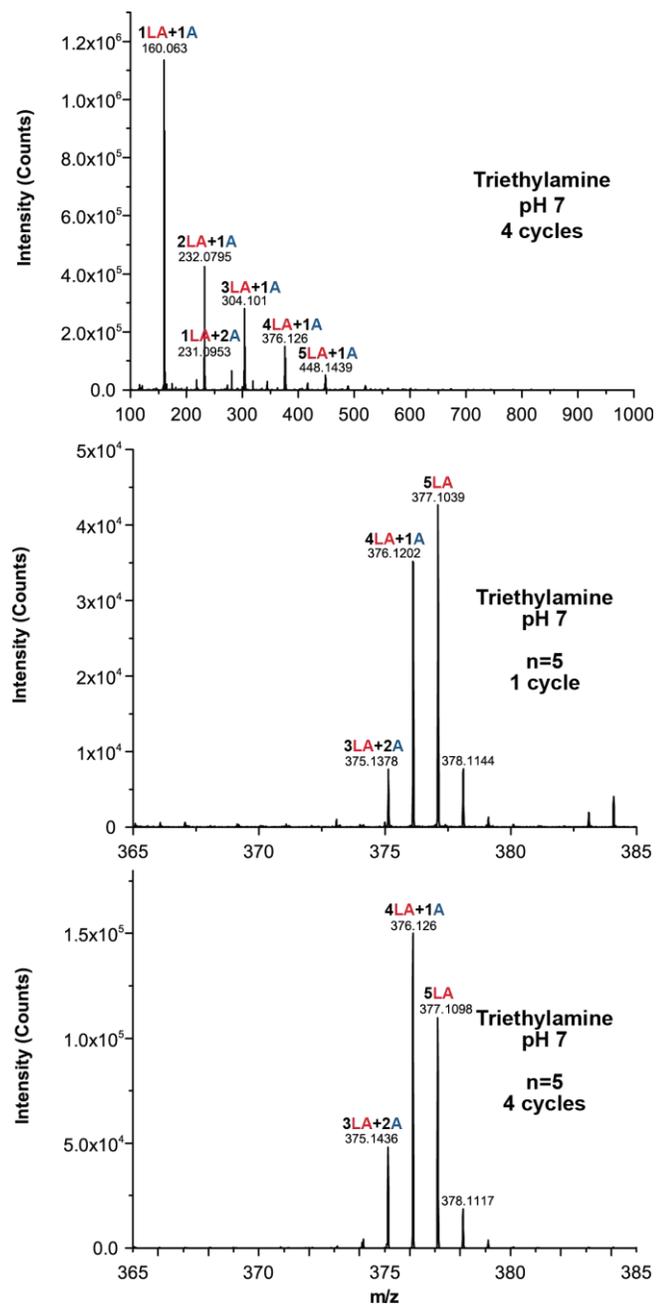


Figure 2.22 – LA/A mixture with triethylamine at pH 7. (top) Full mass spectrum after 4 cycles, adjusted to pH 7 with triethylamine. (middle-bottom) Enrichment of alanine-containing depsipeptides from 1 to 4 cycles. The rate of enrichment may be lower than that of unbuffered solutions. Signals with m/z 378.11 correspond to ¹³C-containing 5LA oligoesters.

2.3.7 *Two-dimensional NMR analysis and quantitative NMR analysis of LA/G oligomers*

In the previous analysis, MS provides detailed information of the oligomer distribution, but quantitation from MS is typically unreliable, especially when no standard compound is available. Here, we further used NMR to measure the yield of oligomers and to test the efficiency of the ester-mediated amide bond formation.

In general, the concentration of certain species in a mixture can be estimated from its peak area in a one-dimensional NMR spectrum. However, the peak position of target species needs to be confirmed first. In our case, the chemical shifts of oligomers were unknown to us due to the complexity of the mixture and the lack of standard compounds. Here, we used heteronuclear single quantum coherence (HSQC) and heteronuclear multiple bond coherence (HMBC) two-dimensional NMR analysis techniques to confirm amide bond formation and assign peaks in the one-dimensional ^1H spectrum.

Here, we used the 16-cycle LA/G sample with the replenishment of LA at the 8th cycle as a representative sample. The ^1H resonances of glycine and lactic acid residues were assigned using HSQC (Figure 2.23). The correlations between methylene and methyl groups were used to confirm the location of lactic acid residues in depsipeptides. In the carbonyl region of HMBC spectra, most of the glycine residues correlate to two carbonyl carbons, with one of the carbonyl carbons correlated to the methyl groups of lactic acid residues. Because HMBC is most sensitive for correlations spanning 2 or 3 covalent bonds, these correlations were used to assign LA and G residues linked amide bonds. Using triglycine as a standard, we confirmed that the C-terminal G-G amide bond is also detected in the HMBC spectrum (Figure 2.24). Due to spectral overlap, we are not able to

differentiate internal G-G amide linkages from internal LA-G amide linkages at this time. Nevertheless, internal G-G amide linkages are observed by tandem MS sequencing. Overlapping internal LA-G and internal G-G amide signals were integrated together in our yield calculations (combined with C-terminal G-G amide signals).

The chemical shifts of resonances in the above discussion are provided in Table 2.1. HSQC and HMBC spectra of the G/LA mixture with different cycles are shown in Figure 2.25 and Figure 2.26, which show similar ^1H , ^{13}C chemical shifts/correlations.

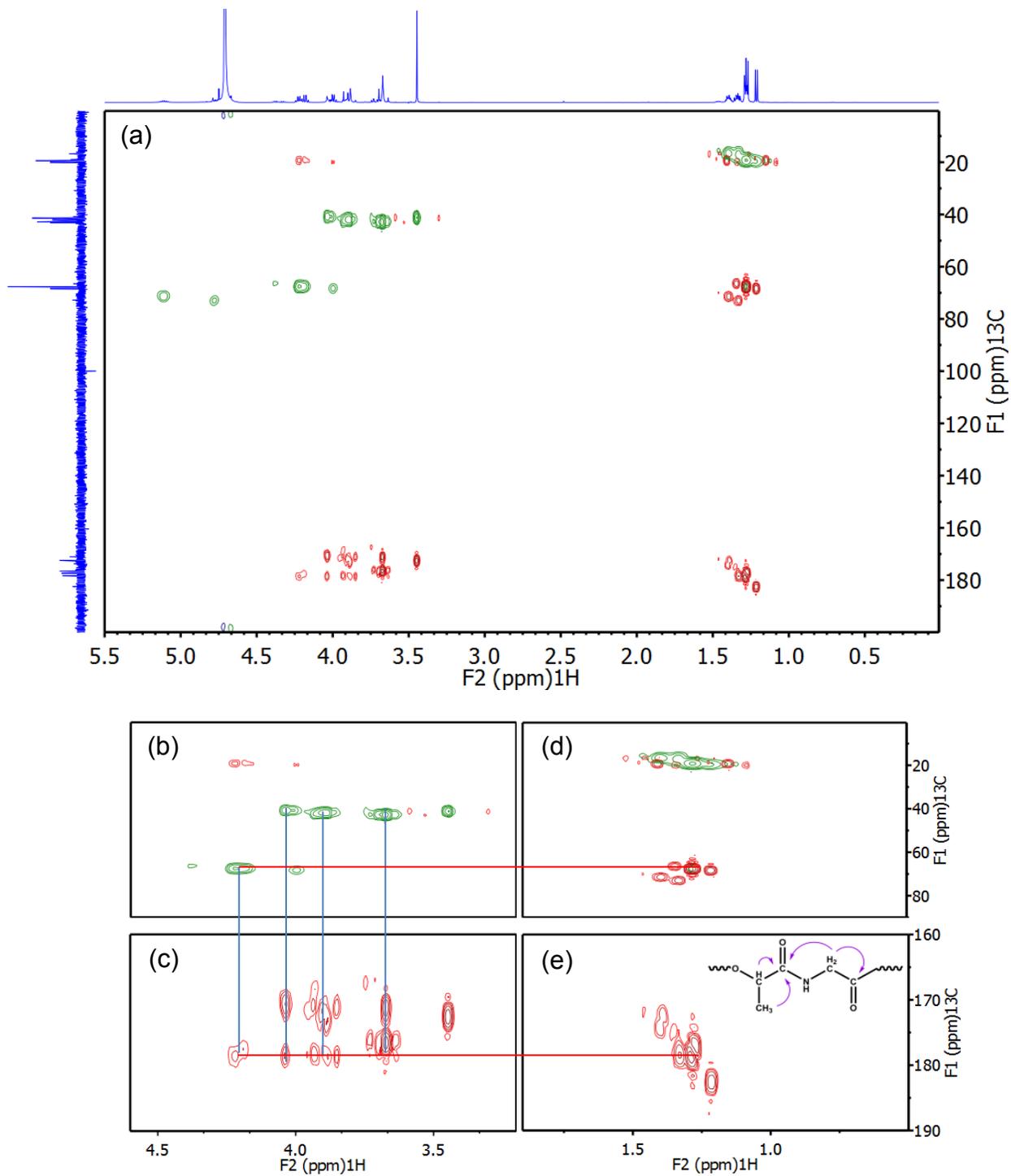


Figure 2.23 – ^1H - ^{13}C HSQC (green) and HMBC (red) spectra of the 16-cycle glycine/lactic acid sample with replenishment of lactic acid at the 8th cycle. (a) Full spectra. (b-e) Enlarged spectral regions of interest. Red lines indicate the correlation of methylene or carbonyl carbon to the methyl protons of lactic acid. Blue lines show the correlation of methylene protons to two different carbonyl carbons.

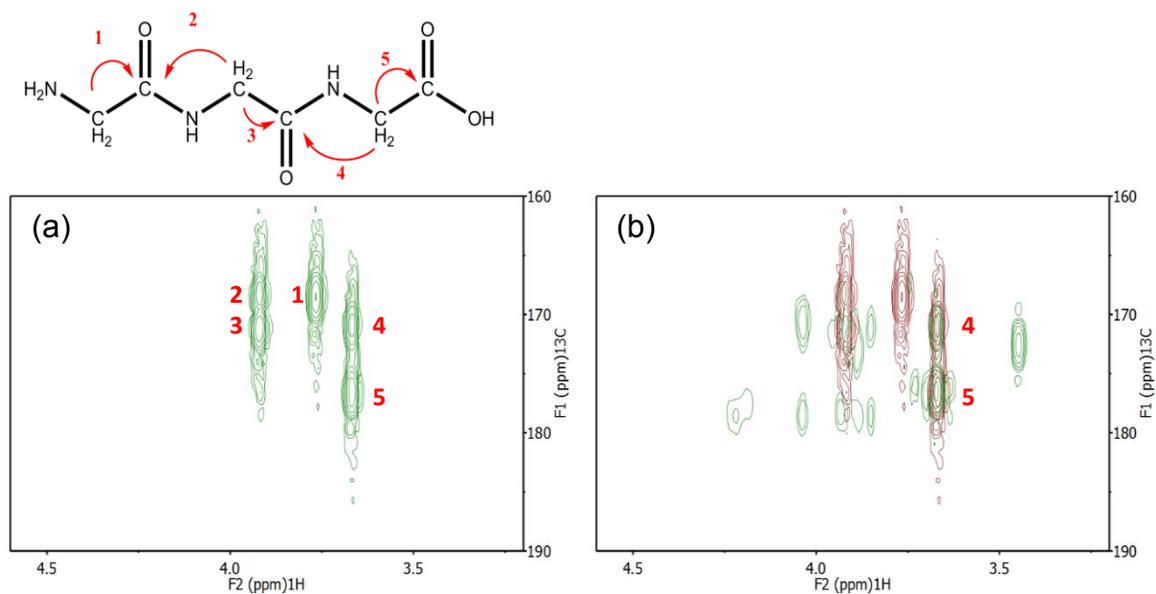


Figure 2.24 – Confirmation of C-terminal amide bond. (a) ^1H - ^{13}C HMBC spectrum of the triglycine standard. Five peaks were assigned based on the correlation between methylene protons and carbonyl carbons. (b) HMBC spectra overlap of the 16 cycles sample (green) and the triglycine standard (red). C-terminus correlations (Peaks 4 and 5) were also observed in the 16-cycle mixture suggested the presence of C-terminal amide bond.

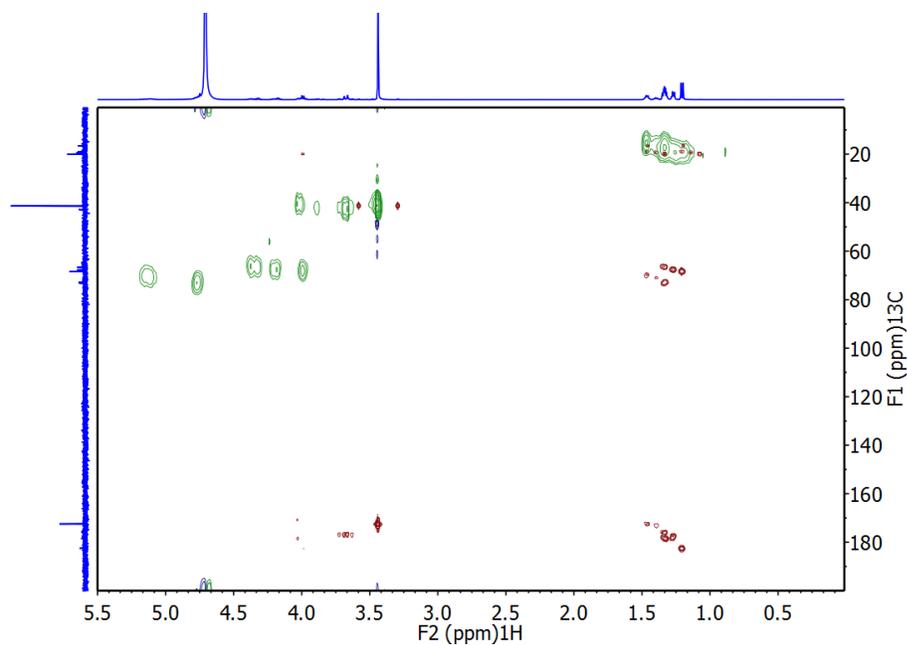


Figure 2.25 – ^1H - ^{13}C HSQC (green) and HMBC (red) spectra of the 1 cycle glycine/lactic acid mixture.

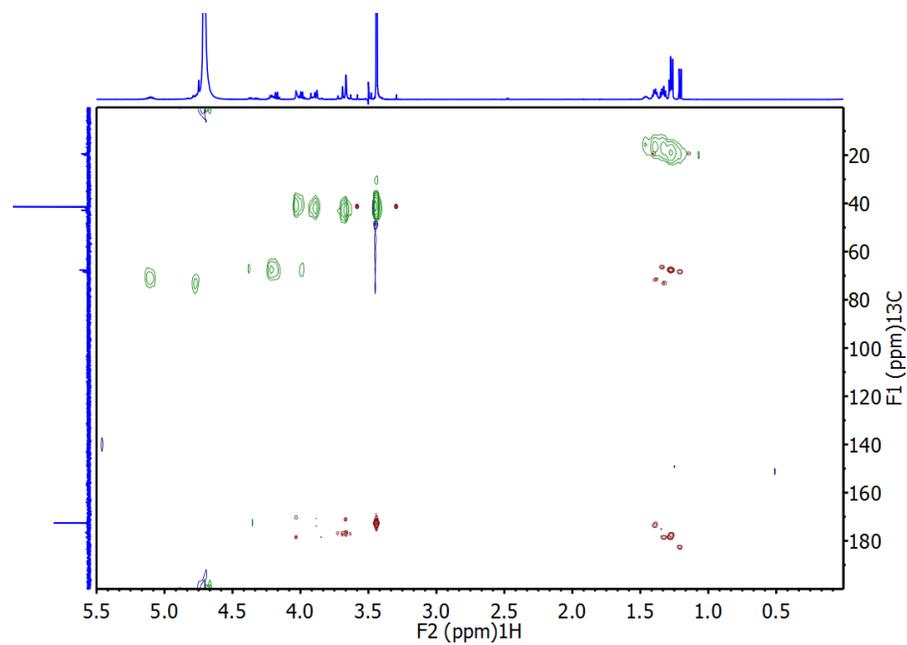


Figure 2.26 – ^1H - ^{13}C HSQC (green) and HMBC (red) spectra of the 8 cycle glycine/lactic acid mixture.

Table 2.1 – Chemical shifts of assigned resonances observed in HSQC and HMBC spectra of the glycine/lactic acid mixture after 16 cycles. Initial amount of lactic acid (0.02 mmol) was added at 8th cycle to reactivate the reaction.

HSQC			HMBC		
Residues	δ (¹ H) (ppm)	δ (¹³ C) (ppm)	Residues	δ (¹ H) (ppm)	δ (¹³ C) (ppm)
Glycine monomer (C-H ₂)	3.45	41.29	Glycine monomer (CH ₂ -C=O)	3.45	172.57
G-G, C-terminus (C-H ₂)	3.67	42.88	G-G, C-terminus (CH ₂ -C=O)	3.67	171.15,176.53
LA-G (C-H ₂)	3.83-3.96	39.17-44.59	LA-G (CH ₂ -C=O)	3.83-3.96	167.51-179.69
	4.04	40.84		4.04	170.63,178.45
LA (C-H ₂)	4.00	68.32	LA (CH ₃ -C=O)	1.29	178.7
	4.22	67.57		1.33	178.45
	4.78	72.75	LA (CH ₃ -CH ₂)	1.28	67.62
	5.11	72.19			
LA (C-H ₃)	1.28	19.28			
	1.40	16.61			

Quantification of free glycine monomer and reacted glycine (incorporated into deipeptides) was accomplished using ¹H NMR. Peaks corresponding to amide bond-linked glycine residues were identified by two-dimensional NMR analyses as discussed previously. One example ¹H spectrum used in our study and its peak assignment are shown in Figure 2.27 and Table 2.2. A 30 seconds relaxation delay was used to ensure full relaxation for accurate integration. Each data point was repeated three times.

As indicated in Eq. (2.1), the integrals of different species in the NMR spectrum are proportional to the products of molar concentration (M_x) and number of protons.

$$\frac{I_x}{I_y} = \frac{M_x N_x}{M_y N_y} \quad (2.1)$$

In this study, potassium hydrogen phthalate (KHP) was used as an external standard because its signals do not overlap with those of interest. The effective concentration of KHP (M_{KHP}) was calibrated by 75 mM glycine standard solution. The integral ratio between KHP and the glycine standard solution can be expressed as Eq. (2.2), which is rearranged to Eq. (2.3) to provide M_{KHP} .

$$\frac{I_{KHP}}{I_{Gly, std}} = \frac{M_{KHP} N_{KHP}}{M_{Gly, std} N_{Gly}} \quad (2.2)$$

$$M_{KHP} = \frac{I_{KHP} M_{Gly, std} N_{Gly}}{I_{Gly, std} N_{KHP}} \quad (2.3)$$

The concentration of a compound of interest (M_x) is calculated using Eq. (2.4) and previously determined M_{KHP} .

$$M_x = \frac{I_x M_{KHP} N_{KHP}}{I_{KHP} N_x} \quad (2.4)$$

Because every sample was diluted three-fold for replicate NMR experiments, the concentration of a compound of interest in the original sample is three times larger than the value calculated using Eq. (2.4). As a result, the concentration of glycine and lactic acid residues were calculated using Eqs. (2.5) and (2.6). All resonances from the methyl group of lactic acid units, including monomer and oligomers, were considered in our calculations.

$$M_{Gly} = \frac{I_{Gly} M_{KHP} N_{KHP}}{I_{KHP} N_{Gly}} \cdot 3 \quad (2.5)$$

$$M_{LA} = \frac{I_{LA} M_{KHP} N_{KHP}}{I_{KHP} N_{LA}} \cdot 3 \quad (2.6)$$

The yields of glycine residues with LA-G amide linkage (Y_{LA-G}) and C-terminal G-G linkage (Y_{G-G}) were calculated by using Eqs. (2.7) and (2.8). The volume of the sample solution (V) and the initial moles of glycine monomer ($n_{Gly, initial}$) were considered to calculate the amount of amide-linked glycine. The ratio between the amount of amide-linked glycine and the initial amount of glycine monomer gives the yield of amide bonds. It was not possible to differentiate internal G-G amide linkages from internal LA-G amide linkages. All signals in the region of LA-G linkages were included in the yield calculation. Both Y_{LA-G} and Y_{G-G} were included in the calculation of overall amide bond yield ($Y_{amide\ bond}$) as shown in Eq. (2.9).

$$Y_{LA-G} = \frac{I_{LA-G} M_{KHP} N_{KHP}}{I_{KHP} N_{LA-G}} \cdot V \cdot \frac{100\%}{n_{Gly, initial}} \quad (2.7)$$

$$Y_{G-G} = \frac{I_{G-G} M_{KHP} N_{KHP}}{I_{KHP} N_{G-G}} \cdot V \cdot \frac{100\%}{n_{Gly, initial}} \quad (2.8)$$

$$Y_{amide\ bond} = Y_{LA-G} + Y_{G-G} \quad (2.9)$$

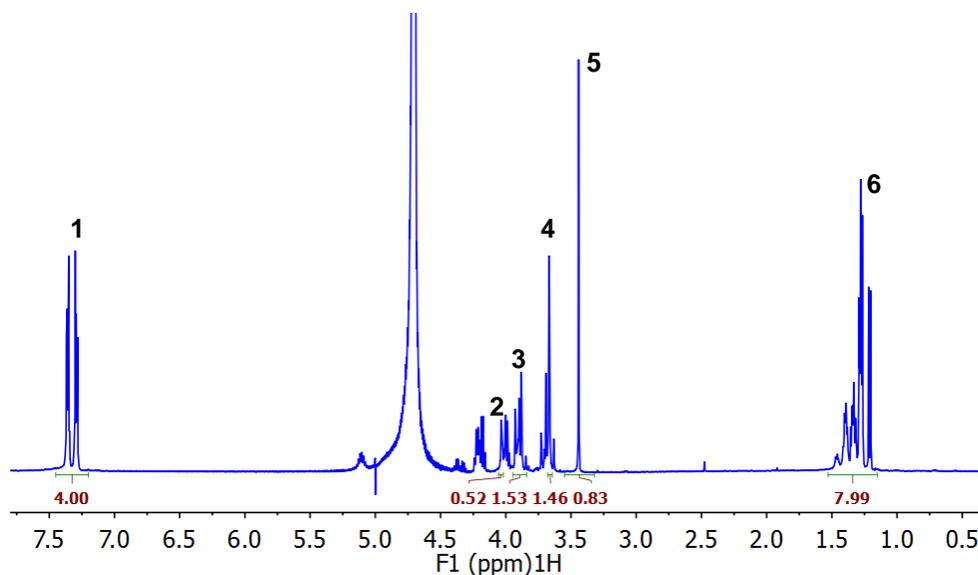


Figure 2.27 – ^1H NMR spectrum of glycine-lactic acid mixture after 16 cycles. Lactic acid was replenished at the 8th cycle.

Table 2.2 – List of peaks used for quantification and their corresponding integrals used in the example shown above.

Peak Number	Residues	Range (ppm)	Integral value (I_x)	Number of protons
1	KHP	7.20-7.45	4.00	4
2	LA-G (CH_2)	4.02-4.06	0.52	2
3	LA-G (CH_2)	3.84-3.95	1.53	2
4	G-G (CH_2)	3.64-3.68	1.46	2
5	Glycine monomer (CH_2)	3.32-3.55	0.83	2
6	LA (CH_3)	1.15-1.53	7.99	3

Based on ^1H NMR analysis, the quantitative analysis of the LA/G reaction is shown in Figure 2.28. After 8 cycles, total amide bond yield is around 10%. Notably, LA monomer appears to deplete with repeated cycling (Figure 2.28(a)). We hypothesize this is due to incorporation into depsipeptides and the evaporation of lactic acid monomer in our open reaction system. With a single replenishment of LA at the 8th cycle, the total amide bond yield achieved after 16 cycles was around 40% and almost all glycine monomer was consumed. Note that even with the two-dimensional NMR analysis, there are still some peaks in the ^1H spectrum that cannot be identified. The actual production rate of amide-linked glycine can be higher than we estimated.

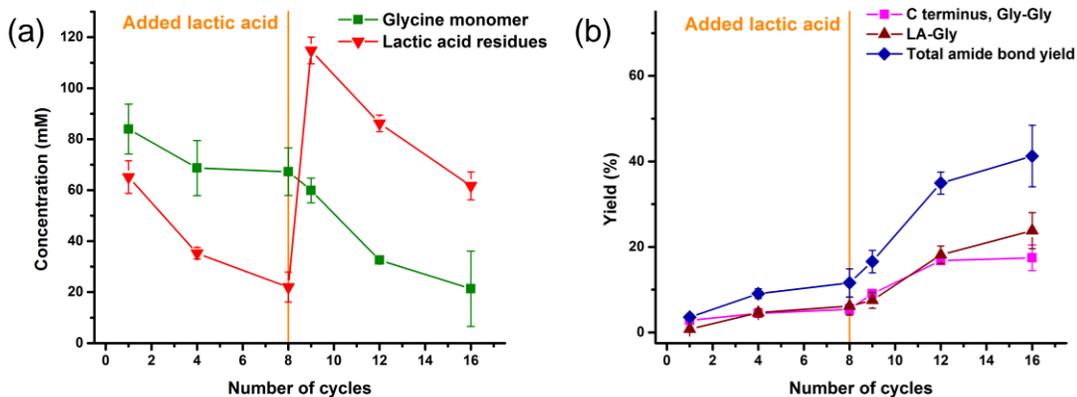


Figure 2.28 – Quantitative analysis of the LA/G reaction across different cycles. (a) The amounts of glycine monomer, total lactic acid (monomer + residues in products). (b) Overall yield of oligomers.

2.3.8 *Exploring the chemical space*

In the previous section, we mainly focus on the copolymerization of lactic acid, glycine and alanine to study the behavior of the reaction. Here, we also include results from amino acids with different side-chains. We have revealed that leucine (L), a more hydrophobic amino acid, also reacts with LA to form depsipeptides similar to glycine and alanine (Figure 2.29). Tandem MS results in Figure 2.30 confirm leucine is also incorporated in the internal sequence of depsipeptides. The presence of the LA-L sequence on the C-terminus suggests the LA-L heterodimer is adding to the larger depsipeptide in addition to LA monomer. When glycine, alanine and leucine were all mixed with lactic acid, leucine seems to be incorporated on apparently an equal basis with A and G in wet-dry cycling reactions and produces oligomers with random sequences (Figure 2.31 and Figure 2.32).

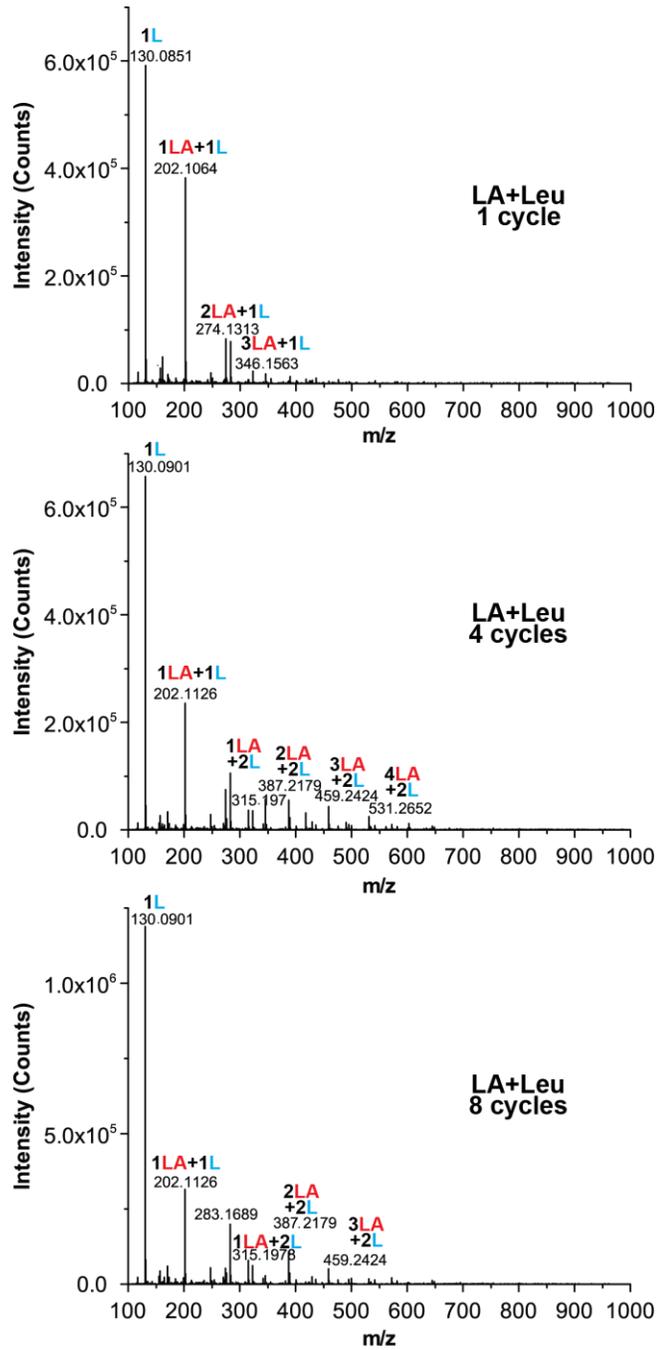


Figure 2.29 – Mass spectra for the 1:1 mol leucine and lactic acid system after 1 cycle, 4 cycles, and 8 cycles.

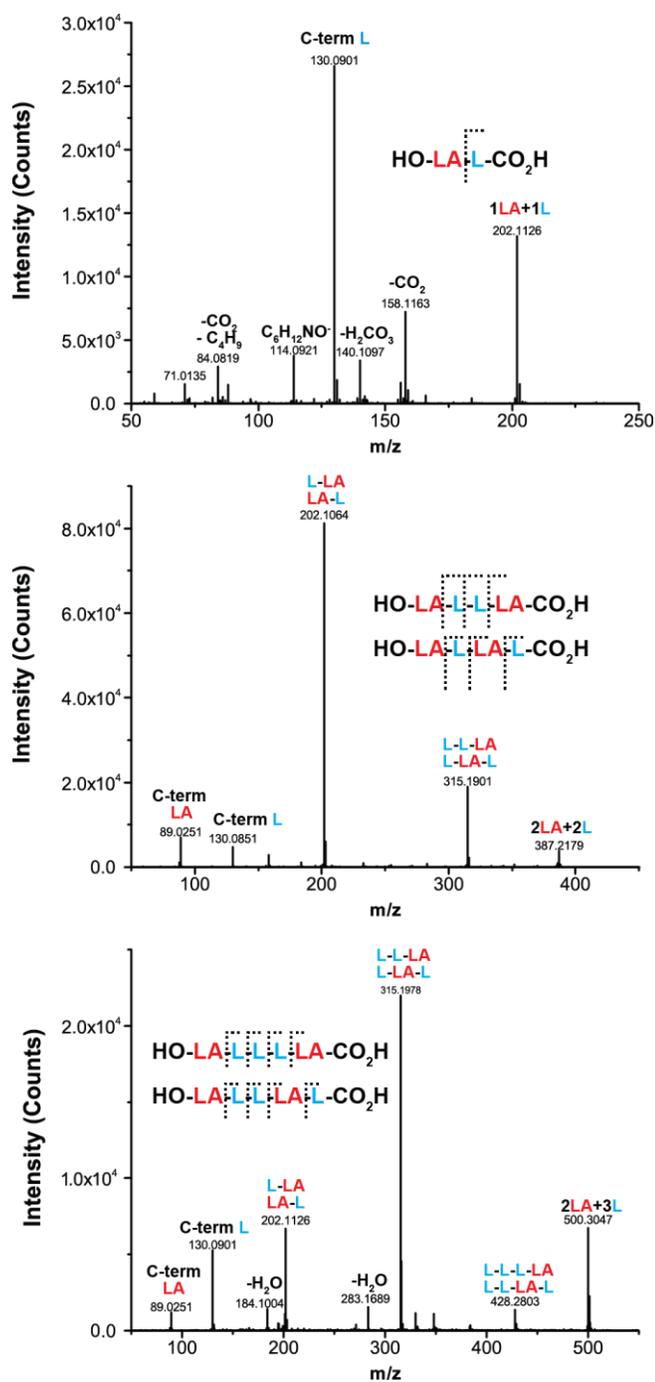


Figure 2.30 – Tandem MS sequencing of leucine-lactic acid depsipeptides. (top) Leucine-lactic acid heterodimer sequence after 8 cycles is LA-L. Collision energy = 15 eV. (middle) Two 2LA+2L sequences are observed after 8 cycles, LA-L-L-LA and LA-L-LA-L. Collision energy = 15 eV. (bottom) Two 2LA+3L sequences are observed after 8 cycles, LA-L-L-L-LA and LA-L-L-LA-L. Collision energy = 18 eV.

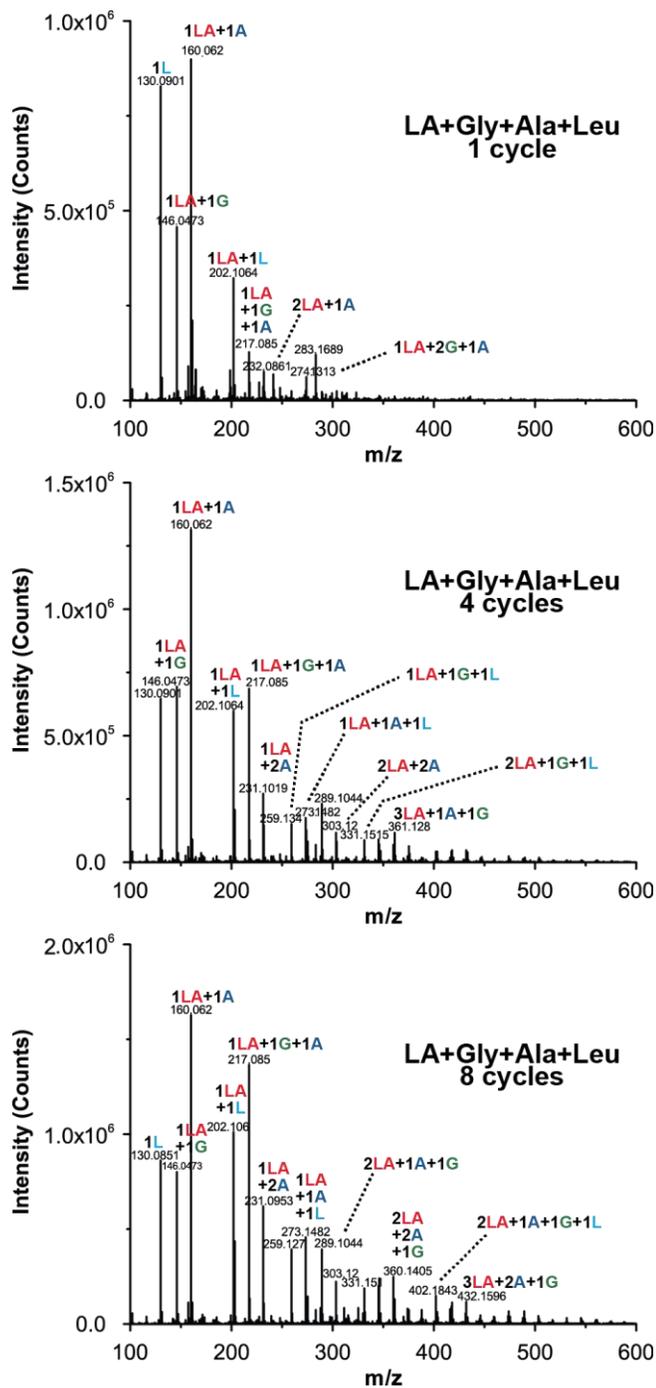


Figure 2.31 – Mass spectra of lactic acid, glycine, alanine, and leucine (3:1:1:1 mol ratio) after 1, 4, and 8 cycles. Depsipeptides containing various mixtures of the three amino acids are observed, and depsipeptides have higher amino acid content as cycling increases.

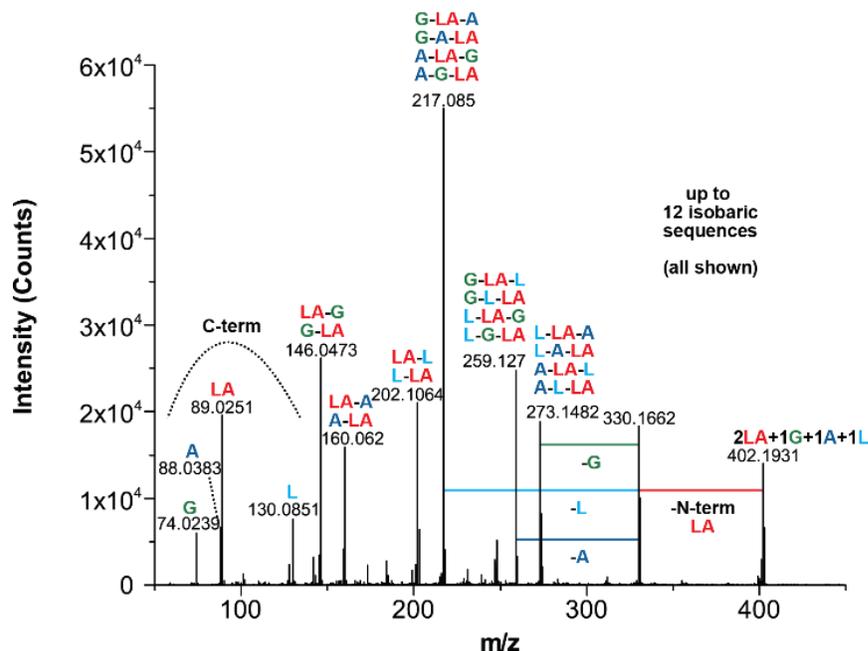


Figure 2.32 – Tandem MS sequencing of isobaric 2LA+1G+1A+1L depsipeptides (theo. parent ion $[M-H]^- = 402.1877$ Da). Formed after 8 cycles. Collision energy = 15 eV. One lactic acid residue is almost exclusively at the N-terminus, and the other lactic acid residue is either at the C-terminus or the residue adjacent to the C-terminus. 12 isobaric sequences could be present: (1) LA-A-G-L-LA, (2) LA-A-G-LA-L, (3) LA-A-L-G-LA, (4) LA-A-L-LA-G, (5) LA-G-A-L-LA, (6) LA-G-A-LA-L, (7) LA-G-L-A-LA, (8) LA-G-L-LA-A, (9) LA-L-A-G-LA, (10) LA-L-A-LA-G, (11) LA-L-G-A-LA, and (12) LA-L-G-LA-A. All potential combinations are shown.

Serine (S), a more polar amino acid, is also incorporated into depsipeptides (Figure 2.33), but the products are more complex, possibly due to the functional side-chain of serine. For example, the peak labeled with # corresponds to loss of CH_2O from serine side-chain in serine-lactic acid dimer, such that serine is converted to glycine in the gas-phase (i.e., this m/z value corresponds to 1LA+1G). This is a common fragmentation pathway for serine in negative-ion mode [72]. After one cycle, serine-lactic acid depsipeptides are observed. However, after increased cycling, these products disappear. There are several potential explanations for this result, which are likely related to the potentially-reactive hydroxyl group on the side-chain of serine. One explanation is the formation of cyclic

products which stunt depsipeptide formation and amino acid enrichment. We do observe a signal with the same mass as 1LA+1A; it is possible this is a by-product of larger cyclic species which are decomposing, but this has not yet been confirmed.

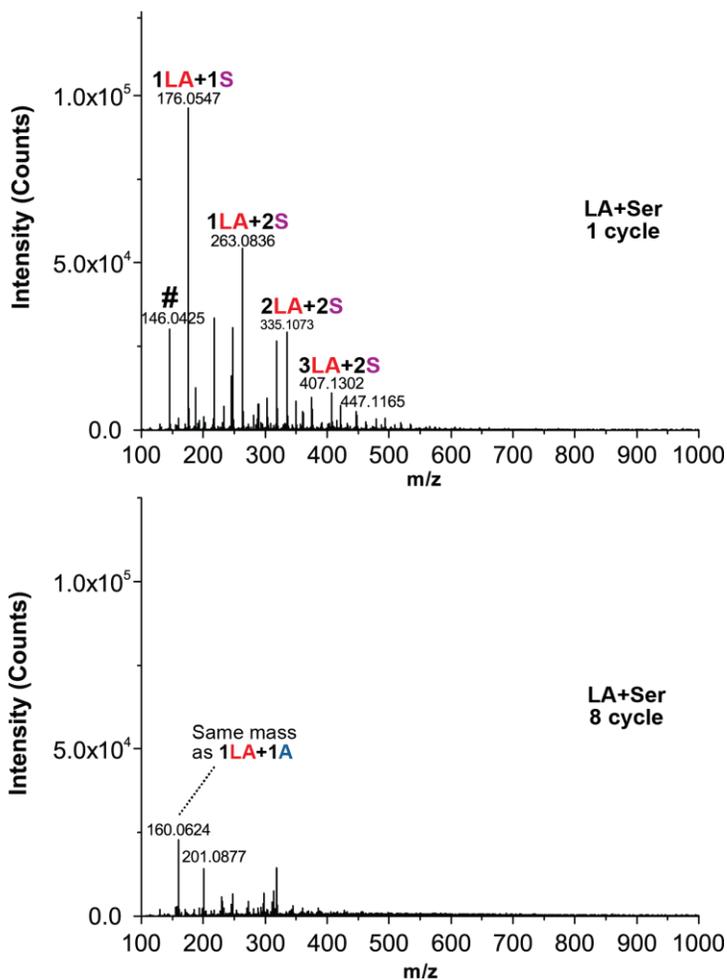


Figure 2.33 – Mass spectra of serine and lactic acid system after 1 and 8 cycles. # corresponds to loss of CH₂O from serine side-chain in serine-lactic acid dimer.

Another more surprising observation is that serine is frequently present on the N-terminus of these depsipeptides. Typically, we observe lactic acid on the N-terminus of depsipeptides (Figure 2.34). This suggests a different mechanism to form polar side-chain depsipeptides vs. nonpolar side-chain depsipeptides. This is also consistent with the full mass spectra obtained, which show depsipeptides are not present after repeated cycling. We hypothesize that unstable ester linkages are formed with the serine side-chain, but this type of sequence isomer cannot be directly confirmed by tandem MS.

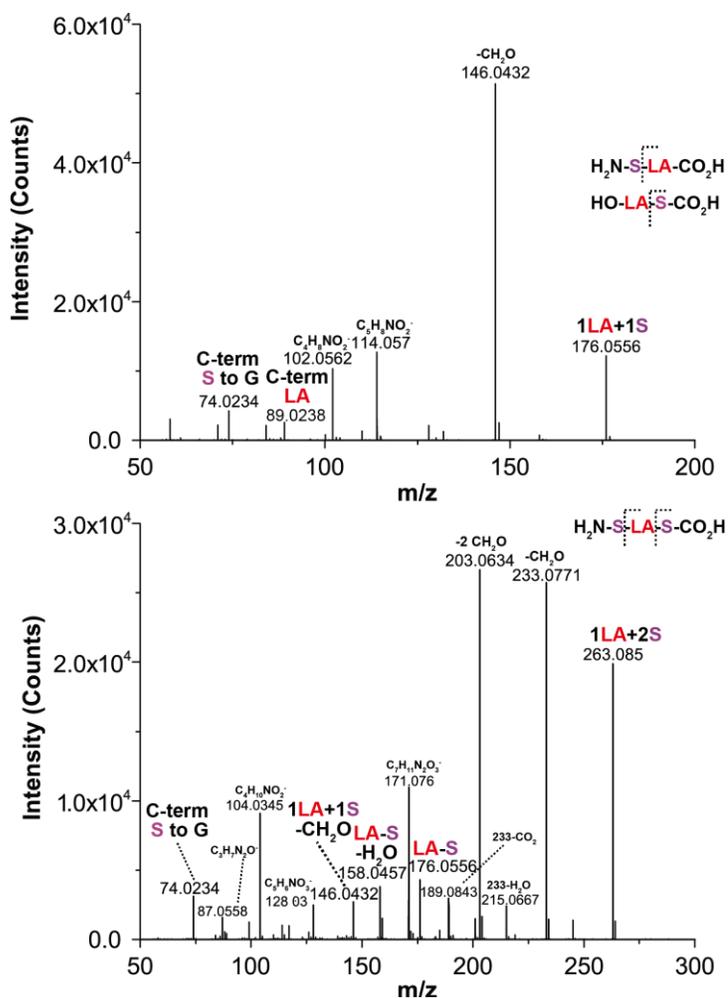


Figure 2.34 – Tandem MS of serine/lactic acid depsipeptides after 1 cycle. The primary sequences observed are S-LA/LA-S and S-LA-S. Collision energy = 20 eV.

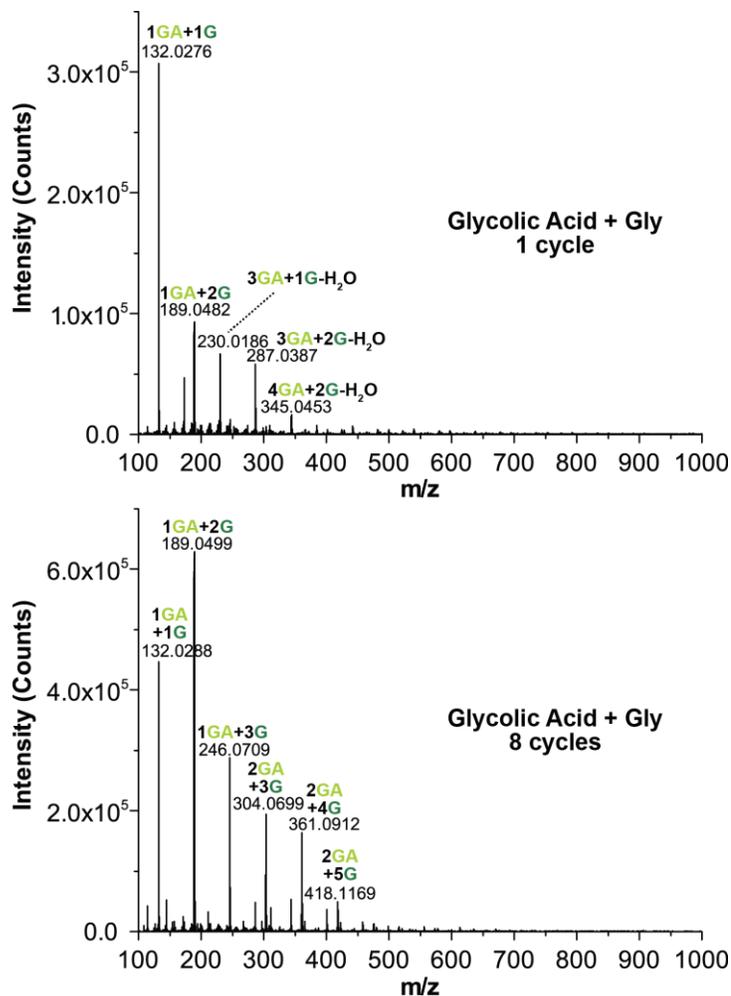


Figure 2.35 – Mass spectra of 1:1 mol GA/G depsipeptides after 1 cycle and after 8 cycles. After 1 cycle, the majority of depsipeptide signals correspond to water-loss species, suggesting they may be cyclic instead of linear. However, after 8 cycles, it appears the depsipeptides preferentially adopt a linear structure over the water-loss species, similar to those formed using lactic acid.

In addition, depsipeptide formation appears to be even more efficient when glycolic acid (GA) is used as the catalyst, in place of lactic acid. Glycolic acid is the simplest alpha-hydroxy acid with an amino acid analog and is found at similar abundance in prebiotic reaction mixtures [47] and meteorites [50] as lactic acid, glycine, and alanine. Depsipeptide with higher abundance of amino acid was found (Figure 2.35). The higher efficiency might come from the simpler structure of glycolic acid relative to lactic acid. The absence of the

methyl side-chain in glycolic acid provides a less hindered environment for esterification and ester-amide exchange reaction. Similar to the LA/A system, the ester-linked heterodimers produced from GA/G gradually disappear and the amide-linked oligomer becomes the dominant species (Figure 2.36).

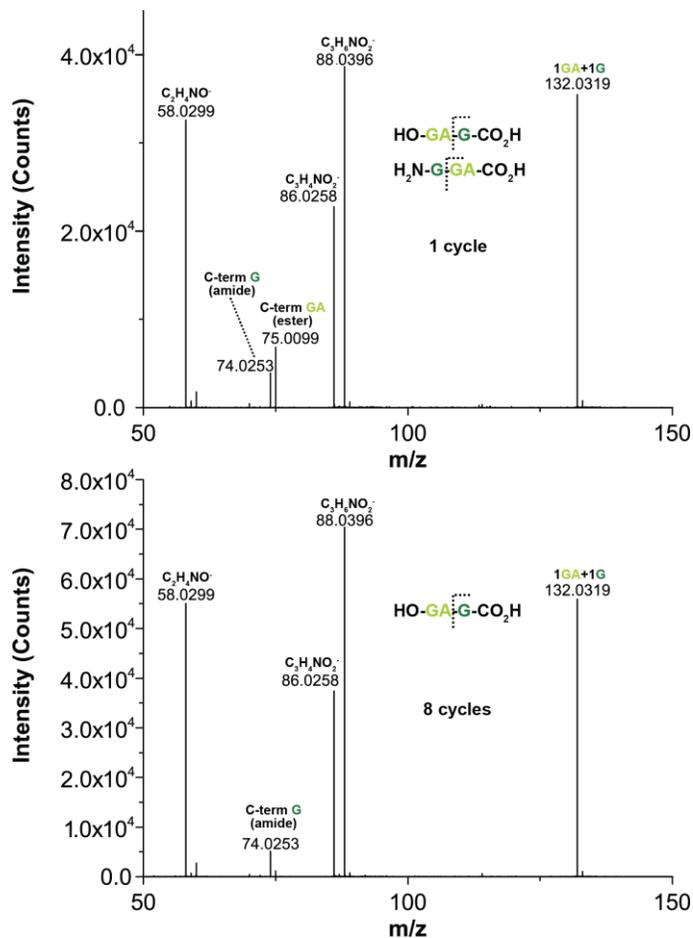


Figure 2.36 – Tandem MS sequencing of GA/G heterodimer after 1 and 8 cycles. Collision energy = 8 eV. After 1 cycle, both the ester and amide-linked heterodimer are present. After 8 cycles, only the amide-linked heterodimer remains.

Almost all amino acids used in life today are L-isomers. Why nature chooses L-amino acid is another unsolved question in the community of prebiotic chemistry [82]. It is worth knowing whether the ester-mediated pathway has a preference for L- or D- amino acids. When D-alanine was mixed with L-lactic acid, similar oligomer distribution forms after a few cycles. The reaction from D-alanine seems to proceed equally well as the reaction from L-alanine, at least for oligomers of length 5 and shorter (Figure 2.37). The m/z 377.13 signal in the bottom spectrum of Figure 2.37 corresponds to ^{13}C -containing depsipeptides and not the 5LA oligoester. As with L-alanine and L-lactic acid, depsipeptides are readily formed and are enriched with peptide sequences. However, a stereochemical preference might occur for longer oligomers [83].

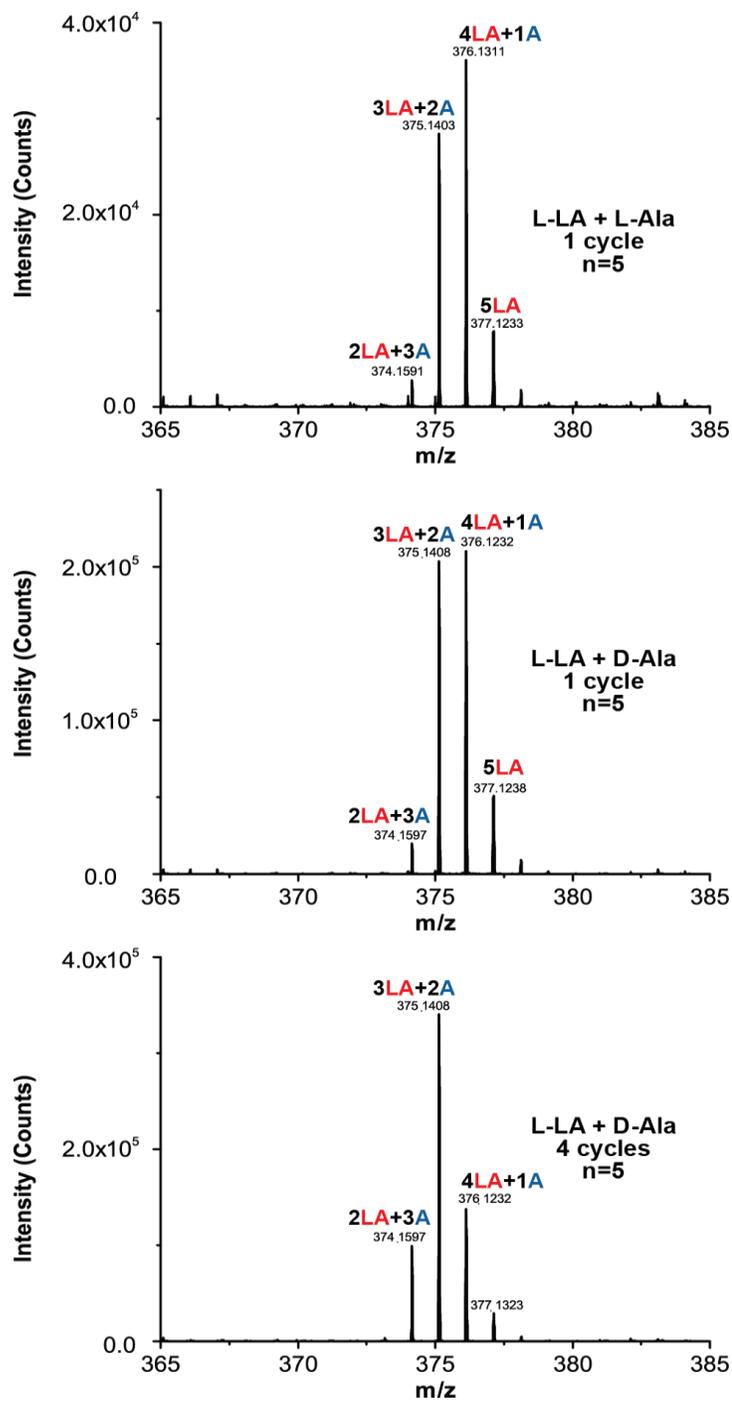


Figure 2.37 – Mass spectra of LA/A oligomers with different chiral configuration. (top) Zoomed-in mass spectrum of L-alanine and L-lactic acid pentamers after 1 cycle; (middle) zoomed-in mass spectrum of D-alanine and L-lactic acid pentamers after 1 cycle; (bottom) zoomed-in mass spectrum of D-alanine and L-lactic acid pentamers after 4 cycles.

2.4 Conclusion

In summary, we have demonstrated a prebiotically plausible mechanism for peptide bond formation enabled by hydroxy acids. Through a combination of ester-amide exchange and ester bond hydrolysis, depsipeptides are enriched with amino acids over time. Quantitative NMR analysis shows the ester-mediated pathway is more efficient than traditional methods.

We also found this reaction happens not only to simple amino acids, but also to other amino acids with more complex side-chains. Clearly, a comprehensive mapping of the chemical space of hydroxy acid-catalyzed peptide formation will require substantial effort, but doing so may provide valuable clues regarding the chemical pressures that shaped the sequences and structures of peptides formed on the prebiotic Earth.

CHAPTER 3. KINETICS OF PREBIOTIC DEPSIPEPTIDE FORMATION FROM THE ESTER-AMIDE EXCHANGE REACTION

3.1 Introduction

Our previous work has demonstrated an alternative pathway for the formation of peptide bonds by mixing amino acids with hydroxy acids. Hydroxy acids readily condense into polyesters, and subsequent ester-amide exchange reactions produce the copolymers of hydroxy acids and amino acids known as depsipeptides. Under repeated wet-dry cycling, depsipeptides emerge with a continuous amino acid backbone. While this reaction has been shown to be quite robust, the underlying mechanism and the kinetics still require more detailed study.

Traditional analysis of depsipeptide copolymers have been performed using NMR spectroscopy and gel permeation chromatography [62]. Using these techniques to analyze complex depsipeptide mixtures, however, is challenging due to extensively spectroscopic and chromatographic peak overlap. To address this issue, Codari *et al.* separated and then quantified lactic acid oligomers by high performance liquid chromatography (HPLC) [84]. Here, we utilize a similar approach to study the kinetics of copolymerization between amino acids and hydroxy acids.

Previous studies on copolymerization modeling applied methods of moments and kinetic Monte Carlo to solve the population balance of copolymers [85-88]. Han *et al.* tracked the kinetics of the copolycondensation with ester-amide exchange reaction, but

their model gave no information of each oligomer [89]. This chapter describes a simple approach that monitors the kinetics of several initial species from the ester-mediated depsipeptide formation.

These results are reprinted from [90] by permission of the PCCP Owner Societies.

3.2 Experimental method

3.2.1 Materials

L-lactic acid, L-valine, formic acid and deuterium oxide (99.9 mol %) were obtained from Sigma-Aldrich. Potassium hydrogen phthalate was from Fisher Chemical. The solvents used in the HPLC analysis were LC-MS grade water and acetonitrile (EMD OmniSolv).

3.2.2 Reaction procedure

The polymerization reaction typically started with 300 μL of 100 mM L-lactic acid and 10 mM L-valine. The solution was placed into a bent glass tube (Figure 3.1) at -15°C for at least 30 minutes to freeze the sample solution. To facilitate the movement of water between the two sides of the bent-tube reactor through evaporation of water, the pressure inside the tube was reduced to 0.33 atm and the tube was flame-sealed while at this reduced pressure. The drying process started when the tubes were inserted through the wall of the oven (65°C to 95°C).

For hydrolysis experiments, lactic acid oligomers were synthesized by drying lactic acid monomer at 100°C for two hours in an open vial. The dry oligoesters were dissolved in deionized water and incubated at temperatures ranging from 65°C to 95°C .

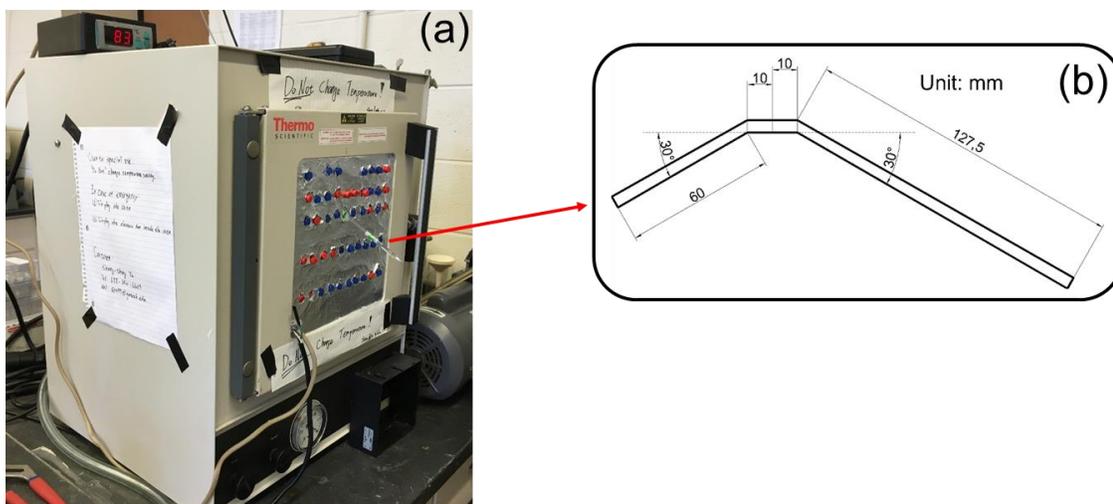


Figure 3.1 – The closed reactor. (a) The oven used to perform the reaction. (b) The blue print of the closed tube reactor.

3.2.3 Polymer characterization

The dry products were dissolved in deionized water and analyzed with an Agilent 1260 HPLC coupled with an Agilent 6130 single quadrupole mass spectrometer and a UV detector (210 nm). Oligomers were separated using a Kinetex XB-C18 column (150 × 2.1 mm, 2.6 μm particle size). The flow rate was 0.5 ml/min and the column temperature was held at 40°C. The mobile phase was water/acetonitrile (ACN) with 0.02 vol % of formic acid. The gradient method started with 2 vol % ACN for the first 2 minutes and ramped to 100 vol % ACN in 15 minutes. The ACN concentration was held at 100 vol % for 10 minutes and returned back to 2 vol % ACN immediately. For MS analysis, all data were obtained in negative mode electrospray ionization with a capillary voltage of 2.0 kV.

¹H NMR spectra were recorded on a Bruker Avance II-500 spectrometer with D₂O as a solvent and with potassium hydrogen phthalate (KHP) as an external standard. All spectra were collected using a 30 degree pulse and a 15 seconds relaxation delay.

3.3 Model development

3.3.1 Modeling of depsipeptide copolymerization

The kinetic model is based on the chain length and the composition of the depsipeptides. The copolymers are first classified by their terminal groups and their number of ester and number of amide bonds, as shown in Table 3.1. $P_{i,j}^O$ represents oligomers with hydroxy termini and $P_{i,j}^N$ with amine termini. Subscripts i and j indicate the number of ester and amide bonds, respectively. Oligomers of both types have a carboxylic acid group at their other termini. The model was developed from the simple step-growth polymerization (A–C) based on the work of Harshe *et al.* [91]. Building upon their model of lactic acid polymerization, the mechanism of copolymerization via ester-amide exchange was developed. The exchange reaction (D) forms two kinds of products: ejection from parent oligomers and growth with a newly attached nucleophile. Finally, we also included the evaporation of lactic acid monomer in our model (E).

Table 3.1 – The notation of species and the list of polymerization events in the model.

A	Esterification between two hydroxy-terminated oligomers	$P_{i,j}^O + P_{r,s}^O \xrightarrow{k_{11}} P_{i+r+1,j+s}^O + W$
B	Esterification between hydroxy-terminated and amine-terminated oligomers	$P_{i,j}^O + P_{r,s}^N \xrightarrow{k_{12}} P_{i+r+1,j+s}^N + W$
C	Ester hydrolysis	$P_{i+r+1,j+s}^X + W \xrightarrow{k_{h1}} P_{i,j}^X + P_{r,s}^O$
D	Ester-amide exchange —Ejection from parent oligomers	$P_{i_o,j_o}^X + P_{r,s}^N \xrightarrow{k_e} P_{(i_o-i-1)+r,(j_o-j)+s+1}^X + P_{i,j}^O$
	Ester-amide exchange —Growth with newly attached nucleophile	$P_{i_o,j_o}^X + P_{r,s}^N \xrightarrow{k_e} P_{i,j}^X + P_{i_o-(i-r)-1,j_o-(j-s-1)}^O$
E	Lactic acid evaporation	$P_{0,0}^O(\text{condensed phase}) \xrightarrow{K_{pLA}} P_{0,0}^O(\text{gas})$

i, r, i_o : Number of ester bonds	W : Water (μmole)
s, j, j_o : Number of amide bonds	$X=O$ or N
$P_{i,j}^O$: Polymer chain terminated with a hydroxy group (μmole)	$P_{r,s}^N$: Polymer chain terminated with an amine group (μmole)

Since the direct formation and hydrolysis of amide bonds are known to be insignificant over the temperature range of our study, these two mechanisms were not included in the model [92, 93]. Because we did not detect cyclic dimers in our previous work or by our current HPLC analysis, the formation of cyclic dimers was not considered. The model assumes that the reaction mixture is homogeneous and all kinetic rate constants are independent of polymer chain length, composition and concentration.

The dynamic change of hydroxy-terminated oligomers is described in Eq. (3.1):

$$\frac{dP_{i,j}^O}{dt} = \frac{1}{V} \{s_{i,j}^O + e_{i,j}^O + h_{i,j}^O\} \quad (3.1)$$

where $P_{i,j}^O$ is the number of moles of hydroxy-terminated oligomers with i ester bonds and j amide bonds. The system volume is V . The effect of volume change is explicitly considered because a significant fraction of water is evaporated during the drying process. The contribution from esterification, exchange reaction and hydrolysis are represented by $s_{i,j}^O$, $e_{i,j}^O$ and $h_{i,j}^O$, respectively.

The contribution from esterification to $P_{i,j}^O$ is expressed as

$$s_{i,j}^O = k_{11} \sum_{r=0}^{i-1} \sum_{s=0}^j P_{r,s}^O P_{i-1-r,j-s}^O - 2k_{11} P_{i,j}^O \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} P_{r,s}^O - k_{12} P_{i,j}^O \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} P_{r,s}^N \quad (3.2)$$

The first two terms correspond to the gain and loss of oligomers with hydroxy termini by ester formation. The last term is ester bond formation between hydroxy-terminated and amine-terminated oligomers, which produces a single amine-terminated oligomer.

The contribution from the exchange reaction is described using Eq. (3.3).

$$\begin{aligned} e_{i,j}^O &= -k_e \left(i P_{i,j}^O \right) \left(\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} P_{r,s}^N \right) \\ &+ k_e \left(\sum_{i_0=i+1}^{\infty} \sum_{j_0=j}^{\infty} \frac{C_{j_0-j}^{(i_0-i-1)+(j_0-j)} C_j^{i+j}}{C_{j_0}^{i_0+j_0}} P_{i_0,j_0}^O \right) \left(\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} P_{r,s}^N \right) \\ &+ k_e \left(\sum_{i_0=i+1}^{\infty} \sum_{j_0=j}^{\infty} \frac{C_{j_0-j}^{(i_0-i-1)+(j_0-j)} C_j^{i+j}}{C_{j_0}^{i_0+j_0}} P_{i_0,j_0}^N \right) \left(\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} P_{r,s}^N \right) \\ &+ k_e \left(\sum_{r=0}^i \sum_{s=0}^{j-1} P_{r,s}^N \right) \left(\sum_{i_0=i-r+1}^{\infty} \sum_{j_0=j-s-1}^{\infty} \frac{C_{j-s-1}^{(i-r)+(j-s-1)} C_{j_0-(j-s-1)}^{[i_0-(i-r)-1]+[j_0-(j-s-1)]}}{C_{j_0}^{i_0+j_0}} P_{i_0,j_0}^O \right) \end{aligned} \quad (3.3)$$

In the exchange reaction, a nucleophilic amine ($P_{r,s}^N$) attacks a parent oligomer (P_{i_o,j_o}^X). The first term in Eq. (3.3) is the loss of $P_{i,j}^O$ by the exchange reaction. The second and third terms in Eq. (3.3) represent the products formed by ejection from hydroxy-terminated and amine-terminated parent oligomers. The parent oligomer might have many different sequences, but only some of them allow the formation of $P_{i,j}^O$, having the specified i ester bonds and j amide bonds. To estimate the rate of forming $P_{i,j}^O$ from P_{i_o,j_o}^O , the total numbers of possible sequences of P_{i_o,j_o}^O and those that allow $P_{i,j}^O$ formation are calculated first. Here, we assume all sequences are equally likely to be present. Therefore, the ratio between the two values gives the fraction of sequences forming $P_{i,j}^O$ from the parent oligomer. The number of possible sequences is the potential arrangements of ester and amide linkages, and it can be calculated by the binomial coefficients ($C_j^{i+j} = (i+j)! / j!i!$). As shown in the first part of Figure 3.2, the numbers of sequences of the parent oligomers ($C_{j_o}^{i_o+j_o}$) and the ejection product (C_j^{i+j}) are calculated using binomial coefficients. For the segment of the parent oligomer in the product with a new amide bond, its number of bonds is calculated by subtracting the number of bonds in the ejection product from that of the parent oligomer. Because one ester bond is consumed by the exchange reaction, it is subtracted from this segment as well. Therefore, this segment has i_o-i-1 ester bonds, j_o-j amide bonds and $C_{j_o-j}^{(i_o-i-1)+(j_o-j)}$ possible sequences. The product of $C_{j_o-j}^{(i_o-i-1)+(j_o-j)}$ and C_j^{i+j} is the number of those sequences that allow the formation of $P_{i,j}^O$. By dividing this number by $C_{j_o}^{i_o+j_o}$, the fraction of P_{i_o,j_o}^O that led to $P_{i,j}^O$ was estimated.

The same approach was used to estimate the reaction rate of growing polymer chains with a new amide bond, as shown in the second part of Figure 3.2. In this case, the number of bonds in the nucleophile and the amide bond that is formed by the exchange reaction are subtracted from the desired product, $P_{i,j}^O$, which gives the number of bonds from the parent oligomer in $P_{i,j}^O$. Then, the difference between the parent oligomer and the segment in the growing polymer chain is calculated and reduced by one to account for the ester bond consumed by the exchange reaction. This simple calculation gives the number of bonds in the ejection product. Using this approach, we counted the numbers of sequences for each segment and calculated the fraction of the exchange reactions that led to the formation of $P_{i,j}^O$. In addition to the theory described above, two examples are included in Appendix A for illustrative purpose.

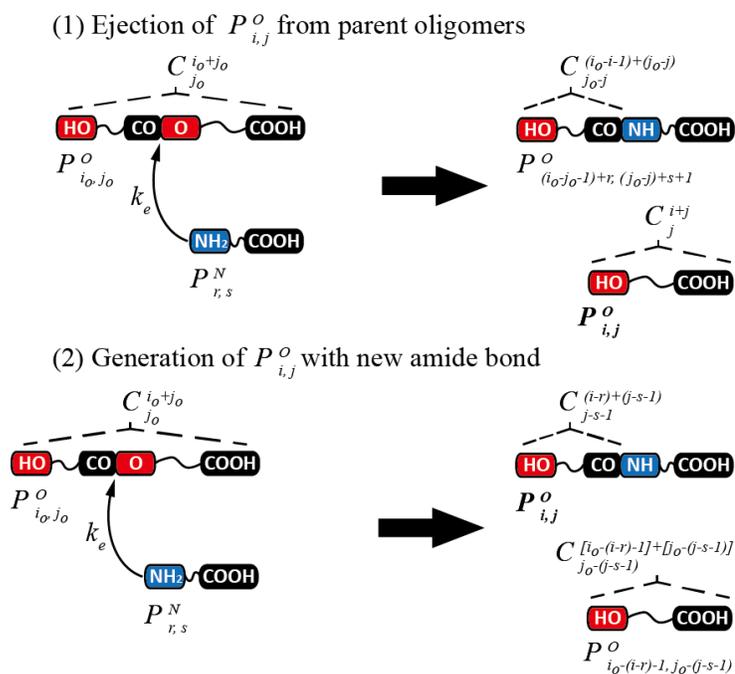


Figure 3.2 – The formation of $P_{i,j}^O$ from the exchange reaction (Reaction D in Table 3.1). The parent oligomer, P_{i_0, j_0}^O , is attacked by the nucleophile ($P_{r,s}^N$). $P_{i,j}^O$ with a specified number of ester and amide bonds forms either from (1) ejection or (2) growth. The numbers of possible polymer sequences were calculated using binomial coefficients.

$$\begin{aligned}
h_{i,j}^O &= -k_{h1} \left(iP_{i,j}^O \right) W \\
&+ 2k_{h1} \left(\sum_{i_o=i+1}^{\infty} \sum_{j_o=j}^{\infty} \frac{C_j^{i+j} C_{j_o-j}^{(i_o-i-1)+(j_o-j)}}{C_{j_o}^{i_o+j_o}} P_{i_o,j_o}^O \right) W \\
&+ k_{h1} \left(\sum_{i_o=i+1}^{\infty} \sum_{j_o=j}^{\infty} \frac{C_j^{i+j} C_{j_o-j}^{(i_o-i-1)+(j_o-j)}}{C_{j_o}^{i_o+j_o}} P_{i_o,j_o}^N \right) W
\end{aligned} \tag{3.4}$$

The hydrolysis rate of $P_{i,j}^O$, shown in Eq. (3.4), is similar to the exchange reaction. Here, W is the number of moles of water, and it replaces the $P_{r,s}^N$ term in Eq. (3.3). The first term represents the loss of oligomers due to hydrolysis of an ester bond. The others represent the increase of the oligomer due to hydrolysis of longer oligomers.

The balance equations for $P_{i,j}^N$ are similar to those for $P_{i,j}^O$:

$$\frac{dP_{i,j}^N}{dt} = \frac{1}{V} \{s_{i,j}^N + e_{i,j}^N + h_{i,j}^N\} \tag{3.5}$$

$$s_{i,j}^N = k_{12} \sum_{r=0}^{i-1} \sum_{s=0}^j P_{r,s}^N P_{i-1-r,j-s}^O - k_{12} P_{i,j}^N \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} P_{r,s}^O \tag{3.6}$$

$$\begin{aligned}
e_{i,j}^N &= -k_e \left(iP_{i,j}^N \right) \left(\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} P_{r,s}^N \right) \\
&+ k_e \left(\sum_{r=0}^i \sum_{s=0}^{j-1} P_{r,s}^N \right) \left(\sum_{i_o=i-r+1}^{\infty} \sum_{j_o=j-s-1}^{\infty} \frac{C_{j-s-1}^{(i-r)+(j-s-1)} C_{j_o-(j-s-1)}^{[i_o-(i-r)-1]+[j_o-(j-s-1)]}}{C_{j_o}^{i_o+j_o}} P_{i_o,j_o}^N \right) \\
&- k_e \left(\sum_{i_o=0}^{\infty} \sum_{j_o=0}^{\infty} i_o P_{i_o,j_o}^O \right) P_{i,j}^N - k_e \left(\sum_{i_o=0}^{\infty} \sum_{j_o=0}^{\infty} i_o P_{i_o,j_o}^N \right) P_{i,j}^N
\end{aligned} \tag{3.7}$$

$$h_{i,j}^N = -k_{h1} \left(i P_{i,j}^N \right) W + k_{h1} \left(\sum_{i_0=i+1}^{\infty} \sum_{j_0=j}^{\infty} \frac{C_j^{i+j} C_{j_0-j}^{(i_0-i-1)+(j_0-j)}}{C_{j_0}^{i_0+j_0}} P_{i_0,j_0}^N \right) W \quad (3.8)$$

In Eq. (3.7), the first term is the loss of $P_{i,j}^N$ when it is attacked by another nucleophilic amine. The second term is the generation rate of $P_{i,j}^N$ with a new amide bond from amine-terminated parent oligomers. The ejection product from amine-terminated parent oligomers are hydroxy-terminated oligomers, so it is included in Eq. (3.3). The last two terms in Eq. (3.7) represent the consumption rate of $P_{i,j}^N$ when it acts as a nucleophile.

The dynamic change for the moles of water comes from both reaction and evaporation during the drying process, which is expressed in Eq. (3.9). The first part in Eq. (3.9) is the change of water because of the esterification (k_{11} and k_{12}) and hydrolysis (k_{h1}). The evaporation rate of water ($R_{v,w}$) in the closed reactor is derived by considering it as a model of water diffusion through a long tube [94]. The flux of water ($N_{w,z}$) is expressed as the summation of diffusion and convection as shown in Eq. (3.10), where K_{pw} is the diffusion rate constant and y_w is the gas phase fraction of water. The flux of air, $N_{b,z}$, is set to zero based on the assumption that the air inside the tubes is stagnant.

$$\frac{dW}{dt} = \frac{1}{V} \left\{ \begin{array}{l} k_{11} \left(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P_{i,j}^O \right) \left(\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} P_{r,s}^O \right) \\ + k_{12} \left(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P_{i,j}^O \right) \left(\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} P_{r,s}^N \right) \\ - k_{h1} W \left(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} i P_{i,j}^O \right) - k_{h1} W \left(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} i P_{i,j}^N \right) \end{array} \right\} - R_{v,w} \quad (3.9)$$

$$N_{w,z} = -K_{pw} \frac{dy_w}{dz} + y_w(N_{w,z} + N_{b,z}) \quad (3.10)$$

After rearrangement, the flux of water is given by

$$N_{w,z} = -\frac{K_{pw}}{1-y_w} \frac{dy_w}{dz} \quad (3.11)$$

Using the quasi-steady state method, the net flux of water through the z-direction is assumed to be zero. That is

$$\frac{d}{dz} \left(-\frac{K_{pw}}{1-y_w} \frac{dy_w}{dz} \right) = 0 \quad (3.12)$$

Here, the following boundary conditions are used

$$z = z_1(t), \quad y_{w,1} = \frac{P_{T_H}^* x_w}{P_{air}(T_H) + P_w^*(T_H) x_w} \quad (3.13)$$

$$z = z_2, \quad y_{w,2} = \frac{P_{T_R}^*}{P_{air}(T_R) + P_w^*(T_R) P_{T_R}^*} \quad (3.14)$$

where $z_1(t)$ is the height of liquid as a function of time and z_2 is the length of the tube inside the oven. Using Raoult's Law, the gas phase fractions of water at the two boundaries, $y_{w,1}$ and $y_{w,2}$, are calculated by Eqs. (3.13) and (3.14). T_H is the oven temperature and T_R is the ambient temperature. The partial pressures of air (P_{air}) at different temperatures are calculated by assuming air is an ideal gas. $P_w^*(T_H)$ and $P_w^*(T_R)$ are the vapor pressures of

water at T_H and T_R , respectively. x_w is the mole fraction of water in the liquid phase. The vapor pressure of water is modeled according to Antoine' Law:

$$P_w^* = \exp\left(A - \frac{B}{T + C}\right) \quad (3.15)$$

with $A = 16.3872$, $B = 3885.70$ and $C = 230.170.28$ The unit of T is $^{\circ}\text{C}$, with P_w^* in kPa.

Solving Eq. (3.12) gives the concentration profile of water inside the tube:

$$\frac{1 - y_w}{1 - y_{w,1}} = \left(\frac{1 - y_{w,2}}{1 - y_{w,1}}\right)^{\frac{z - z_1}{z_2 - z_1}} \quad (3.16)$$

We obtained the rate of mass transfer at the liquid-gas interface by differentiating Eq. (3.16).

$$R_{v,w} = S_A N_{w,z} \Big|_{z_1} = \left(\frac{S_A K_{pw}}{1 - y_{w,1}}\right) \left(\frac{dy_w}{dz} \Big|_{z_1}\right) = \left(\frac{S_A K_{pw}}{z_2 - z_1}\right) \ln\left(\frac{1 - y_{w,2}}{1 - y_{w,1}}\right) \quad (3.17)$$

where S_A is the cross-sectional area of the tube.

The same mass transport model was used for the evaporation rate of the lactic acid monomer. Because of the lower concentration of lactic acid compared to water, the convective flux was assumed to be zero. Also, the temperature dependence of the lactic acid vapor pressure is unknown, so a general mass transport coefficient (K_{pLA}) that includes the pressure effect was used.

$$R_{v,LA} = K_{pLA} S_A \left(\frac{x_{LA}}{z_2 - z_1(t)} \right) \quad (3.18)$$

The numerical values of the rate constants and mass transfer constants were estimated at each temperature, using the maximum likelihood method [95]. Following the method proposed by Bard, we also calculated the 95% confidence interval on each optimized parameter from its parameter covariance matrix [95].

All simulations and calculations were performed in MATLAB with the *patternsearch* function for minimization and the *ode15s* function for integration of differential equations.

3.3.2 Calculation of heats and entropies of activation

The activation parameters for each reaction were determined based on kinetic data and the method used by Wynne-Jones and Eyring [96]. The activation energy, E_a , was calculated from the slope of the fitted Arrhenius equation:

$$\ln(k) = \frac{-E_a}{R} \frac{1}{T} + \ln(A) \quad (3.19)$$

where R is the universal gas constant, T is temperature (K) and A is the prefactor in the Arrhenius equation.

The enthalpy (ΔH^\ddagger) and free energy (ΔG^\ddagger) of activation for each reaction were calculated from the following two equations:

$$\Delta H^\ddagger = E_a - RT \quad (3.20)$$

$$\Delta G^\ddagger = -RT \ln \left(\frac{k h}{k_B T} \right) \quad (3.21)$$

Here, k is the rate constant for the particular reaction at temperature T , h is Planck's constant and k_B is Boltzmann's constant.

Finally, the activation entropy ($T\Delta S^\ddagger$) is the difference between the activation enthalpy and activation free energy.

$$T\Delta S^\ddagger = \Delta H^\ddagger - \Delta G^\ddagger \quad (3.22)$$

3.4 Results and discussion

3.4.1 Characterization of depsipeptides by HPLC method

Several amino acid and hydroxy acid pairs were studied. It was found that most of the initial depsipeptides between L-valine (V) and L-lactic acid (LA) were well-separated using HPLC. In addition, both of these materials have been proposed as prebiotically plausible compounds [47, 97]. Therefore, this study focuses on the kinetics of valine and lactic acid copolymerization. The loss of lactic acid from the dry solid during wet-dry cycling was observed in our previous study. We hypothesized that this loss was due to the evaporation of lactic acid during the drying stage, but the evaporation mechanism was not investigated or quantified. Because the amount of lactic acid present in these reactions impacts the rate and yield of peptide bond formation, we developed a bent-tube reactor to prevent loss of

lactic acid during sample heating (Figure 3.3). Specifically, the bent-tube reactor allows us to contain all species and to monitor the evaporation of water and lactic acid. Here, we combine both experimental and model-based approaches to achieve a comprehensive analysis of the formation of depsipeptides in a model prebiotic scenario.

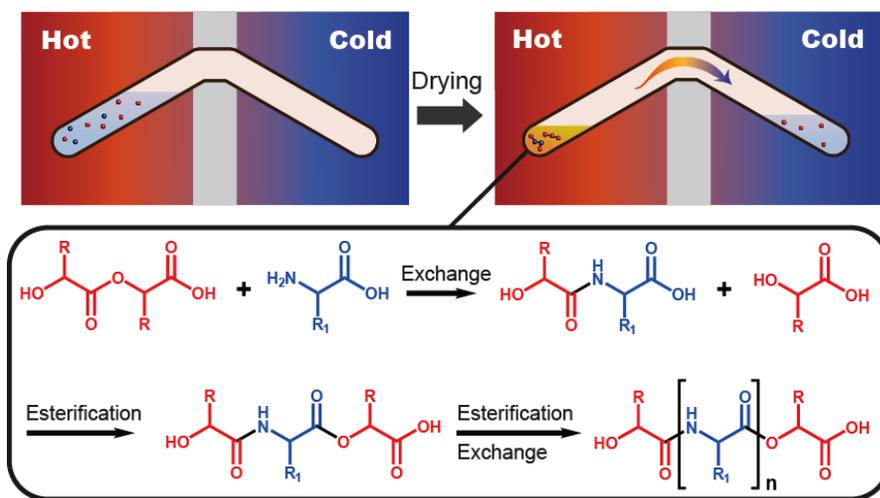


Figure 3.3 – The formation of lactic acid/valine depsipeptides in the bent-tube closed reactor. R is the methyl group for lactic acid and R₁ is the isopropyl group for valine. The solution was sealed in a bent glass tube and inserted into an oven through an insulating wall. During the drying process, evaporated water and lactic acid were collected on the cold side of the tube outside the oven. The copolymerization between valine and lactic acid proceeded in the drying mixture (hot side) via esterification and exchange reactions.

Following HPLC separation, negative-ion mode mass spectrometry was used to identify each peak in the chromatogram as shown in Figure 3.4. We focused on the following species: 1V, 1LA, 1V-1LA, 2LA, 1LA-1V, 3LA, 4LA and 5LA. The notation for 1V-1LA and 1LA-1V is based on their sequences from the N-terminus. The mass spectrum for each of these chromatographic peaks has only one dominant species, except for 1V and 4LA, which contained other products (Figure 3.5). We also found that some species tended to form non-covalent clusters. This phenomenon was observed specifically for 1LA-1V and 3LA, which formed $[2M-H]^-$ and $[2(3M)+Na-2H]^-$, respectively.

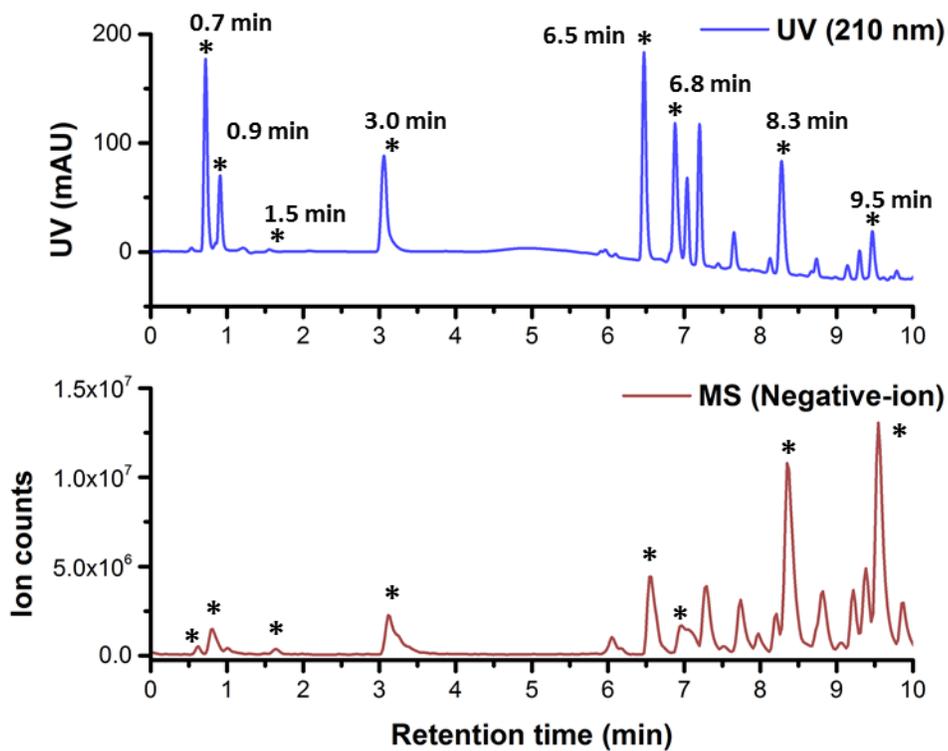


Figure 3.4 – HPLC-UV/MS analysis of the LA/V depsipeptide mixture. The sample was prepared from 33.34 mM valine and 100 mM lactic acid, 300 μ l solution. The solution was dried in the closed reactor for 24 hours.

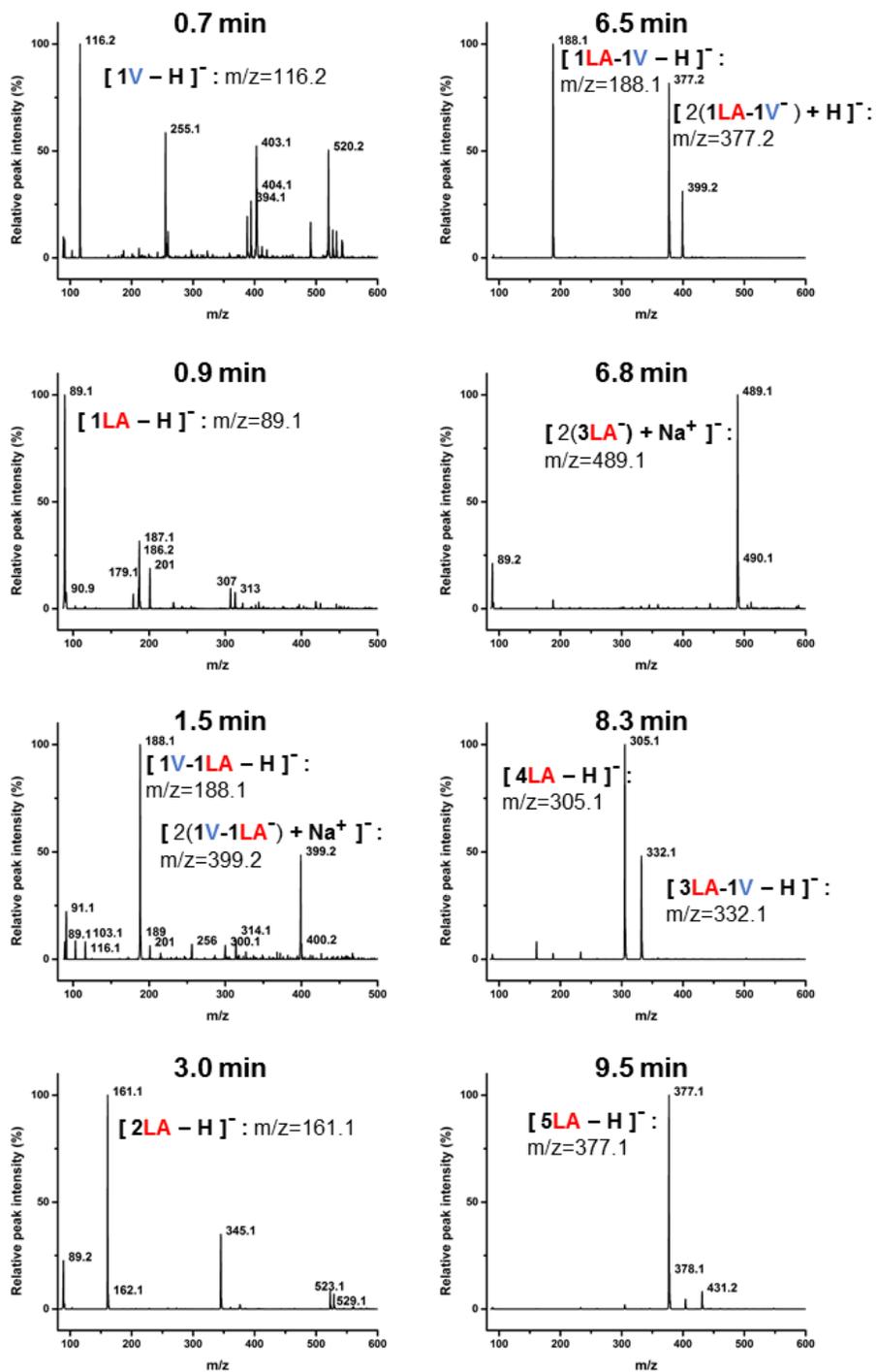


Figure 3.5 – Mass spectra of LA/V depsipeptides after HPLC separation.

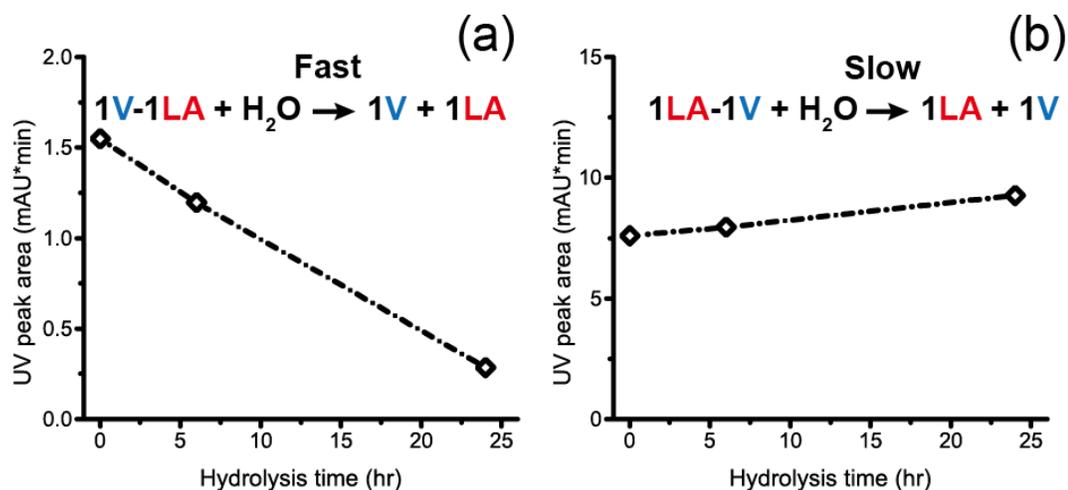


Figure 3.6 – Changes of UV peak area for the two hetero-dimers during hydrolysis. (a) The mixed dimer eluted at 1.5 min. (b) The mixed dimer eluted at 6.5 min. Sample were prepared by drying the monomer mixture at 85°C in the closed reactor for 6 hours.

The mixed dimers, 1V-1LA and 1LA-1V, have identical masses, but have different retention times in the HPLC trace. A hydrolysis experiment was performed to discern their linkages as seen in Figure 3.6. The area for the peak eluting at 1.5 min decreased significantly after 24 hours of hydrolysis, while the complementary species at 6.5 min increased slightly, due the hydrolysis from longer oligomers. Susceptibility to hydrolysis indicates whether the mixed dimers are linked by a labile ester bond or by a more stable amide bond.

Following the method proposed by Codari *et al.* [84], the response factor of 2LA was estimated by monitoring the hydrolysis from lactide to the lactic acid dimer. It was shown that the extinction coefficients increase linearly with the number of chromophores in the polymers (i.e., carbonyl groups). Therefore, the response factors of longer lactic acid oligomers were approximated by extrapolating the extinction coefficient of 2LA against the chain length. The UV response factor of the LA-V dimer was estimated by calculating the conversion of valine after a short-time reaction to the LA-V dimer. The response factor

of 2LA was also used for V-LA due to the similarity between their chromophores. The UV response factors used in this study are summarized in Table 3.2. Quantitative NMR analysis was used to determine the concentration of the valine monomer because LC-MS analysis indicated that other products co-eluted with this monomer.

Table 3.2 – UV response factors at 210 nm for LA/V oligomers.

Compound	UV response factor, k [$\mu\text{mole/mAU}\cdot\text{sec}$]
1LA	2.64×10^{-5}
2LA	8.29×10^{-6}
3LA	5.53×10^{-6}
4LA	4.15×10^{-6}
5LA	3.32×10^{-6}
1V-1LA	8.29×10^{-6}
1LA-1V	2.48×10^{-6}

3.4.2 Estimation of rate constants

The ester hydrolysis rate constants (k_{h1}) were determined by monitoring the oligomer distribution during hydrolysis of polylactic acid. The initial lactic acid oligomers were prepared by drying lactic acid at 100°C for two hours. This reaction produced lactic acid oligomers up to tetramers with the reaction products being fully soluble in water. The solution was incubated at temperatures ranging from 65°C to 95°C for 8 hours (Figure 3.7). The forward polymerization rate in this aqueous solution was negligible considering the low concentration of hydroxy acid (100 mM total, including free monomer and monomer in polymer). Therefore, the hydrolysis rate constants were estimated independently using Eq. (3.4). To estimate the rate and mass transfer constants for polymerization, drying experiments in the closed reactor at temperatures ranging from 65°C to 95°C were

performed. During the drying experiment of pure water, we monitored the amount of water that was collected on the cold side of the bent-tube reactor (Figure 3.8) and fit the data with Eq. (3.17) to estimate the mass transfer constant of water (K_{pw}). The rate constants of ester hydrolysis and water evaporation both follow the linear Arrhenius relation (Figure 3.9 and Figure 3.10).

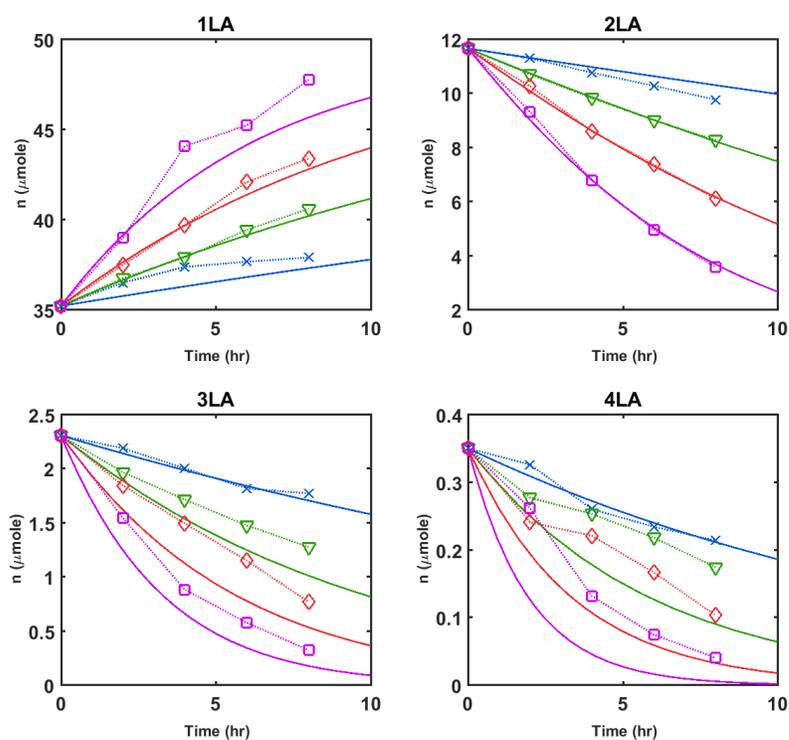


Figure 3.7 – Monomer and oligomer distributions versus time profile in the hydrolysis stage. The reaction temperatures are denoted as different colors: purple (95°C), red (85°C), green (75°C) and blue (65°C). Solid lines are the model predictions. The experimental data points are represented by \square , \diamond , ∇ and \times symbols, respectively.

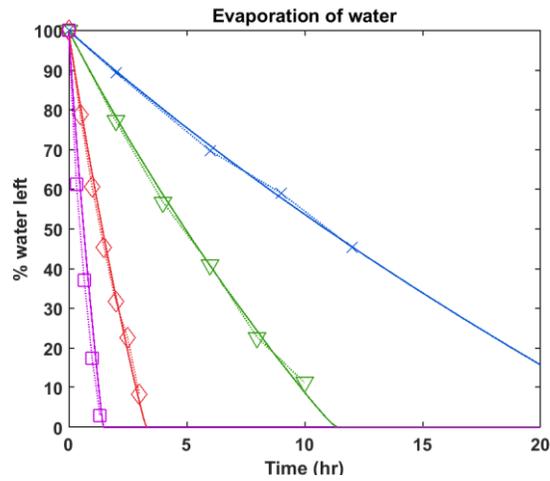


Figure 3.8 – The amount of water left in the drying side of the reaction at different temperatures. Purple (95°C), red (85°C), green (75°C) and blue (65°C). Solid lines are the model predictions. The experimental data points are represented by \square , \diamond , ∇ and \times symbols, respectively.

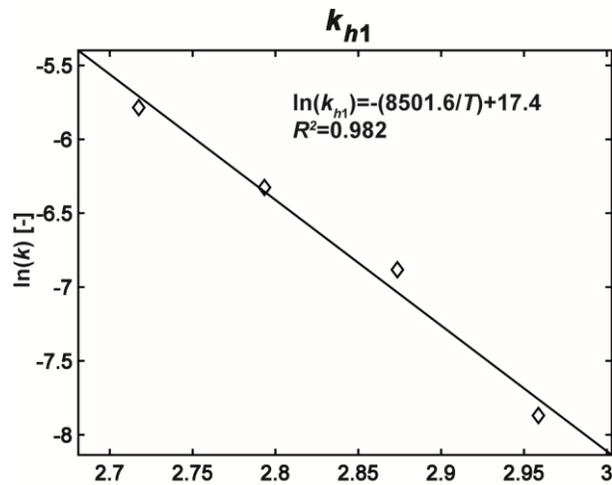


Figure 3.9 – Arrhenius plot of the hydrolysis rate constants evaluated from the lactic acid hydrolysis experiments.

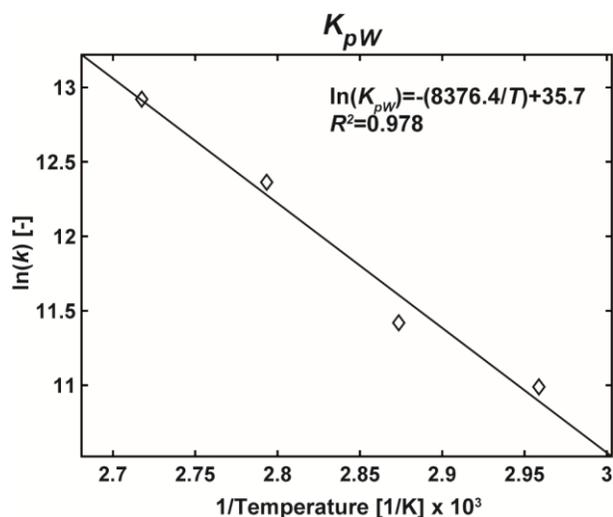


Figure 3.10 – Arrhenius plot of the water evaporation rate constants.

To estimate the four constants in the copolymerization reaction, the model was fit to the data from the single dry-down reaction in the closed reactor at different temperatures. The predictions of the model with these optimized rate constants and the experimental data are shown in Figure 3.11. The theoretical results and the experimental data achieved agreement qualitatively and quantitatively.

Figure 3.12 shows the trend of these optimized rate constants under different temperatures in Arrhenius plots. In most cases, the 95% confidence interval of each optimized parameter was narrow, except for k_{12} , which had a relatively large confidence interval. We speculate that the accuracy of k_{12} estimation can be improved if more copolymer products can be quantified.

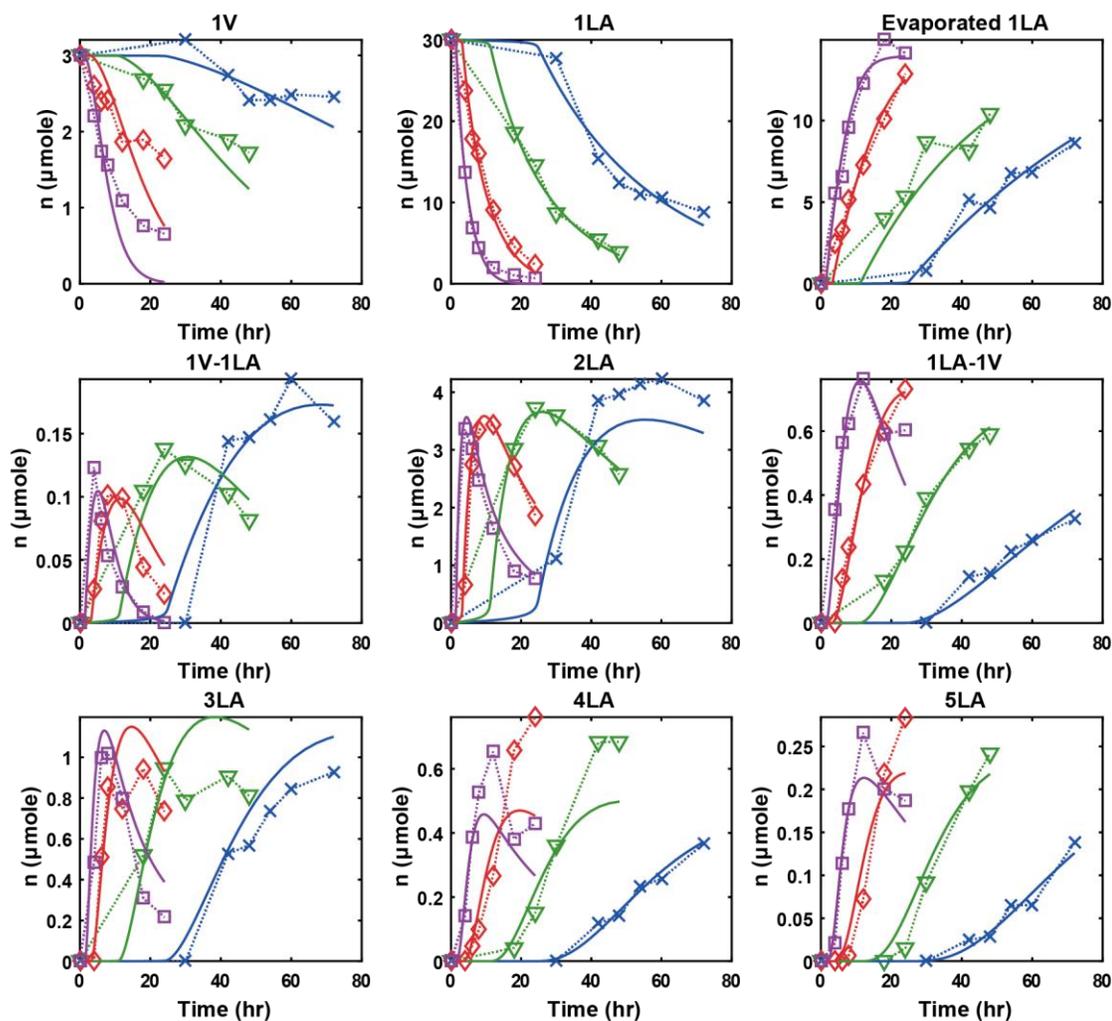


Figure 3.11 – Oligomer distributions versus time profile in the drying stage. The reaction temperatures are denoted as different colors: purple (95°C), red (85°C), green (75°C) and blue (65°C). Solid lines are the model predictions. The experimental data points are represented by \square , \diamond , ∇ and \times symbols, respectively. Molar ratio (LA/V) is 100/10.

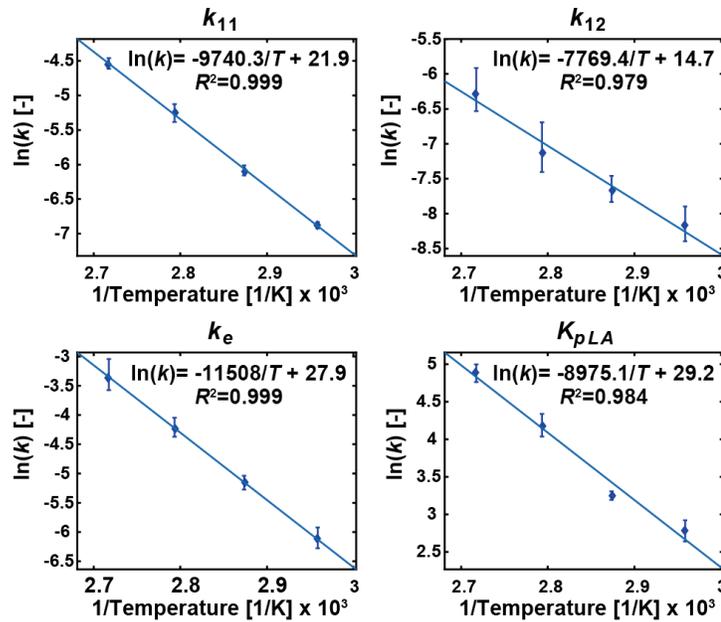


Figure 3.12 – The Arrhenius plot of the four rate constants evaluated from the drying experiments. The estimated rate constants at each temperature are plotted along with their 95% confidence intervals.

All rate constants displayed a linear trend between $\ln(k)$ and $1/T$, with the lowest R^2 observed being 0.979 for k_{12} . It is worth noting that the empirical Arrhenius equation still represents the reaction under a nearly dry state, which is a less homogeneous environment in general. The observation is consistent with what has been previously reported for the melt polymerization of lactic acid, where the melted monomer acts as a solvent to enhance the diffusion rate of each reactant [91]. These fits allow us to predict the behavior of the system at other temperatures and provide more insight on the activation energy of each step.

The predicted values of the valine monomer and the lactic acid tetramer deviated from the experimental measurements when the reaction time increased. The model predicted a higher consumption rate of the valine monomer. It is possible that the viscosity of the mixture starts to play a greater role at the later stage of the reaction and limits the

ability of valine to participate in the ester-amide exchange reaction. The exchange rate can also depend on the chain length of oligomers that were attacked by the valine monomer. The deviation of 4LA is not surprising since the UV signal of 4LA also contains other depsipeptide products, and the measured value overestimates the actual value.

We also investigated the copolymerization for different molar ratios (LA/V) ranging from 100/0 to 100/50. For samples with lower loading of valine, the developed kinetic model continued to fit the experimental data well with the same rate constants obtained from the 100/10 experiment (Figure 3.13 and Figure 3.14). We also evaluated the rate constants of pure lactic acid polymerization by fitting the 100/0 data with our model. Compared to the polymerization of pure lactic acid, it is worth noting that the forward esterification rate constant (k_{11}) is significantly higher when a small amount of valine was added. At 85°C, the k_{11} values are 0.0026 and 0.00518 ($\text{L}\cdot\text{mol}^{-1}\cdot\text{h}^{-1}$) for the 100/0 and 100/0 experiments, respectively. We speculate that the additional carboxylic acid groups from the amino acid promote the esterification by acting as an acid catalyst [98]. When the amount of valine increased, additional formation of the two hetero-dimers was observed (Figure 3.15). However, the 100/20 sample gives similar results as the 100/50 one and our model no longer fits the experimental data. Significant deviations from experimental data were observed in the concentration changes of the two hetero-dimers. We attribute this limitation of our model to the assumption of a homogeneous mixture. When the amount of valine increased beyond its solubility limit in the oligomer mixture, some valine monomers were observed to precipitate as solid and may not participate in the reaction. Therefore, even when the initial amount of valine monomer increased dramatically, the progress of the reaction seemed to be limited by the solubility of valine.

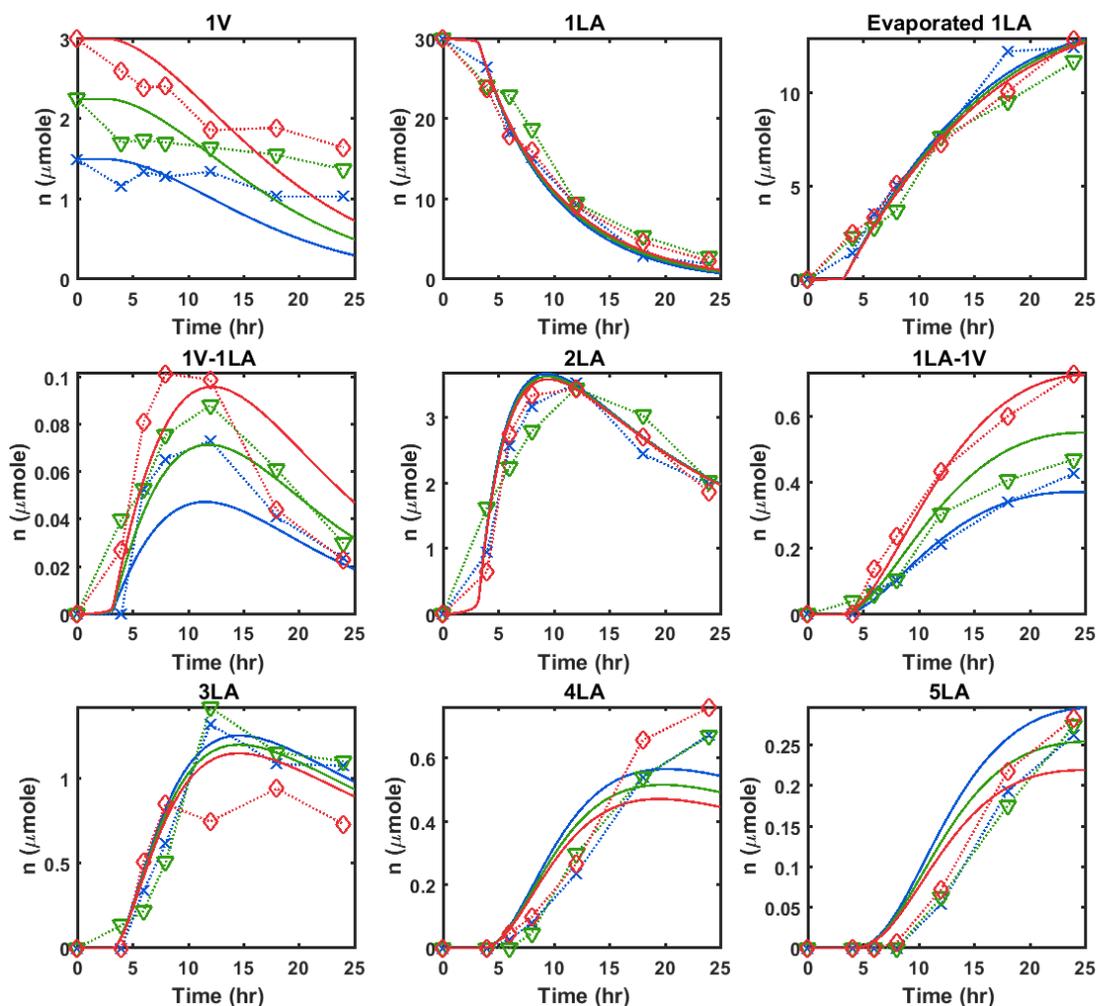


Figure 3.13 – Experimental results and model prediction for low valine loading. The results for different LA/V ratios are denoted as different colors: blue (100/5), green (100/7.5) and red (100/10). Solid lines are the model predictions. The experimental data points are represented by the same colors and by the \times , ∇ and \diamond symbols, respectively. All simulations were calculated by using the rate constants obtained from the 100/10 experiment.

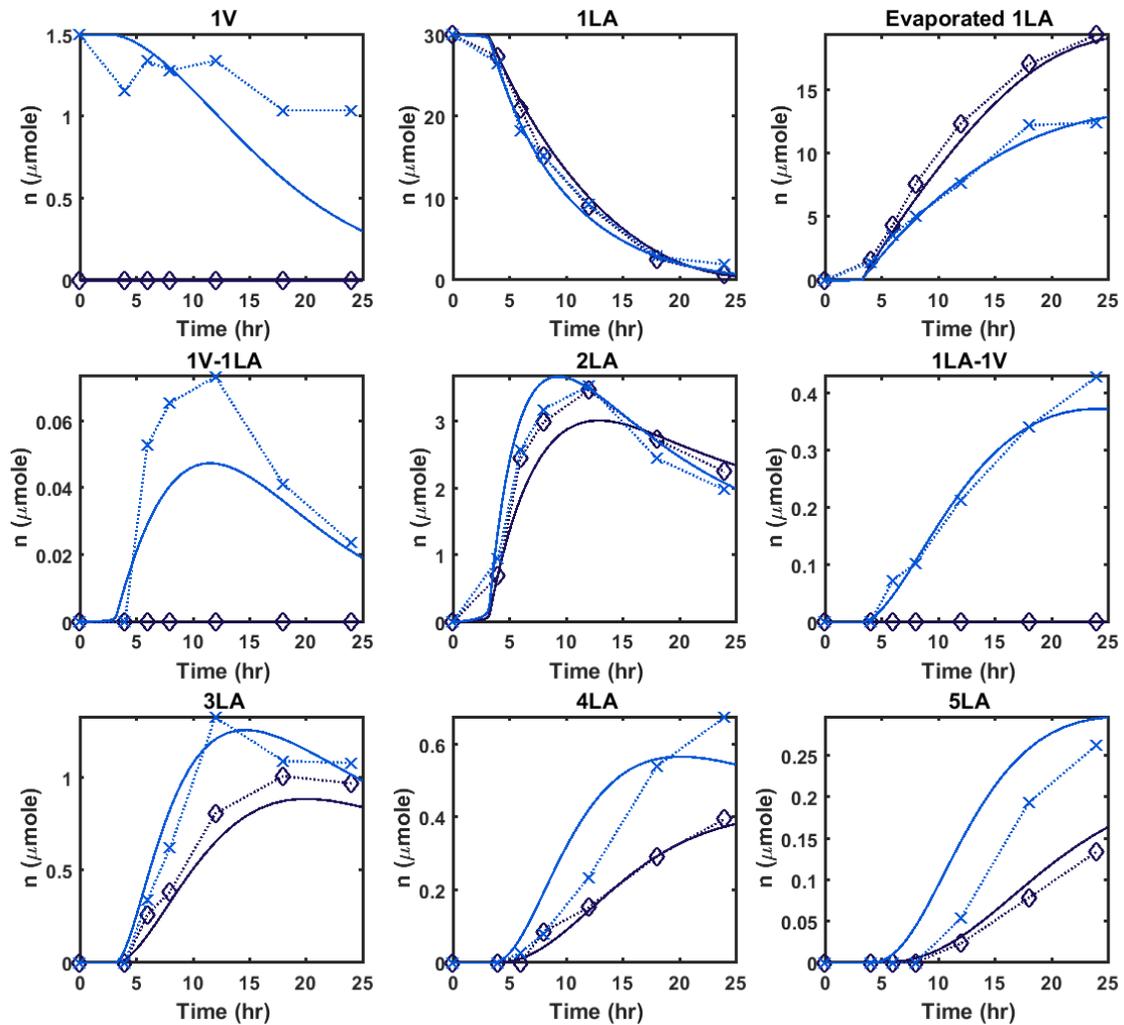


Figure 3.14 – Experimental results and model prediction for the 100/0 and 100/5 experiment. The results for different LA/V ratios are denoted as different colors: black (100/0) and blue (100/5). Solid lines are the model predictions. The experimental data points are represented by the same colors and by the \diamond and \times symbols, respectively. Optimized rate constants from the data was used for the model prediction of the 100/0 experiment. The simulation for 100/5 ratio experiment was calculated by using the rate constants obtained from the 100/5 experiment.

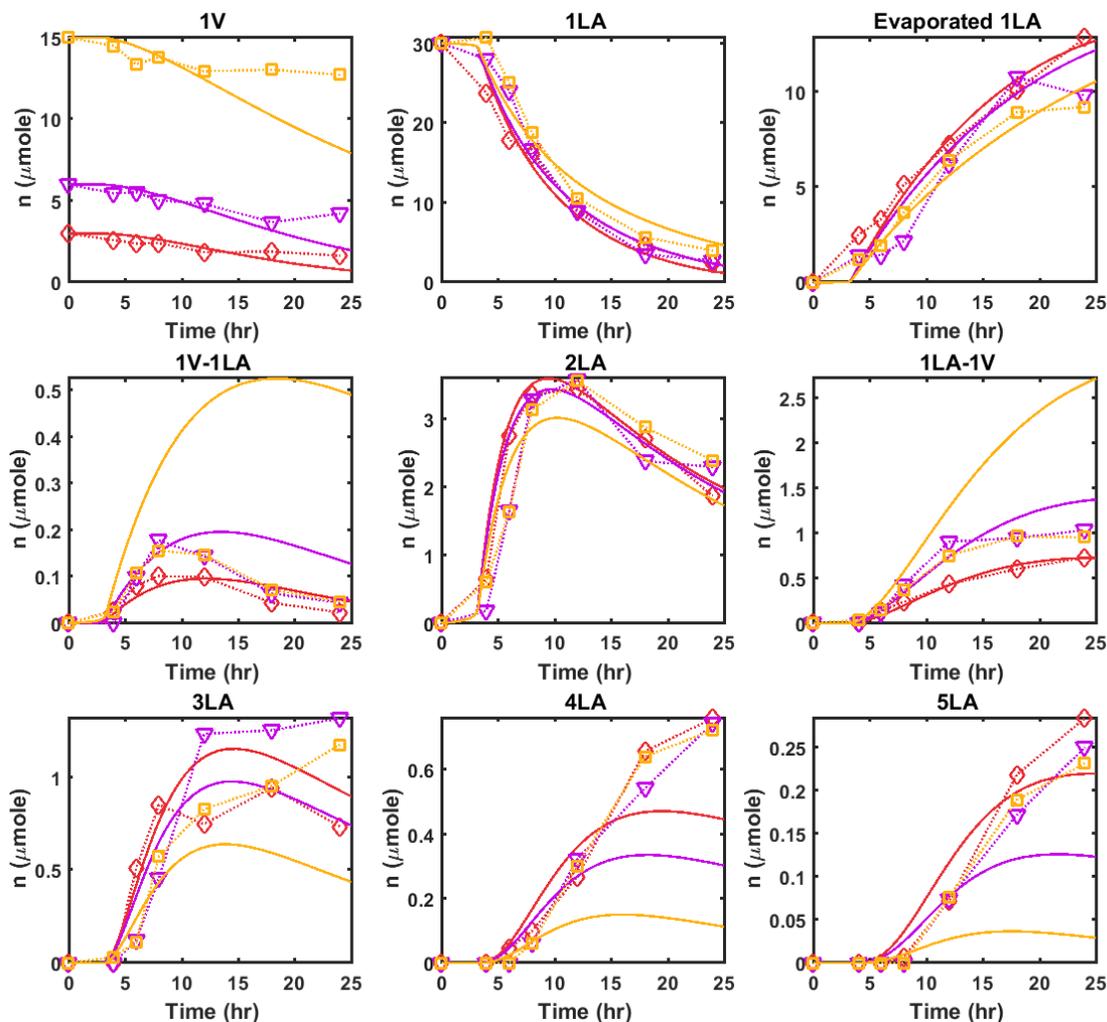


Figure 3.15 – Experimental results and model prediction for high valine loading. The results for different LA/V ratios are denoted as different colors: red (100/10), purple (100/20) and orange (100/50). Solid lines are the model predictions. The experimental data points are represented by the same colors and by the \diamond , ∇ and \square symbols, respectively. All simulations were calculated by using the rate constants obtained from the 100/10 experiment.

Our kinetic model for the copolymerization between valine and lactic acid helped us reveal the rates for each reaction and their temperature dependencies. In this section, we further discuss why some reactions might be more favorable than others and why the ester-mediated pathway is more advantageous than the direct peptide bond formation. From our analysis, we found the esterification rate constant between two hydroxy-terminated oligomers (k_{11}) is significantly larger than that between a hydroxy-terminated oligomer and an amine-terminated oligomer (k_{12}).

The difference between the two rate constants can possibly be explained by the acidities of their carboxylic acid groups. The acidity of the valine carboxylic acid ($\text{pK}_a=2.29$) is higher than that of lactic acid ($\text{pK}_a=3.86$). Therefore, it is much easier for lactic acid to access the protonated and reactive form for esterification. It is also possible to use the same theory to explain the deviation between the experimental data and the model prediction for the valine consumption. Due to the presence of the relatively stronger electron-withdrawing amide group when compared to the ester group, the pK_a of the carboxylic acid on LA-V should be lower than that of 2LA. Thus, the carboxylic acid group on the LA-V should be less reactive for additional esterification. Nevertheless, the esterification rate constants of these two dimers (2LA and LA-V) were set to equal to k_{11} for the sake of simplicity in our model. Because the esterification rate for LA-V might be slower, we anticipate that the consumption of valine monomer by the exchange reaction should be slower in the real reaction scenario.

3.4.3 Flux analysis

To help elucidate the major pathway in the copolymerization, a flux analysis was performed based on the simulation of the LA/V system at 85°C with the initial molar ratio of 100/10, as illustrated in Figure 3.16. Inspection of concentration profile alone does not indicate the dominant pathway, since some important pathways may be fast and thus give low concentration for some intermediates. One possible route is the esterification between valine and lactic acid (F1). This reaction eventually leads to pure peptide dimer (F2), but the flux of formation is low and the product can soon be consumed by the exchange reaction. From our simulation results, the dominant reaction pathway starts from the exchange reaction on those oligomers that contain ester bonds (F6). The LA-V dimer can then be further esterified by other oligomers (F7). Based on this analysis, the key species are the oligomers with lactic acids linked to their N-termini (actually, it is the O-termini in that case), which is consistent with our proposed mechanism in Chapter 2 (Figure 2.19 and Figure 2.20). The formation of pure peptides from this reaction is still possible, but the abundance might be too low to be detected.

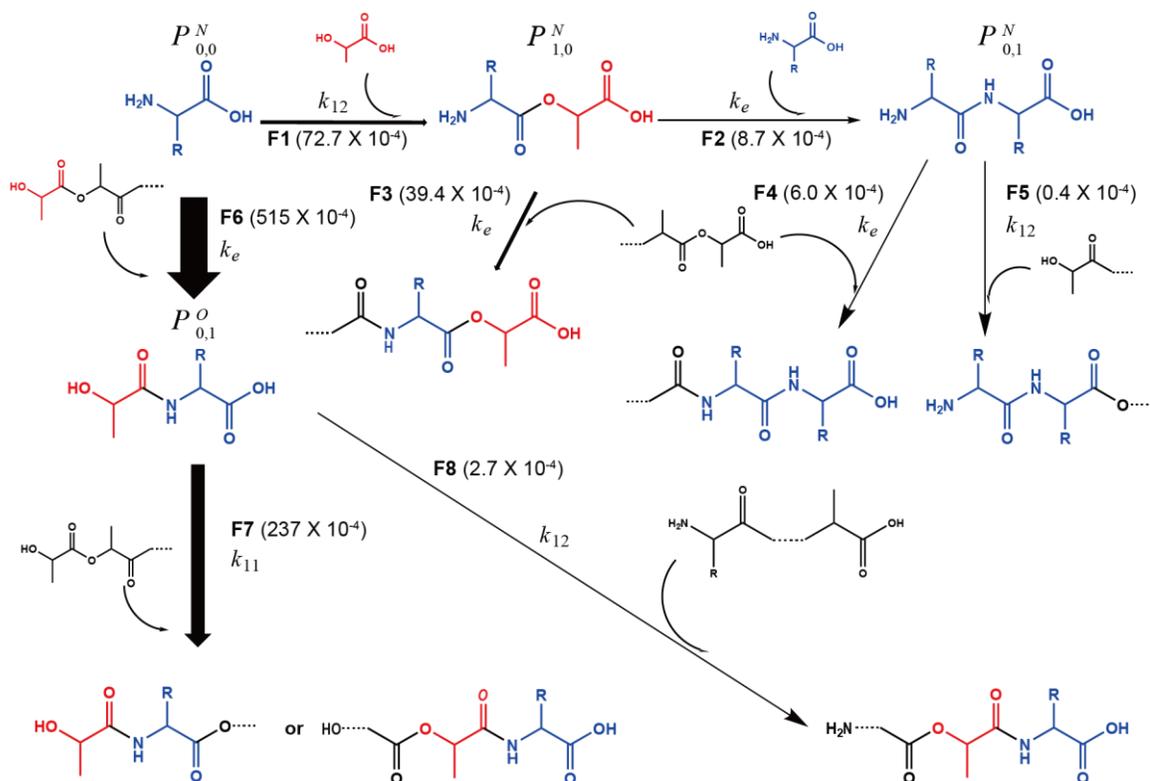


Figure 3.16 – Flux analysis of the studied reactions. The fluxes were calculated from the simulation of 24 hours reaction of 100/10 LA/V mixture at 85°C. The magnitudes of fluxes ($\mu\text{mole/hr}$) are the average rates over 24 hours. The width of each arrow indicates the relative magnitudes of each flux. Only the fluxes of forward reactions are included because the fluxes of hydrolysis and reverse reactions are negligible in the dry state.

The magnitude of each flux changes when using different ratios of valine and lactic acid. The difference between the two pathways become less significant when the amounts of valine and lactic acid are similar. However, based on our previous analysis of copolymerization experiments with different ratios, only a fraction of valine monomer can actually participate in the reaction when a high loading of valine was used. Under such conditions, the flux through F6 and F7 would still be the major pathway. We also investigated the initial fluxes of each species (Figure 3.17). The flux of lactic acid dimer formation is significantly larger than that of the subsequent ester-amide exchange reaction to form the LA-V dimer. However, the rate constant of the exchange reaction is actually

greater than that of lactic acid esterification. It is the concentration of valine monomer that limits the flux of the ester-amide exchange reactions. Therefore, we conclude that the major pathway is actually gated by the solubility of valine in the reaction mixture. It is possible some of the valine monomers separated into a different phase and had fewer chances to participate the reaction. If simpler amino acids like glycine are used, we anticipate that higher fluxes can be achieved.

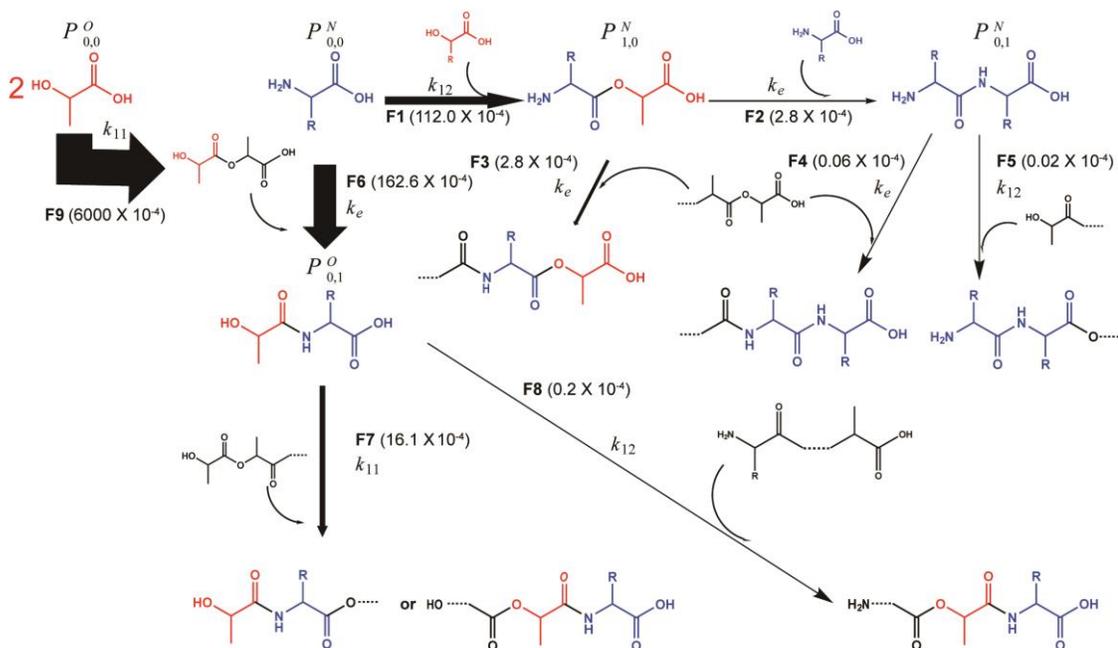


Figure 3.17 – Flux analysis for the reactions in 6 hours. The average fluxes ($\mu\text{mole/hr}$) were calculated from the simulation of 6 hr reaction of 100/10 LA/V mixture at 85°C . The width of each arrow indicates the relative magnitudes of each flux. Only fluxes of forward reactions are included because the fluxes of hydrolysis and reverse reactions are negligible in the dry state.

3.4.4 Activation parameters

We also calculated the activation parameters of the ester-amide exchange pathway and compared these with reported values of direct peptide bond formation, using Eqs. (3.20), (3.21) and (3.22). The result is shown in Figure 3.18. The reaction free energies of the exchange reaction were calculated based on the fitted results of the Arrhenius plot. The

activation energies of glycine dimerization under hydrothermal condition are used for the direct peptide bond formation between LA and V because the actual kinetic data is not available at this point [99]. The free energy of the final product was assumed to be the value of glycine dimer and was calculated from the data of Shock *et al.* [100].

The activation free energy is the overall energy barrier required for the reactants to overcome the transition state. It is the combination of activation enthalpy and activation entropy. Here, we first found that the free energy of esterification is slightly negative, which is consistent with previous reports [54]. Second, the activation free energies of the lactic acid esterification and the exchange reaction are roughly 3 kcal/mol lower than that of the direct amide bond formation. The difference in activation energies means the amide formation rate is approximately 100-fold faster if hydroxy acid participates in the reaction. This result confirms our initial hypothesis that the ester-amide exchange pathway is an easier route leading to the peptide bond formation.

Interestingly, the advantage of the ester-mediated pathway does not totally come from the decrease of activation enthalpies, but also from the decrease of activation entropies. This result is similar to what has been found in ribosome-catalyzed peptide synthesis [101]. It has been suggested that ribosomes create a water-poor environment to reduce the entropic penalty of zwitterionic intermediates [102]. In our experiment, the drying process removed water by evaporation and potentially provided a similar environment that avoids the entropic penalty of zwitterionic intermediates and thus, enhances the formation rate of the amide bond.

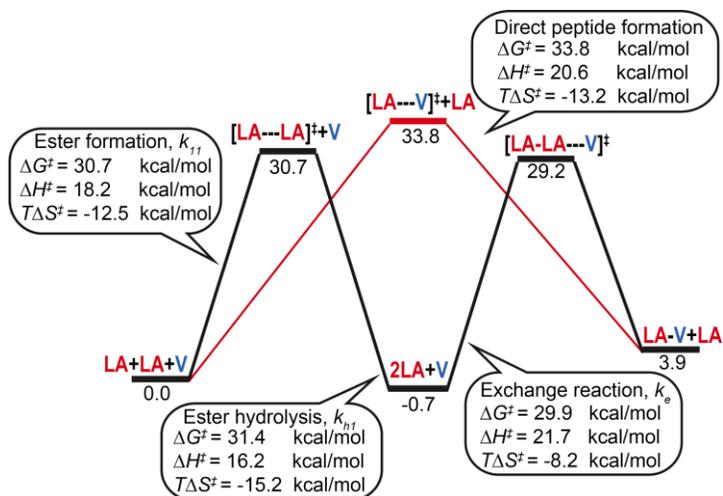


Figure 3.18 – Activation parameters of the ester-amide exchange pathway and the pathway for direct peptide bond formation. All parameter values were estimated for 85°C.

So far, our results suggest the copolymerization between valine and lactic acid can be explained by our model with a limited amount of information and a small number of parameters. Although the presence of other oligomers was observed through MS analysis, it is difficult to determine quantitative amounts of each specific compound via LC-MS when no commercial standards are available. Here, the simulation provides an alternative way to estimate the growth of polymer chains and their compositional evolution. In addition, most of the past simulation work for copolymerization was done by kinetic Monte Carlo simulation to include the sequence information [87, 88]. In the future, the rate constants estimated in this work can be coupled with a kinetic Monte Carlo method to provide details of the copolymer sequences.

3.5 Conclusion

In this work, we developed an analytical procedure and a deterministic model to investigate the formation of depsipeptides, a potential ancestor of polypeptides on the Earth. Our results demonstrate that the complicated copolymerization can be described by only having

the chain length and composition for the reaction conditions considered here. A future model will also include the possible effect of viscosity and the sequence information. Although the reaction rate is feasible enough to be observed on the laboratory time scale, a scenario where long chain depsipeptides with a high peptide bond content can form has not been found yet. With the above work, we aim to search for a scenario to generate a polymer system that can assemble and exhibit rudimentary properties leading to selection and evolution.

CHAPTER 4. GROWTH OF PROTO-PEPTIDES BY CONTINUOUS FEEDING OF MONOMERS

4.1 Introduction

In previous chapters, we have confirmed the robustness of the ester-mediated reaction, but there are still some challenges in synthesizing long oligomers. Although oligomers up to 14 units have been identified, their abundances were low and they contained only small amounts of amino acids. From the tandem MS experiment, we have found oligomers with peptide backbone gradually emerged as the cycling continued. Those species with continuous peptide backbones are the major products of interest. However, it is difficult to use tandem MS to identify and to quantify all peptide fragments.

Similar to a living polymerization, the ester-mediated reaction gradually adds each amino acid to the C-terminus of oligomers one by one without termination. Since the N-terminus of depsipeptides usually has a hydroxy group that allows further esterification, the polymerization is controlled by the amounts of unreacted amino acids. Therefore, the average chain length should be determined by the ratio of hydroxy acids and amino acids. Also, the livingness of the ester-mediated reaction should allow new monomers to elongate the existing oligomers. It is possible that new building blocks (amino acids) formed on other parts of early Earth mixed with existing depsipeptides by raining or tides.

The objective of this chapter is to identify key scenarios that enable a continuous growth of oligomers with repeated amino acid backbone. The copolymerization of glycolic acid (GA) and glycine (G) is chosen as a model reaction because of their higher abundance

in the prebiotic mixtures and their higher reactivity. The first part of this chapter provides details about how to analyze the GA/G oligomer mixture. The second part is a comparison between two feeding policies and their impacts on the oligomer distribution.

4.2 Experimental method

4.2.1 Materials

Glycolic acid, glycine, glycine anhydride, diglycine, triglycine, methyl glycolate, formic acid, 2,5-dihydroxybenzoic acid and phosphoric acid were obtained from Sigma-Aldrich. Methanol, HPLC grade water and acetonitrile were from EMD Millipore. 99% of 1,1,3,3-tetramethylguanidine came from Alfa-Aesar.

4.2.2 Day-night machine

The cycling experiment with feeding was conducted by a custom-built instrument, as shown in Figure 4.1. 100 μ L of the sample solution containing 100 mM of glycolic acid and 100 mM glycine was heated to 95⁰C by a Bio-Rad MyCycler 96 well thermal cycler. Every 24 hours, a pipette controlled by a Biomek 3000 Automation Workstation rehydrated the dry samples with 100 μ L of an aqueous solution containing small amounts of fresh monomers, typically 20 mM of glycolic acid and glycine, respectively.



Figure 4.1 – The day-night machine used for the feeding experiment.

4.2.3 *Oligomer characterization*

Before characterization, each sample was first dissolved in 200 μL of water. The solution was sealed in a vial and incubated at 95°C for 48 hours to remove all ester linkages. After the hydrolysis treatment, the aqueous solution was diluted with acetonitrile to make a 50 vol % acetonitrile/water solution. Samples were analyzed with an Agilent 1260 HPLC coupled with an Agilent 6130 single quadrupole mass spectrometer and a UV detector (210 nm). Oligomers were separated using a SeQuant ZIC-HILIC column (150 \times 2.1 mm, 3.5 μm particle size, 100 Å). The flow rate was 0.2 ml/min and the column temperature was held at 40°C. The mobile phase was water/acetonitrile with 0.5 vol % formic acid. The gradient method started with 5 vol % water and ramped to 50 vol % water in 25 minutes. The water concentration was held at 50 vol % for 5 minutes and returned back to 5 vol % immediately. For LC-MS analysis, all data were obtained in negative-ion mode

electrospray ionization with a capillary voltage of 2.0 kV. The log-ratio analysis was performed using MZmine 2 software and plotted by OriginPro 9.0.

For sequence analysis, tandem MS experiments were performed on a Waters Xevo G2 mass spectrometer, equipped with a quadrupole/time-of-flight (qTOF) mass analyzer. Samples were diluted to 0.039 mg mL⁻¹ of starting monomers and directly infused into the mass spectrometer by electrospray ionization in negative-ion mode. The capillary was set at 2.0 kV, the sampling cone voltage was 30, the extraction cone voltage was 3.0, and the source temperature was 90°C. The desolvation gas flow was 600 L hr⁻¹ and the desolvation temperature was 250°C.

Matrix-assisted laser desorption/ionization (MALDI) experiment was performed using Bruker AutoFlex III mass spectrometer. 2,5-dihydroxybenzoic acid matrix and the dry oligomer mixture were dissolved in a 50 vol % acetonitrile/water solution with 0.1 vol % formic acid. 1 µL of the solution was deposited on the MALDI plate and allow to dry for the MS analysis.

4.2.4 *Synthesis of standard compounds*

For the synthesis of GA-G dimer standard, 2 mmoles of glycine and 2 mmoles of methyl glycolate were mixed in 500 µL of 1,1,3,3-tetramethylguanidine and 100 µL of methanol for methanolysis. The solution was stirred at 100°C for 8 hours. The crude products were first purified by an anion exchange column packed with 25 g of QAE Sephadex[™] A-25 resin (GE Healthcare) with a pH 5 ammonium formate buffer running a gradient from 50 mM to 500 mM. Products were then passed through a cation exchange column packed with 15 g of Dowex 50WX8 hydrogen form (Sigma-Aldrich) charged with 0.1 vol % formic

acid solution. Final isolation of product was achieved by flowing products through a Teledyne Combiflash-Rf+ flash chromatography system using a RediSep C18aq 150 g Gold column and a UV-Vis detector. Identification of desired product was verified by an Agilent 6130 single quadrupole mass spectrometer with a UV detector using a 2.0 kV capillary voltage. The expected $[M-H]^-$ value is 132.0295 Da for the 1GA-1G dimer. Species with m/z values equal to 132.1 Da and 265.0 Da (common $[2M-H]^-$ ESI artifact) were observed. In a similar procedure, diglycine was combined with methyl glycolate in the initial reaction for the synthesis of the GA-G-G trimer. For the trimer, the expected $[M-H]^-$ is 189.0509 Da. Signals of 189.0 Da and 379.0 Da (common $[2M-H]^-$ ESI artifact) were detected.

4.3 Results & discussion

4.3.1 *Quantitative analysis of the GA/G depsipeptides*

To study the effect of feeding on the distribution of oligomers, we used a custom-built day-night machine to feed the samples with small amounts of fresh glycolic acid and glycine every 24 hours at 95°C. This drying temperature was chosen to enhance the reaction rate and to reduce the experimental time.

To reduce the complexity of the oligomer mixtures and reveal the abundance of those oligomers with repeated amino acid backbone, we utilized the higher stability of amide bonds compared to ester bonds. The procedure is conceptualized in Figure 4.2. Each sample was hydrolyzed at 95°C for 48 hours to remove all ester linkages. This procedure allows us to reduce the peak overlapping in chromatograms and to focus on those oligomers containing stable amide bonds.

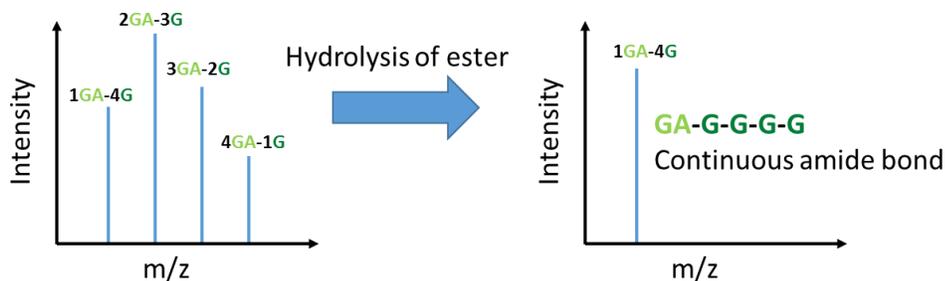


Figure 4.2 – Using hydrolysis to simplify the oligomer mixture.

In Chapter 3, we used a reverse-phase column to separate the LA/V oligomer mixtures. However, the same approach was not successful here due to the low hydrophobicity of the GA/G depsipeptides. Hao *et al.* have used the hydrophilic interaction chromatography (HILIC) to separate similar mixtures containing glycine, diglycine and triglycine from the Maillard reaction [103]. Based on their procedure, we successfully used an HILIC column to separate the GA/G depsipeptides. The quality of separation is evaluated by our LC-MS analysis. Figure 4.3(a) shows the total ion chromatogram of the sample without the hydrolysis treatment. The HILIC column could separate some species in the oligomer mixture, but further examination of the mass spectrum of each peak revealed many oligomers co-eluted together (Figure 4.3(b), (c), (d), and (e)). For example, the 1GA-2G trimer was found to eluted at 10 minute, but 3GA-3G and 4GA-4G also eluted at the same time.

The peak overlapping was reduced by hydrolyzing the sample at 95°C for 48 hours. The total ion chromatogram shows more resolved peaks, compared to the sample without the hydrolysis treatment (Figure 4.4(a)). As shown in Figure 4.4(b), (c), (d) and (e), the mass spectrum for each peak contains only one dominant species. Signals for oligomers containing multiple glycolic acid disappeared after the 48 hours hydrolysis reaction.

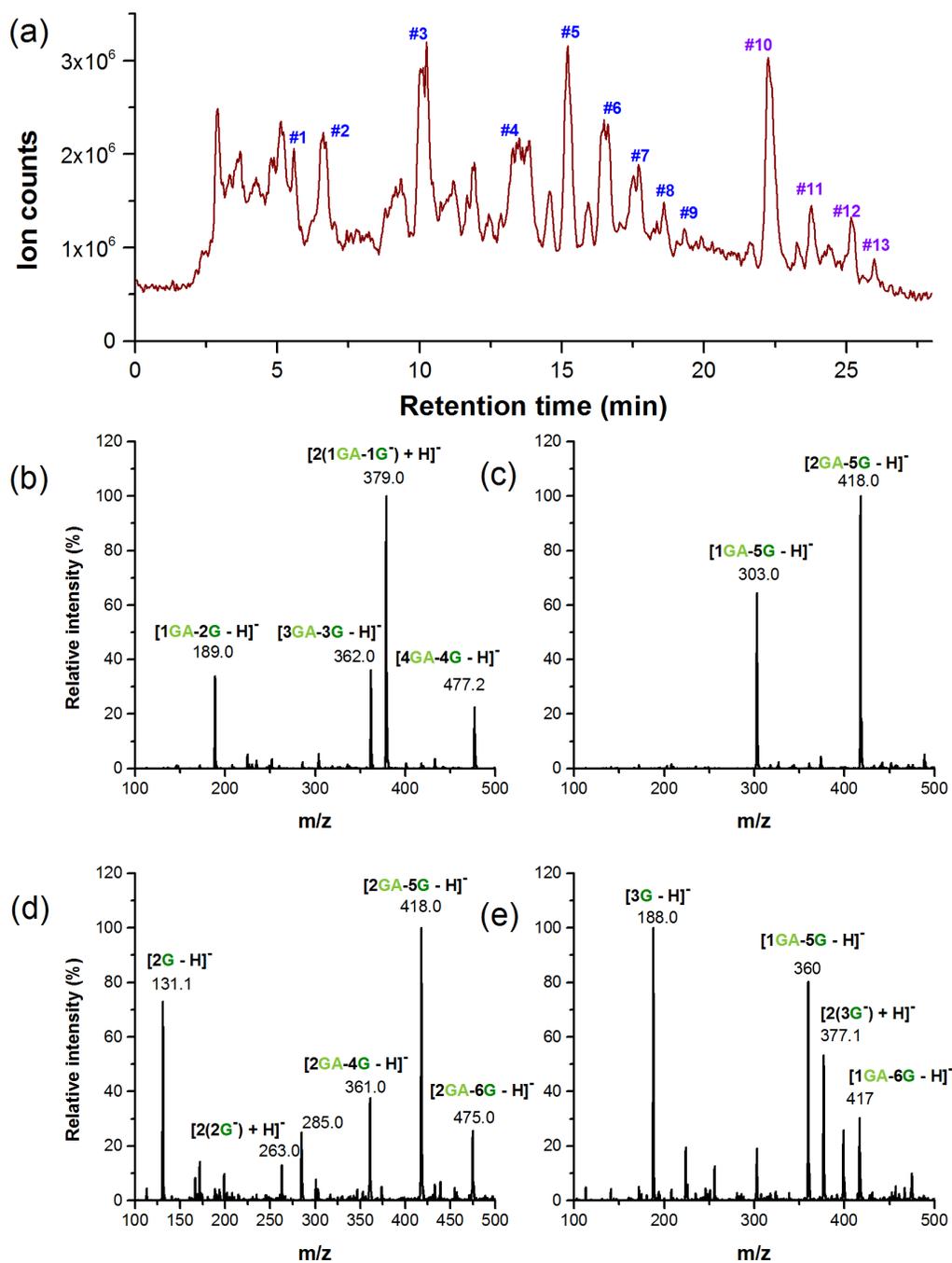


Figure 4.3 – HPLC-MS analysis of the GA/G oligomer mixture before the hydrolysis treatment. (a) The full ion chromatogram. (b) Mass spectrum of Peak 3. (c) Mass spectrum of Peak 5. (d) Mass spectrum of Peak 11. (e) Mass spectrum of Peak 12. The sample was prepared by drying a mixture of glycolic acid (100 mM) and glycine (100 mM) at 95°C for 16 days. 20 mM solution of glycolic acid and glycine was added to the dry mixture every 24 hours.

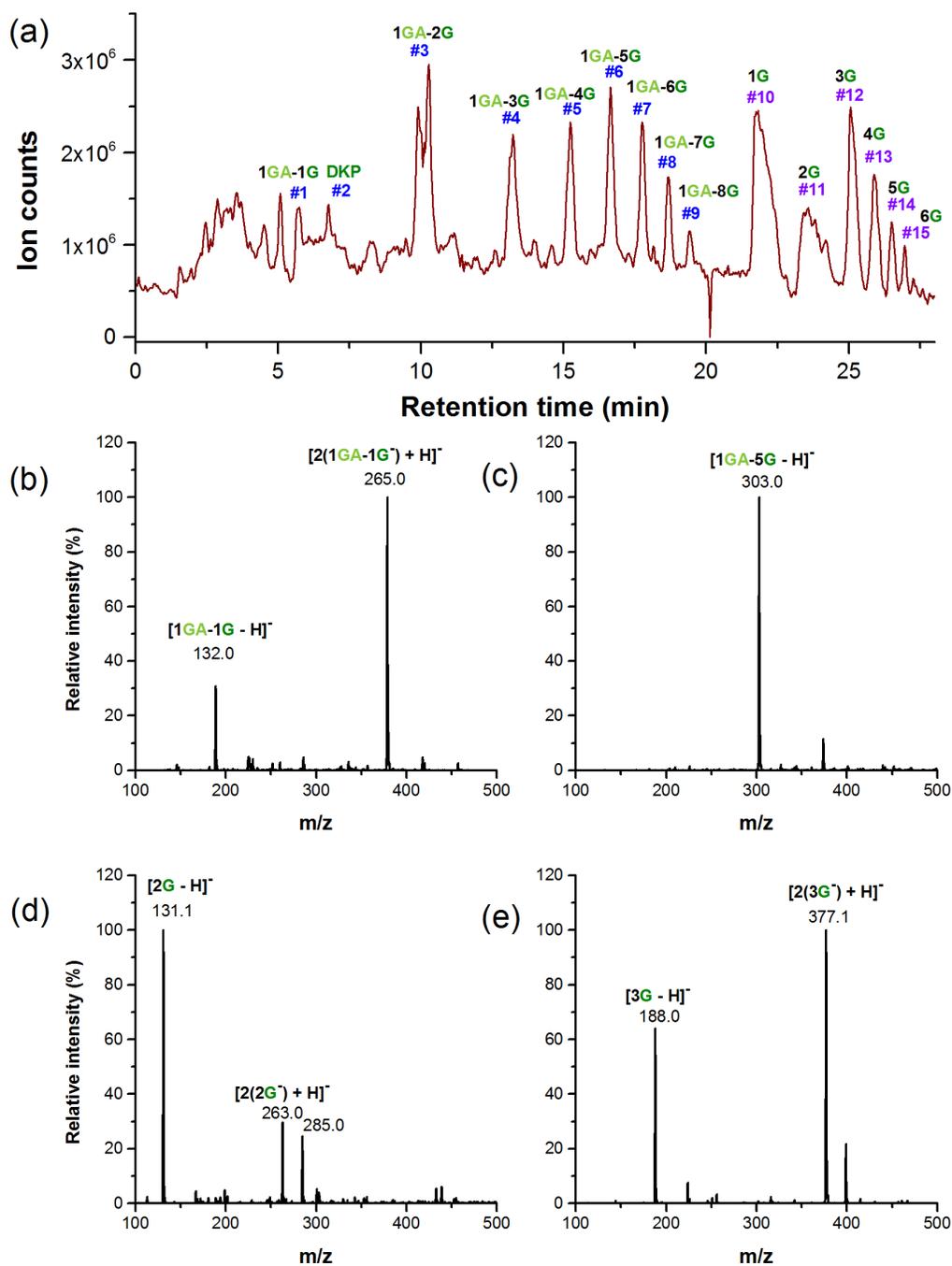


Figure 4.4 – HPLC-MS analysis of the GA/G oligomer mixture after the hydrolysis treatment. (a) The full ion chromatogram. (b) Mass spectrum of Peak 3. (c) Mass spectrum of Peak 5. (d) Mass spectrum of Peak 11. (e) Mass spectrum of Peak 12. The sample was prepared by drying a mixture of glycolic acid (100 mM) and glycine (100 mM) at 95°C for 16 days. 20 mM solution of glycolic acid and glycine was added to the dry mixture every 24 hours. The sample was hydrolyzed at 95°C for 48 hours prior to the LC-MS experiment.

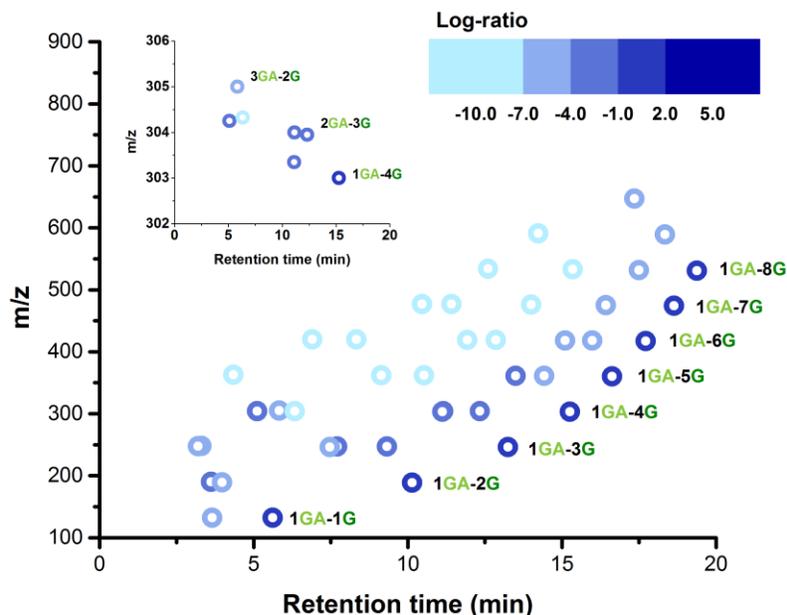


Figure 4.5 – Log-ratio analysis of the LC-MS data before and after the hydrolysis treatment. The sample was prepared by drying a mixture of glycolic acid (100 mM) and glycine (100 mM) at 95°C for 16 days. 20 mM solution of glycolic acid and glycine was added to the dry mixture every 24 hours.

The effect of the hydrolysis treatment is further illustrated by a two-dimensional log-ratio plot (Figure 4.5). The color strength of each species indicates the ratio of peak areas in the MS chromatograms after and before hydrolysis as defined in Eq. (4.1).

$$\text{Log-ratio} = \ln\left(\frac{\text{Peak area after hydrolysis}}{\text{Peak area before hydrolysis}}\right) / \ln(2) \quad (4.1)$$

Although the peak areas in the MS chromatograms usually are not reliable without standards, this approach provides a semi-quantitative way to compare two LC-MS experiments. Before the hydrolysis treatment, species with similar masses were found to elute at different times. For example, the pentamer series (Figure 4.5, inset) ranges from 3GA-2G to 1GA-4G. For the same chain length, oligomers containing more glycolic acid

units tend to elute earlier. This is consistent with the working principle of the HILIC column. Since glycolic acid esters should be less hydrophilic compared to glycine amide bonds, the affinity of glycolic acid ester to the column is weaker. Those oligomers containing multiple glycolic acid residues co-elute with other oligomers as well. After the hydrolysis treatment, the peak areas of those glycolic acid-enriched oligomers decrease and those of the 1GA-nG depsipeptides series increase. Since the log ratio values of those species containing multiple glycolic acid units are mostly below -4 (less than 6% of their initial values), we conclude the hydrolysis treatment has successfully removed nearly all ester bonds. The remaining 1GA-nG series should have only one dominant series: 1GA-(1G)_n. All units are linked by amide bonds with the glycolic acid residues locate at the N-termini.

In addition to those depsipeptides with N-capped glycolic acid, we also found pure glycine peptides (nG) without glycolic acid units (Peak 11 to Peak 15 in Figure 4.3 and Figure 4.4). Those pure peptides can be the result of both glycine polycondensation and hydrolysis of N-terminal amide bond in depsipeptides. Because we also detected those oligomers in the chromatogram before the hydrolysis treatment, it is likely that some of them were produced from glycine polycondensation. However, we are unable to estimate the amounts of nG formed from polycondensation directly because of the intense peak overlapping. The possible cause of glycine polycondensation is discussed in later part of this work.

We attempted to identify the presence of DKP in the reaction mixtures. Due to its low ionization efficiency, no distinct signal could be detected in the negative-ion mode chromatogram. Figure 4.6(a) and (b) show the UV and positive-ion mode chromatograms

of the DKP. Although no clear signal was observed in the MS chromatogram, the position of eluted DKP was confirmed by injecting DKP standard compound and comparing the positions of their UV signals. As shown in Figure 4.6(c) and (d), the MS of signals of DKP from the reaction mixture and from the standard compound can be observed in positive-ion mode, but it is generally mixed with MS background signals.

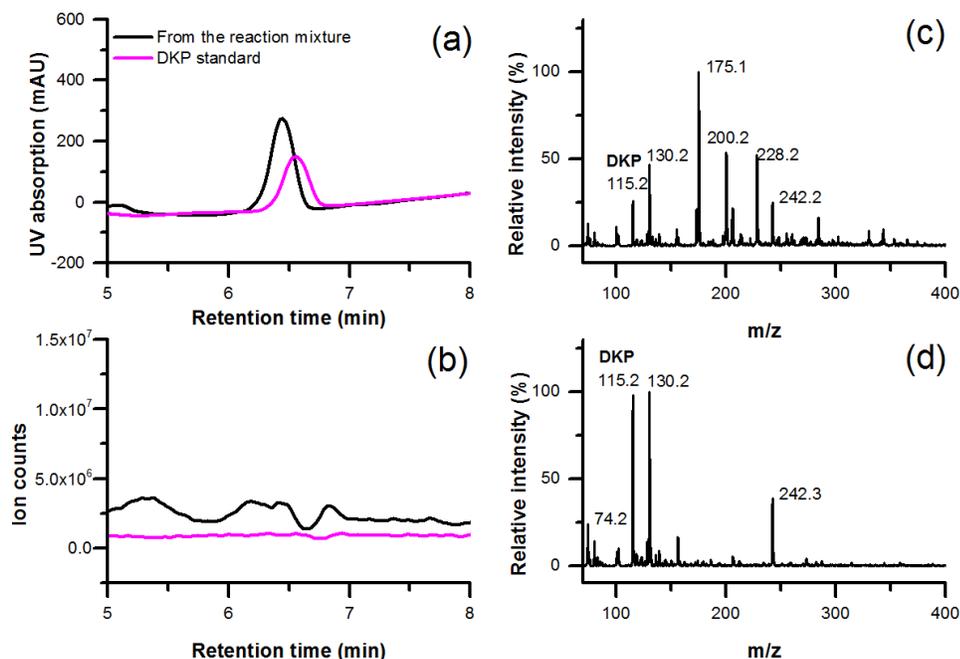


Figure 4.6 – Confirmation of glycine DKP. (a) UV traces. (b) MS positive ion trace. (c) MS signal of DKP from reaction mixture. (d) MS signal of DKP from standard.

The UV extinction coefficient of each oligomer was determined by injecting standard compounds with known concentration. While glycine monomer, DKP and short peptides are commercially available, depsipeptides are not. We synthesized the 1GA-1G and 1GA-G-G oligomers from the reaction of methyl glycolate with glycine and diglycine, respectively. The UV responses of 1GA-1G, 1GA-G-G and DKP are shown in Figure 4.7. Although the peaks of these three species do not have ideal Gaussian shapes, their UV

responses are linear with R^2 values higher than 0.99. Therefore, their UV response factors were used directly for quantitation. The standards of 1G, 2G and 3G with different concentrations were used to obtain the UV response factors of pure peptides (Figure 4.8).

To estimate the extinction coefficients of longer oligomers, the extinction coefficients of amide bonds and carboxylic acid groups were calculated from the results of short oligomers. Assuming the extinction coefficients of the chromophores in the longer oligomers are as same as those in the short oligomers, we calculated the UV extinction coefficients of 1GA-nG and nG oligomer series by counting the numbers of amide and carboxylic acid groups. The full list of UV response factors (1/extinction coefficient) used in this work is shown in Table 4.1.

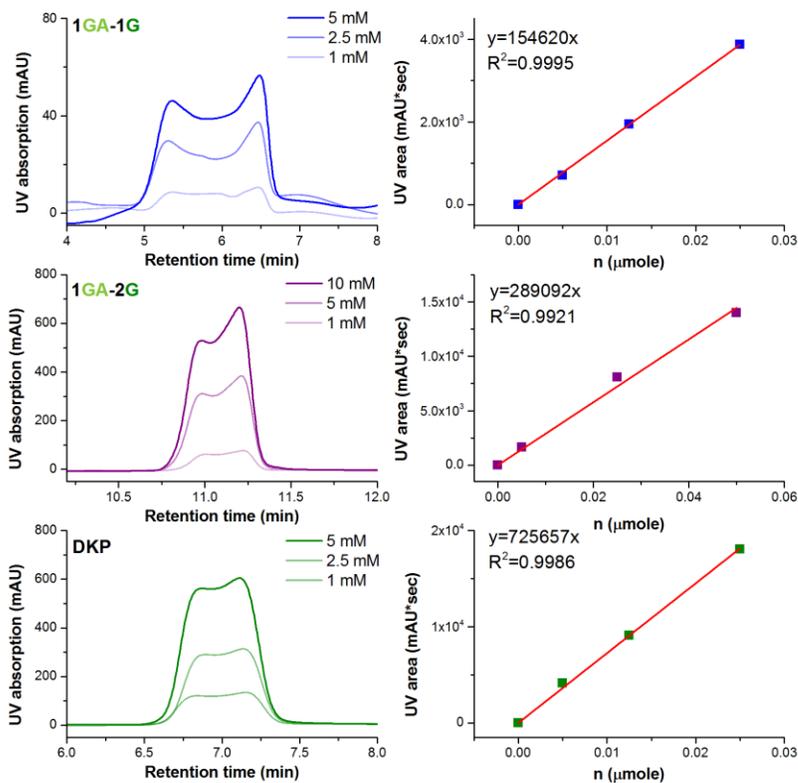


Figure 4.7 – UV responses curves of 1GA-1G, 1GA-2G and DKP.

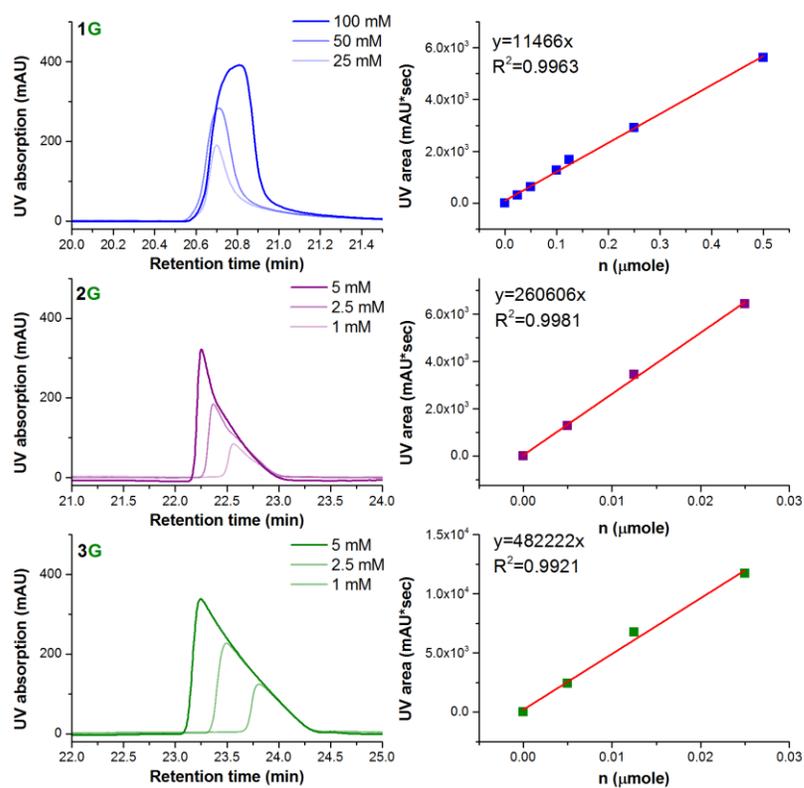


Figure 4.8 – UV responses curves of 1G, 2G and 3G.

Table 4.1 – UV response factors at 210 nm of GA/G oligomers.

Compound	UV response factor, k [$\mu\text{mole}/\text{mAU}\cdot\text{sec}$]	Compound	UV response factor, k [$\mu\text{mole}/\text{mAU}\cdot\text{sec}$]
1GA-1G	6.47×10^{-6}	1G	7.98×10^{-5}
1GA-2G	3.47×10^{-6}	2G	3.84×10^{-6}
1GA-3G	2.36×10^{-6}	3G	2.07×10^{-6}
1GA-4G	1.79×10^{-6}	4G	1.42×10^{-6}
1GA-5G	1.44×10^{-6}	5G	1.08×10^{-6}
1GA-6G	1.21×10^{-6}	6G	8.72×10^{-7}
1GA-7G	1.04×10^{-6}	7G	7.31×10^{-7}
1GA-8G	9.12×10^{-7}	8G	6.29×10^{-7}
DKP	5.52×10^{-7}		

We further used tandem MS to analyze the sequences of oligomers in the hydrolyzed samples as shown in Figure 4.9. All 1GA-nG depsipeptides indeed have only one dominant sequence: one glycolic acid unit is in the N-terminus, followed by multiple glycine units. The longest depsipeptide available for tandem MS experiment is 1GA-9G. The intensities of even longer depsipeptides are too low for tandem MS analysis.

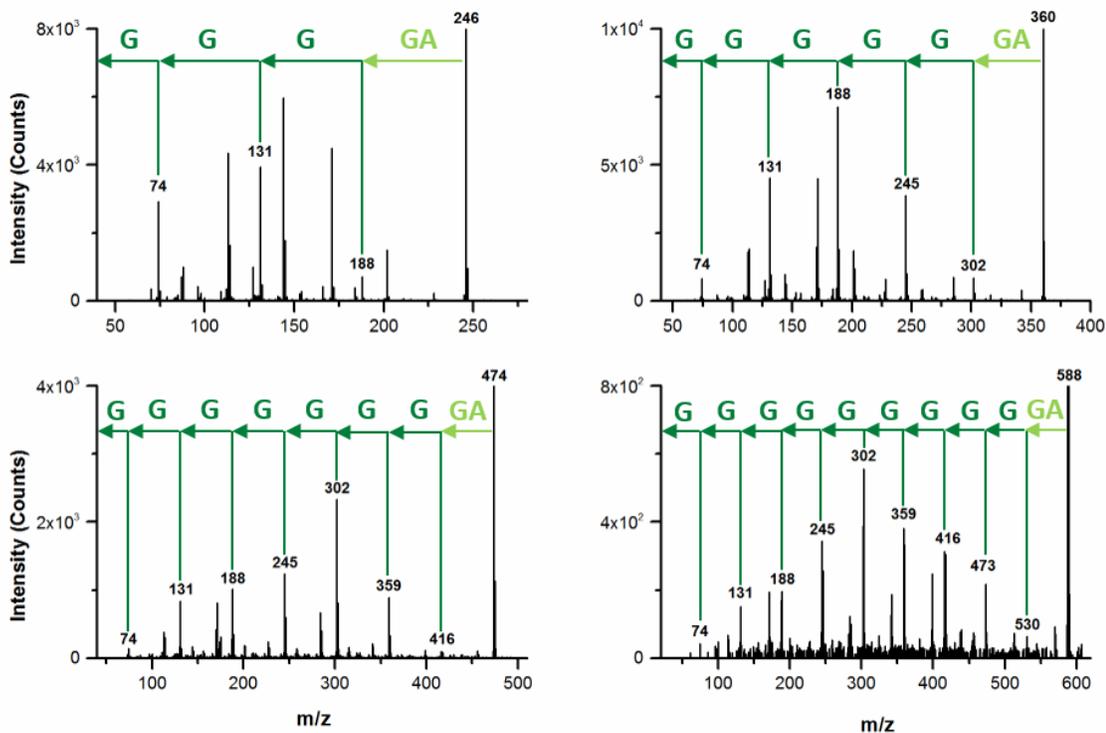


Figure 4.9 – Tandem MS sequencing analysis of the hydrolyzed GA/G mixture. Sample was prepared by drying a mixture of glycolic acid (100 mM) and glycine (100 mM) at 95°C for 16 days. 20 mM solution of glycine was added to the dry mixture every 24 hours.

4.3.2 Oligomer distribution under different feeding policies

In this section, we will discuss the oligomer distribution under two different feeding compositions: (1) Feeding with equal amounts of glycolic acid and glycine. (2) Feeding

with glycine only. In both experiments, the reaction started with 1:1 mixture of glycolic acid and glycine. We first evaluated the progress of the reaction by the following equations:

$$\text{Instantaneous yield} = \left(\frac{\sum_{i=1}^{\infty} i \cdot n_i(t)}{n_{1G}(t_0) + F_{1G} \cdot t} \right) \times 100\% \quad (4.2)$$

$$\text{Instantaneous conversion} = \left(1 - \frac{n_{1G}(t)}{n_{1G}(t_0) + F_{1G} \cdot t} \right) \times 100\% \quad (4.3)$$

where $n_i(t)$ is the amount of oligomer (1GA-nG or nG) containing i glycine units at time point t . The summation of $i \cdot n_i(t)$ is the total amount of glycine units in the oligomers. F_{1G} is the feeding rate of fresh glycine monomer and n_{1G} is the amount of unreacted glycine monomer at given time point. The denominators in both equations give the total amount of glycine units in the mixture. The results of Eq. (4.2) and Eq. (4.3) are the instantaneous yield and conversion, respectively. There are generally used for semi-batch reactors, which are similar to our day-night machine with continuous feeding of materials.

Figure 4.10(a) and (b) show the instantaneous yields of three different products: 1GA-nG, nG and DKP from the two feeding experiments. In both cases, the major product is the 1GA-nG depsipeptide. Its yield is typically three to four times higher than the nG pure peptide series. During the hydrolysis treatment, certain amounts of peptide bond could be hydrolyzed along with ester bonds. The hydrolysis of depsipeptides produces both shorter depsipeptides and pure peptides, but the hydrolysis of nG peptides does not produce depsipeptides. Therefore, the difference between the yield of 1GA-nG and that of nG might be higher before the hydrolysis treatment. This result confirms that even under the high

temperature of 95°C, when both ester-amide exchange and the direct amide bond formation can proceed, the ester-mediated pathway is the favorable reaction.

In both cases, small amounts of glycine DKP formed. DKP could be produced via the cyclization of diglycine, or the hydrolysis of longer peptides [104]. Since pure glycine peptides are observed, it is not surprising to detect DKP. Note that the ester-mediated reaction can also produce cyclized 1GA-1G dimer, but it might be converted back to linear dimer through the hydrolysis of ester bonds.

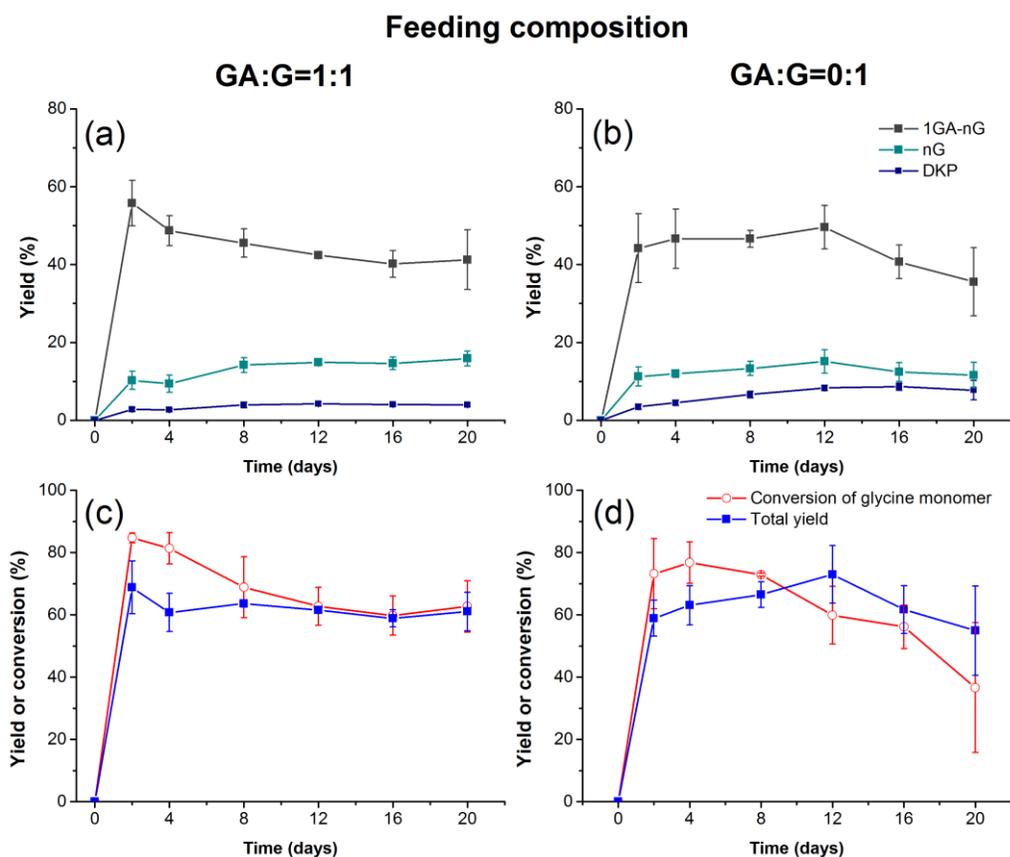


Figure 4.10 – Instantaneous yield and conversion in different feeding scenarios. (a) and (b): yields of 1GA-nG, nG (cyan) and DKP (dark blue) over time. (c) and (d): Instantaneous conversion of unreacted glycine monomer (red) and the total yield of products (blue), the sum of 1GA-nG, nG and DKP.

Figure 4.10(c) and (d) are the conversion of glycine monomers and the total yield of products, the sum of 1GA-nG, nG and DKP. In both experiments, the yield and the conversion follow each other closely, indicating our quantitation method can account for most of the products formed in the reaction. For both experiments, the instantaneous yield and conversion generally remain stable over time. This suggests the ester-mediated reaction is efficient in the time period of our study. Additional monomers can be converted into oligomers without substantial obstacles. However, in the case of glycine only feeding (Figure 4.10(d)), it appears that the reaction may be slowing down toward the end, possibly due to a reduction in the availability of hydroxy groups to promote ester amide exchange. No downturn can be seen in Figure 4.10(c) when glycolic acid is also added.

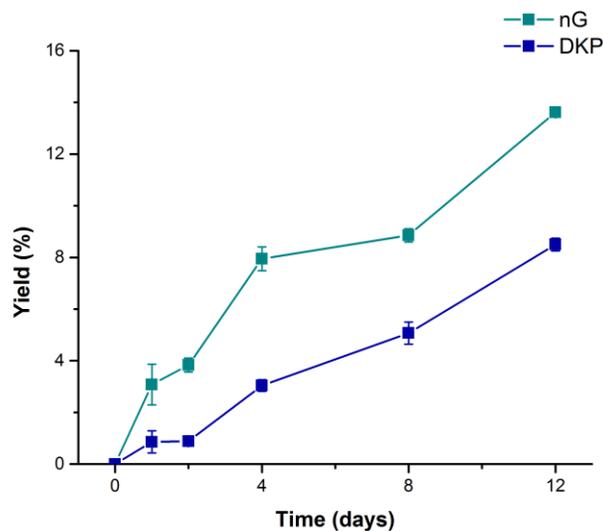


Figure 4.11 – Yields of glycine peptide (nG) and DKP in a cycling experiment without glycolic acid. The initial solution was acidified by phosphoric acid.

For the formation of pure peptides and DKP, Rodriguez-Garcia *et al.* have reported the condensation of glycine in an acidic condition. Due to the presence of glycolic acid, the pH of the initial solution is about 2.8. It is possible that the acidic environment caused by glycolic acid also enables the polycondensation of glycine. In a control experiment, the pH of the solution containing only glycine was lowered to 2.8 by phosphoric acid, and then was subjected to the same 95°C drying cycles. Glycine peptides and DKP indeed formed after a few cycles (Figure 4.11). The yields of nG peptides and DKP are similar to the those in the feeding experiment. We conclude that hydroxy acids enable the amide bond formation not only via the ester-amide exchange reaction, but also via the direct polycondensation by creating an acidic environment.

We further evaluate the effect of feeding compositions by comparing the distribution of 1GA-nG depsipeptides from the two feeding policies, after 2 days of hydrolysis at 95°C (Figure 4.12). When both glycolic acid and glycine are used, the distribution of oligomers remains unchanged over time. Although the abundance of longer oligomers gradually increases, the dominant species appears to be the 1GA-2G trimer containing only two amide bonds. In the second case, when only glycine is fed, the initial distribution is similar to the first case, but it shifts to higher molecular weight as time passes.

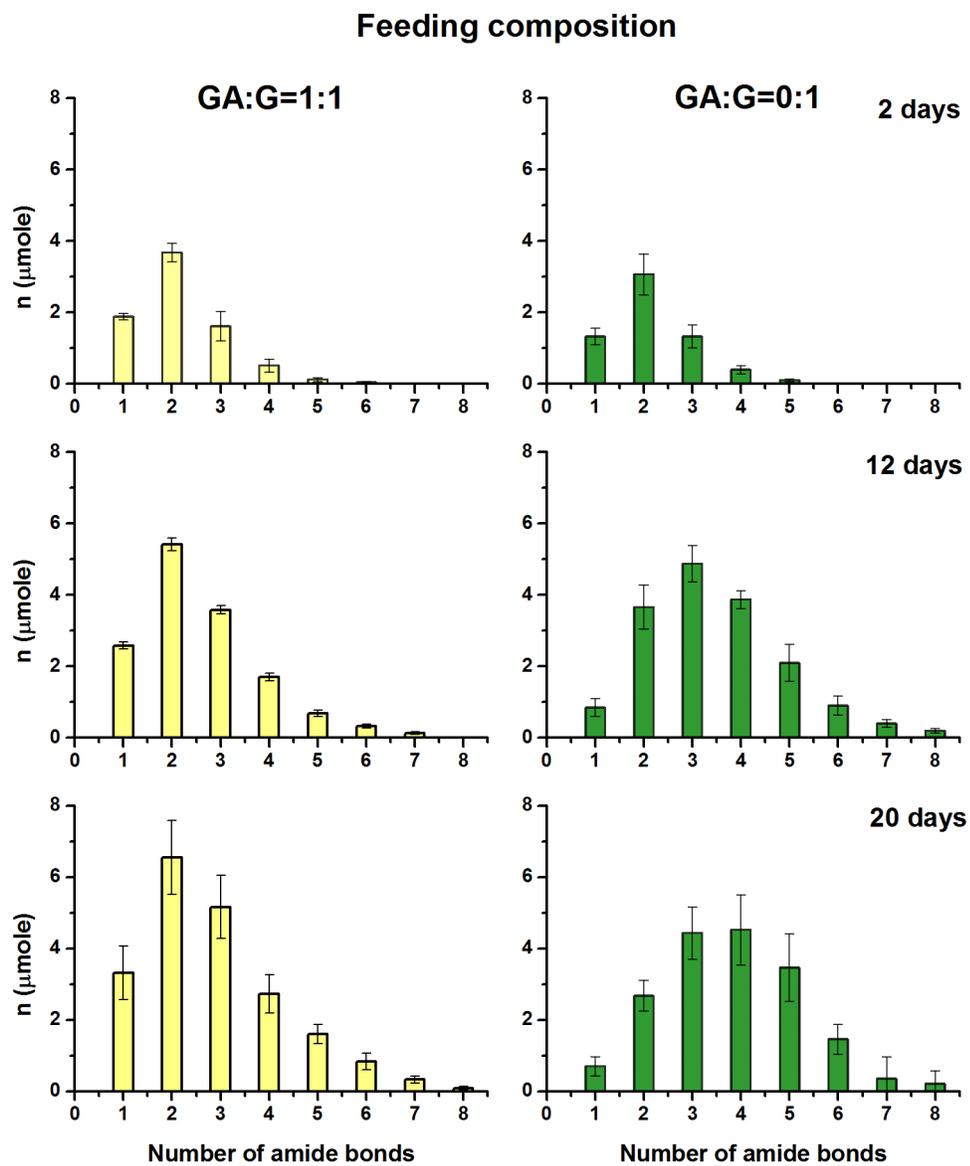


Figure 4.12 – Chain length distribution of 1GA-nG under different feeding policies. The y-axis is the number of glycine units in each oligomer.

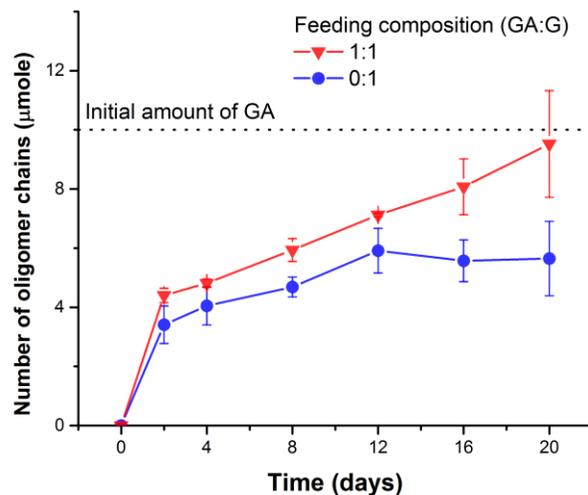


Figure 4.13 – The number of oligomer chains in 1GA-nG.

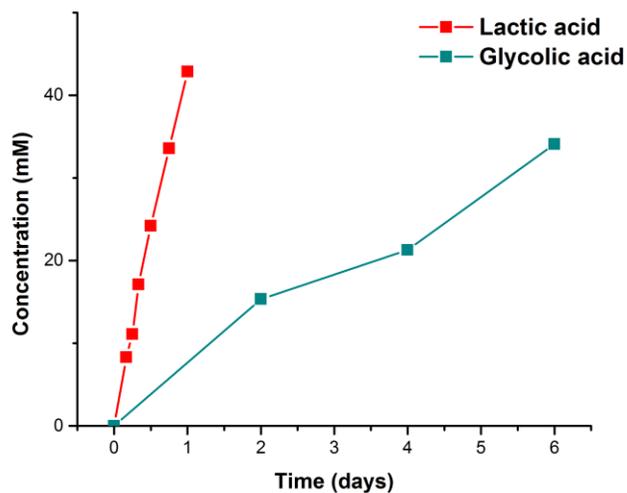


Figure 4.14 – Evaporation rate of lactic acid and glycolic acid at 85°C. The evaporation rates were determined by using the closed reactor as described in Chapter 3. The concentrations of lactic acid and glycolic acid were measured by ^1H NMR.

Figure 4.13 shows the total number of 1GA-nG chains over time. When no additional glycolic acid is used for feeding, the final chain number is only 56% of the initial amount of glycolic acid. This might be the result of glycolic acid evaporation, similar to what we observed when using lactic acid. The evaporation rate of glycolic acid was monitored in the closed reactor with the procedure described in Chapter 3. At the same drying temperature (85°C), the evaporation of glycolic acid appears to be much slower than lactic acid (Figure 4.14). This might be one of the reasons why glycolic acid is more efficient than lactic acid. Compared to lactic acid, more glycolic acid can be retained in the mixture and continue to promote the reaction.

When feeding with glycolic acid, the number of chains indeed increases over time, suggesting more oligomer chains are initiated by additional glycolic acid. The increasing number of active chains may consume free glycine more rapidly, but not necessarily contribute to elongation of existing chains. Most of the free amino acids are then consumed by making new short oligomers, instead of extending long oligomers. When no additional glycolic acid is used, the total number of chains remains constant and monomers are added to one end of chains step by step. The behavior of the ester-mediated pathway is similar to the chain-growth living polymerization [105], where no termination step is present to deactivated oligomer chains. The average molecular weight is determined by the ratio between monomers and initiators, and the active chains have the ability to connect new monomers sequentially.

So far, the longest oligomer found by the UV detector is 1GA-8G. From our MS analysis, we were able to find even longer oligomers such as 1GA-13G (Figure 4.15), but their UV responses were too low to obtain accurate measurements. However, oligoglycines longer than five residues are insoluble in water by forming polyglycine I structure, similar to antiparallel beta-sheets, and polyglycine II, an extended helix [106-108]. Therefore, the abundance of long oligomers we measured by HPLC-UV analysis might be less than their actual values. We also attempted to identify longer oligomers by using MALDI-TOF-MS experiment. As shown in Figure 4.16, oligomers up to 17-mer were detected in the sample without the hydrolysis treatment, but some of them containing two glycolic acid units. For the hydrolysed sample, we anticipate the results obtained from MALDI experiment will be similar to that from the ESI experiment (Figure 4.15). Note that the quality of MS signals in the MALDI experiment depends on whether samples can be mixed with matrix homogeneously. Yalcin *et al.* studied the effect of solvent composition on the MALDI-MS signals of polystyrene [109]. They found MS signals of long polymer might be lost if the sample is prepared by a polymer nonosolvent. Due to the low solubility of our products in common solvents (water, acetonitrile and methanol), it is possible that MALDI cannot capture the signals of some longer oligomers.

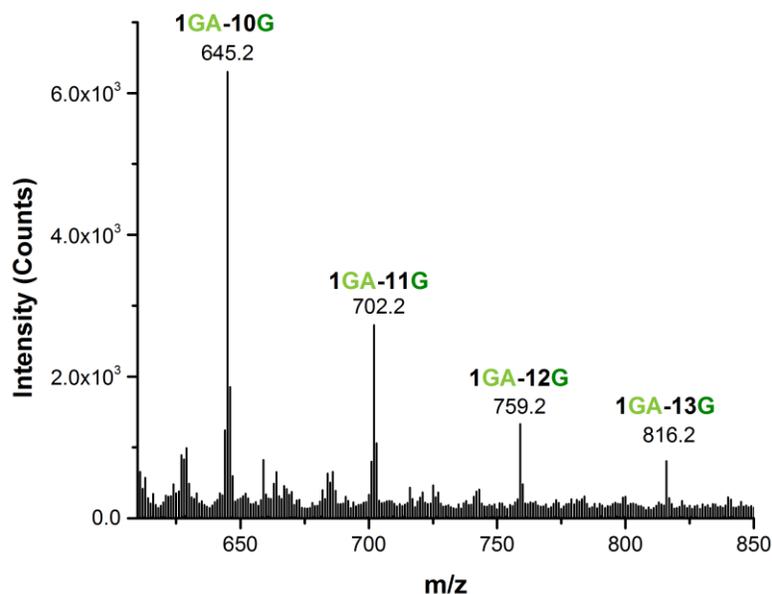


Figure 4.15 – The largest oligomers observed. The sample was prepared by drying a mixture of glycolic acid (100 mM) and glycine (100 mM) at 95°C for 16 days. 20 mM solution of glycine was added to the dry mixture every 24 hours.

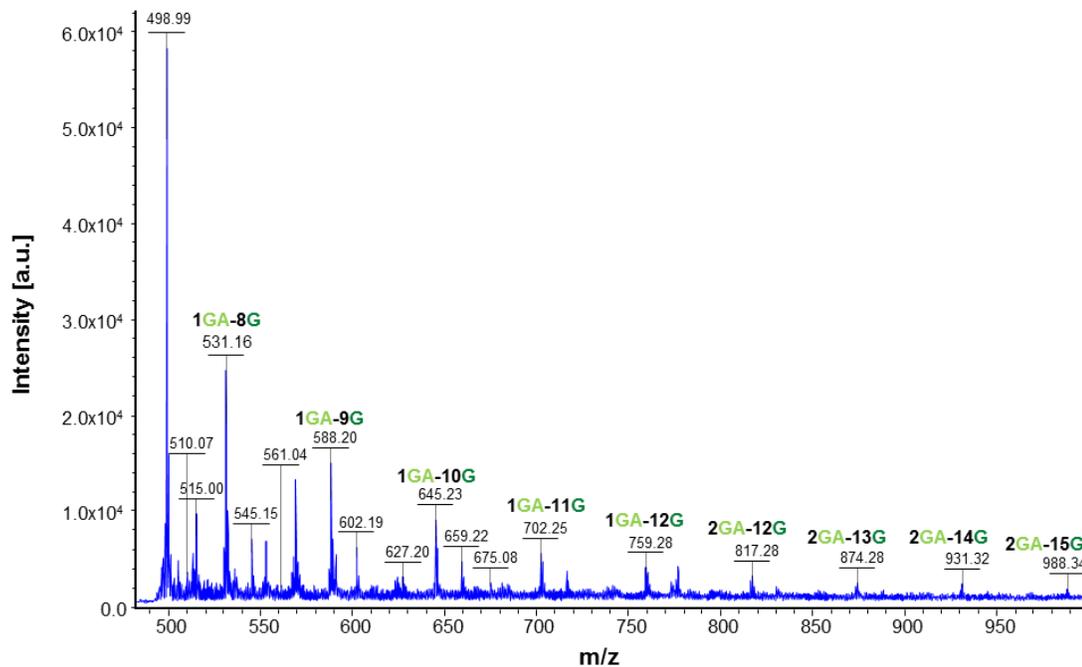


Figure 4.16 – MALDI-MS analysis. The sample was prepared by drying a mixture of glycolic acid (100 mM) and glycine (100 mM) at 95°C for 24 days. 20 mM solution of glycine was added to the dry mixture every 24 hours.

4.3.3 The effect of hydrolysis treatment on amide bonds

Although the hydrolysis treatment successfully removes all ester bonds and reduces the complexity of the mixtures, certain amounts of amide bonds might be cleaved at the same time. To estimate the loss of amide bonds from the hydrolysis treatment, we monitored the hydrolysis rate of the 1GA-2G standard under the same condition (95°C, pH 2.8). The hydrolysis of 1GA-2G oligomer produces short oligomers such as 1GA-1G, 2G, 1GA, 1G and DKP.

In Chapter 3, we established a model for the copolymerization and hydrolysis of depsipeptides. Here, a simplified version was used to study the kinetics of amide bond hydrolysis. P_i^O represents oligomers with hydroxy termini and P_i^N with amine termini. Subscript i indicates the number of amide bonds. For example, P_2^O is the 1GA-2G trimer and P_1^N is the 2G dimer. The hydrolysis of P_i^O is represented by Eq. (4.4). The first term is the loss of oligomers due to the amide bond hydrolysis. The second term represents the generation of oligomers from the hydrolysis of longer oligomers. The hydrolysis of P_i^N is similar to the one for P_i^O , but Eq. (4.5) includes a third term for the hydrolysis of the P_i^N oligomer series. Because the formation of DKP is observed in our experiment, we also include the cyclization of diglycine in Eq. (4.6) and (4.7). P_c is the amount of DKP. We assume it can only be generated from the cyclization of the glycine dimer (P_1^N). Therefore, the rate equation for P_1^N has one additional term for the cyclization.

$$\frac{dP_i^O}{dt} = -ik_{ha}WP_i^O + k_{ha}W \sum_{j=i+1}^{\infty} P_j^O \quad (4.4)$$

$$\frac{dP_i^N}{dt} = -ik_{ha}WP_i^N + k_{ha}W \sum_{j=i+1}^{\infty} P_j^O + 2k_{ha}W \sum_{k=i+1}^{\infty} P_k^N \quad (4.5)$$

$$\frac{dP_c}{dt} = k_c P_1^N \quad (4.6)$$

$$\frac{dP_1^N}{dt} = -k_{ha}WP_1^N + k_{ha}W \sum_{j=2}^{\infty} P_j^O + 2k_{ha}W \sum_{k=2}^{\infty} P_k^N - k_c P_1^N \quad (4.7)$$

The rate constants for amide bond hydrolysis and diglycine cyclization (k_c) were determined by fitting the data with Eqs. (4.4) to (4.7). The experimental data and simulation with optimized rate constants are shown in Figure 4.17. At 95°C, k_{ha} is 6.03×10^{-5} (L·mol⁻¹·h⁻¹) and k_c is 1.75×10^{-3} (h⁻¹). Previously determined ester hydrolysis rate constants at 95°C are 3.07×10^{-3} for polylactic acid and 1.35×10^{-2} (L·mol⁻¹·h⁻¹) for polymalic acid. The amide bonds are indeed more stable than ester bonds under similar hydrolysis condition.

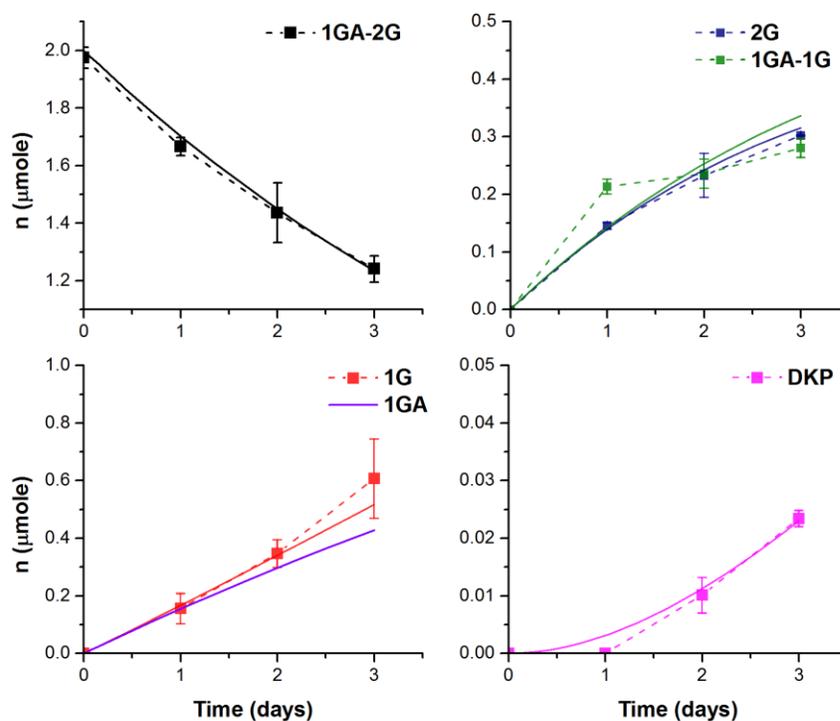


Figure 4.17 – Hydrolysis rate of 1GA-2G at 95°C. Solid lines are the model predictions. The experimental data are represented by the ■ symbols. Experimental data for 1GA is unavailable due to its low UV response. Only its model prediction is shown and it is not included in the fitting process.

To estimate the initial distribution of the 1GA-nG series without amide bond hydrolysis, we solved Eqs. (4.4) to (4.7) backward in time, using the experimental data as the initial condition and assuming all amide bonds have the same hydrolysis rate. The estimated distributions between the two experiments are shown in Figure 4.18. We are able to observe more long chain oligomers in the reconstructed distribution, compared to the experimental distribution. However, the most abundant oligomer is still the 1GA-2G when the feeding contains both glycolic acid and glycine. In the experiment fed with only glycine, the 1GA-1G dimer disappears in the 12 and 20 days experiments, suggesting it has been converted into longer oligomers. We further calculated the average number of amide bonds

in each distribution and summarized the results in Figure 4.19. The estimated distributions have higher values than the experimental distribution as expected. When feeding with both glycolic acid and glycine, the average chain length stays around 2.5. Feeding with only glycine gives a faster growth of average chain length in the first few days. However, the growth rate becomes slow after 16 days of reaction. As we mentioned before, it is possible that longer oligomers are not soluble in water and acetonitrile. Since we are only able to measure the concentrations of soluble oligomers, the growth rate might be higher than what we estimated in here. Those longer oligomers might also stop growing when they become insoluble.

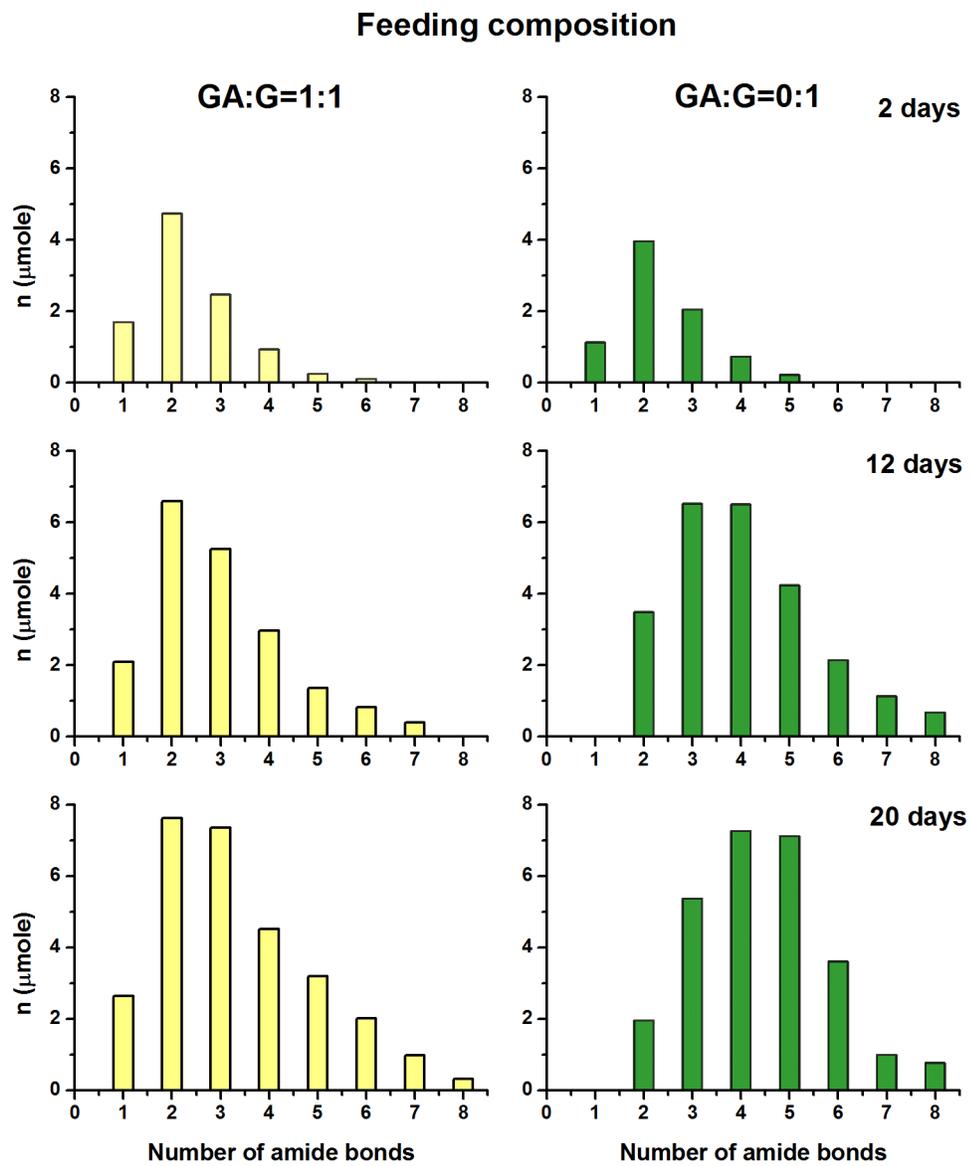


Figure 4.18 – Estimated oligomer distribution before the hydrolysis treatment.

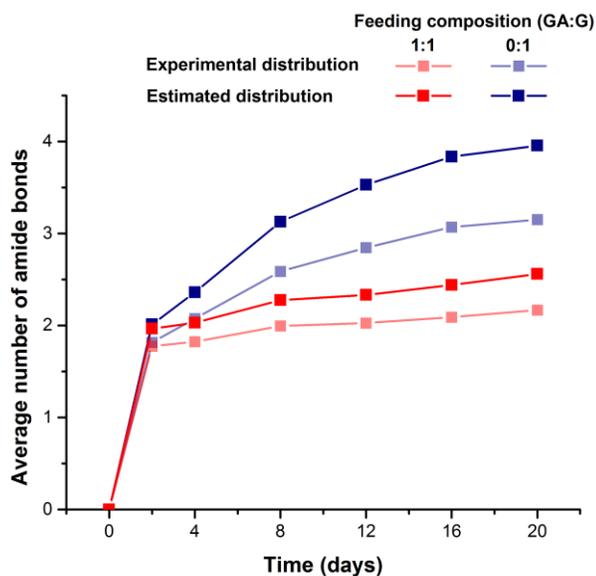


Figure 4.19 – Average number of amide bonds in each oligomer distribution. The results of the two feeding policies are represented by red (1:1) and blue (0:1). Lines with light and strong colors represent the experimental and estimated distribution, respectively.

This study focuses on the polymerization of glycine mainly because of its higher reactivity and its high abundance in the prebiotic mixture. Due to the absence of side-chains on the alpha-carbon, glycine usually is not considered as a key amino acid to form common secondary structures like alpha-helices and beta-sheets. As mentioned before, oligoglycines are insoluble in water by forming structures similar to beta-sheets and helices [106-108]. Along with polyproline, polyglycine forms basic motifs in collagens [110]. Considering the high abundance of glycine in prebiotic reaction mixtures, polyglycine I and II might be some of the earliest secondary structures. In the future, it is worth investigating if long chain depsipeptides from the ester-mediated reaction are able to form similar secondary structures, as polyglycine.

4.4 Conclusion

In this chapter, we established a quantitative method to monitor the progress of the GA/G copolymerization and studied the effect of feeding composition on the oligomer distribution. The combination of the condensation and the ester-amide exchange reaction works similar to the chain-growth living polymerization that is used to synthesize commercial polymers. Fresh monomers can be added to the C-termini of oligomers continuously without termination. Adding more glycine eventually leads to longer oligomers compared to our previous studies. Faster growth of average chain length was found when only glycine was fed to the dry mixture every cycle. Oligomers containing up to 13 repeated glycine residues have been identified by MS, much longer than is needed to form secondary structures. These results provide insights on how to synthesize longer proto-peptides in future studies.

CHAPTER 5. CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

The main goal of this thesis is to develop a new strategy that polymerizes amino acids into peptides in prebiotically-plausible scenarios. This simple task has been a challenge in the field of prebiotic chemistry for more than sixty years. In general, the polycondensation of amino acids is slow and instead of forming long chain peptides, stable cyclic dimer, DKP, forms as a side-product.

Apparently, direct polycondensation is probably not the best way to synthesize peptides. In this thesis, we present another route that forms amide bonds more efficiently. Starting from a mixture of hydroxy acids and amino acids in a wet-dry cycling environment, hydroxy acids form oligoesters easily, and then undergo a series of ester-amide exchange reactions to produce depsipeptides with peptide backbones.

In our first demonstration of this new routes, this idea was tested in many different conditions, such as concentration, reaction temperature, pH and with different pairs of hydroxy acids/ amino acids. The formation of amide bonds was confirmed by MS, FT-IR and NMR. When multiple amino acids mix together, the tandem MS analysis reveals that this reaction produces a large library of oligomers with random sequences.

To develop a theoretical framework of the ester-mediated reaction, I applied our knowledge in polymer science and chemical engineering. A model based on the population balance was used to study the kinetics of the ester-mediated copolymerization. This simple model is able to describe the behavior of the copolymerization and to provide rate constants

for each step. Even in the dry state, each reaction still follows the classic Arrhenius relation, possibly because of a more homogeneous environment made by melt hydroxy acids. Further calculation reveals that the activation energies of the ester-mediated pathway are lower than for direct amide formation.

Next, we interrogated the ability of the ester-mediated reaction, trying to find a strategy enables the formation long chain polymers with peptide backbones. We realized that this reaction is similar to the chain-growth living polymerization: new monomers add to one end of chains step by step without termination. Like operating a semi-batch reactor, unreacted monomers were fed to the dry mixture every cycle by an automated day-night machine. The feeding composition was found to affect the oligomer distribution. Longer oligomers can be produced more efficiently when only amino acids are fed because of a smaller number of active oligomer chains. For the copolymerization of glycolic acid and glycine, the ester-mediated reaction produces more amide bonds under the same condition, compared to the direct polycondensation of glycine.

Along with the success of our proposed pathway, we also realize several challenges still remain. The first one is the difficulty in analysis, especially when two or more amino acids are used. This reaction produces many oligomers with various chain lengths, compositions and sequences. Even with state of the art mass spectrometers, it is hard to track and quantitate all products. Second, amino acids with different side-chains have different reactivity to be incorporated. The presence of side-chains seems to increase the steric hindrance and lower the solubility of oligomers. Although the copolymerization of amino acids with bulky side-chains is possible, their lower reaction rate and low solubility might restrict further study in a reasonable time scale. Finally, the model developed in this

thesis considers only two different monomers. Although it would be interesting to probe the behavior of the system by adding other monomers, it will increase the complexity of the model dramatically and require more computational resources. Also, it will be more difficult to quantitate all possible copolymers for model parameterization and validation.

In summary, the above work provides a comprehensive study of the ester-mediated reaction to synthesize amide bonds. This simple reaction has been proved to be a robust pathway to polymerize amino acids in mild conditions. I hope that the knowledge gained from this study helps us finish at least one piece of the puzzle in the mystery of the origin of life on our Earth.

5.2 Recommendations

Moving forward, it is important to examine whether some properties of modern proteins can be generated from the depsipeptides. If the formation of peptide bonds is no longer a problem, what is the role of these oligomers in the next step of evolution? Here, I list a few recommendations that might be promising for future study.

5.2.1 Secondary structures and assemblies of depsipeptides

One important aspect of modern proteins is their ability to form secondary structures like alpha-helices and beta-sheets. Those special structures might enhance their stability and promote the self-replication of oligomers [111]. From the results of this thesis, forming long chain depsipeptides from simple building blocks like glycolic acid and glycine is relatively easy. However, incorporating more complex amino acids into long chain oligomers is still challenging. For lab-scale study, it seems to be more reasonable to focus

on the properties of short oligomers. It has been shown that some of the short peptides are able to form different nano-structures via self-assembly. For example, diphenylalanine and its derivatives are able to self-assemble into highly ordered structures such as nanotube, fiber and particles [112, 113]. Frederix *et al.* used the coarse-grain molecular dynamics simulation to search for candidates in a tripeptides library for hydrogel formation. Using this approach, they have found three tripeptides are able to form stable hydrogels in water[114].

For water soluble peptides, the secondary structures can be easily identified by circular dichroism. Usually, it is more difficult to confirm secondary structures of water-insoluble peptides. Recently, Greenwald *et al.* used carbonyl sulfide to synthesize peptides in an aqueous solution [115]. They found some longer oligomers made by alanine and valine precipitated from the solution. Using X-ray diffraction technique, they found those water-insoluble oligopeptides are able to assemble into amyloid fibers with beta-sheet structures.

The challenging part is how to design an experiment to explore the wide chemical space of amino acids. I recommend starting with those amino acids that have been identified in the prebiotic mixtures, such as glycine, alanine and aspartic acid. Repeating units of GAGA is the core structure of silk fibroin [116]. In addition, other studies have shown that even glycine peptides are able to form structures like vesicles if polar amino acids like aspartic acid are connected to the C-termini [117, 118]. Although the ester-mediated pathway produces a large library of oligomers with random compositions and sequences, it is worth investigating if some of them are able to adopt similar structures as modern proteins.

5.2.2 Catalytic activity of depsipeptides

Another important function of modern proteins is their catalytic activity. Some simple peptides have been shown to catalyze the hydrolysis of esters [119, 120]. The hydrolysis reaction is important as a tool to select certain products based on their stability, but it is more intriguing to see if some depsipeptides can enable the formation of new covalent bonds.

One potential model reaction is the aldol reaction. It combines two carbonyl compounds (aldehydes or ketone) to beta-hydroxy carbonyl compounds, also known as the aldols. This simple reaction is one possible route to the formation of polysaccharides on the early Earth [121]. The uncatalyzed aldol reaction is not only slow, but also produces racemic mixtures. Ideal catalysts for the aldol reaction should enhance the reaction rate and select one type of enantiomers as the main product. Numbers of studies have shown short peptides containing proline residues are able to catalyze the aldol reaction with enantioselectivities [122, 123]. A more recent study by Tena-Solsona *et al.* used a hydrogel environment formed by short peptides to promote the aldol reaction [124].

The model reaction to study the aldol reaction is the reaction of cyclohexanone and 4-nitro-benzaldehyde. Due to their distinct chemical shifts, the progress of the aldol reaction can be monitored easily by ^1H NMR. The enantiomer ratio can be determined by chiral HPLC. While NMR and HPLC provide useful information about the kinetic of aldol reaction, they might not be efficient tools to investigate a large pool of potential catalysts. Robbins and Hartwig have used GC-MS as a tool to survey an array of samples containing different substrates and catalysts [125]. A similar approach can be used to screen the

catalytic activities of various depsipeptides. Once identified, detailed kinetics of those catalysts can be studied by ^1H NMR and chiral HPLC. Combining with the day-night machine used in this thesis, a large library of depsipeptides can be generated and tested more efficiently.

5.2.3 *The rise of homochirality*

Almost all amino acids used in our life today are in their L-configuration. However, most of the prebiotic reactions produce racemic mixtures. L- and D-forms were likely to coexist on the early Earth [82]. Another unsolved problem in the field of prebiotic chemistry is why our life adopted one configuration preferentially. In Chapter 2, we briefly tested the reactivity of D-alanine in the ester-mediated reaction and found it worked similar to L-alanine. This result suggests the preference of one configuration over the other might not be distinctive initially.

Byrne *et al.* have studied the effect of peptide secondary structures on the enantioselectivity of peptide coupling [83]. When new amino acids are incorporated to the existing peptides, the helix structures of the peptides prefer one configuration over the other. We can test if the oligopeptides with secondary structures have a similar effect in the ester-mediated reaction. A simple experiment is to mix L-alanine oligomers with hydroxy acids and a racemic mixture of alanine. In the past, we have a few attempts to evaluate the efficiency of L- and D- amino acids in this chain-extension reaction by MS, but the MS signals were not reliable enough to give a concrete answer. From what we have learned so far, a better approach is first removing all ester linkages by hydrolysis, and then using a combination of reverse-phase and chiral chromatography to separate oligomers with

different chain lengths and configurations. This should yield a more robust result and confirm if the enantioselectivity caused by secondary structures also happening in the ester-mediated reaction.

5.2.4 *Some other prebiotic conditions*

Although the day-night cycling procedure used in the thesis is effective to synthesize the desired depsipeptides, there are other factors that can be included to improve the efficiency of the polymerization.

The ester-mediated pathway is a two-step process: esterification and ester-amide exchange reaction. The exchange reaction happens when the amine groups of amino acids are deprotonated. While the esterification is more favorable in an acidic condition, the exchange might happen faster in a neutral or a basic environment. The effect of pH is only briefly investigated in this work. The dry stage can be highly acidic to promote the esterification. To perform the exchange reaction, bases like ammonia can be added to the mixture during the rehydration process, similar to the pH-swing scenario used in this thesis.

To demonstrate the effectiveness of the ester-mediated reaction, we used the aqueous solution containing only the hydroxy acids and amino acids. However, the prebiotic mixture can be more complex. A direction worth investigating is the interaction between depsipeptides and some metal ions. Ferrous (Fe^{2+}) is the co-factor for the catalytic activities of some proteins [126]. Some peptides are known to form a complex with copper ions [127]. It is possible that the presence of metal ions can enhance the production of depsipeptides, similar to the salt-induced peptide synthesis [93].

In addition to pH and metal ions, the effect of UV radiation might be relevant as well. The early atmosphere did not have UV-shielding compounds like O₂ and O₃. Therefore, UV light might be more common than what we have today. UV light is known to activate molecules to their excited states [128]. This approach has been used as a prebiotically plausible way to synthesize nucleotides [129]. The additional energy from UV-radiation might also promote the polymerization of depsipeptides.

5.2.5 Application of the ester-mediated amide bond formation

As mentioned before, the solid-phase peptides synthesis requires a high amount of organic solvents and condensing agents. The yield is typically low when long chain peptides are desired. The ester-mediated pathway discussed in this thesis requires only low-cost starting materials in aqueous solution. The sequence of oligomers can be controlled by adding amino acids at different times like the feeding strategy in Chapter 4. The ester-mediated reaction can be a cheap and environmentally friendly method to synthesize polypeptides in high yield. The challenge is how to increase the reaction rate, so it can be competitive against solid-phase peptide synthesis. There are other factors in the cycling process that can potentially enhance the reaction rate. In the future, the knowledge we learned in the prebiotic chemistry might lead to new applications and achieve broader impacts.

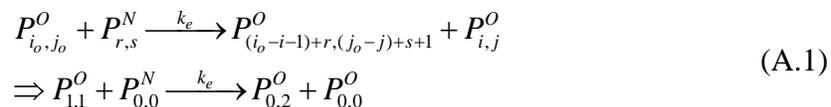
APPENDIX A. ILLUSTRATIVE EXAMPLES OF THE ESTER-AMIDE EXCHANGE MODEL

A.1 Simple example

The first example is the reaction between $P_{1,1}^O$ (parent oligomer) and $P_{0,0}^N$ (nucleophilic attacker) to form $P_{0,0}^O$ (ejection product) and $P_{0,2}^O$ (growth product).

A.1.1 Ejection of $P_{0,0}^O$ from parent oligomers $P_{1,1}^O$

As shown in Figure A.1, the parent oligomer in this case has two possible sequences ($C_1^{1+1} = 2$), but only one of them leads to $P_{0,0}^O$ by the ejection reaction.



To calculate the possibility of forming $P_{0,0}^O$ by the ejection reaction, we first count the numbers of ester and amide bonds in the two products that are originated from the parent oligomer. The ejection product does not have any ester or amide bond and it can only be arranged in one way ($C_0^{0+0} = 1$). The other product, $P_{0,2}^O$, has two amide bonds, but one of them is made from the exchange reaction. Only one amide bond is from the parent oligomer, so this segment has one possible sequence ($C_{1-0}^{(1-0-1)+(1-0)} = 1$). Therefore, only one sequence of the parent oligomer allows this reaction to happen. Averagely speaking, half of the ejection reaction produces $P_{0,0}^O$.

$$\frac{dP_{0,0}^O}{dt} = \frac{C_{1-0}^{(1-0-1)+(1-0)} C_0^{0+0}}{C_1^{1+1}} k_e P_{1,1}^O P_{0,0}^N = \frac{1 \cdot 1}{2} k_e P_{1,1}^O P_{0,0}^N \quad (\text{A.2})$$

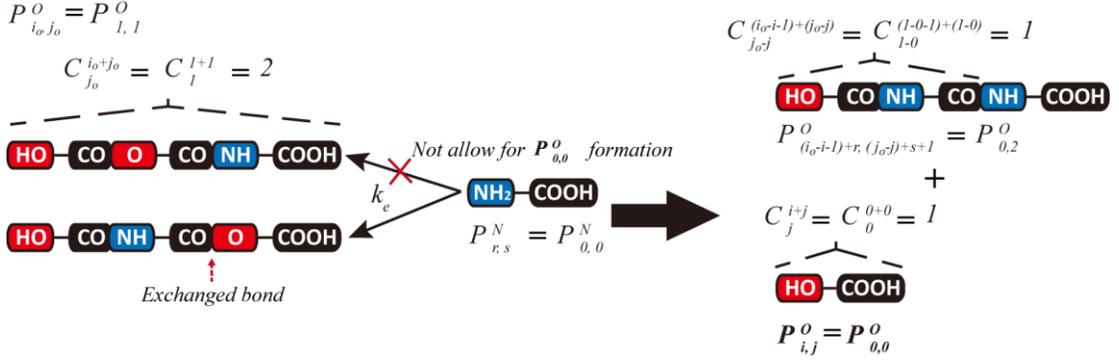
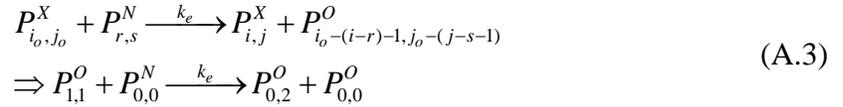


Figure A.1 – The ejection formation of $P_{0,0}^O$ from $P_{1,1}^O$ by a $P_{0,0}^N$.

A.1.2 Growth of $P_{0,2}^O$ with new amide bond from $P_{1,1}^O$

In the second case, we consider the growth rate of the oligomer $P_{0,2}^O$ with a newly attached amide bond for the same reaction (Figure A.2).



Again, we need to calculate the numbers of ester and amide bonds in the two products that belong to the parent oligomer. The product with the nucleophile has two amide bonds, but the segment from the parent oligomer has only one amide bond, so it has one possible arrangement ($C_{2-0-1}^{(0-0)+(2-0-1)} = 1$). Because one amide bond of the parent oligomer has been used to the product with the nucleophile and its ester bond has been consumed by the exchange reaction, the ejection product contains zero linkage and has

only one sequence ($C_{1-(2-0-1)}^{[1-(0-0)-1]+[1-(2-0-1)]} = 1$). Therefore, only half of the reaction from $P_{1,1}^O$ and $P_{0,0}^N$ yields $P_{0,2}^O$.

$$\frac{dP_{0,2}^O}{dt} = \frac{C_{2-0-1}^{(0-0)+(2-0-1)} C_{1-(2-0-1)}^{[1-(0-0)-1]+[1-(2-0-1)]}}{C_1^{1+1}} k_e P_{1,1}^O P_{0,0}^N = \frac{1 \cdot 1}{2} k_e P_{1,1}^O P_{0,0}^N \quad (\text{A.4})$$

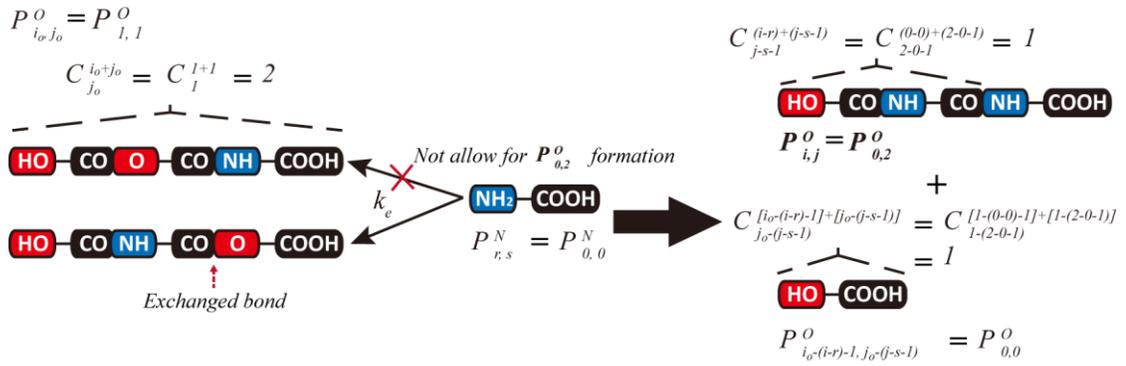


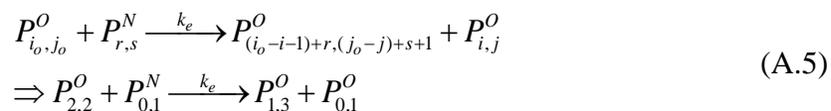
Figure A.2 – The growth of $P_{0,2}^O$ from $P_{1,1}^O$ by a $P_{0,0}^N$.

A.2 Advanced example

The second example is a more complex case. It is the reaction between $P_{2,2}^O$ (parent oligomer) and $P_{0,1}^N$ (nucleophilic attacker) to form $P_{0,1}^O$ (ejection product) and $P_{1,3}^O$ (growth product).

A.2.1 Ejection of $P_{0,1}^O$ from parent oligomers $P_{2,2}^O$

As shown in Figure A.3, the parent oligomer in this case has six possible sequences ($C_2^{2+2} = 6$), but only two of them can form $P_{0,0}^O$ by the exchange reaction.



To calculate the possibility of forming $P_{0,1}^O$ by the ejection reaction, we count the numbers of ester and amide bonds in the two products that are originated from the parent oligomer. The ejection product has one amide bond and it can only be arranged in one way ($C_1^{0+1} = 1$). The other product, $P_{1,3}^O$, has one ester bond and three amide bonds, but one of amide is made from the exchange reaction and another amide bond comes for the nucleophilic attacker. Only one ester bond and one amide bond are from the parent oligomer and this segment has two possible sequences ($C_{2-1}^{(2-0-1)+(2-1)} = 2$). Therefore, two sequences of the parent oligomer allow this reaction to happen. Averagely speaking, one-third of the ejection reaction produces $P_{0,1}^O$.

$$\frac{dP_{0,1}^O}{dt} = \frac{C_{2-1}^{(2-0-1)+(2-1)} C_1^{0+1}}{C_2^{2+2}} k_e P_{2,2}^O P_{0,1}^N = \frac{2 \cdot 1}{6} k_e P_{2,2}^O P_{0,1}^N
\tag{A.6}$$

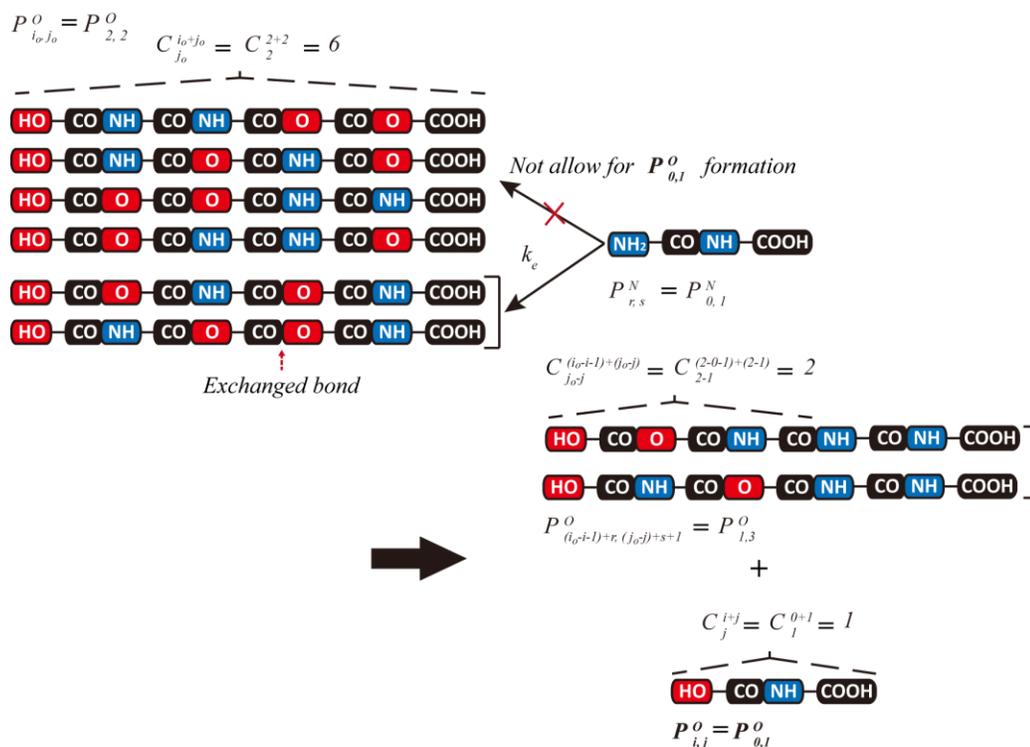
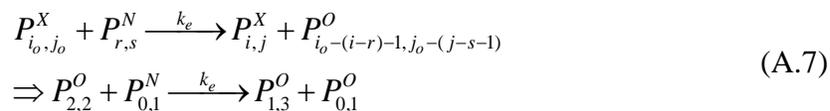


Figure A.3 – The ejection formation of $P_{0,1}^O$ from $P_{2,2}^O$ by a $P_{0,1}^N$.

A.2.2 Growth of $P_{1,3}^O$ with new amide bond from $P_{2,2}^O$

Here, we consider the growth rate of the oligomer $P_{1,3}^O$ with a newly attached amide bond for the same reaction (Figure A.4).



The product with the nucleophile has one ester bond and three amide bonds, but the segment from the parent oligomer has only one ester and one amide bond. Therefore, it has two possible arrangements ($C_{3-1-1}^{(1-0)+(3-1-1)} = 2$). Because two bonds of the parent oligomer have been used in the growth product and one ester bond is consumed by the exchange

reaction, the ejection product has one amide bond and one possible sequence

($C_{2-(3-1-1)}^{[2-(1-0)-1]+[2-(3-1-1)]} = 1$). Therefore, only one-third of the reaction from $P_{2,2}^O$ and $P_{0,1}^N$ yields

$P_{1,3}^O$.

$$\frac{dP_{1,3}^O}{dt} = \frac{C_{3-1-1}^{(1-0)+(3-1-1)} C_{2-(3-1-1)}^{[2-(1-0)-1]+[2-(3-1-1)]}}{C_2^{2+2}} k_e P_{2,2}^O P_{0,1}^N = \frac{2 \cdot 1}{6} k_e P_{2,2}^O P_{0,1}^N \quad (\text{A.8})$$

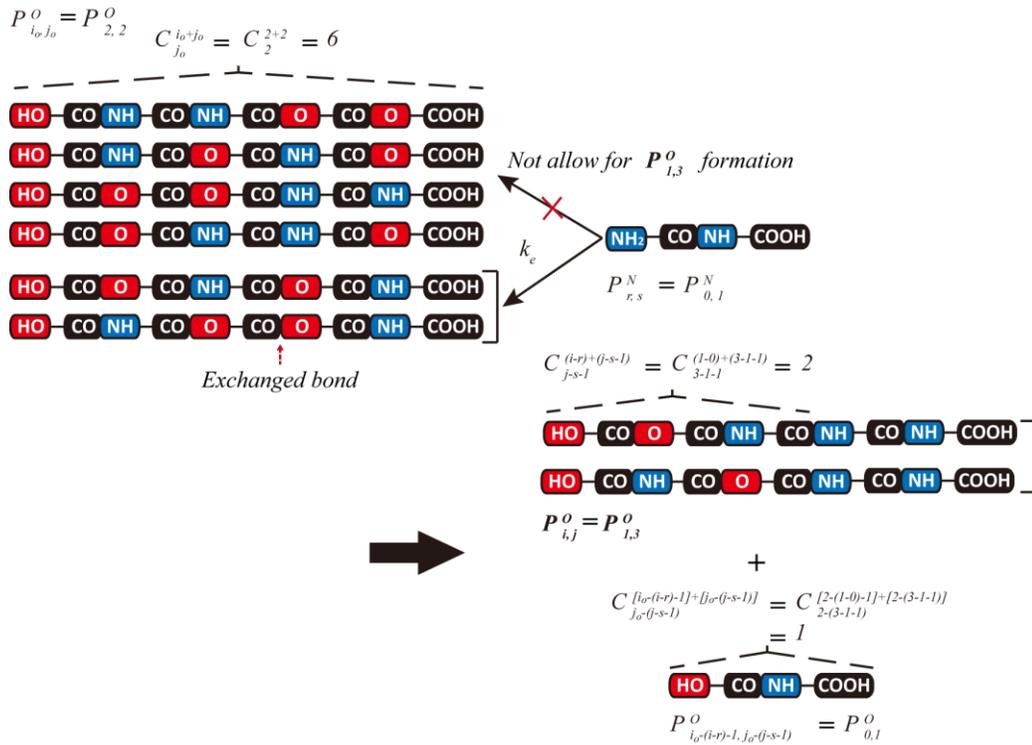


Figure A.4 – The growth of $P_{1,3}^O$ from $P_{2,2}^O$ by a $P_{0,1}^N$.

REFERENCES

1. Oró, J., Miller, S.L., and Lazcano, A., *The origin and early evolution of life on earth*. Annual Review of Earth and Planetary Sciences, 1990. **18**(1): p. 317-356.
2. Miller, S.L., *A production of amino acids under possible primitive earth conditions*. Science, 1953. **117**(3046): p. 528-529.
3. Bada, J.L. and Lazcano, A., *Prebiotic soup--revisiting the miller experiment*. Science, 2003. **300**(5620): p. 745-746.
4. Orgel, L.E., *Prebiotic chemistry and the origin of the RNA world*. Critical Reviews in Biochemistry and Molecular Biology, 2004. **39**(2): p. 99-123.
5. Bada, J.L. and Lazcano, A., *Some like it hot, but not the first biomolecules*. Science, 2002. **296**(5575): p. 1982-1983.
6. Gilbert, W., *Origin of life: The RNA world*. Nature, 1986. **319**(6055): p. 618-618.
7. Orgel, L.E., *Evolution of the genetic apparatus*. Journal of Molecular Biology, 1968. **38**(3): p. 381-393.
8. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R., *Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena*. Cell, 1982. **31**(1): p. 147-157.
9. Fuller, W.D., Sanchez, R.A., and Orgel, L.E., *Studies in prebiotic synthesis*. Journal of Molecular Biology, 1972. **67**(1): p. 25-33.
10. Bean, H.D., Sheng, Y., Collins, J.P., Anet, F.A.L., Leszczynski, J., and Hud, N.V., *Formation of a beta-pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction*. Journal of the American Chemical Society, 2007. **129**(31): p. 9556-9557.
11. Goncarencu, A. and Berezovsky, I.N., *Protein function from its emergence to diversity in contemporary proteins*. Physical biology, 2015. **12**(4): p. 045002.
12. Smith, R.M. and Hansen, D.E., *The pH-rate profile for the hydrolysis of a peptide bond*. Journal of the American Chemical Society, 1998. **120**(35): p. 8910-8913.
13. Grover, M.A., He, C., Hsieh, M.-C., and Yu, S.-S., *A chemical engineering perspective on the origins of life*. Processes, 2015. **3**(2): p. 309.

14. Bada, J.L., *New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments*. Chemical Society Reviews, 2013. **42**(5): p. 2186-2196.
15. Parker, E.T., Cleaves, H.J., Dworkin, J.P., Glavin, D.P., Callahan, M., Aubrey, A., Lazcano, A., and Bada, J.L., *Primordial synthesis of amines and amino acids in a 1958 Miller H₂S-rich spark discharge experiment*. Proceedings of the National Academy of Sciences, 2011. **108**(14): p. 5526-5531.
16. Parker, E.T., Zhou, M., Burton, A.S., Glavin, D.P., Dworkin, J.P., Krishnamurthy, R., Fernández, F.M., and Bada, J.L., *A Plausible Simultaneous Synthesis of Amino Acids and Simple Peptides on the Primordial Earth*. Angewandte Chemie International Edition, 2014. **53**(31): p. 8132-8136.
17. Kvenvolden, K., Lawless, J., Pering, K., Peterson, E., Flores, J., Ponnampereuma, C., Kaplan, I.R., and Moore, C., *Evidence for extraterrestrial amino acids and hydrocarbons in the murchison meteorite*. Nature, 1970. **228**(5275): p. 923-926.
18. Burton, A.S., Stern, J.C., Elsila, J.E., Glavin, D.P., and Dworkin, J.P., *Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites*. Chemical Society Reviews, 2012. **41**(16): p. 5459-5472.
19. Nick Pace, C. and Martin Scholtz, J., *A helix propensity scale based on experimental studies of peptides and proteins*. Biophysical Journal, 1998. **75**(1): p. 422-427.
20. Schmeing, T.M. and Ramakrishnan, V., *What recent ribosome structures have revealed about the mechanism of translation*. Nature, 2009. **461**(7268): p. 1234-1242.
21. Merrifield, R.B., *Solid phase peptide synthesis. I. The synthesis of a tetrapeptide*. Journal of the American Chemical Society, 1963. **85**(14): p. 2149-2154.
22. Bruice, P.Y., *Organic chemistry*. 2007: Pearson Prentice Hall.
23. Martin, R.B., *Free energies and equilibria of peptide bond hydrolysis and formation*. Biopolymers, 1998. **45**(5): p. 351-353.
24. Fox, S.W. and Harada, K., *Thermal copolymerization of amino acids to a product resembling protein*. Science, 1958. **128**(3333): p. 1214.
25. Fox, S.W. and Harada, K., *The thermal copolymerization of amino acids common to protein*. Journal of the American Chemical Society, 1960. **82**(14): p. 3745-3751.
26. Lahav, N., White, D., and Chang, S., *Peptide formation in the prebiotic era: Thermal condensation of glycine in fluctuating clay environments*. Science, 1978. **201**(4350): p. 67-69.

27. Orgel, L.E., *The origin of polynucleotide-directed protein synthesis*. Journal of Molecular Evolution, 1989. **29**(6): p. 465-474.
28. Bada, J.L., Miller, S.L., and Zhao, M., *The stability of amino acids at submarine hydrothermal vent temperatures*. Origins of Life and Evolution of the Biosphere, 1995. **25**(1): p. 111-118.
29. Rodriguez-Garcia, M., Surman, A.J., Cooper, G.J.T., Suarez-Marina, I., Hosni, Z., Lee, M.P., and Cronin, L., *Formation of oligopeptides in high yield under simple programmable conditions*. Nature Communications, 2015. **6**.
30. Bujdák, J. and Rode, B., *Glycine oligomerization on silica and alumina*. Reaction Kinetics and Catalysis Letters, 1997. **62**(2): p. 281-286.
31. Bujdák, J. and Rode, B., *Silica, alumina and clay catalyzed peptide bond formation: Enhanced efficiency of alumina catalyst*. Origins of Life and Evolution of the Biosphere, 1999. **29**(5): p. 451-461.
32. Lambert, J.-F., *Adsorption and polymerization of amino acids on mineral surfaces: A review*. Origins of Life and Evolution of Biospheres, 2008. **38**(3): p. 211-242.
33. Saetia, S., Liedl, K., Eder, A., and Rode, B., *Evaporation cycle experiments — a simulation of salt-induced peptide synthesis under possible prebiotic conditions*. Origins of Life and Evolution of the Biospheres, 1993. **23**(3): p. 167-176.
34. Rode, B.M. and Schwendinger, M.G., *Copper-catalyzed amino-acid condensation in water - a simple possible way of prebiotic peptide formation*. Origins of Life and Evolution of the Biosphere, 1990. **20**(5): p. 401-410.
35. Rode, B., Son, H., Suwannachot, Y., and Bujdak, J., *The combination of salt induced peptide formation reaction and clay catalysis: A way to higher peptides under primitive earth conditions*. Origins of Life and Evolution of the Biosphere, 1999. **29**(3): p. 273-286.
36. Imai, E.-i., Honda, H., Hatori, K., Brack, A., and Matsuno, K., *Elongation of oligopeptides in a simulated submarine hydrothermal system*. Science, 1999. **283**(5403): p. 831-833.
37. Cleaves, H.J., Aubrey, A.D., and Bada, J.L., *An evaluation of the critical parameters for abiotic peptide synthesis in submarine hydrothermal systems*. Origins of Life and Evolution of Biospheres, 2009. **39**(2): p. 109-126.
38. Griffith, E.C. and Vaida, V., *In situ observation of peptide bond formation at the water–air interface*. Proceedings of the National Academy of Sciences, 2012. **109**(39): p. 15697-15701.
39. Ferris, J.P., Hill, A.R., Liu, R., and Orgel, L.E., *Synthesis of long prebiotic oligomers on mineral surfaces*. Nature, 1996. **381**(6577): p. 59-61.

40. Leman, L., Orgel, L., and Ghadiri, M.R., *Carbonyl sulfide-mediated prebiotic formation of peptides*. *Science*, 2004. **306**(5694): p. 283-286.
41. Commeyras, A., Collet, H., Boiteau, L., Taillades, J., Vandenabeele-Trambouze, O., Cottet, H., Biron, J.-P., Plasson, R., Mion, L., Lagrille, O., Martin, H., Selsis, F., and Dobrijevic, M., *Prebiotic synthesis of sequential peptides on the Hadean beach by a molecular engine working with nitrogen oxides as energy sources*. *Polymer International*, 2002. **51**(7): p. 661-665.
42. Pascal, R., Boiteau, L., and Commeyras, A., *From the prebiotic synthesis of alpha-amino acids towards a primitive translation apparatus for the synthesis of peptides*, in *Prebiotic Chemistry*, P. Walde, Editor. 2005, Springer Berlin Heidelberg. p. 69-122.
43. Bean, H., Anet, F.L., Gould, I., and Hud, N., *Glyoxylate as a Backbone Linkage for a Prebiotic Ancestor of RNA*. *Origins of Life and Evolution of Biospheres*, 2006. **36**(1): p. 39-63.
44. Nelson, K.E., Levy, M., and Miller, S.L., *Peptide nucleic acids rather than RNA may have been the first genetic molecule*. *Proceedings of the National Academy of Sciences*, 2000. **97**(8): p. 3868-3871.
45. Orgel, L.E., *Some consequences of the RNA world hypothesis*. *Origins of Life and Evolution of the Biosphere*, 2003. **33**(2): p. 211-218.
46. Mamajanov, I., MacDonald, P.J., Ying, J., Duncanson, D.M., Dowdy, G.R., Walker, C.A., Engelhart, A.E., Fernández, F.M., Grover, M.A., Hud, N.V., and Schork, F.J., *Ester formation and hydrolysis during wet-dry cycles: Generation of far-from-equilibrium polymers in a model prebiotic reaction*. *Macromolecules*, 2014. **47**(4): p. 1334-1343.
47. Miller, S.L. and Urey, H.C., *Organic compound synthesis on the primitive earth: Several questions about the origin of life have been answered, but much remains to be studied*. *Science*, 1959. **130**(3370): p. 245-251.
48. Parker, E.T., Cleaves, H.J., Bada, J.L., and Fernández, F.M., *Quantitation of alpha-hydroxy acids in complex prebiotic mixtures via liquid chromatography/tandem mass spectrometry*. *Rapid Communications in Mass Spectrometry*, 2016. **30**(18): p. 2043-2051.
49. Cleaves, H.J., Chalmers, J.H., Lazcano, A., Miller, S.L., and Bada, J.L., *A reassessment of prebiotic organic synthesis in neutral planetary atmospheres*. *Origins of Life and Evolution of Biospheres*, 2008. **38**(2): p. 105-115.
50. Peltzer, E.T. and Bada, J.L., *alpha-Hydroxycarboxylic acids in the Murchison meteorite*. *Nature*, 1978. **272**(5652): p. 443-444.

51. Pizzarello, S. and Shock, E., *The organic composition of carbonaceous meteorites: The evolutionary story ahead of biochemistry*. Cold Spring Harbor Perspectives in Biology, 2010. **2**(3): p. a002105.
52. Cooper, G., Kimmich, N., Belisle, W., Sarinana, J., Brabham, K., and Garrel, L., *Carbonaceous Meteorites as a Source of Sugar-related Organic Compounds for the Early Earth*. Nature, 2001. **414**(6866): p. 879-883.
53. Huber, C. and Wächtershäuser, G., *alpha-Hydroxy and alpha-amino acids under possible hadean, volcanic origin-of-life conditions*. Science, 2006. **314**(5799): p. 630-632.
54. Jencks, W.P. and Gilchrist, M., *The free energies of hydrolysis of some esters and thiol esters of acetic acid*. Journal of the American Chemical Society, 1964. **86**(21): p. 4651-4654.
55. Mehta, R., Kumar, V., Bhunia, H., and Upadhyay, S.N., *Synthesis of poly(lactic acid): A review*. Journal of Macromolecular Science, Part C: Polymer Reviews, 2005. **45**(4): p. 325-349.
56. Garlotta, D., *A literature review of poly(lactic acid)*. Journal of Polymers and the Environment, 2001. **9**(2): p. 63-84.
57. Sabir, M., Xu, X., and Li, L., *A review on biodegradable polymeric materials for bone tissue engineering applications*. Journal of Materials Science, 2009. **44**(21): p. 5713-5724.
58. Ailianou, A., Ramachandran, K., Kossuth, M.B., Oberhauser, J.P., and Kornfield, J.A., *Multiplicity of Morphologies in Poly(L-lactide) Bioresorbable Vascular Scaffolds*. Proceedings of the National Academy of Sciences, 2016. **113**(42): p. 11670-11675.
59. Kumari, A., Yadav, S.K., and Yadav, S.C., *Biodegradable polymeric nanoparticles based drug delivery systems*. Colloids and Surfaces B: Biointerfaces, 2010. **75**(1): p. 1-18.
60. Dorresteyjn, R., Ragg, R., Rago, G., Billecke, N., Bonn, M., Parekh, S.H., Battagliarin, G., Peneva, K., Wagner, M., Klapper, M., and Müllen, K., *Biocompatible polylactide-block-polypeptide-block-polylactide nanocarrier*. Biomacromolecules, 2013. **14**(5): p. 1572-1577.
61. Dorresteyjn, R., Haschick, R., Klapper, M., and Müllen, K., *Poly(L-lactide) nanoparticles via ring-opening polymerization in non-aqueous emulsion*. Macromolecular Chemistry and Physics, 2012. **213**(19): p. 1996-2002.
62. Cao, H., Feng, Y., Wang, H., Zhang, L., Khan, M., and Guo, J., *Synthesis of depsipeptides from L-amino acids and lactones*. Frontiers of Chemical Science and Engineering, 2011. **5**(4): p. 409-415.

63. Croll, T.I., O'Connor, A.J., Stevens, G.W., and Cooper-White, J.J., *Controllable surface modification of Poly(lactic-co-glycolic acid) (PLGA) by hydrolysis or aminolysis I: Physical, chemical, and theoretical aspects*. *Biomacromolecules*, 2004. **5**(2): p. 463-473.
64. Holmes, S.A., *Aminolysis of poly(ethylene terephthalate) in aqueous amine and amine vapor*. *Journal of Applied Polymer Science*, 1996. **61**(2): p. 255-260.
65. Wang, B., Waters, A.L., Valeriote, F.A., and Hamann, M.T., *An efficient and cost-effective approach to Kahalalide F N-terminal modifications using a nuisance algal bloom of Bryopsis Pennata*. *Biochimica et Biophysica Acta - General Subjects*, 2015. **1850**(9): p. 1849-1854.
66. Hamann, M.T. and Scheuer, P.J., *Kahalalide F: A bioactive depsipeptide from the Sacoglossan Mollusk Elysia Rufescens and the green alga Bryopsis sp.* *Journal of the American Chemical Society*, 1993. **115**(13): p. 5825-5826.
67. Sun, H., Cheng, R., Deng, C., Meng, F., Dias, A.A., Hendriks, M., Feijen, J., and Zhong, Z., *Enzymatically and reductively degradable alpha-amino acid-based poly(ester amide)s: Synthesis, cell compatibility, and intracellular anticancer drug delivery*. *Biomacromolecules*, 2015. **16**(2): p. 597-605.
68. Ouchi, T., Nozaki, T., Ishikawa, A., Fujimoto, I., and Ohya, Y., *Synthesis and enzymatic hydrolysis of lactic acid-depsipeptide copolymers with functionalized pendant groups*. *Journal of Polymer Science Part A: Polymer Chemistry*, 1997. **35**(2): p. 377-383.
69. Takahashi, A., Umezaki, M., Yoshida, Y., Kuzuya, A., and Ohya, Y., *A macromolecular prodrug-type injectable polymer composed of poly(depsipeptide-co-lactide)-g-PEG for sustained release of drugs*. *Polymers for Advanced Technologies*, 2014. **25**(11): p. 1226-1233.
70. Sun, H., Meng, F., Dias, A.A., Hendriks, M., Feijen, J., and Zhong, Z., *alpha-amino acid containing degradable polymers as functional biomaterials: Rational design, synthetic pathway, and biomedical applications*. *Biomacromolecules*, 2011. **12**(6): p. 1937-1955.
71. Forsythe, J.G., Yu, S.-S., Mamajanov, I., Grover, M.A., Krishnamurthy, R., Fernández, F.M., and Hud, N.V., *Ester-mediated amide bond formation driven by wet-dry cycles: A possible path to polypeptides on the prebiotic earth*. *Angewandte Chemie International Edition*, 2015. **54**(34): p. 9871-9875.
72. Bowie, J.H., Brinkworth, C.S., and Dua, S., *Collision-induced fragmentations of the (M-H)⁻ parent anions of underivatized peptides: An aid to structure determination and some unusual negative ion cleavages*. *Mass Spectrometry Reviews*, 2002. **21**(2): p. 87-107.

73. Bokatzian-Johnson, S.S., Stover, M.L., Dixon, D.A., and Cassady, C.J., *Gas-phase deprotonation of the peptide backbone for tripeptides and their methyl esters with hydrogen and methyl side chains*. The Journal of Physical Chemistry B, 2012. **116**(51): p. 14844-14858.
74. Parella, T., *Pulse program catalogue*. NMR Guide v4. 0. BrukerBiospin, 2004.
75. Suzuki, S., Iwashita, Y., Shimanouchi, T., and Tsuboi, M., *Infrared spectra of isotopic polyglycines*. Biopolymers, 1966. **4**(3): p. 337-350.
76. Cronin, J.R. and Moore, C.B., *Amino acid analyses of the Murchison, Murray, and Allende carbonaceous chondrites*. Science, 1971. **172**(3990): p. 1327-1329.
77. Lahav, N., White, D., and Chang, S., *Peptide formation in prebiotic era - thermal condensation of glycine in fluctuating clay environments*. Science, 1978. **201**(4350): p. 67-69.
78. Orgel, L.E., *The origin of polynucleotide-directed protein-synthesis*. Journal of Molecular Evolution, 1989. **29**(6): p. 465-474.
79. Nagayama, M., Takaoka, O., Inomata, K., and Yamagata, Y., *Diketopiperazine-mediated peptide formation in aqueous-solution*. Origins of Life and Evolution of the Biosphere, 1990. **20**(3-4): p. 249-257.
80. Shanker, U., Bhushan, B., Bhattacharjee, G., and Kamaluddin, *Oligomerization of glycine and alanine catalyzed by iron oxides: Implications for prebiotic chemistry*. Origins of Life and Evolution of Biospheres, 2012. **42**(1): p. 31-45.
81. Kua, J. and Bada, J.L., *Primordial ocean chemistry and its compatibility with the RNA world*. Origins of Life and Evolution of Biospheres, 2011. **41**(6): p. 553-558.
82. Blackmond, D.G., *The origin of biological homochirality*. Cold Spring Harbor Perspectives in Biology, 2010. **2**(5).
83. Byrne, L., Sola, J., and Clayden, J., *Screw sense alone can govern enantioselective extension of a helical peptide by kinetic resolution of a racemic amino acid*. Chemical Communications, 2015. **51**(54): p. 10965-10968.
84. Codari, F., Moscatelli, D., Storti, G., and Morbidelli, M., *Characterization of low-molecular-weight PLA using HPLC*. Macromolecular Materials and Engineering, 2010. **295**(1): p. 58-66.
85. Zargar, A. and Schork, F.J., *Copolymer sequence distributions in controlled radical polymerization*. Macromolecular Reaction Engineering, 2009. **3**(2-3): p. 118-130.
86. Schork, J., *Control of polymerization reactors*. 1993: Taylor & Francis.

87. Regatte, V.R., Gao, H., Konstantinov, I.A., Arturo, S.G., and Broadbelt, L.J., *Design of copolymers based on sequence distribution for a targeted molecular weight and conversion*. Macromolecular Theory and Simulations, 2014. **23**(9): p. 564-574.
88. Wang, L. and Broadbelt, L.J., *Tracking explicit chain sequence in kinetic Monte Carlo simulations*. Macromolecular Theory and Simulations, 2011. **20**(1): p. 54-64.
89. Han, M.J., Kang, H.C., and Choi, K.B., *Kinetics of the copolycondensation by aminolysis, alcoholysis, and interchange reactions in the synthesis of poly(ester amide)*. Macromolecules, 1986. **19**(6): p. 1649-1652.
90. Yu, S.-S., Krishnamurthy, R., Fernandez, F.M., Hud, N.V., Schork, F.J., and Grover, M.A., *Kinetics of prebiotic depsipeptide formation from the ester-amide exchange reaction*. Physical Chemistry Chemical Physics, 2016.
91. Harshe, Y.M., Storti, G., Morbidelli, M., Gelosa, S., and Moscatelli, D., *Polycondensation kinetics of lactic acid*. Macromolecular Reaction Engineering, 2007. **1**(6): p. 611-621.
92. Radzicka, A. and Wolfenden, R., *Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases*. Journal of the American Chemical Society, 1996. **118**(26): p. 6105-6109.
93. Saetia, S., Liedl, K.R., Eder, A.H., and Rode, B.M., *Evaporation cycle experiments — a simulation of salt-induced peptide synthesis under possible prebiotic conditions*. Origins of Life and Evolution of the Biospheres, 1993. **23**(3): p. 167-176.
94. Bird, R.B., Stewart, W.E., and Lightfoot, E.N., *Transport phenomena*. 2nd ed. 2007: Wiley.
95. Bard, Y., *Nonlinear parameter estimation*. 1974: Academic Press.
96. Wynne-Jones, W.F.K. and Eyring, H., *The absolute rate of reactions in condensed phases*. Journal of Chemical Physics, 1935. **3**(8): p. 492-502.
97. Ring, D., Wolman, Y., Friedmann, N., and Miller, S.L., *Prebiotic synthesis of hydrophobic and protein amino acids*. Proceedings of the National Academy of Sciences, 1972. **69**(3): p. 765-768.
98. Filachione, E.M. and Fisher, C.H., *Lactic acid condensation polymers*. Industrial & Engineering Chemistry, 1944. **36**(3): p. 223-228.
99. Sakata, K., Kitadai, N., and Yokoyama, T., *Effects of pH and temperature on dimerization rate of glycine: Evaluation of favorable environmental conditions for chemical evolution of life*. Geochimica et Cosmochimica Acta, 2010. **74**(23): p. 6841-6851.

100. Shock, E.L., *Stability of peptides in high-temperature aqueous solutions*. *Geochimica et Cosmochimica Acta*, 1992. **56**(9): p. 3481-3491.
101. Sievers, A., Beringer, M., Rodnina, M.V., and Wolfenden, R., *The ribosome as an entropy trap*. *Proceedings of the National Academy of Sciences*, 2004. **101**(21): p. 7897-7901.
102. Trobro, S. and Åqvist, J., *Mechanism of peptide bond synthesis on the ribosome*. *Proceedings of the National Academy of Sciences*, 2005. **102**(35): p. 12395-12400.
103. Hao, Z., Lu, C.-Y., Xiao, B., Weng, N., Parker, B., Knapp, M., and Ho, C.-T., *Separation of amino acids, peptides and corresponding Amadori compounds on a silica column at elevated temperature*. *Journal of Chromatography A*, 2007. **1147**(2): p. 165-171.
104. Wolfenden, R., *Degrees of difficulty of water-consuming reactions in the absence of enzymes*. *Chemical Reviews*, 2006. **106**(8): p. 3379-3396.
105. Odian, G., *Principles of polymerization*. 2004: John Wiley & Sons.
106. Bykov, S. and Asher, S., *Raman studies of solution polyglycine conformations*. *The Journal of Physical Chemistry B*, 2010. **114**(19): p. 6636-6641.
107. Crick, F.H.C. and Rich, A., *Structure of polyglycine II*. *Nature*, 1955. **176**(4486): p. 780-781.
108. Lotz, B., *Crystal structure of polyglycine I*. *Journal of Molecular Biology*, 1974. **87**(2): p. 169-180.
109. Yalcin, T., Dai, Y., and Li, L., *Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for polymer analysis: solvent effect in sample preparation*. *Journal of the American Society for Mass Spectrometry*, 1998. **9**(12): p. 1303-1310.
110. Voet, D. and Voet, J.G., *Biochemistry, 4th Edition*. 2010: John Wiley & Sons.
111. Lee, D.H., Granja, J.R., Martinez, J.A., Severin, K., and Ghadiri, M.R., *A self-replicating peptide*. *Nature*, 1996. **382**(6591): p. 525-528.
112. Reches, M. and Gazit, E., *Casting metal nanowires within discrete self-assembled peptide nanotubes*. *Science*, 2003. **300**(5619): p. 625-627.
113. Mahler, A., Reches, M., Rechter, M., Cohen, S., and Gazit, E., *Rigid, self-assembled hydrogel composed of a modified aromatic dipeptide*. *Advanced Materials*, 2006. **18**(11): p. 1365-1370.
114. FrederixPim, W.J.M., Scott, G.G., Abul-Haija, Y.M., Kalafatovic, D., Pappas, C.G., Javid, N., Hunt, N.T., Ulijn, R.V., and Tuttle, T., *Exploring the sequence space for*

- (tri-)peptide self-assembly to design and discover new hydrogels. *Nature Chemistry*, 2015. **7**(1): p. 30-37.
115. Greenwald, J., Friedmann, M.P., and Riek, R., *Amyloid aggregates arise from amino acid condensations under prebiotic conditions*. *Angewandte Chemie International Edition*, 2016. **55**(38): p. 11609-11613.
 116. Jonker, A.M., Löwik, D.W.P.M., and van Hest, J.C.M., *Peptide- and protein-based hydrogels*. *Chemistry of Materials*, 2011. **24**(5): p. 759-773.
 117. Santoso, S., Hwang, W., Hartman, H., and Zhang, S., *Self-assembly of surfactant-like peptides with variable glycine tails to form nanotubes and nanovesicles*. *Nano Letters*, 2002. **2**(7): p. 687-691.
 118. Vauthey, S., Santoso, S., Gong, H., Watson, N., and Zhang, S., *Molecular self-assembly of surfactant-like peptides to form nanotubes and nanovesicles*. *Proceedings of the National Academy of Sciences*, 2002. **99**(8): p. 5355-5360.
 119. Friedmann, M.P., Torbeev, V., Zelenay, V., Sobol, A., Greenwald, J., and Riek, R., *Towards prebiotic catalytic amyloids using high throughput screening*. *PLOS ONE*, 2015. **10**(12): p. e0143948.
 120. Rufo, C.M., Moroz, Y.S., Moroz, O.V., Stöhr, J., Smith, T.A., Hu, X., DeGrado, W.F., and Korendovych, I.V., *Short peptides self-assemble to produce catalytic amyloids*. *Nature Chemistry*, 2014. **6**(4): p. 303-309.
 121. Weber, A.L., *The sugar model: Catalysis by amines and amino acid products*. *Origins of Life and Evolution of the Biosphere*, 2001. **31**(1): p. 71-86.
 122. Lei, M., Shi, L., Li, G., Chen, S., Fang, W., Ge, Z., Cheng, T., and Li, R., *Dipeptide-catalyzed direct asymmetric aldol reactions in the presence of water*. *Tetrahedron*, 2007. **63**(33): p. 7892-7898.
 123. Revell, J.D., Gantenbein, D., Krattiger, P., and Wennemers, H., *Solid-supported and pegylated H-Pro-Pro-Asp-NHR as catalysts for asymmetric aldol reactions*. *Peptide Science*, 2006. **84**(1): p. 105-113.
 124. Tena-Solsona, M., Nanda, J., Díaz-Oltra, S., Chotera, A., Ashkenasy, G., and Escuder, B., *Emergent catalytic behavior of self-assembled low molecular weight peptide-based aggregates and hydrogels*. *Chemistry – A European Journal*, 2016. **22**(19): p. 6687-6694.
 125. Robbins, D.W. and Hartwig, J.F., *A simple, multidimensional approach to high-throughput discovery of catalytic reactions*. *Science*, 2011. **333**(6048): p. 1423-1427.
 126. Thomson, A.J. and Gray, H.B., *Bio-inorganic chemistry*. *Current Opinion in Chemical Biology*, 1998. **2**(2): p. 155-158.

127. Bi, X. and Yang, K.-L., *Complexation of copper ions with histidine-containing tripeptides immobilized on solid surfaces*. Langmuir, 2007. **23**(22): p. 11067-11073.
128. Ranjan, S. and Sasselov, D.D., *Influence of the UV environment on the synthesis of prebiotic molecules*. Astrobiology, 2016. **16**(1): p. 68-88.
129. Powner, M.W., Gerland, B., and Sutherland, J.D., *Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions*. Nature, 2009. **459**(7244): p. 239-242.