이학박사학위논문

Discrete and Sequence-Defined Polyesters: Synthesis and Their Applications in Information Storage

분자량이 단일하고 서열이 정의된 폴리에스터의 합성과 정보 저장 매체로의 활용

2022 년 8 월

서울대학교 대학원

화학부 고분자화학전공

이 정 민

Ph. D. Dissertation

Discrete and Sequence-Defined Polyesters: Synthesis and Their Applications in Information Storage

Jeongmin Lee

Research Advisor: Prof. Kyoung Taek Kim

Department of Chemistry

Seoul National University

Discrete and Sequence-Defined Polyesters: Synthesis and Their Applications in Information Storage

분자량이 단일하고 서열이 정의된 폴리에스터의 합성과 정보 저장 매체로의 활용

지도교수 김 경 택

이 논문을 이학박사 학위논문으로 제출함.
2022년 8월

서울대학교 대학원
화학부 고분자화학 전공
이 정 민

이 정 민의 이학박사 학위논문을 인준함
2022년 8월

위 원 장	최 태 림 (인)
부 위 원 장	김 경 택 (인)
위 원	손 병 혁 (인)
위 원	이 연 (인)
위 원	서 명 은 (인)

Abstract

Discrete and Sequence-Defined Polyesters: Synthesis and Their Applications in Information Storage

Jeongmin Lee
Major in polymer chemistry
Department of Chemistry
Seoul National University

The synthesis of sequence-defined polymers with perfect control over the molecular weight, monomer sequence, and stereoconfiguration remains a challenge in polymer chemistry. Sequence-defined polymers can be used in extensive applications, such as foldamers, catalysis, and antimicrobials. Particularly, these polymers can serve as information storage media by converting their monomer sequences into digital information such as binary code. Several strategies to synthesize uniform macromolecules, such as stepwise iterative synthesis and iterative exponential growth, have been developed, but they have limitations in terms of scalability, polymer length, and periodic sequence. In this study, the synthesis of sequence-defined polyesters using a cross-convergent method and information storage using the aperiodic sequence of the polymers are demonstrated.

In this dissertation, high molecular weight sequence-defined poly(phenyllactic-*co*-lactic acid)s (PcLs) were synthesized using a cross-convergent method combined

with preparative size exclusion chromatography (prep-SEC) for purification. This method facilitates scalable synthesis of binary-encoded PcLs with minimal chemical reactions. The stored information in a 64-bit PcL could be decoded via a single measurement using MALDI-TOF tandem mass spectrometry. Furthermore, a degradative sequencing method was used for the analysis of the monomer sequence of a high molecular weight 128-bit PcL.

Despite the step-economical synthesis of sequence-defined polymers via the cross-convergent method, large information storage requires the synthesis of a multitude of sequence-defined polymers accompanied by massive chemical reactions. This challenge could be overcome by the semiautomated synthesis of sequence-defined poly(L-lactic-*co*-glycolic acid)s (PLGAs) using continuous flow chemistry. This accelerated synthesis allowed the encoding of a bitmap image (896-bit) in 14 PLGAs incurring only a fraction of time compared to batch reactions. Moreover, introducing an 8-bit address code as the PLGA chain identifier enabled direct tandem mass sequencing of the mixture of several PLGA chains.

Most sequencing methods for sequence-defined polymers, such as tandem mass and self-immolative sequencing, which inevitably consume polymers in each analysis, are destructive. Therefore, a nondestructive sequencing method for sequence-defined enantiopure oligoesters, oligo(L-lactic-*co*-glycolic acid)s (oLGs) and oligo(L-mandelic-*co*-D-phenyllactic acid)s (oMPs), was developed using ^{13}C NMR spectroscopy. The sequence of a mixture of oLG and oMP was deciphered via a single ^{13}C NMR measurement because of the non-overlapping chemical shift region of the sequence-indicating peaks. A bitmap image (192-bit) was encoded in enantiopure octameric oLG and oMP through semi-automated flow synthesis, and 12 equimolar mixtures of oLG and oMP were decoded completely.

Although biodegradable polyhydroxyalkanoates (PHAs) are an attractive alternative to hydrocarbon-based plastics, their application is inhibited by the limited chemical structure of PHAs derived by biological and chemical synthesis. Therefore, a synthetic procedure for a library of hydroxyalkanoates (HAs) with desired atomic compositions, stereochemical configurations, and substituent chemistry from accessible terminal epoxides was established. The defined HAs served as building blocks for sequence-defined PHAs with controlled molecular weight, monomer sequence, stereoconfiguration, and functional moieties. Furthermore, macromolecular engineering of sequence-regulated PHAs allowed for controlling crystallinity and thermal properties of the polymers.

From the above experiments, a new method for the scalable synthesis of high molecular weight sequence-defined polyesters was established, and their efficiency was improved by introducing flow chemistry. Furthermore, it was demonstrated that the sequence-defined polymers could serve as information storage media, and encoding and decoding strategies were developed. Finally, the potential of macromolecular engineering using sequence-regulated polymers was suggested. These results will contribute to the unlimited diversity of polymers in the discovery and understanding of structure-function relationships. Furthermore, molecular engineering of synthetic polymers will also contribute to the development of molecular media, paving way for a broad range of potential applications.

Keywords: sequence-defined polymer, cross-convergent method, continuous-flow synthesis, information storage, sequencing, macromolecular engineering

Student Number: 2017-23953

Table of Contents

Chapter 1. Introduction

Chapter 2. High-Density Information Storage in an Absolutely Defined Aperiodic Sequence of Monodisperse Copolyester

Chapter 3. Semiautomated Synthesis of Sequence-Defined Polymers for Information Storage

Chapter 4. Nondestructive Sequencing of Enantiopure Oligoesters by Nuclear Magnetic Resonance Spectroscopy

Chapter 5. Synthesis of Enantiomeric ω-Substituted Hydroxyalkanoates from Terminal Epoxides: Building Blocks for Sequence-defined polyesters and Macromolecular Engineering

Appendix

List of Schemes

List of Tables

List of Figures

Abbreviations

DNA	Deoxyribonucleic acid
DMT	Dimethoxytrityl
ASCII	American standard code for information interchange
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
TEMPO	(2,2,6,6-Tetramethylpiperidin-1-yl)oxyl
QR code	Quick response code
ESI-MS	Electrospray ionization-mass spectrometry
MS/MS	Tandem mass spectrometry
LC-MS	Liquid chromatography-mass spectrometry
PcL	Poly(phenyllactic-*co*-lactic acid)
P	*rac*-Phenyllactic acid
L	*rac*-Lactic acid
PAH	Poly(α-hydroxy acid)
TBDMS	*tert*-Butyldimethylsilyl
$BF_3 \cdot Et_2O$	Boron trifluoride etherate
Da	Dalton
NMR	Nuclear magnetic resonance
Prep-SEC	Preparative size-exclusion chromatography
GPC	Gel permeation chromatography
Bz	Benzyl
PA	Poly(*rac*-phenyllactic acid)
PLGA	Poly(L-lactic-*co*-glycolic acid)
L	L-Lactic acid
G	Glycolic acid
THF	Tetrahydrofuran
Et_3SiH	triethylsilane
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide

V_{rl}	Volume of reaction loop
ppm	Parts per million
oLG	Oligo(L-lactic-*co*-glycolic acid)
oMP	Oligo(L-mandelic-*co*-D-phenyllactic acid)
$Pd(PPh_3)_4$	Tetrakis(triphenylphosphine)palladium(0)
DPTS	4-(Dimethylamino) pyridinium 4-toluenesulfonate
M	L-Mandelic acid
P	D-Phenyllactic acid
HA	Hydroxyalkanoate
PHA	Polyhydroxyalkanoate
oHA	Oligo(hydroxyalkanoate)
P3HB	Poly(3-hydroxybutyrate)
T_m	Melting temperature
HKR	Hydrolytic kinetic resolution
MCL	Medium-chain-length
DIC	*N,N′*-Diisopropylcarbodiimide
$Đ$	Polydispersity index
M_n	Number-average molecular weight
T_g	Glass transition temperature

Chapter 1. Introduction

1.1 Overview

Since the notation of "polymerization" was proposed by Hermann Staudinger in 1932, polymer chemistry has extensively grown.[1] In turn, various commercial plastics have been developed and produced enormously to the extent that polymeric materials have become an inevitable part of our lives.[2] Conventional polymerization methods, such as free radical polymerization and step-growth polymerization, are straightforward but limited in that the obtained polymers have monotonous microstructures. This issue has been overcome by developing living polymerization, in which chain-breaking reactions, such as termination and transfer, do not exist, and all chains are instantaneously initiated and simultaneously grown, allowing controlled composition of monomers, chain length, molecular weight distribution, topology, and functionality.[4,5] Consequently, synthetic polymers can be used in an extensive range of applications, such as hydrogels, conductors, vitrimers, and nanomaterials.[6,7]

However, inherent statistical uncertainties in terms of the number and sequence of monomers[8,9] hinder the use of synthetic polymers in highly advanced areas in which biopolymers, such as DNA and proteins, with absolutely defined molecular weights and sequences, are used. Thus, there are perpetual attempts to devise an innovative strategy for the synthesis of uniform polymers.[10–14] For example, sequence-regulated polymers resulting from step-growth polymerization[15,16] or metathesis polymerization[17,18] of sequence-controlled monomers facilitate control over the sequence; however, the molar mass distribution remains a challenge. Another viable method has been developed in which single unit monomer insertion

through a radical addition reaction mechanism with monomer additions occurs one at a time; however, this approach has limitations on the yield of monomer insertion and incompleteness.[19,20]

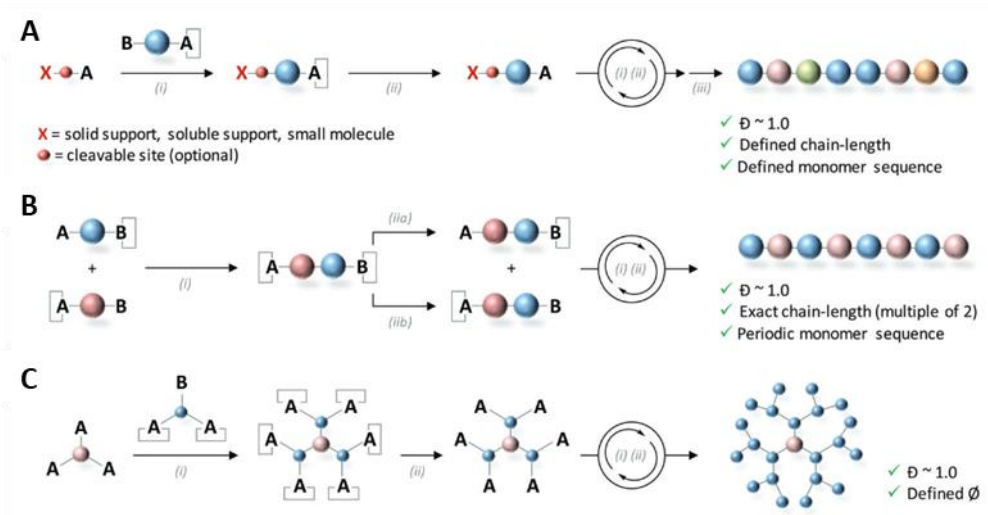


Figure 1-1. Classical multistep growth mechanisms: (A) stepwise iterative synthesis, (B) iterative exponential growth, (C) divergent dendrimer synthesis. Function A reacts with function B. Step (i) coupling step. Step (ii) deprotection step. A final cleavage step (iii) is required in some cases.[21]

Eventually, multistep growth synthesis can resolve the issue of synthesizing uniform polymers, wherein reaction and purification are separately performed.[21,22] There are three types of multistep growth synthesis: stepwise iterative synthesis,[23,24] iterative exponential growth,[25] and dendrimer synthesis.[26,27] The stepwise iterative strategy involves linear growth of monomers by successive iterations of deprotecting the terminal end, followed by coupling with a monoprotected monomer. Iterative exponential growth, meanwhile, enables linear growth by doubling the repeating units by successive iterations of orthogonal deprotections of two terminal ends,

followed by coupling between monoprotected reagents. Dendrimers are branched, precise macromolecules resulting from repetitive reactions, including deprotection of the terminal ends of multifunctional building blocks and coupling reactions with monoactivated dendritic monomers.

Uniform polymers obtained from the aforementioned strategies, which are superior to conventional polymerization, have controlled absolute molecular weight, monomer sequence, and stereoconfiguration. Therefore, they enable whole new suite of applications, such as efficient platform for catalysis, antibacterial biomaterials, and molecular recognition.[28] Particularly, digital information can be stored in sequence-defined polymer by converting monomer sequence into binary code.[29] This topic has developed in recent years, and consequently several "writing" and "reading" techniques were devised and improved for large-scale information storage.[30] In addition, these polymers can be used in cryptography,[31] anti-counterfeiting,[32] and product-identification.[33]

1.2 Stepwise iterative synthesis

One of the most prevalent methods for sequence-defined polymers is solid-phase synthesis because of their easy purification. Although this technique was initially developed for the chemical synthesis of biopolymers, such as proteins and nucleic acids,[34] it has been applied extensively in the development of sequence-defined oligomers and polymers. This method involves a coupling reaction of a polymer precursor fixed to an insoluble solid with a monoprotected monomer and a deprotection reaction at the end of the polymer. The coupling and deprotection steps yield sequence-defined polymers with a step-wise increase in the number of repeating units.

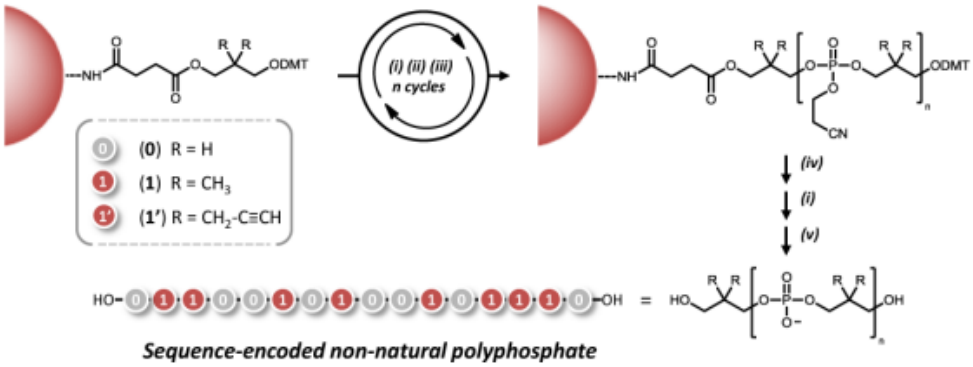


Figure 1-2. General strategy for synthesis of sequence-defined non-natural polyphosphate. Iterative phosphoramidite protocol constitutes five steps. (i) DMT deprotection; (ii) Coupling; (iii) Oxidation; (iv) Cyanoethyl deprotection; (v) Cleavage.[35]

Lutz et al. (2015)[35] reported a synthetic method for sequence-defined non-natural polyphosphates using iterative phosphoramidite protocols on a solid polystyrene support. As shown in Figure 1-2, phosphoramidite coupling between the monomer bound to the solid support and the dimethoxytrityl (DMT)-protected monomer yields a dimer. The resulting dimer underwent successive oxidation, DMT deprotection, and coupling to obtain the desired lengths of the polymers. Finally, cleavage of the solid support yielded sequence-defined oligomers. Moreover, the quantitative modification of the alkyne side chains by copper-catalyzed azide-alkyne cycloaddition was achieved. Lutz et al. (2015)[36] further developed a synthetic method using 1000 Å pore glass as a solid support and introduced capping steps, obtaining sequence-defined polymers with more than 100 repeating units.

Several strategies for protecting-group-free iterative synthesis have been developed. Du Prez and coworkers (2013)[37] reported a synthetic strategy for

sequence-defined oligomers using thiolactone-based chemistry (Figure 1-3). This protecting-group-free approach comprises two steps: selective aminolysis of thioesters releasing a thiol group and coupling of a monomer containing thiolactone. Repetitive reaction cycles and cleavage from the solid support yielded sequence-defined oligomers with several functional groups. The same group demonstrated the automated synthesis of multifunctional sequence-defined oligomers using a peptide synthesizer with similar thiolactone chemistry.[38]

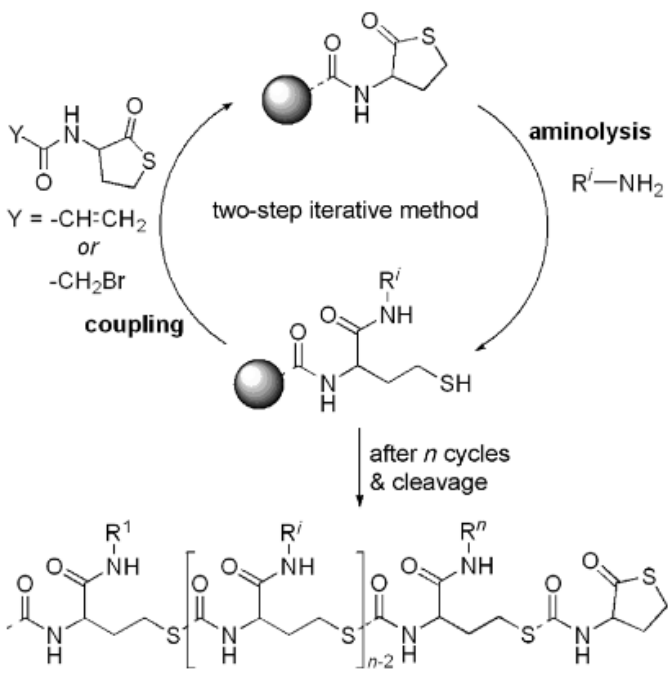


Figure 1-3. Two-step iterative method for the synthesis of sequence-defined oligomers on solid support based on thiolactone-based chemistry.[38]

While solid-phase synthesis is advantageous in terms of its facile purification, milligram scale reactions limit the scope of potential applications. Alabi et al. (2014)[39] introduced a liquid-phase approach using soluble fluorous supports instead of solid supports. They synthesized multifunctional sequence-defined polymers on fluorous tags via sequential phosphine-catalyzed Michael addition and photoinitiated thiol-ene click addition. The use of a fluorous support allows homogeneous reaction conditions in common organic solvents, which results in rapid solution kinetics. Additionally, sequence-defined polymers can be selectively purified via rapid fluorous solid-phase extraction. Meier et al. (2016)[40] demonstrated the synthesis of sequence-defined oligomers at a high yield and multigram scale using an iterative Passerini three-component reaction. Livingston et al. (2019)[41] demonstrated liquid-phase synthesis with molecular sieving, which facilitated the synthesis of sequence-defined polyethers with structural and functional diversities.

1.3 Iterative exponential growth

Stepwise iterative synthesis is a powerful strategy to control the composition and sequence of monomers, but the chain length is limited owing to deletion errors owing to indistinguishable products and unreacted precursors. For example, if the reaction yield is 99.5%, the proportion of the desired 10-mer is 95%; however, a sequence-defined polymer without any deletion errors is only 60% after 100 repetitive reactions. Therefore, another strategy was proposed by Whiting et al. (1982)[42], which involved the synthesis of discrete polyethylene with up to 192 carbon atoms by repeated conversion of the chain end and coupling reaction. This elegant molecular doubling approach is based on iterative exponential growth.[43]

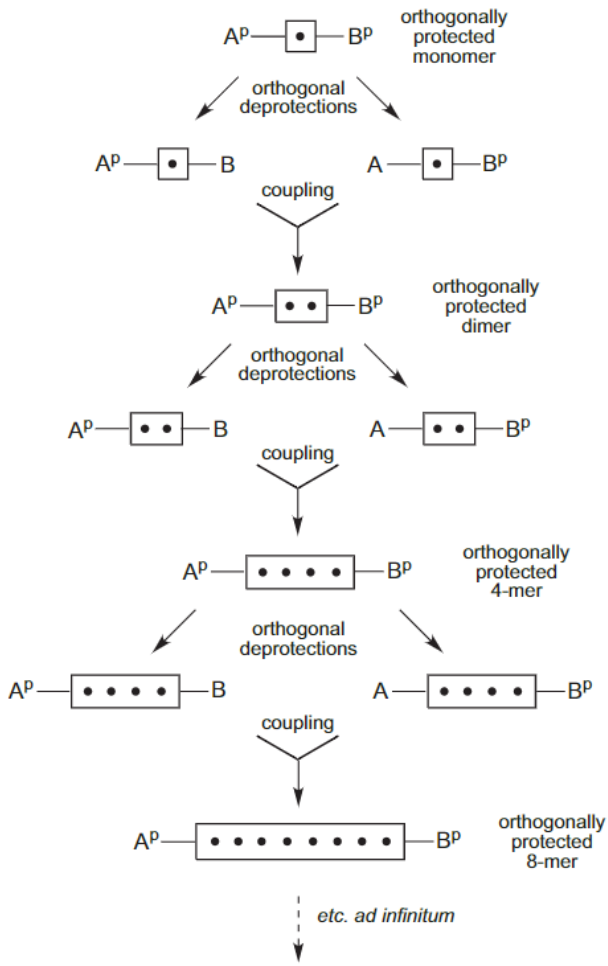


Figure 1-4. Scheme for synthesis of discrete polymers through iterative exponential growth.[44]

As shown in Figure 1-4, the orthogonally protected monomer A^P-M-B^P was selectively deprotected to yield two monoprotected intermediates: A-M-B^P and A^P-M-B. These intermediates were coupled to prepare the orthogonally protected dimer A^P-MM-B^P. These orthogonal deprotection and coupling steps constitute a single iteration. In this process, the coupling product serves as a constituent unit for the next iteration, thus inducing molecular doubling.[44] Due to good efficiency and

scalability, several discrete conjugated macromolecules, such as oligo(phenylene ethynylene)s,[45] oligo(thiophene)s,[46] oligo(*p*-phenylene)s,[47] and oligo(*m*-aniline)s,[48] were reported but they were limited to short chain lengths due to poor solubility.

Furthermore, there are other strategies to achieve discrete commodity polymers, such as polyethylene,[49] polyether,[50] and polyester.[51,52] These discrete polymers could serve as a standard reference in analytical instruments, such as size exclusion chromatography or mass spectrometry, and model compounds for degradability or crystallinity studies. For example, Hawker et al. (2008)[53] reported an iterative exponential growth of discrete ε-caprolactone oligomers and polymers with up to 64 repeating units. In this study, *tert*-butyldimethylsilyl ether and benzyl ester served as the protecting groups for the hydroxyl and carboxylic acid groups, respectively. The monoprotected precursors were coupled using 1,3-dicyclohexylcarbodiimide and 4-(dimethylamino)pyridinium-*p*-toluenesulfonate to afford the ester backbone. The quantitative yields of each deprotection and coupling reaction facilitated the synthesis of discrete high molecular weight polymers. Hawker et al. also reported the synthesis of discrete poly(lactic acid)s using a similar chemistry approach and investigated the correlation between the number of repeating units and their distinct physical properties.[54]

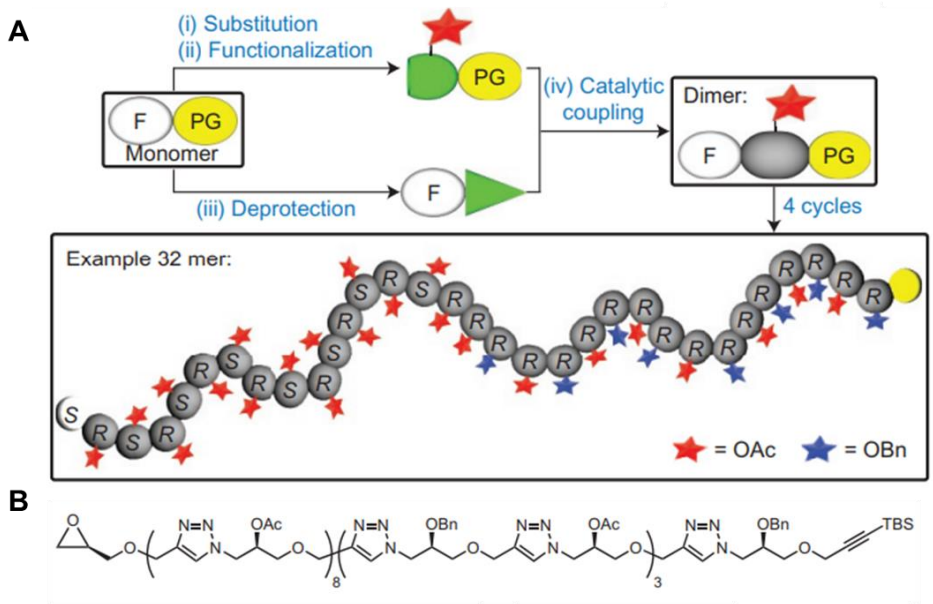


Figure 1-5. (A) Illustration of the iterative exponential growth plus side-chain functionalization (IEG+). (B) An example of actual chemical structure resulting from IEG+.[55]

In addition to the ability to control molecular weight and molar mass distribution, iterative exponential growth enables the control of other structural parameters of polymers, including stereoregularity. Johnson et al. (2015)[55] reported iterative exponential growth plus side-chain functionalization by employing chiral monomers containing functional groups (Figure 1–5). Repetition of orthogonal deprotection and functionalization was followed by coupling afforded sequence- and stereo-controlled polymers with acetyl and benzyl functional groups. Scalable synthesis of functionalized sequence-defined polymers using flow chemistry was demonstrated by the same group.[56] Furthermore, this strategy could be applied to the assembly of unimolecular block copolymers with varying side chain chemical structures and stereochemical sequences.[57,58] Recently, the same group investigated

biological properties influenced by the control of stereochemistry and conformational rigidity of bottlebrush polymers resulting from ring-opening metathesis polymerization of sequence-defined norbornene-terminated macromonomers.[59]

1.4 Digital information storage in synthetic macromolecules

With the exponential growth in the demand for digital information storage, new methods for writing and storing information are required.[60–62] Owing to their excellent durability and high information density, synthetic DNAs can serve as an alternative to silicon-based devices.[63,64] Davis (1988)[65] initially demonstrated synthetic DNAs containing a coded image. A novel study of large-scale information storage in DNA was reported by Church et al. (2012)[66] in which 5.27 megabits of an entire book's draft, including 53,426 words, 11 images, and 1 JavaScript program were split into DNA nucleotides, comprising address code, data block, and common sequence for amplification and sequencing. Because encoding and decoding strategies have been developed over time, numerous studies on data storage with respect to higher information density and lower error rates have been reported.[67,68]

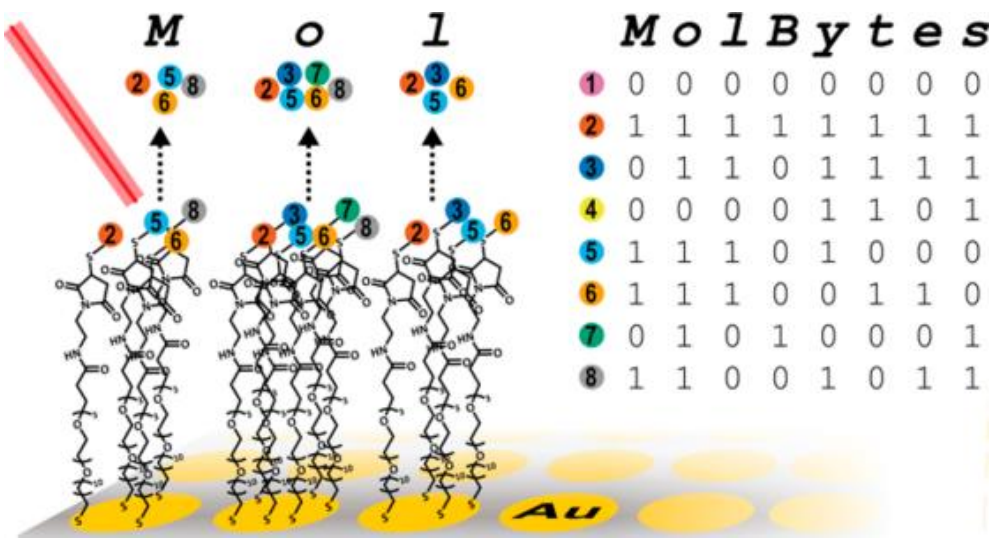


Figure 1-6. Schematic illustration of an example of information storage in immobilized oligopeptides. In the first spot, four different oligopeptides are immobilized, representing 01001101, which indicates the character "M."[70]

In addition to synthetic DNA, proteins, small molecules, and polymers have been used for alternative data storage.[69,70] Whitesides et al. (2019)[71] showed information storage in mixtures of readily available and stable oligopeptides (Figure 1-6). Each polypeptide immobilized on a monolayer has a different molecular weight, which enables us to distinguish the presence of the peptides representing "1" in the binary system. The combination of eight oligopeptides indicates a byte message written in the ASCII code. A total of 32 oligopeptides with different molecular weights were used to encode four bytes in one spot. The information stored in each spot was decoded using a MALDI-TOF mass spectrometer. Consequently, massive amounts of information could be stored on an array plate and decoded at a rate of 20 bits/s. The same group demonstrated another study on information storage using a similar approach.[72] They converted oligopeptides with distinguishable molecular

weights into dye molecules with different fluorescent emission maximum wavelengths. This strategy facilitates the storage and reading of information at higher rates than those in previous studies.

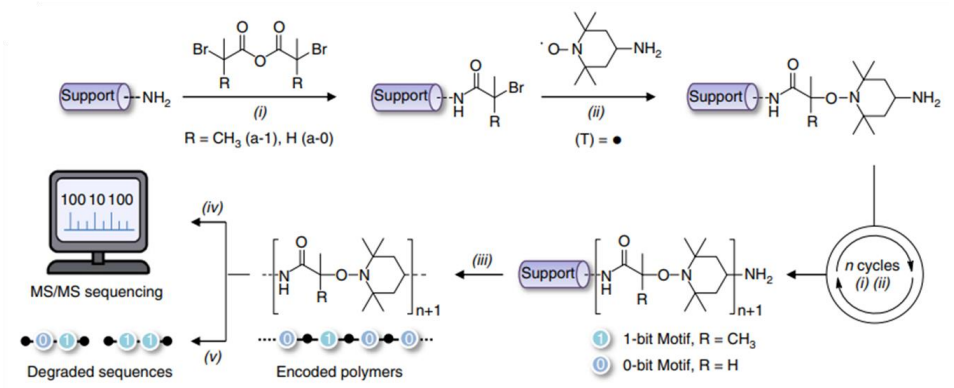


Figure 1-7. General concept of the synthesis of oligo(alkoxyamine amide)s via solid phase synthesis. A TEMPO spacer between primary amine allows easy decoding and erasing of polymers.[74]

Sequence-defined synthetic polymers have attracted significant attention as an alternative to synthetic DNA as information storage media because of the low cost of monomers, higher information density, and stability of synthesized polymer.[73] The most widely used writing method to synthesize sequence-defined polymers is stepwise iterative synthesis because of their superiority in terms of control over the monomer order. Lutz et al. (2015)[74] introduced sequence-encodable polymers resulting from the repetition of two chemoselective reactions: the reaction of a primary amine with a bromo-functionalized anhydride, and the radical coupling of a carbon-centered radical with a nitroxide (Figure 1-7). The two different amino-TEMPO comonomers serve as one-bit information representing 0 or 1 in the binary

code, according to the methyl or dimethyl moieties. The low-energy homolytic cleavage attributed to the TEMPO spacer enables the binary-coded polymers to be easily decoded, using tandem mass spectrometry, and destroys information via thermal degradation.

To store large amounts of information, several strategies have been developed.[75] The most facile approach to increase the information density of uniform polymers is employing numerous monomers with diverse functionalities.[76] When 2^n different monomers are used for encoding, a storage density is n bits per monomer. For example, the maximum storage density of DNA, consisting of four different nucleotides, is 2 bits per nucleotide. Therefore, the amount of information stored in a pentamer constituting 15 different monomers is equivalent to that stored in a 16-mer synthesized using two different monomers. Du Prez et al. (2018)[77] used multifunctional sequence-defined macromolecules for chemical data storage via the synthesis of amide-urethane oligomers using an automated protecting-group-free two-step iterative method based on thiolactone chemistry. A 33 × 33 QR code was encoded in 71 sequence-defined oligomers, using a library of 15 acrylate monomers.

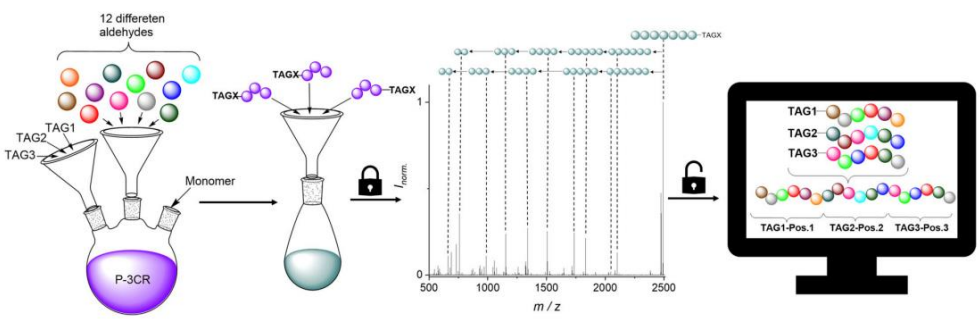


Figure 1-8. Concept of the decoding mixtures in sequence-defined oligomers to increase data storage by introducing distinguishable molecular tags.[78]

However, because of the complexity of the resulting mass spectra, every oligomer had to be separately analyzed, implying that every macromolecule has to be stored in separate vials, which in turn hindered the practicality of the approach to increase information density. Therefore, reading mixtures of sequence-defined macromolecules is a more efficient approach than analyzing isolated oligomers to increase data storage capacity. Meier et al. (2020)[78] introduced molecular tags suitable for unambiguous identification and distinction between different oligomers (Figure 1-8). Twelve different sequence-defined tetramers and three hexamers were synthesized via iterative Passerini three-component reactions and subsequent deprotection steps. Consequently, three different halogenated mass markers were incorporated into the oligomers, providing the molecule with a characteristic isotopic pattern. The oligomers were distinguished using electrospray ionization mass spectrometry (ESI-MS), and specific isotopic patterns, along with the sequence of each molecule, were decoded by tandem ESI-MS/MS.

1.5 Decoding methods of sequence-defined polymers

It is essential to analyze the monomer sequence to utilize sequence-defined polymers as information storage media. Universal techniques for molecular sequencing include electrospray ionization (ESI) and matrix-assisted laser desorption/ionization time-of-flight tandem mass spectrometry.[79] The uniform macromolecules are ionized and separated based on their mass-to-charge ratio, followed by additional fragmentation, separation, and detection. The resulting fragmentation pattern can be used to reconstruct the precursor ions and the ultimate polymer sequence. The obtained fragmentation pattern strongly depends on the nature of the polymer backbone.[80] Lutz et al. are constantly studying data storage in

synthetic macromolecules, and they developed sequencing methods using tandem mass spectrometry. Recently, they achieved MS sequencing of long-coded polymers through careful macromolecular design.[81] The sequence-coded polymers prepared by automated phosphoramidite chemistry comprised phosphoramidite monomers representing bit information, mass tags, and cleavable spacers. Under collision-induced dissociation conditions, the weak alkoxyamine bonds in the cleavable spacer were selectively cleaved first. Subsequently, each fragment, which could be distinguished owing to mass tags, was deciphered by C-O bond cleavage of the phosphodiester linkage. Consequently, large quantities of information stored in long polymers could be decoded.

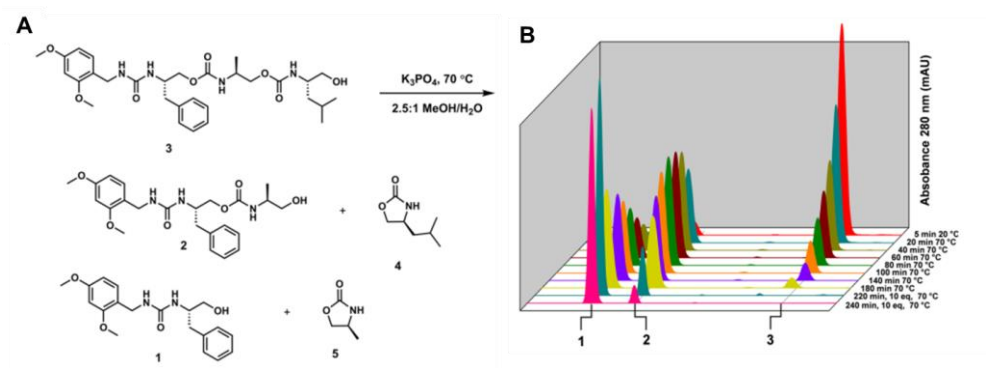


Figure 1-9. (A) Scheme for the self-immolation of sequence-defined oligourethanes. (B) LC-MS traces at the denoted time intervals.[82]

Anslyn et al. (2020)[82] proposed an elegant sequencing technique without tandem mass spectrometry (MS/MS). They designed a sequence-defined oligourethane, wherein a terminal alcohol induces favorable intramolecular cyclization that releases 2-oxazolidinone and a new terminal alcohol, which can

generate successive cyclizations (Figure 1-9). Sequence-defined oligourethanes were subjected to self-immolation conditions with K_3PO_4 at 70 °C. Consequently, the monomer sequence could be deciphered by analyzing each iteration of the truncated oligomers using a single LC/MS trace.

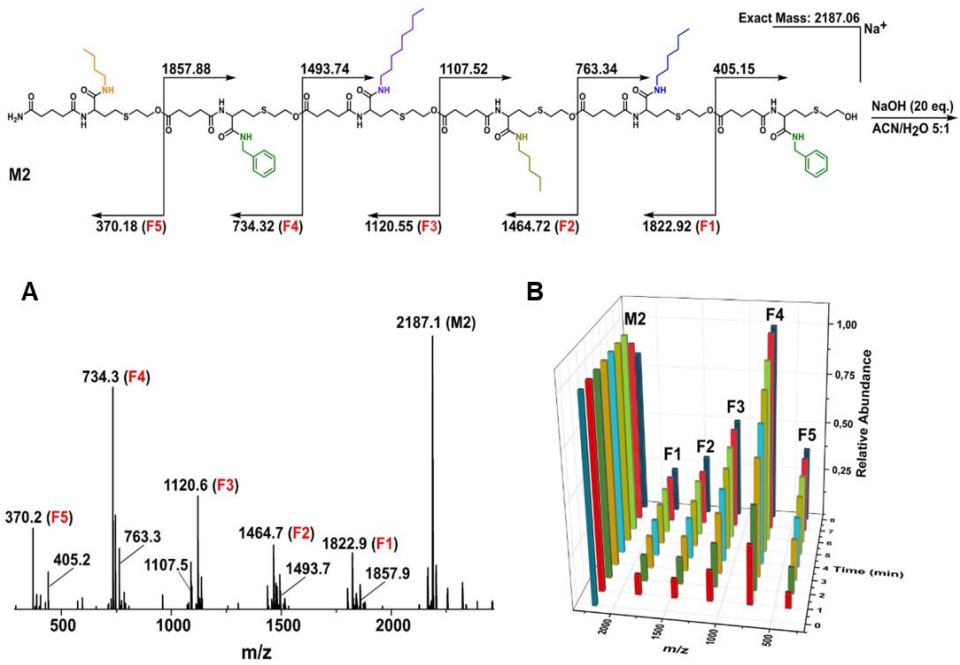


Figure 1-10. (A) ESI-MS analysis after hydrolysis of sequence-defined oligoesters at 2 min. (B) ESI-MS traces at the denoted time intervals.[83]

Du Prez et al. (2022)[83] reported a similar strategy, which relies on the random cleavage of ester units (Figure 1-10). The ester moiety in the backbone of a sequence-defined oligo(thioether ester)s was readily hydrolyzed in a basic environment. Therefore, the macromolecules were subjected to hydrolysis using sodium hydroxide,

followed by quenching by the addition of acetic acid. Analysis of the ESI-MS spectrum of the treated fragments enabled the determination of the monomer sequence. Additionally, introducing a long-wavelength chromophore at the chain end enabled LC-MS sequencing by analyzing a single series of truncated structures.

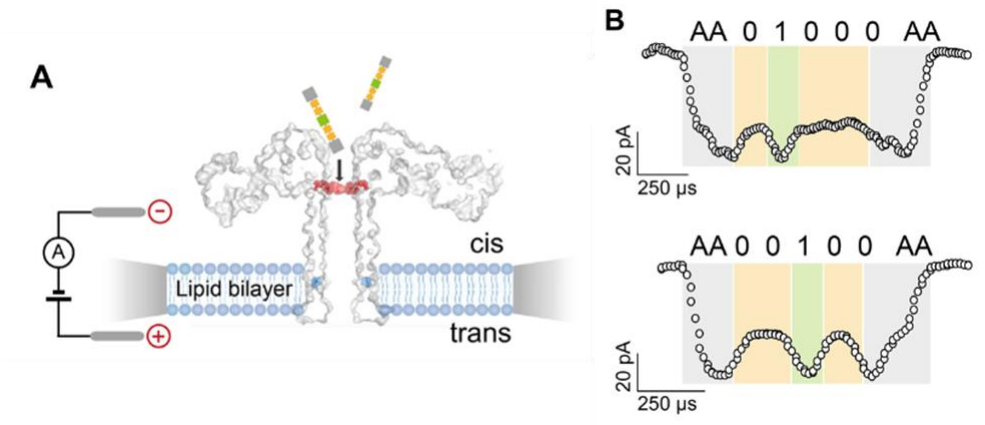


Figure 1-11. (A) Illustration of aerolysin nanopore sequencing setup. (B) Translocation events of the uniform polymers with different sequence.[84]

Unlike in the case of DNA, which can be replicated, unlimited decoding of sequence-defined polymers is impossible. This is because, upon analysis by degradation sequencing, such as tandem mass spectrometry, the samples are consumed. Peraro et al. (2020)[84] demonstrated a noble approach by employing nanopore sequencing, which was developed for DNA sequencing (Figure 1-11). Monitoring the fluctuation of the ion current arising from the translocation of bio-hybrid macromolecules into narrow aerolysin mutant channels allows sequencing with single-bit resolution. Moreover, decoding of mixed samples cannot be attained

without compromising information density.

1.6 Summary of thesis

The following four chapters describe different research topics, and each chapter is in the form of a publication. The authors wholly or partially contributed to the experiments and characterizations in this study. Portions of this thesis have been published and submitted for publication.

Chapter 2: "High-density information storage in an absolutely defined periodic sequence of monodisperse copolyesters" Lee, J. M.; Koo, M. B.; Lee, S. W.; Kwon, J.; Shim, Y. H.; Kim, So. Y.; Kim, K. T. *Nat. Commun.* **2020**, *11*, 56.

Chapter 3: "Semiautomated synthesis of sequence-defined polymers for information storage" Lee, J. M.; Kwon, J.; Lee, S. J.; Jang, H.; Kim. D.; Song, J.; Kim, K. T. *Sci. Adv.* **2022**, *8*, eabl8614.

Chapter 4: "Nondestructive sequencing of enantiopure oligoesters by nuclear magnetic resonance spectroscopy. Lee, J. M.; Jang, H.; Lee, S. W.; Kim, K. T. J. Am. *JACS Au in revision*.
L.J.M. and J.H. contributed equally to this work.

Chapter 5: "Synthesis of enantiomeric ω-substituted hydroxyalkanoates from terminal epoxides and alkenes: building blocks for discrete and sequence-defined polyesters" Kim. D.; Lee, J. M.; Lee, S. W.; Kim, K. T. *Macromolecules Accepted*.
K.D. and L.J.M. contributed equally to this work.

1.7 References

1. Staudinger, H.; Ochiai, E. Über hochpolymere Verbindungen. *Z. Phys. Chem.* **1932**, *158*, 35–55.
2. Seymour, R. B. Polymers are everywhere. *J. Chem. Educ.* **1988**, *65*, 327–334.
3. Feldman, D. Polymer history. *Des. Monomers Polym.* **2008**, *11*, 1–15.
4. Matyjaszewski, K. Architecturally Complex Polymers with Controlled Heterogeneity. *Science* **2011**, *333*, 1104–1105.
5. Grubbs, R. B.; Grubbs, R. H. 50th Anniversary Perspective: Living Polymerization-Emphasizing the Molecule in Macromolecules. *Macromolecules* **2017**, *50*, 6979–6997.
6. Matyjaszewski, K.; Tsarevsky, N. V. Macromolecular engineering by atom transfer radical polymerization. *J. Am. Chem. Soc.* **2014**, *136*, 6513–6533.
7. Lodge, T. P. Celebrating 50 Years of Macromolecules. *Macromolecules* **2017**, *50*, 9525–9527.
8. Odian, G. (2004) *Principle of polymerization.* (4th ed.). Wiley.
9. Lutz, J.-F. Aperiodic copolymers. *ACS Macro Lett.* **2014**, *3*, 1020–1023.
10. Lutz, J.-F.; Lehn, J.-M.; Meijer, E. W.; Matyjaszewski, K. From precision polymers to complex materials and systems. *Nat. Rev. Mater.* **2016**, *1*, 16024.
11. Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-controlled polymers. *Science* **2013**, *341*, 1238149.
12. Ouchi, M.; Badi, N.; Lutz, J.-F.; Sawamoto, M. Single-chain technology using discrete synthetic macromolecules. *Nat. Chem.* **2011**, *3*, 917–924.
13. ten Brummelhuis, N. Controlling monomer-sequence using supramolecular templates. *Polym. Chem.* **2015**, *6*, 654–667.
14. Satoh, K.; Matsuda, M.; Nagai, K.; Kamigaito, M. AAB-sequence living radical chain copolymerization of naturally occurring limonene with maleimide: an end-to-end sequence-regulated copolymer. *J. Am. Chem. Soc.* **2010**, *132*, 10003–10005.
15. Satoh, K.; Ozawa, S.; Mizutani, M.; Nagai, K.; Kamigaito, M. Sequence-regulated vinyl copolymers by metal-catalysed step-growth radical

polymerization. *Nat. Commun.* **2010**, *1*, 1–6.

16. Li, J.; Rothstein, S. N.; Little, S. R.; Edenborn, H. M.; Meyer, T. Y. The effect of monomer order on the hydrolysis of biodegradable poly (lactic-co-glycolic acid) repeating sequence copolymers. *J. Am. Chem. Soc.* **2012**, *134*, 16352–16359.
17. Zhang, J.; Matta, M. E.; Hillmyer, M. A. Synthesis of sequence-specific vinyl copolymers by regioselective ROMP of multiply substituted cyclooctenes. *ACS Macro Lett.* **2012**, *1*, 1383–1387.
18. Nowalk, J. A.; Fang, C.; Short, A. L.; Weiss, R. M.; Swisher, J. H.; Liu, P.; Meyer, T. Y. Sequence-controlled polymers through entropy-driven ring-opening metathesis polymerization: theory, molecular weight control, and monomer design. *J. Am. Chem. Soc.* **2019**, *141*, 5741–5752.
19. Xu, J. Single unit monomer insertion: a versatile platform for molecular engineering through radical addition reactions and polymerization. *Macromolecules* **2019**, *52*, 9068–9093.
20. Xu, J.; Fu, C.; Shanmugam, S.; Hawker, C. J.; Moad, G.; Boyer, C. Synthesis of Discrete Oligomers by Sequential PET-RAFT Single-Unit Monomer Insertion. *Angew. Chem. Int. Ed.* **2017**, *56*, 8376–8383.
21. Nerantzaki, M.; Lutz, J.-F. Multistep Growth "Polymerizations". *Macromol. Chem. Phys.* **2021**, 2100368.
22. Solleder, S. C.; Schneider, R. V.; Wetzel, K. S.; Boukis, A. C.; Meier, M. A. Recent progress in the design of monodisperse, sequence-defined macromolecules. *Macromol. Rapid Commun.* **2017**, *38*, 1600711.
23. Merrifield, R. B. Solid phase synthesis (Nobel lecture). *Angew. Chem. Int. Ed. Engl.* **1985**, *24*, 799–810.
24. Gravert, D. J.; Janda, K. D. Organic synthesis on soluble polymer supports: liquid-phase methodologies. *Chem. Rev.* **1997**, *97*, 489–510.
25. Takizawa, K.; Tang, C.; Hawker, C. J. Molecularly defined caprolactone oligomers and polymers: synthesis and characterization. *J. Am. Chem. Soc.* **2008**, *130*, 1718–1726.
26. Grayson, S. M.; Frechet, J. M. Convergent dendrons and dendrimers: from

synthesis to applications. *Chem. Rev.* **2001**, *101*, 3819–3868.

27. Tomalia, D. A.; Fréchet, J. M. Discovery of dendrimers and dendritic polymers: a brief historical perspective. *J. Polym. Sci. A: Polym. Chem.* **2002**, *40*, 2719–2728.
28. Aksakal, R.; Mertens, C.; Soete, M.; Badi, N.; Du Prez, F. Applications of Discrete Synthetic Macromolecules in Life and Materials Science: Recent and Future Trends. *Adv. Sci.* **2021**, *8*, 2004038.
29. Colquhoun, H.; Lutz, J.-F. Information-containing macromolecules. *Nat. Chem.* 2014, 6, 455–456.
30. Rutten, M. G. T. A.; Vaandrager, F. W.; Elemans, J. A. A. W.; Nolte, R. J. M. Encoding information into polymers. *Nat. Rev. Chem.* **2018**, *2*, 365–381.
31. Boukis, A. C.; Reiter, K.; Frölich, M.; Hofheinz, D.; Meier, M. A. Multicomponent reactions provide key molecules for secret communication. *Nat. commun.* **2018**, *9*, 1–10.
32. Holloway, J. O.; Van Lijsebetten, F.; Badi, N.; Houck, H. A.; Du Prez, F. E. From sequence-defined macromolecules to macromolecular pin codes. *Adv. Sci.* **2020**, *7*, 1903698.
33. Gunay, U. S.; Petit, B. E.; Karamessini, D.; Al Ouahabi, A.; Amalian, J. A.; Chendo, C.; Bouquey, M.; Gigmes, D.; Charles, L.; Lutz, J.-F. Chemoselective synthesis of uniform sequence-coded polyurethanes and their use as molecular tags. *Chem* **2016**, *1*, 114–126.
34. Merrifield, R. B. Solid phase peptide synthesis. I. the synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154.
35. Al Ouahabi, A.; Charles, L.; Lutz, J.-F. Synthesis of non-natural sequence-encoded polymers using phosphoramidite chemistry. *J. Am. Chem. Soc.* **2015**, *137*, 5629–5635.
36. Al Ouahabi, A.; Kotera, M.; Charles, L.; Lutz, J.-F. Synthesis of monodisperse sequence-coded polymers with chain lengths above DP100. *ACS Macro Lett.* **2015**, *4*, 1077–1080.
37. Espeel, P.; Carrette, L. L.; Bury, K.; Capenberghs, S.; Martins, J. C.; Du Prez, F.

E.; Madder, A. Multifunctionalized sequence-defined oligomers from a single building block. *Angew. Chem. Int. Ed.* **2013**, *52*, 13261–13264.

38. Martens, S.; Van den Begin, J.; Madder, A.; Du Prez, F. E.; Espeel, P. Automated Synthesis of Monodisperse Oligomers, Featuring Sequence Control and Tailored Functionalization. *J. Am. Chem. Soc.* **2016**, *138*, 14182–14185.
39. Porel, M.; Alabi, C. A. Sequence-defined polymers via orthogonal allyl acrylamide building blocks. *J. Am. Chem. Soc.* **2014**, *136*, 13162–13165.
40. Solleder, S. C.; Zengel, D.; Wetzel, K. S.; Meier, M. A. R. A scalable and high-yield strategy for the synthesis of sequence-defined macromolecules. *Angew. Chem. Int. Ed.* **2016**, *55*, 1204–1207.
41. Dong, R.; Liu, R.; Gaffney, P. R. J.; Schaepertoens, M.; Marchetti, P.; Wiliams, C. M.; Chen, R.; Livingston, A. G. Sequence-defined multifunctional polyethers via liquid-phase synthesis with molecular sieving. *Nat. Chem.* **2019**, *11*, 136–145.
42. Paynter, O. I.; Simmonds, D. J.; Whiting, M. C. The synthesis of long-chain unbranched aliphatic compounds by molecular doubling. *J. Chem. Soc., Chem. Commun.* **1982**, 1165–1166.
43. Igner, E.; Paynter, O. I.; Simmonds, D. J.; Whiting, M. C. Studies on the synthesis of linear aliphatic compounds. Part 2. The realisation of a strategy for repeated molecular doubling. *J. Chem. Soc., Perkin Trans. 1* **1987**, 2447–2454.
44. Binauld, S.; Damiron, D.; Connal, L. A.; Hawker, C. J.; Drockenmuller, E. Precise synthesis of molecularly defined oligomers and polymers by orthogonal iterative divergent/convergent approaches. *Macromol. Rapid commun.* **2011**, *32*, 147–168.
45. Zhang, J.; Moore, J. S.; Xu, Z.; Aguirre, R. A. Nanoarchitectures. 1. Controlled synthesis of phenylacetylene sequences. *J. Am. Chem. Soc.* **1992**, *114*, 2273–2274.
46. Sakurai, S. I.; Goto, H.; Yashima, E. Synthesis and chiroptical properties of optically active, regioregular oligothiophenes. *Org. Lett.* **2001**, *3*, 2379–2382.
47. Read, M. W.; Escobedo, J. O.; Willis, D. M.; Beck, P. A.; Strongin, R. M.

Convenient iterative synthesis of an octameric tetracarboxylate-functionalized oligophenylene rod with divergent end groups. *Org. Lett.* **2000**, *2*, 3201–3204.

48. Sadighi, J. P.; Singer, R. A.; Buchwald, S. L. Palladium-catalyzed synthesis of monodisperse, controlled-length, and functionalized oligoanilines. *J. Am. Chem. Soc.* **1998**, *120*, 4960–4976.
49. Bidd, I.; Whiting, M. C. The synthesis of pure n-paraffins with chain-lengths between one and four hundred. *J. Chem. Soc., Chem. Commun.* **1985**, 543–544.
50. Hawker, C. J.; Malmström, E. E.; Frank, C. W.; Kampf, J. P. Exact linear analogs of dendritic polyether macromolecules: design, synthesis, and unique properties. *J. Am. Chem. Soc.* **1997**, *119*, 9903–9904.
51. Lengweiler, U. D.; Fritz, M. G.; Seebach, D. Synthese monodisperser linearer und cyclischer Oligomere der (R)-3-Hydroxybuttersaure mit bis zu 128 Einheiten. *Helv. Chim. Acta* **1996**, *79*, 670–701.
52. Duan, S.; Yang, X.; Yang, Z.; Liu, Y.; Shi, Q.; Yang, Z.; Wu, H.; Han, Y.; Wang, Y.; Shen, H.; Huang, Z.; Dong, X.-H. Zhang, Z. A Versatile Synthetic Platform for Discrete Oligo-and Polyesters Based on Optimized Protective Groups Via Iterative Exponential Growth. *Macromolecules* **2021**, *54*, 10830–10837.
53. Takizawa, K.; Tang, C.; Hawker, C. J. Molecularly defined caprolactone oligomers and polymers: synthesis and characterization. *J. Am. Chem. Soc.* **2008**, *130*, 1718–1726.
54. Takizawa, K.; Nulwala, H.; Hu. J.; Yoshinaga, K.; Hawker, C. J. Molecularly defined (L)-lactic acid oligomers and polymers: synthesis and characterization. *J. Polym. Sci. A: Polym. Chem.* **2008**, *46*, 5877–5990.
55. Barnes, J. C.; Ehrlich, D. J. C.; Gao, A. X.; Leibfarth, F. A.; Jiang, Y.; Zhou, E.; Jamison, T. F.; Johnson, J. A. Iterative exponential growth of stereo- and sequence-controlled polymers. *Nat. Chem.* **2015**, *7*, 810–815.
56. Leibfarth, F. A.; Johnson, J. A.; Jamison, T. F. Scalable synthesis of sequence-defined, unimolecular macromolecules by Flow-IEG. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 10617–10622.
57. Jiang. Y.; Golder, M. R.; Nguyen, H. V.-T.; Wang, Y.; Zhong, M.; Barnes, J. C.;

Ehrlich, D. J. C.; Johnson, J. A. Iterative exponential growth synthesis and assembly of uniform diblock copolymers. *J. Am. Chem. Soc.* **2016**, *138*, 9369–9372.

58. Golder, M. R.; Jiang, Y.; Teichen, P. E.; Nguyen, H. V. T.; Wang, W.; Milos, N.; Freedman, S. A.; Willard, A. P.; Johnson, J. A. Stereochemical sequence dictates unimolecular diblock copolymer assembly. *J. Am. Chem. Soc.* **2018**, *140*, 1596–1599.
59. Nguyen, H. V. T.; Jiang, Y.; Mohapatra, S.; Wang, W.; Barnes, J. C.; Oldenhuis, N. J.; Chen, K. K.; Axelrod, S.; Chen, Q.; Golder, M. R.; Young, K.; Suvlu, D.; Shen, Y.; Willard, A. P.; Hore, M. J. A.; Gomez-Bombarelli, R.; Johnson, J. A. Bottlebrush polymers with flexible enantiomeric side chains display differential biological properties. *Nat. Chem.* **2022**, *14*, 85–93.
60. Lloyd, S. Ultimate physical limits to computation. *Nature* **2000**, *406*, 1047–1054.
61. Salahuddin, S.; Ni, K.; Datta, S. The era of hyper-scaling in electronics. *Nat. Electron.* **2018**, *1*, 442–450.
62. Zhirnov, V.; Zadegan, R. M.; Sandhu, G. S.; Church, G. M.; Hughes, W. L. Nucleic acid memory. *Nat. Mater.* **2016**, *15*, 366–370.
63. Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; LeProust, E. M.; Sipos, B.; Birney, E. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **2013**, *494*, 77–80.
64. Grass, R. N.; Heckel, R.; Puddu, M.; Paunescu, D.; Stark, W. J. Robust chemical preservation of digital information on DNA in silica with error-correcting codes. *Angew. Chem. Int. Ed.* **2015**, *54*, 2552–2555.
65. Davis, J. Microvenus. *Art Journal* **1996**, *55*, 70–74.
66. Church, G. M.; Gao, Y.; Kosuri, S. Next-generation digital information storage in DNA. *Science* **2012**, *337*, 1628–1628.
67. Kosuri, S.; Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **2014**, *11*, 499–507.
68. Shendure, J.; Balasubramanian, S.; Church, G. M.; Gilbert, W.; Rogers, J.; Schloss, J. A.; Waterston, R. H. DNA sequencing at 40: past, present and future.

Nature **2017**, *550*, 345–353.

69. Green, J. E.; Choi, J. W.; Boukai, A.; Bunimovich, Y.; Johnston-Halperin, E.; DeIonno, E.; Luo, Y.; Sheriff, B. A.; Xu, K.; Shin, Y. S.; Tseng, H.-R.; Stoddart, J. F.; Heath, J. R. A 160-kilobit molecular electronic memory patterned at 1011 bits per square centimetre. *Nature* **2007**, *445*, 414–417.
70. Arcadia, C. E.; Kennedy, E.; Geiser, J.; Dombroski, A.; Oakley, K.; Chen, S.- L.; Sprague, L.; Ozmen, M.; Sello, J.; Weber, P. M.; Reda, S.; Rose, C.; Kim, E.; Rubenstein, B. M.; Rosenstein, J. K. Multicomponent molecular memory. *Nat. Commun.* **2020**, *11*, 1–8.
71. Cafferty, B. J.; Ten, A. S.; Fink, M. J.; Morey, S.; Preston, D. J.; Mrksich, M.; Whitesides, G. M. Storage of information using small organic molecules. *ACS Cent. Sci.* **2019**, *5*, 911–916.
72. Nagarkar, A. A.; Root, S. E.; Fink, M. J.; Ten, A. S.; Cafferty, B. J.; Richardson, D. S.; Mrksich, M.; Whitesides, G. M. Storing and Reading Information in Mixtures of Fluorescent Molecules. *ACS Cent. Sci.* **2021**, *7*, 1728–1735.
73. Charles, L.; Lutz, J.-F., Design of Abiological Digital Poly (phosphodiester)s. *Acc. Chem. Res.* **2021**, *54*, 1791–1800.
74. Roy, R. K.; Meszynska, A.; Laure, C.; Verchin, C.; Lutz, J.-F. Design and synthesis of digitally encoded polymers that can be decoded and erased. *Nat. Commun.* **2015**, *6*, 7237.
75. Laurent, E.; Amalian, J.-A.; Parmentier, M.; Oswald, L.; Al Ouahabi, A.; Dufour, F.; Launay, K.; Clément, J.-L.; Gigmes, D.; Delsuc, M.-A.; Charles, L.; Lutz, J.-F. High-capacity digital polymers: storing images in single molecules. *Macromolecules* **2020**, *53*, 4022–4029.
76. Boukis, A. C.; Meier, M. A. Data storage in sequence-defined macromolecules via multicomponent reactions. *Eur. Polym. J.* **2018**, *104*, 32–38.
77. Martens, S.; Landuyt, A.; Espeel, P.; Devreese, B.; Dawyndt, P.; Du Prez, F. Multifunctional sequence-defined macromolecules for chemical data storage. *Nat. Commun.* **2018**, *9*, 4451.
78. Frölich, M.; Hofheinz, D.; Meier, M. A. Reading mixtures of uniform sequence-

defined macromolecules to increase data storage capacity. *Commun. Chem.* **2020**, *3*, 184.

79. Zydziak, N.; Konrad, W.; Feist, F.; Afonin, S.; Weidner, S.; Barner-Kowolik, C. Coding and decoding libraries of sequence-defined functional copolymers synthesized via photoligation. *Nat. Commun.* **2016**, *7*, 13672.
80. Wesdemiotis, C.; Solak, N.; Polce, M. J.; Dabney, D. E.; Chaicharoen, K.; Katzenmeyer, B. C. Fragmentation pathways of polymer ions. *Mass Spectrom. Rev.* **2011**, *30*, 523–559.
81. Al Ouahabi, A.; Amalian, J.-A.; Charles, L.; Lutz, J.-F. Mass spectrometry sequencing of long digital polymers facilitated by programmed inter-byte fragmentation. *Nat. Commun.* **2017**, *8*, 976.
82. Dahlhauser, S. D.; Escamilla, P. R.; VandeWalle, A. N.; York, J. T.; Rapagnani, R. M.; Shei, J. S.; Glass, S. A.; Coronado, J. N.; Moor, S. R.; Saunders, D. P.; Anslyn, E. V. Sequencing of sequence-defined oligourethanes via controlled self-immolation. *J. Am. Chem. Soc.* **2020**, *142*, 2744–2749.
83. Soete, M.; Du Prez, F. E. Sequencing of Uniform Multifunctional Oligoesters via Random Chain Cleavages. *Angew. Chem. Int. Ed.* **2022**, e202202819.
84. Cao, C.; Krapp, L. F.; Al Ouahabi, A.; König, N. F.; Cirauqui, N.; Radenovic, A.; Lutz, J.-F.; Peraro, M. D. Aerolysin nanopores decode digital information stored in tailored macromolecular analytes. *Sci. Adv.* **2020**, *6*, eabc2661.

Chapter 2

High-Density Information Storage in an Absolutely Defined Aperiodic Sequence of Monodisperse Copolyester

(L.S.W. and L.H. contributed to the synthesis of PcLs)

2.1 Abstract

Synthesis of a polymer composed of a large discrete number of chemically distinct monomers in an absolutely defined aperiodic sequence remains a challenge in polymer chemistry. The synthesis has largely been limited to oligomers having a limited number of repeating units due to the difficulties associated with the step-by-step addition of individual monomers to achieve high molecular weights. Here we report the copolymers of α-hydroxy acids, poly(phenyllactic-*co*-lactic acid) (PcL) built via the cross-convergent method from four dyads of monomers as constituent units. Our proposed method allows scalable synthesis of sequence-defined PcL in a minimal number of coupling steps from reagents in stoichiometric amounts. Digital information can be stored in an aperiodic sequence of PcL, which can be fully retrieved as binary code by mass spectrometry sequencing. The information storage density (bit/Da) of PcL is 50% higher than DNA, and the storage capacity of PcL can also be increased by adjusting the molecular weight (~ 38 kDa).

2.2 Introduction

Unlike DNAs and proteins, the formation of synthetic long-chain molecules involves statistical uncertainties in terms of the number and sequence of monomers that join during the polymerization.[1] Synthetic polymers without statistical uncertainty in their molecular weights and sequences can store information in their chemical structures, which makes them low-cost alternatives to DNAs as molecular medium for a large-scale storage of digital information.[2,3] However, only polymers and copolymers with narrow molecular weight distributions are produced via the living or controlled polymerization.[4] The composition of the monomers of a

copolymer can be controlled, but the spatial distribution of the monomers within the copolymer is either random or completely segregated. As a result, the scalable synthesis of polymers with absolutely defined molecular weights and sequences remains a long-standing challenge in polymer chemistry.[5–8]

The solid-phase synthesis has been a method to synthesize polymers with defined aperiodic sequences.[9–12] Although sequence-defined polymers with more than 100 repeating units have been synthesized,[13,14] synthesis has largely been limited to sequence-defined oligomers with a limited number of repeating units due to the difficulties associated with the repeated step-by-step addition of individual monomers to achieve high molecular weights.[15–26] The iterative convergent method has been widely adopted for the synthesis of dendritic and linear macromolecules with precisely defined chemical structures.[27–30] In this process, the coupling product becomes a constituent unit for the next iteration of the coupling reaction. Therefore, the number of repeating units grows exponentially as does the molecular weight of the resulting polymer without distribution.[31–37] Although high molecular-weight polymers without molecular weight distribution can be synthesized efficiently by the convergent method, the self-iterative nature of the coupling step hinders the convergent synthesis of monodisperse oligomers and polymers with perfectly defined aperiodic sequences composed of two or more chemically distinct monomers. For this reason, only oligomers and block co-oligomers with repetitive or palindromic sequences have been synthesized by the convergent pathway.[37–39]

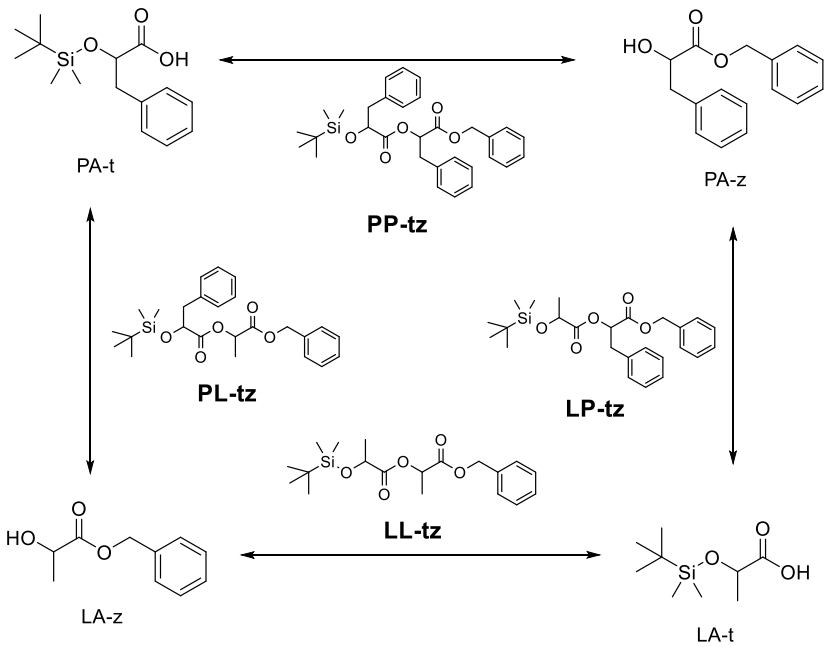
Here we report the synthesis of poly(α-hydroxy acid) (PAH) composed of a large discrete number of monomers in an absolutely defined aperiodic sequence. We demonstrate that digital information can be stored in the aperiodic sequence of poly(phenyllactic-*co*-lactic acid) (PcL), built via the direct translation of binary code to the chemical structure built with four dyads of monomers as constituent units. This

process was named to the 'cross-convergent' method. The stored 64-bit word (SEQUENCE) can be fully retrieved as binary code in a single tandem mass spectrometry sequencing. To expand the ability to read the stored information, we also devised a sequencing method for PcLs with a large number of repeating units (128 units) by measuring the mass of fragments produced by the hydrolysis of the polymer backbone. PcL can be synthesized in a minimal number of coupling steps from reagents in stoichiometric amounts, and its digital information storage density (bit/Da) is 50% higher than that of DNA.[40–43] The molecular weights and production quantities of the reported PAHs are scalable, and no statistical uncertainty is associated with either molecular weight or sequence. The reported PcL could serve as a molecular medium for the storage of digital information. In addition, sequence-defined monodisperse polymers could contribute to the exploration of new properties and functions of polymers arising from the unlimited diversity of their chemical structures.

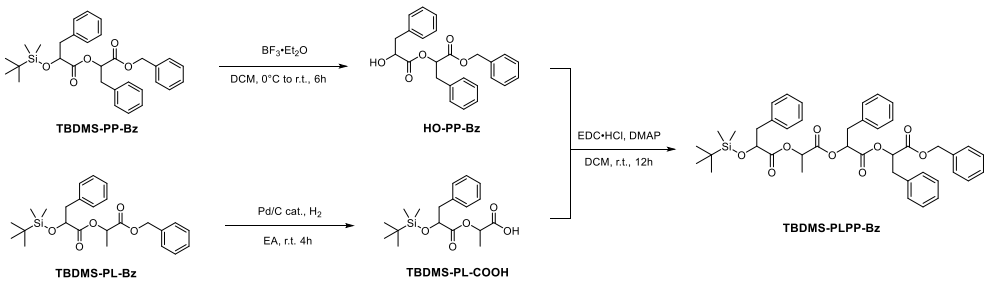
2.3 Results and discussion

Synthesis of PcL. To solve the inability of the iterative convergent method to address an aperiodic sequence of a polymer, we devised a cross-convergent pathway to define an aperiodic sequence of monomers comprised of a copolymer. The chemical sequence of a copolymer consisting of two monomers can be converted directly to a digital binary code. We used a binary code indicating a word ‘SEQUENCE’ to construct a copolymer with an absolutely defined aperiodic sequence of two monomer. In this work, we used *rac*-phenyllactic acid (P) and *rac*-lactic acid (L) as the monomers representing 0 and 1, respectively. Any sequence of a copolyester, poly(phenyllactic-*co*-lactic acid) (PcL), can be expressed as a combination of the permutations of its constituent monomers. For a binary sequence, the permutations

of 0 and 1 give four combinations (00, 01, 10, and 11), which can be translated to the chemical structures composed of two monomers (dyads).



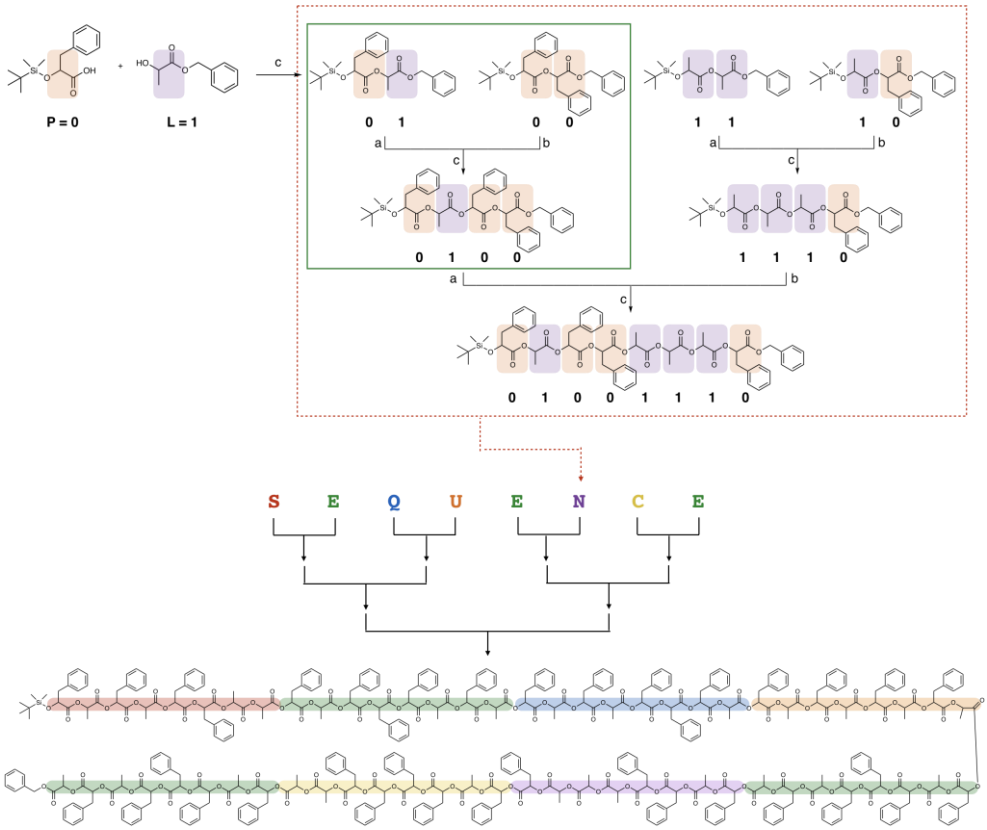
Scheme 2-1. Scheme for preparation of dyads with all permutations.



Scheme 2-2. Synthesis of a sequence-defined tetrad, PLPP, by cross-convergent method.

Four dyads encompassing all possible permutations of the binary sequences (PP, PL, LP, and LL) were synthesized by coupling α-hydroxy acids orthogonally protected by benzyl ester and *tert*-butyldimethylsilyl (TBDMS) ether (Scheme 2-1) and they were characterized by NMR spectroscopy and ESI-MS.[33–35,44] The dyad PP was quantitatively converted to HO-PP-Bz by the selective removal of the TBDMS group with boron trifluoride etherate ($BF_3 \cdot Et_2O$) at room temperature. Hydrogenation of the benzyl ester in the dyad PL with Pd/C generated TBDMS-PL-COOH in high yield (> 95%). The equimolar mixture of TBDMS-PL-COOH and HO-PP-Bz was subsequently coupled to form a tetrad, PLPP by esterification (Scheme 2-2). These orthogonal deprotection and coupling steps constituted one iteration of the convergent growth of PcL (indicated by the green box in Scheme 2-3). The same procedure was repeated to synthesize all five tetrads by cross-converging the required dyads and they were confirmed by NMR spectroscopy. The tetrads were sequentially converged to form 8-bit characters (S, E, Q, U, N, C), and joined to form the full 64-bit word consisting of 8 ASCII characters (SEQUENCE) (Scheme 2-3).

The chromatographic separability of PcL and its constituent units on silica stationary phases diminishes as the molecular weights of the compounds increase.[45] This renders a mixture of constituent units and the desired high molecular weight PcL inseparable (Figure 2-1C). However, the difference between the molecular weight of a PcL and its constituent units persists through all iterations of convergent coupling due to the exponential growth of the molecular weight of the coupled product (Figure 2-1E). Therefore, the monodisperse PcLs with greater than 16 repeating units were purified by preparative size-exclusion chromatography (prep-SEC) in a recycling mode with a loading of up to 1 g per separation. This purification method allowed us to obtain monodisperse PcLs in high yield with stoichiometric



Scheme 2-3. Schematic illustration of the cross-convergent strategy to synthesize PcL with a sequence corresponding to the binary code for the 64-bit word SEQUENCE.

amounts of reagents in a scalable manner with regard to both molecular weight (> 38000 Da) and quantity (> 1 g). This purification method can also be applied to any set of monodisperse polymers with a difference in a hydrodynamic volume, for example, linear and cyclic polymers having the same chemical composition and molecular weight as the intramolecular cyclization of a polymer chain reduces the hydrodynamic volume, which translates to a lower molecular weight on SEC.[46]

Taking the redundancy of the tetrads and the frequency of the letter E into account, 18 convergent coupling steps with stoichiometric amounts of reagents were required to synthesize sequence-defined PcL from four dyad constituent units. The successful synthesis of monodisperse PcL storing a 128-bit binary code was confirmed by ^{1}H and ^{13}C NMR, GPC and MALDI-TOF MS (Figure 2-2, Figure 2-3A and 2-3B). Detailed characterization of all of PcLs is described in Appendix (A.2.1). Due to the exponential growth of PcL via the cross-convergent pathway, no deletion errors, accumulation of polymer chains missing terminal residues, were detected by MALDI-TOF MS analysis of PcL composed of 64 or 128 repeating units (Figure 2-3C). Data from each analysis, in particular MALDI-TOF MS, indicated that the synthesized PcLs were monodisperse in molecular weight and free of lower-molecular-weight impurities.

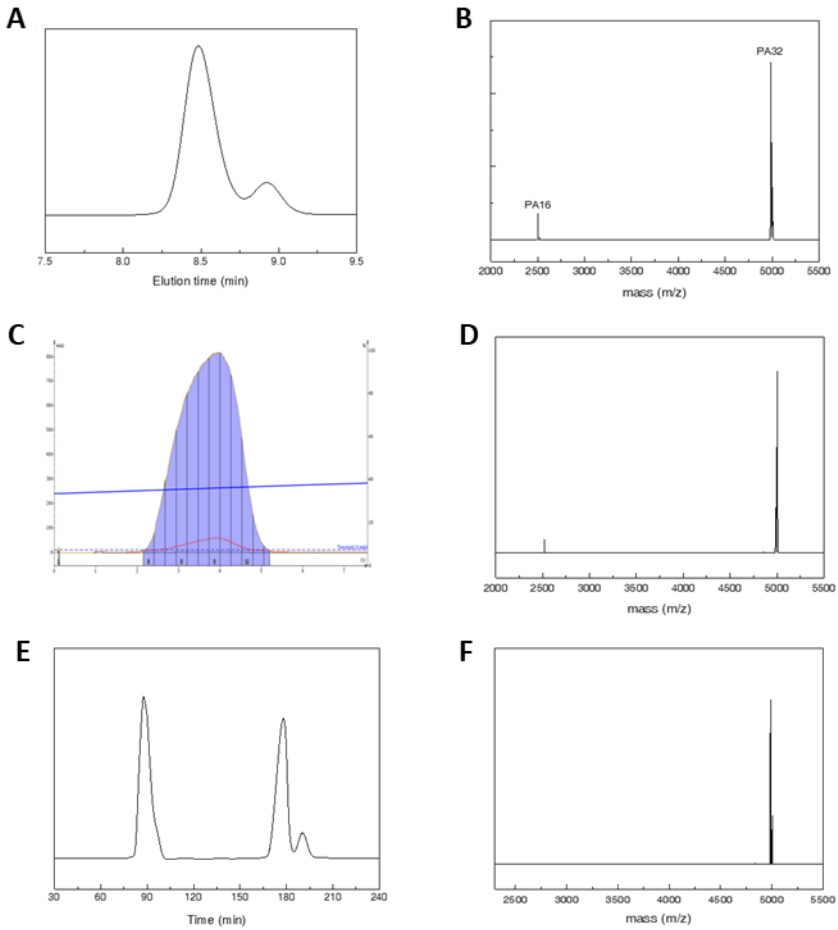


Figure 2-1. (A) Gel permeation chromatography and (B) MALDI-TOF mass spectra of the crude reaction mixture of PA32. (C) UV-detector trace of the automated column chromatography of the crude mixture. (D) MALDI-TOF mass spectrum after purification. (E) SEC traces of the crude mixture during the recycling. (F) MALDI-TOF mass spectrum of PA32 after SEC.

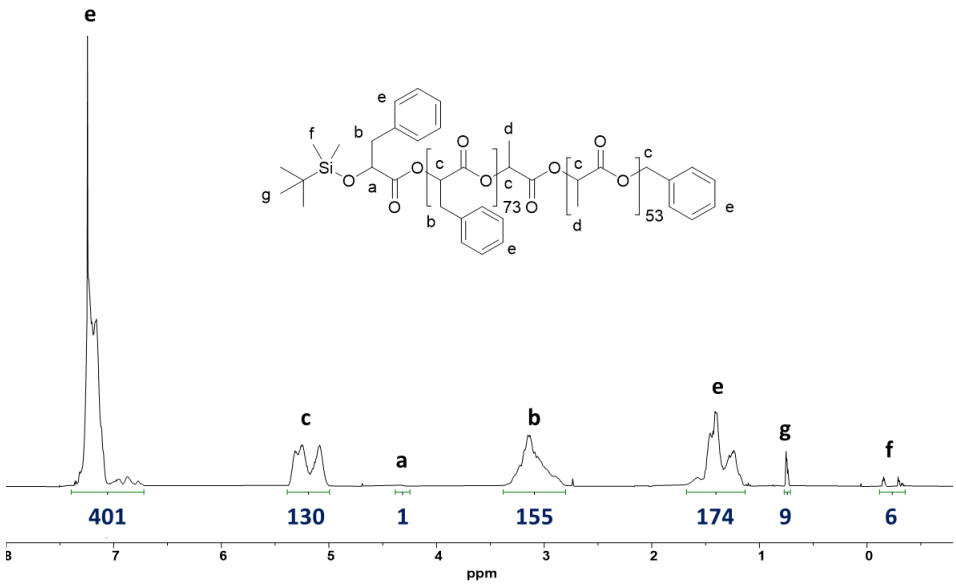


Figure 2-2. ^{1}H NMR spectrum of the 128-bit PcL.

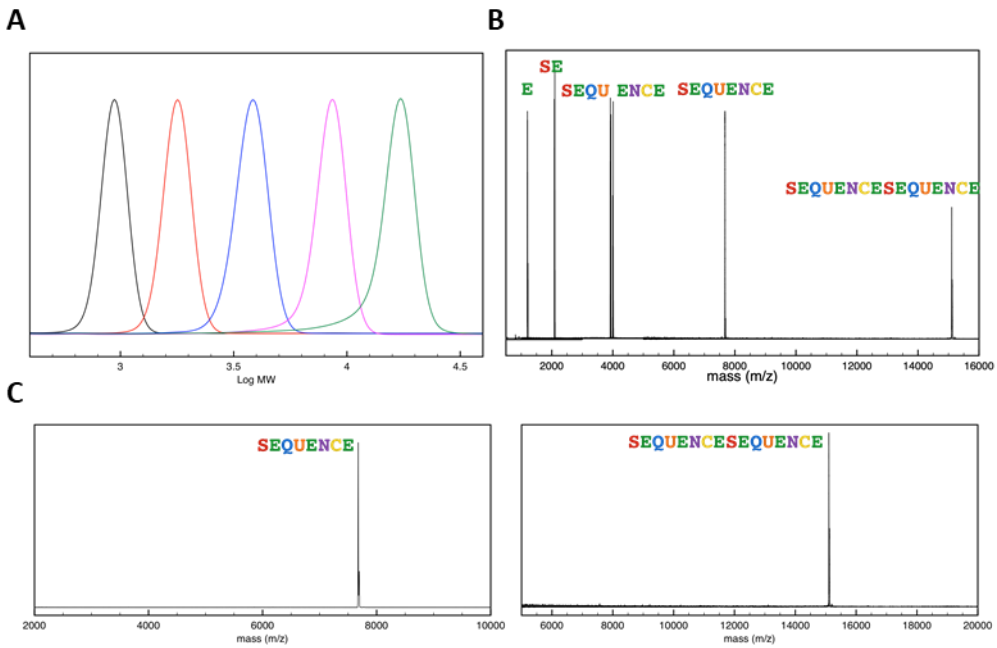


Figure 2-3. (A) Gel-permeation chromatography (GPC) analysis of PcLs with 8, 16, 32, 64, 128 repeating units. (B) Combined MALDI-TOF mass spectra of PcLs encoding the letter (**E**, 1201.5 Da), two-letter word (**SE**, 2082.9 Da), four-letter words (**SEQU**, 3920.4 Da and **ENCE**, 3996.5 Da), 64-bit word (**SEQUENCE**, 7674.5 Da), and 128-bit word (**SEQUENCESEQUENCE**, 15105.3 Da). (C) MALDI-TOF mass spectra of 64- and 128-bit PcL showing no deletion errors or contamination of lower molecular weight fragments.

MALDI-TOF mass sequencing of PcL. The PcL sequence was decoded by a tandem mass technique using MALDI-TOF MS/MS in a positive-ion mode because of its high signal-to-noise ratio and its ability to detect the fragmentation patterns of high molecular weight PcLs (<10 kDa). Fragmentation occurred at the C(α)-O bond of the PcL backbone, which produced a*i* fragments that contain the original alpha group (TBDMS) and a new carboxylic acid terminus and y*i* fragments that contain the original omega group (benzyl ester) and a new alkene terminus.[47] Each fragment series had mass difference by 72 Da (residue L) or 148 Da (residue P) in decreasing order relative to the peak corresponding to $[M+Na]^+$ ion (Figure 2-4A). A single mass spectrum was sufficient for decoding all the information in the PcL (Figure 2-3C). This was because a single mass spectrum showed a series of fragments that can be read in opposite directions, one from the TBDMS-terminus to the benzyl-terminus, and the other from the benzyl-terminus to the TBDMS-terminus. This also enhanced precision, because two retrieved sequences containing the same information could be compared. For example, as the consequence of the analysis of tandem mass spectrum of 8-bit PcL, the monomer sequence from TBDMS-terminus to the benzyl-terminus could be decoded Si-PLPPL by analyzing a series of y*i* fragments and the sequence in opposite direction was PLLLP-Bz, which is determined by a series of a*i* fragments. Because the number of repeating units of the PcL is 8, the sequence could be retrieved by Si-PLPPLLLP-Bz, which indicate a character **N** when the sequence is converted into binary code (Figure 2-4B and Table 2-1). In same principle, the chemical sequence of the 64-bit PcL was directly converted to binary code designating the word **SEQUENCE** by MALDI-TOF/TOF tandem mass sequencing (Figure 2-5 and Table 2-2). Sequencing results of other PcLs are shown in Appendix (A.2.2).

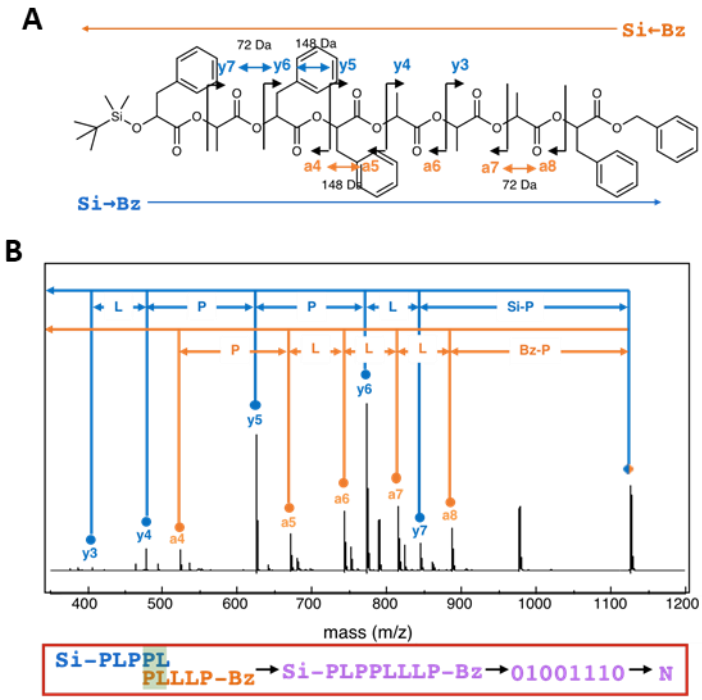


Figure 2-4. (A) Fragmentation of 8-bit PcL under MALDI-TOF MS/MS experiments showing a series of a*i* and y*i* fragments. (B) MALDI-TOF MS/MS spectrum of a PcL, in which 8-bit information corresponding to the letter N was stored.

Table 2-1. Decoding table of PcL (8-bit, **N**)

N (PLPPLLLP)									
Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	1125.4	1125.6			$[M+Na]^+$	1125.4	1125.6		
y7	845.3	845.5	280.1	Si-P	a8	887.3	887.5	238.1	Bz-P
y6	773.3	773.6	71.9	L	a7	815.3	815.6	71.9	L
y5	625.2	625.5	148.1	P	a6	743.3	743.6	72.0	L
y4	477.2	477.4	148.1	P	a5	671.3	671.6	72.0	L
y3	405.1	405.4	72.0	L	a4	523.2	523.5	148.1	P
y2	333.1	333.3	72.1	L	a3	375.2	375.4	148.1	P
y1	261.1	261.3	72.0	L	a2	303.1	303.3	72.1	L

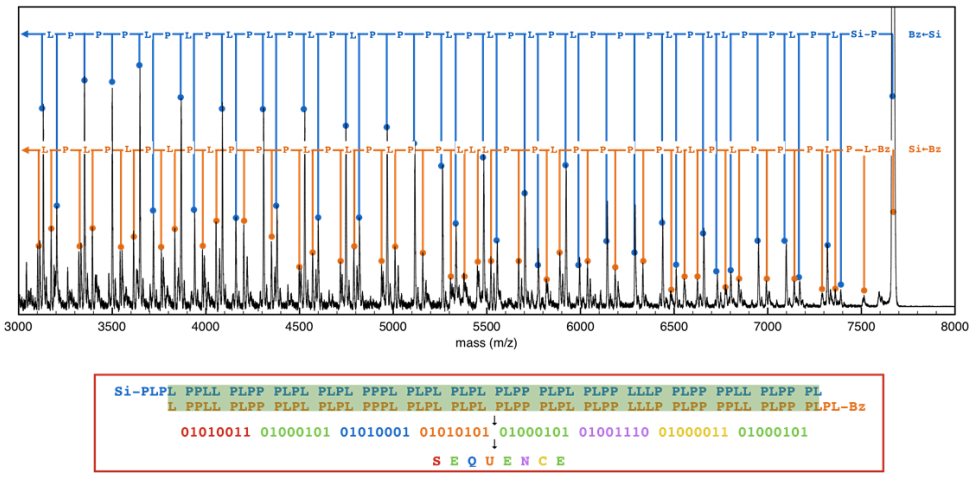


Figure 2-5. Tandem mass sequencing of the entire 64-bit information stored in the PcL. The entire chemical sequence was decoded using a single mass spectrum, followed by conversion to digital information to read the word **SEQUENCE**.

Table 2-2. Decoding table of PcL (64-bit, **SEQUENCE**)

Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	7667.6	7670.2			$[M+Na]^+$	7667.6	7670.2		
y63	7387.5	7394.6	275.6	Si-P	a64	7505.6	7514.3	155.9	Bz-L
y62	7315.5	7322.6	72.0	L	a63	7357.5	7365.3	149.0	P
y61	7167.4	7174.2	148.4	P	a62	7285.5	7294.0	71.3	L
y60	7095.4	7102.1	72.1	L	a61	7137.4	7145.7	148.3	P
y59	6947.4	6955.5	146.6	P	a60	6989.4	6997.2	148.5	P
y58	6799.3	6806.4	149.1	P	a59	6841.3	6849.3	147.9	P
y57	6727.3	6734.3	72.1	L	a58	6769.3	6776.1	73.2	L
y56	6655.3	6662.3	72.0	L	a57	6621.3	6627.5	148.6	P
y55	6507.2	6515.2	147.1	P	a56	6549.2	6556.9	70.6	L
y54	6435.2	6442.5	72.7	L	a55	6477.2	6484.8	72.1	L
y53	6287.1	6295.5	147.0	P	a54	6329.2	6336.1	148.7	P
y52	6139.1	6146.5	149.0	P	a53	6181.1	6189.1	147.0	P
y51	5991.0	5998.5	148.0	P	a52	6033.1	6040.4	148.7	P
y50	5919.0	5926.0	72.5	L	a51	5885.0	5893.3	147.1	P
y49	5771.0	5778.4	147.6	P	a50	5813.0	5820.6	72.7	L
y48	5698.9	5706.6	71.8	L	a49	5664.9	5672.1	148.5	P
y47	5550.9	5558.8	147.8	P	a48	5516.9	5524.6	147.5	P
y46	5478.9	5486.3	72.5	L	a47	5444.9	5453.1	71.5	L

y45	5330.8	5339.3	147.0	P	a46	5372.8	5380	73.1	L
y44	5259.8	5266.0	73.3	L	a45	5300.8	5308.8	71.2	L
y43	5110.7	5118.8	147.2	P	a44	5152.8	5160.8	148.0	P
y42	4962.7	4970.5	148.3	P	a43	5004.7	5012.6	148.2	P
y41	4814.6	4822.7	147.8	P	a42	4932.7	4940.8	71.8	L
y40	4742.6	4750.3	72.4	L	a41	4784.6	4792.5	148.3	P
y39	4594.6	4602.6	147.7	P	a40	4712.6	4720.5	72.0	L
y38	4522.5	4530.3	72.3	L	a39	4564.6	4573.1	147.4	P
y37	4374.5	4382.9	147.4	P	a38	4492.6	4501.0	72.1	L
y36	4302.5	4310.2	72.7	L	a37	4344.5	4352.1	148.9	P
y35	4154.4	4162.7	147.5	P	a36	4196.4	4204.6	147.5	P
y34	4082.4	4090.8	71.9	L	a35	4048.4	4056.5	148.1	P
y33	3934.3	3942.3	148.5	P	a34	3976.4	3984.5	72.0	L
y32	3862.3	3870.0	72.3	L	a33	3828.3	3836.1	148.4	P
y31	3714.3	3722.4	147.6	P	a32	3756.3	3764.3	71.8	L
y30	3642.2	3649.6	72.8	L	a31	3608.2	3615.9	148.4	P
y29	3494.2	3501.5	148.1	P	a30	3536.2	3544.3	71.6	L
y28	3346.1	3353.6	147.9	P	a29	3388.2	3396.3	148.0	P
y27	3198.1	3205.5	148.1	P	a28	3316.2	3323.8	72.5	L
y26	3126.1	3133.4	72.1	L	a27	3168.1	3176.2	147.6	P
y25	2978.0	2985.2	148.2	P	a26	3096.1	3103.4	72.8	L
y24	2906.0	2913.4	71.8	L	a25	2948.0	2955.4	148.0	P
y23	2757.9	2765.5	147.9	P	a24	2876.0	2883.8	71.6	L
y22	2685.9	2693.1	72.4	L	a23	2728.0	2735.4	148.4	P
y21	2537.9	2545.2	147.9	P	a22	2579.9	2587.2	148.2	P
y20	2389.8	2396.8	148.4	P	a21	2430.8	2439.1	148.1	P
y19	2317.8	2324.6	72.2	L	a20	2359.8	2367.0	72.1	L
y18	2245.8	2252.3	72.3	L	a19	2211.8	2218.1	148.9	P
y17	2173.8	2180.4	71.9	L	a18	2139.8	2146.3	71.8	L
y16	2025.7	2031.9	148.5	P	a17	1991.7	1997.9	148.4	P
y15	1877.6	1883.6	148.3	P	a16	1919.7	1926.0	71.9	L
y14	1805.6	1811.8	71.8	L	a15	1771.6	1777.8	148.2	P
y13	1657.6	1663.4	148.4	P	a14	1699.6	1705.6	72.2	L
y12	1509.5	1515.2	148.2	P	a13	1551.6	1557.4	148.2	P
y11	1361.5	1367.0	148.2	P	a12	1403.5	1409.1	148.3	P
y10	1213.4	1218.4	148.6	P	a11	1255.4	1260.5	148.6	P
y9	1141.4	1145.8	72.6	L	a10	1183.4	1188.3	72.2	L
y8	1069.4	1074.1	71.7	L	a9	1035.4	1039.9	148.4	P
y7	921.3	925.5	148.6	P	a8	963.4	967.7	72.2	L
y6	849.3	853.3	72.2	L	a7	891.3	895.5	72.2	L
y5	701.2	704.7	148.6	P	a6	743.3	746.9	148.6	P
y4	553.2	555.9	148.8	P	a5	595.2	598.3	148.6	P
y3	405.1	406.5	149.4	P	a4	523.2	525.7	72.6	L

PcLs with high molecular weights (> 64 repeating units) could not be directly sequenced under our tandem mass condition with MALDI-TOF MS/MS presumably due to the inability to analyze high molecular-weight parent ions by the instrument. To overcome this limitation, we devised a degradative sequencing method using MALDI-TOF mass spectrometry because of higher molecular-weight limit (~ 20 kDa) for the desorption of polymers from the matrix. Poly(α-hydroxy acid) (PAH) could easily be degraded to their constituting monomers via hydrolysis of ester groups in the polymer backbone. The random hydrolysis of ester groups along the polymer backbone would create the fragments of the parent PcL, which could be directly detected by MALDI-TOF mass spectrometry. Therefore, the sequence of monomers in PcL could be decoded by reading the mass difference between adjacent fragments from the mass spectrum.

Utilizing facile degradability of PcL,[48–50] we chemically degraded a 128-bit PcL containing a 16-letter word (**SEQUENCESEQUENCE**) via the hydrolysis of the ester groups in the main chain. In the presence of NaOH (0.3 equivalent to the ester groups), 128-bit PcL was incubated in THF for 30 min at 70 ℃. Generation of low molecular weight fragments resulting from hydrolysis could be confirmed by GPC analysis (Figure 2-6). MALDI-TOF mass spectra of the hydrolyzed PcL showed a series of mass peaks in decreasing order from the mass of the molecular ion (Figure 2-7A). The mass spectrum could be decoded to binary code by reading the mass difference between the peaks in a direction from the TBDMS-terminus to the benzyl terminus. The full sequence except ultimate 8 residues (residue 1 to 120) could be deciphered utilizing a series of MALDI-TOF mass spectra of the degraded PcL covering the entire molecular weight range (1000–16000 Da). The last 8 residues (residue 121 to 128) close to the Bz-terminus of the PcL were subsequently sequenced by a tandem mass spectrum of one of the lower molecular-weight

fragments (x22, m/z of a parent ion = 2703 Da in MALDI-TOF) in order to remove the noise from the matrix molecules in MALDI-TOF mass. The decoded sequence of the 128-bit PcL coincided with the chemical structure of the PcL without any error (Figure 2-7B). The detailed decoding data is represented in Appendix (A.2.3).

The density of information storage, defined as the number of bits per unit mass, was 0.009 bit/Da in PcL. This was 50% higher than the storage density of DNA (0.006 bit/Da).[40–44] Due to the simplicity in chemical structures, our results suggest that the cost of synthesizing sequence-specific PcLs could be substantially lower than the cost of writing information on DNAs. Thus, PcLs could provide an alternative to DNA and molecular media for archival storage of large amounts of information.

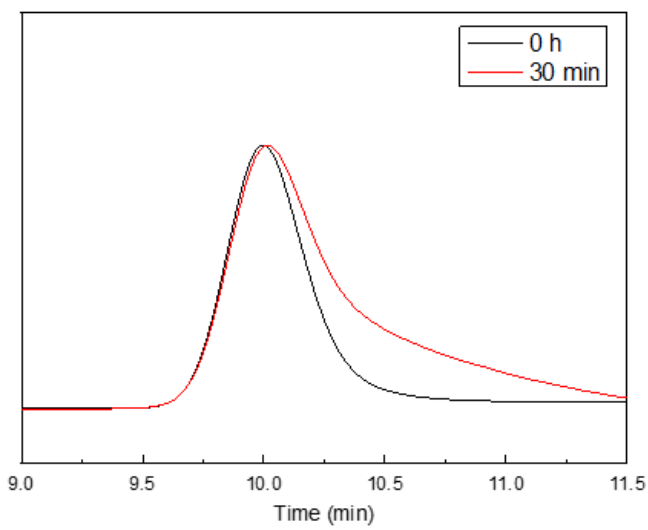


Figure 2-6. GPC result showing hydrolysis of sequence-defined polyester, 128-bit PcL, in basic condition.

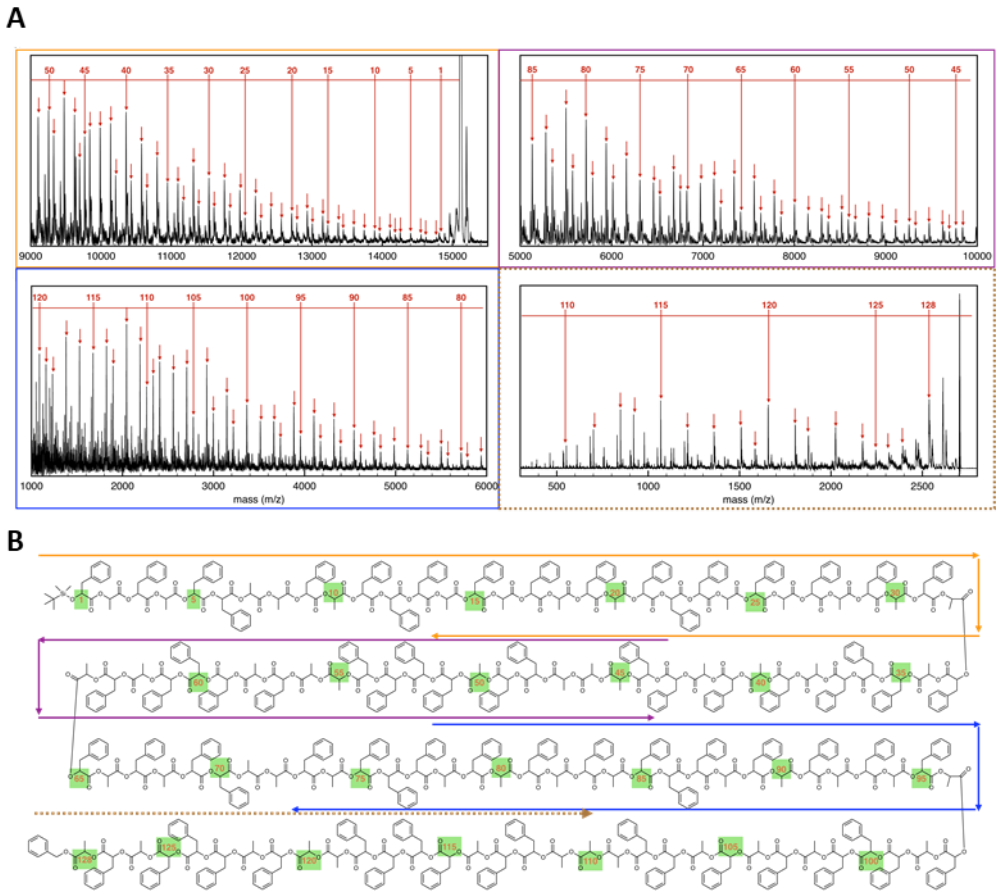


Figure 2-7. (A) A series of MALDI-TOF mass spectra of chemically degraded PcL via hydrolysis. The assigned peaks are marked by arrows. The mass spectra of x*i* fragments were used for MALDI-TOF sequencing. The sequence of the last 8 repeating units at the Bz terminus was decoded by MALDI-TOF MS/MS to avoid the noise from the signal of the matrix molecules. (B) The chemical structure of 128-bit PcL drawn with the decoded sequence. The repeating units are numbered in an increasing order from the first repeating unit (P) at the TBDMS terminus.

Scalability of the convergent synthesis of PAH. The storage of large information in molecular media demands the original bit-information to be divided to a unit size (for example, 64 or 128 bit), so that the divided information can be stored in multiple polymer chains. Each polymer chain is required to store the unit data along with the address information, so that the original information is fully restored by the integration of decoded sequences of all polymer chains. Therefore, one of the requirements for synthetic polymers as information storage media is a sequencing method for polymer chains having a number of repeating units that is sufficiently large to meet the required capacity of information storage.

To demonstrate the scalability of the convergent method in terms of molecular weight, we synthesized monodisperse poly(*rac*-phenyllactic acid) (PA*n*) composed of a large number of repeating units (Figure 2-8A). Here, *n* represents the number of repeating units. Starting with tetramers PA4 (10 g, 12.27 mmol), seven iterations of convergent growth yielded PA256 with molecular weight of 38175 Da (experimentally found at 38191 Da) in an overall yield of 15% (1.07 g, 0.028 mmol) (Figure 2-8C and 2-8D). PA512 with a molecular weight greater than 70 kDa could only be synthesized in low yields (< 5%) even after an extended period of time for the coupling reaction. This could be a result from the difficulties of finding the chain end of high molecular weight PA*n* under the conditions we employed for the esterification. The convergent growth strategy also allowed us to synthesize monodisperse PAHs with any number of repeating units by changing the constituent units in the final coupling step. For example, PA80 and PA96 were synthesized by selecting the constituents that afforded the desired number of repeating units (Figure 2-8B). Our results indicated that the number of repeating units of PAH could be optimized to store information with a desired length and format suitable for a large-scale storage of digital information.

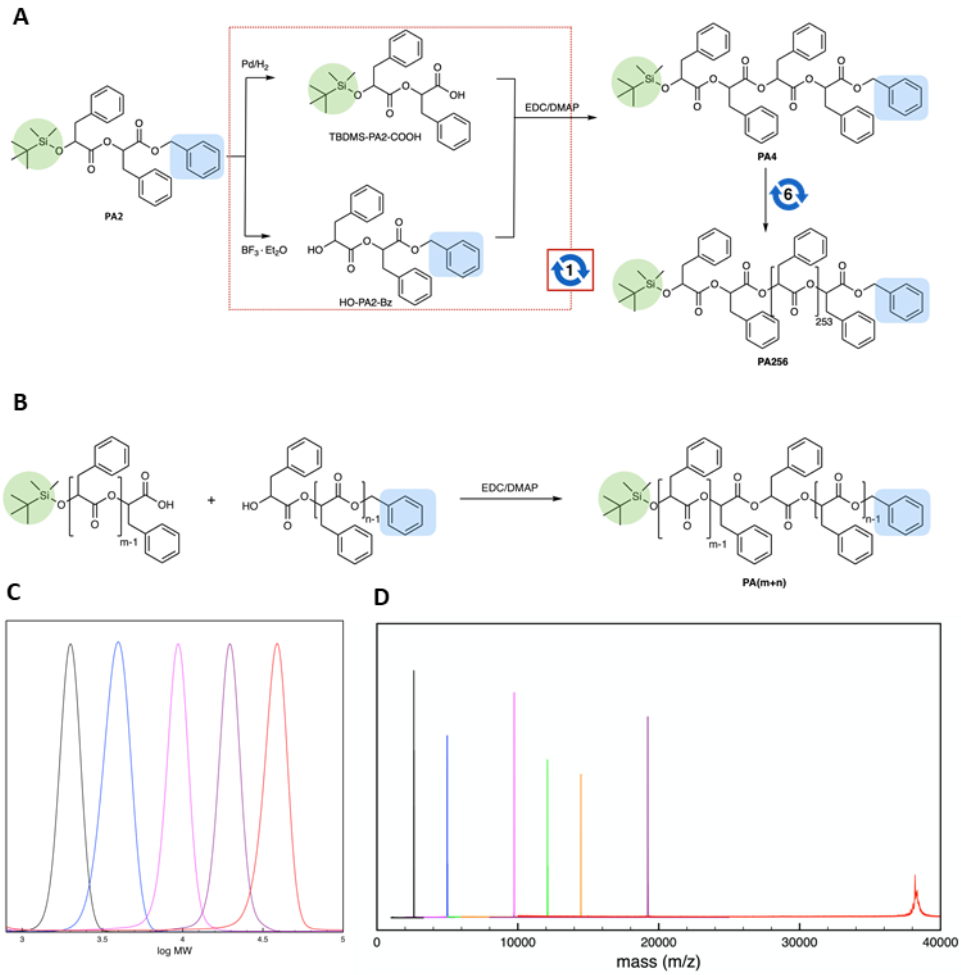


Figure 2-8. (A) The iterative convergent synthesis of poly(*rac*-phenyllactic acid) (PA*n*). The deprotection and subsequent esterification reactions shown in the red dotted box constitute a convergent growth step. The number shown with circular arrows represent an iteration of the convergent growth step. (B) Method to obtain the monodisperse PAHs with desired number of repeating units via combination of constituent units. (C) Gel-permeation chromatography of PAs with 16, 32, 64, 128, and 256 repeating units. (D) Combined MALDI-TOF mass spectra of PA16 (black, 2615.2 Da), PA32 (blue, 4986.4 Da), PA64 (magenta, 9728.4 Da), PA80 (green, 12100.7 Da), PA96 (orange, 14474.0 Da), PA128 (purple, 19220.6 Da), and PA256 (red, 38191 Da). MALDI-TOF MS of PA256 was measured in a linear mode.

2.4 Conclusion

In summary, we synthesized monodisperse poly(α-hydroxy acid)s (PAHs) composed of two chemically distinct monomers in absolutely defined aperiodic sequences via the cross-convergent method. Our proposed method allowed to synthesize polymers with any aperiodic sequence by utilizing the permutation of two monomers (dyads) as constituent units for the convergent synthetic pathway. The synthesis of monodisperse and sequence-defined PAHs with large numbers of repeating units was accomplished with notably fewer synthetic steps than were required for solid-phase synthesis of the same polymers. In addition, only stoichiometric amounts of constituent units were needed to synthesize a monodisperse product. This is in contrast to solid-phase synthesis, which requires a large excess of reagents to prevent incomplete coupling during the introduction of individual monomers. We demonstrated that digital information can be stored in a sequence of poly(phenyllactic-*co*-lactic acid) (PcL), which can be fully retrieved by a single measurement of MALDI-TOF tandem mass spectrometry. Utilizing facile degradability of ester groups in the PcL, the MALDI-TOF mass sequencing of high molecular weight PcL allowed a sequential read of the stored information. Information storage density of PcL (bit/Da) is 50% higher than that of DNA, which render the PcL to be an alternative molecular media for storing digital information.

Given the wide availability of α-hydroxy acids with different substituents and stereochemical configurations, the monodisperse and sequence-defined PAHs reported here should provide exciting opportunities to explore an unlimited diversity of chemical structures and the consequent properties and functions of synthetic polymers.[51–53] Our results also suggest that large-scale preparation of monodisperse polymers with precisely defined sequences and desired molecular weights is possible if executed in an automated and parallel fashion.[42]

2.5 Experimental

Materials and Methods. *rac*-lactide (Alfa Aesar) was recrystallized from MeOH/EtOAc mixture prior to use. All other reagents and chemicals were purchased from commercial sources and used without further purification. Dichloromethane (DCM), dimethylformamide (DMF), and toluene were distilled over CaH_2 under N_2. ^{1}H NMR spectra were recorded on a Varian 500 MHz using $CDCl_3$ as solvent. ^{13}C NMR spectra were recorded on an Agilent 400-MR DD2 Magnetic Resonance System using $CDCl_3$ as solvent. Gel permeation chromatography (GPC) was performed on an Agilent 1260 Infinity equipped with a PL gel 5 μm mixed D column and differential refractive index detectors. THF was used as an eluent with a flow rate of 0.3 mL min^{-1} at 35 °C. A polystyrene standard kit (Agilent Technologies) was used for calibration. Automated column chromatography was performed on a Biotage SP1 flash chromatography purification system equipped with a silica column cartridge (KP-Sil 100g and 50g). *n*-hexane and ethyl acetate were used as eluent.

SEC of monodisperse PAH. Size exclusion chromatography (SEC) of PAHs was conducted by injecting 5 mL of a PAH solution in $CHCl_3$ (100 mg mL^{-1}) to a Recycling Preparative HPLC (LC-9260 NEXT, Japan Analytical Industry) system equipped with JAIGEL-2.5H/2H/3H columns and a differential refractometer. Chloroform was used as an eluent with a flow rate of 3.5 mL min^{-1}. Before injection, the solution was filtered through a PTFE syringe filter (Whatman, 0.2 μm pore). The SEC was performed under a cycling mode until the coinciding peaks were separated. The desired fraction was collected using a fraction collector. Two prep-SEC systems ran in parallel, giving the maximum capacity of separation of 1 g.

MADLI-TOF and Tandem mass sequencing of PcL. Molecular weights of PcLs and their fragments were measured on a Bruker Ultraflex TOF/TOF mass

spectrometer equipped with a smartbeam 2 (Nd:YAG laser) at 2000 Hz (MALDI-MS) or 1000 Hz (MALDI-MS/MS). For MALDI-MS analysis, the instrument was operated in a positive reflector mode. Voltage for ion source and reflector is controlled depending on the molecular weight of a polymer. PAHs with molecular weight above 20 kDa were analyzed in a positive linear mode. External calibration was based on peptide and protein (ProteoMass Peptide/Protein MADLI-MS Calibration Kit (mass to charge ratio from 750 to 66000 Da), Sigma). Tandem mass sequencing (MS/MS) was performed in positive reflector mode without gas option. The precursor ion was used as internal calibration. For MALDI and MS/MS analysis, 2-(4-Hydroxyphenylazo)benzoic acid (HABA) or trans-2-[3-(4-tert-Butylphenyl)-2-methyl-2-propenylidene]malononitrile (DCTB) was used as a matrix. A polymer sample and matrix were dissolved in THF at 5mg mL^{-1} and 30mg mL^{-1}, respectively, and, these solutions were mixed in 1:1 to 1:5 ratio depending on the molecular weight of the analyte. 0.8 μL of the mixed solution was spotted on a MALDI plate, and dried in the air.

General procedure for deprotection of the benzyl group by hydrogenation. A compound protected with TBDMS and benzyl groups was dissolved in ethyl acetate. Palladium on activated charcoal (10% Pd/C, 0.03~0.2 eq.) was added to the solution, and the suspension was purged with argon for 15 minutes. The argon atmosphere was then replaced with hydrogen atmosphere, and the reaction mixture was stirred at room temperature. The reaction was monitored by thin layer chromatography (TLC) analysis. Upon completion of the reaction, the suspension was filtered through a Celite cake to remove Pd/C. The product was obtained by removing the solvent from the filtrates under reduced pressure.

General procedure for deprotection of the TBDMS group with fluoride. A compound protected with TBDMS and benzyl groups was dissolved in dry DCM.

The solution was cooled to 0°C on an ice bath, and BF_3•Et_2O was added dropwise. The reaction mixture was stirred at room temperature for 4 h. The reaction was monitored by TLC analysis. Upon completion of the reaction, the reaction was quenched with saturated $NaHCO_3$ followed by dilution with water. The organic layer was separated and washed with brine. The combined organic layer was dried over $MgSO_4$, and the solvent was removed under reduced pressure. The crude product was purified by automated column chromatography.

General procedure for esterification reactions. Alcohol and carboxylic acid (1.05 eq.) were dissolved in dry DCM, and the mixture was cooled to 0°C on an ice bath. To the mixture, 4-(dimethylamino)pyridine (DMAP, 0.2 eq.) and 1.4 eq. EDC•HCl were added. The reaction mixture was stirred overnight at room temperature, and the reaction was monitored by TLC analysis. Upon completion of the reaction, the reaction mixture was washed with water and brine. The combined organic layer was dried over $MgSO_4$, and the solvent was removed under reduced pressure. The crude product was purified by automated column chromatography. Products with 32 repeating units or more were purified by prep-SEC with a series of columns using chloroform as an eluent.

Chemical degradation of PcL. A vial was charged with 128-bit PcL (25 mg, 1.6 μmol) and THF (4 mL). To this solution, 0.5 M NaOH solution (150 μL, 47 eq. to the PcL) was added. The vial was tightly sealed and heated to 70°C. A portion (0.6 mL) of the solution was taken at a 30 min interval, which was subsequently diluted with ethyl acetate (10 mL). The diluted solution was washed with brine (5 mL). The combined organic layer was dried with $MgSO_4$ and concentrated in vacuo. The chemical degradation of PcL was confirmed by GPC using DMF as an eluent. Upon hydrolysis, PcL is dissociated into two fragments. One is the c*i* fragment that contains the original alpha group and a new carboxylic acid terminus. The other is

the x*i* fragment that contains the original omega group and a new hydroxyl terminus. The mass peaks of c*i* fragments were found to exhibit low intensities in MALDI-TOF. Therefore, the peaks corresponding to xi fragments were used for sequencing.

2.6 References

1. Odian, G. Principles of Polymerization, 4th Ed., Wiley (2004).
2. Extance, A. Could the molecule known for storing genetic information also store the world's data? *Nature* **2016**, *537*, 22–24.
3. Rutten, M. G. T. A.; Vaandrager, F. W.; Elemans, J. A. A. W.; Nolte, R. J. M. Encoding information into polymers. *Nat. Rev. Chem.* **2018**, *2*, 365–381.
4. Matyjaszewski, K. Atom Transfer Radical Polymerization (ATRP): Current Status and Future Perspectives. *Macromolecules* **2012**, *45*, 4015–4039.
5. Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-controlled polymers. *Science* **2013**, *341*, 1238149.
6. Ouchi, M.; Badi, N.; Lutz, J.-F.; Sawamoto, M. Single-chain technology using discrete synthetic macromolecules. *Nat. Chem.* **2011**, *3*, 917–924.
7. Lutz, J.-F.; Lehn, J.-M.; Meijer, E. W.; Matyjaszewski, K. From precision polymers to complex materials and systems. *Nat. Rev. Mater.* **2016**, *1*, 16024.
8. Lodge, T. P. Celebrating 50 Years of Macromolecules. *Macromolecules* **2017**, *50*, 9525–9527.
9. Merrifield, R. B. Solid phase peptide synthesis. I. the synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154.
10. Al Ouahabi, A.; Charles, L.; Lutz, J.-F. Synthesis of non-natural sequence-encoded polymers using phosphoramidite chemistry. *J. Am. Chem. Soc.* **2015**, *137*, 5629–5635.
11. Roy, R. K.; Meszynska, A.; Laure, C.; Verchin, C.; Lutz, J.-F. Design and synthesis of digitally encoded polymers that can be decoded and erased. *Nat. Commun.* **2015**, *6*, 7237.

12. Li, T.; Liu, L.; Wei, N.; Yang, J.-Y.; Chapla, D. G.; Moremen, K. W.; Boons, G.-J. An automated platform for the enzyme-mediated assembly of complex oligosaccharides. *Nat. Chem.* **2019**, *11*, 229–236.
13. Al Ouahabi, A.; Kotera, M.; Charles, L.; Lutz, J.-F. Synthesis of monodisperse sequence-coded polymers with chain lengths above DP100. *ACS Macro Lett.* **2015**, *4*, 1077–1080.
14. Al Ouahabi, A.; Amalian, J.-A.; Charles, L.; Lutz, J.-F. Mass spectrometry sequencing of long digital polymers facilitated by programmed inter-byte fragmentation. *Nat. Commun.* **2017**, *8*, 976.
15. Lewandowski, B.; De Bo, G.; Wardm, J. W.; Papmeyer, M.; Kuschel, S.; Aldegunde, M. J.; Gramlich, P. M. E.; Heckmanm, D.; Goldup, S. M.; D'Souze, D. M.; Fernandes, A. E.; Leigh, D. A. Sequence-specific peptide synthesis by an artificial small-molecule machine. *Science* **2013**, *339*, 189–193.
16. Huang, Z.; Noble, B. B.; Corrigan, N.; Chu, Y.; Satoh, K.; Thomas, D. S.; Hawker, C. J.; Moad, G.; Kamigaito, M.; Coote, M. L.; Boyer, C.; Xu, J. Discrete and stereospecific oligomers prepared by monomer insertion. *J. Am. Chem. Soc.* **2018**, *140*, 13392–13406.
17. Hibi, Y.; Ouchi, M.; Sawamoto, M. A strategy for sequence control in vinyl polymers via iterative controlled radical cyclization. *Nat. Commun.* **2016**, *7*, 11064.
18. Dong, R.; Liu, R.; Gaffney, P. R. J.; Schaepertoens, M.; Marchetti, P.; Wiliams, C. M.; Chen, R.; Livingston, A. G. Sequence-defined multifunctional polyethers via liquid-phase synthesis with molecular sieving. *Nat. Chem.* **2019**, *11*, 136–145.
19. Martens, S.; Landuyt, A.; Espeel, P.; Decreese, B.; Dawyndt, P.; Du Prez, F. Multifunctional sequence-defined macromolecules for chemical data storage. *Nat. Commun.* **2018**, *9*, 4451.
20. Solleder, S. C.; Zengel, D.; Wetzel, K. S.; Meier, M. A. R. A scalable and high-yield strategy for the synthesis of sequence-defined macromolecules. *Angew. Chem. Int. Ed.* **2016**, *55*, 1204–1207.
21. Porel, M.; Alabi, C. A. Sequence-defned polymers via orthogonal allyl

acrylamide building blocks. *J. Am. Chem. Soc.* **2014**, *136*, 13162–13165.

22. Huang, Z.; Shi, Q.; Guo, J.; Meng, F.; Zhang, Y.; Lu, Y.; Qian, Z.; Li, X.; Zhou, N.; Zhang, Z.; Zhu, X. Binary tree-inspired digital dendrimer. *Nat. Commun.* **2019**, *10*, 1918.
23. Bootwicha, T.; Feilner, J. M.; Myers, E. L.; Aggarwal, V. K. Iterative assembly line synthesis of polypropionates with full stereocontrol. *Nat. Chem.* **2017**, *9*, 896–902.
24. Solleder, S. C.; Schneider, R. V.; Wetzel, K. S.; Boukis, A. C.; Meier, M. A. R. Recent progress in the design of monodisperse, sequence-defned macromolecules. *Macromol. Rapid Commun.* **2017**, *38*, 1600711.
25. Cafferty, B. J.; Ten, A. S.; Fink, M. J.; Morey, S.; Preston, D. J.; Mrksich, M.; Whitesides, G. M. Storage of information using small organic molecules. *ACS Cent. Sci.* **2019**, *5*, 911–916.
26. Konrad, W.; Fengler, C.; Putwa, S.; Barner-Kowollik, C. Protection-group-free synthesis of sequence-defined macromolecules via precision λ-orthogonal photochemistry. *Angew. Chem. Int. Ed.* **2019**, *58*, 7133–7137.
27. Hawker, C. J.; Fréchet, J. M. J. Preparation of polymers with controlled molecular architectecture. A new convergent approach to dendritic macromolecules. *J. Am. Chem. Soc.* **1990**, *112*, 7638–7647.
28. Hawker, C. J.; Malmström, E. E.; Frank, C. W.; Kampf, J. P. Exact linear analogs of dendritic polyether macromolecules: design, synthesis, and unique properties. *J. Am. Chem. Soc.* **1997**, *119*, 9903–9904.
29. Binauld, S.; Damiron, D.; Connal, L. A.; Hawker, C. J.; Drockenmuller, E. Precise synthesis of molecularly defined oligomers and polymers by orthogonal iterative divergent/convergent approaches. *Macromol. Rapid. Commun.* **2011**, *32*, 147–168.
30. Grayson, S. M.; Fréchet, J. M. J. Convergent dendrons and dendrimers: from synthesis to applications. *Chem. Rev.* **2001**, *101*, 3819–3868.
31. Bidd, I.; Whiting, M. C. The synthesis of pure n-paraffins with chain-lengths between one and four hundred. *J. Chem. Soc. Chem. Commun.* **1985**, 543–544.

32. Barnes, J. C.; Ehrlich, D. J. C.; Gao, A. X.; Leibfarth, F. A.; Jiang, Y.; Zhou, E.; Jamison, T. F.; Johnson, J. A. Iterative exponential growth of stereo- and sequence-controlled polymers. *Nat. Chem.* **2015**, *7*, 810–815.

33. Takizawa, K.; Nulwala, H.; Hu. J.; Yoshinaga, K.; Hawker, C. J. Molecularly defined (L)-lactic acid oligomers and polymers: synthesis and characterization. *J. Polym. Sci. A: Polym. Chem.* **2008**, *46*, 5877–5990.

34. van Genebeck, B.; de Waal, B. F. M.; Gosens, M. M. J.; Pitet, L. M.; Palmans, A. R. A.; Meijer, E. W. Synthesis and self-assembly of discrete dimethylsiloxane–lactic acid diblock co-oligomers: the dononacontamer and its shorter homologues. *J. Am. Chem Soc.* **2016**, *138*, 4210–4218.

35. Takizawa, K.; Tang, C.; Hawker, C. J. Molecularly defined caprolactone oligomers and polymers: synthesis and characterization. *J. Am. Chem. Soc.* **2008**, *130*, 1718–1726.

36. Amir, F.; Jia, Z.; Monteiro, M. J. Sequence control of macromers via iterative sequential and exponential growth. *J. Am. Chem. Soc.* **2016**, *138*, 16600–16603.

37. Huang, Z.; Zhao, J.; Wang, Z.; Meng, F.; Ding, K.; Pan, X.; Zhou, N.; Li, X.; Zhang, Z.; Zhu, X. Combining orthogonal chain-end deprotections and thiol-maleimide Michael coupling: engineering discrete oligomers by an iterative growth strategy. *Angew. Chem. Int. Ed.* **2017**, *56*, 13612–13617.

38. Jiang. Y.; Golder, M. R.; Nguyen, H. V.-T.; Wang, Y.; Zhong, M.; Barnes, J. C.; Ehrlich, D. J. C.; Johnson, J. A. Iterative exponential growth synthesis and assembly of uniform diblock copolymers. *J. Am. Chem. Soc.* **2016**, *138*, 9369–9372.

39. Leibfarth, F. A.; Johnson, J. A.; Jamison, T. F. Scalable synthesis of sequence-defined, unimolecular macromolecules by Flow-IEG. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 10617–10622.

40. Church, G. M.; Gao, Y.; Kosuri, S. Next-generation digital information storage in DNA. *Science* **2012**, *337*, 1628.

41. Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; LeProust, E. M.; Sipos, B.; Birney, E. Towards practical, high-capacity, low-maintenance information storage

in synthesized DNA. *Nature* **2013**, *494*, 77–80.

42. Erlich, Y.; Zielinski, D. DNA Fountain enables a robust and efficient storage architecture. *Science* **2017**, *355*, 950–954.
43. Organick, L.; Ang, S. D.; Chen. Y.-J.; Lopez, R.; Yekhanin, S.; Makarychev, K.; Racz, M. Z.; Kamath, G.; Gopalan, P.; Nguyen, B.; Takahashi, C. N.; Newman, S.; Paker, H.-Y.; Rashtchian, C.; Stewart, K.; Gupta, G.; Carlson, R.; Mulligan, J.; Carmean, D.; Seelig, G.; Ceze, L.; Strauss, K. Random access in large-scale DNA data storage. *Nat. Biotechnol.* **2018**, *36*, 242–248.
44. Stayshich, R. M.; Meyer, T. Y. New insights into poly(lactic-co-glycolic acid) microstructure: using repeating sequence copolymers to decipher complex NMR and thermal behavior. *J. Am. Chem. Soc.* **2010**, *132*, 10920–10934.
45. Lawrence, J.; Lee, S.-H.; Abdilla, A.; Nothling, M. D.; Ren, J. M.; Knight, A. S.; Fleischmann, C.; Li, Y.; Abrams, A. S.; Schmidt, B. V. K. J.; Hawker, M. C.; Connal, L. A.; McHrath, A. J.; Clark, P. G.; Gutekunst, W. R.; Hawker, C. J. A versatile and scalable strategy to discrete oligomers. *J. Am. Chem. Soc.* **2016**, *138*, 6306–6310.
46. Hoskins, H. N.; Grayson, S. M. Cyclic polyesters: synthetic approaches and potential applications. *Polym. Chem.* **2011**, *2*, 289–299.
47. Wesdemiotis, C.; Solak, N.; Polce, M. J.; Dabney, D. E.; Chaicharoen, K.; Katzenmeyer, B. C. Fragmentation pathways of polymer ions. *Mass Spectrom. Rev.* **2011**, *30*, 523–559.
48. Albertsson, A.-C.; Varma, I. K. Recent developments in ring opening polymerization of lactones for biomedical applications. *Biomacromolecules* **2003**, *4*, 1466–1486.
49. Cummins, C.; Mokarian-Tabari, P.; Holmes, J. D.; Morris, M. A. Selective Etching of Polylactic Acid in Poly(styrene)-Block-Poly(D,L)Lactide Diblock Copolymer for Nanoscale Patterning. *J. Appl. Polym. Sci.* **2014**, DOI: 10.1002/APP.40798.
50. Jung, J. H.; Ree, M.; Kim, H. Acid- and base-catalyzed hydrolyses of aliphaticpolycarbonates and polyesters. *Catal. Today* **2006**, *115*, 283–287.

51. Kalelkar, P. P.; Alas, G. R.; Collard, D. M. Synthesis of an alkene-containing copolylactide and its facile modification by the addition of thiols. *Macromolecules* **2016**, *49*, 2609–2617.
52. Zhang, Q.; Ren, H.; Baker, G. L. Synthesis of a Library of Propargylated and PEGylated α-Hydroxy Acids Toward "Clickable" Polylactides via Hydrolysis of Cyanohydrin Derivatives. *J. Org. Chem.* **2014**, *79*, 9546–9555.
53. Jiang, X.; Vogel, E. B.; Smith, M. R.; Baker, G. L. "Clickable" Polyglycolides: Tunable Synthons for Thermoresponsive, Degradable Polymers. *Macromolecules* **2008**, *41*, 1937–1944.

Chapter 3

Semiautomated Synthesis of Sequence-defined Polymers for Information Storage

(K.J., L.S.J., and J.H. contributed to the synthesis of PLGAs)

3.1 Abstract

Accelerated and parallel synthesis of sequence-defined polymers is an utmost challenge for realizing ultra-high-density storage of digital information in molecular media. Here we report step-economical synthesis of sequence-defined poly(L-lactic-*co*-glycolic acid)s (PLGAs) using continuous flow chemistry. A reactor performed the programmed coupling of the two-bit storing building blocks to generate a library of their permutations in a single continuous flow, followed by their sequential convergences to a sequence-defined PLGA storing 64 bits in four successive flows. We demonstrate that a bitmap image (896 bits) can be encoded and decoded in 14 PLGAs using only a fraction of the time required for an equivalent synthesis by conventional batch processes. The bundle of 14 PLGAs could be decoded in a single tandem mass sequencing.

3.2 Introduction

DNA stores genetic information as a sequence of four monomers having different pendant nucleobases constituting a polymer chain. Chemical synthesis of DNA based on the solid-phase synthesis enables encoding of non-biological information at a far greater density than any existing magnetic, optical, and electronic media without a periodic rewriting owing to the deterioration of media over storage.[1–3] Well-established sequencing technologies can efficiently decode the information stored in DNA.[4,5] Consequently, in recent decades, DNA has been pursued as a macromolecular medium to store digital information, which could circumvent the impending shortage of storage capacity due to the accelerated production of digital information in the age of internet and mobile devices.[6,7]

However, information encoding in DNA primarily relies on the repetitive addition of individual monomers to define non-biological sequences. Considering that one repeating unit of DNA of an average molecular weight of 309 Da stores less than two bits,[8] this encoding based on repetitive monomer additions consumes time and cost, and, thus, remains a significant obstacle for realizing large-scale information storage in molecular media.[9,10]

Sequence-defined synthetic polymers can serve as an alternative to DNA as information-storing molecular media because of their efficient and low-cost synthesis through streamlined chemical reactions with simpler monomers at the expense of losing biological functions such as replication and enzyme-based sequencing.[11–15] We recently reported that the sequence of a copolyester built with two monomers, representing 0 and 1, could be unambiguously defined by a step-economical pathway, that is, the cross-convergent approach.[16] Compared to the solid-phase synthesis for sequence-defined macromolecules,[17,18] the cross-convergent method facilitates encoding a binary code up to the length of up to 256 bits as a sequence of monomers with a minimal number of synthetic steps and without using a large excess of reagents.[19] The copolyester, poly(L-lactic acid-*co*-glycolic acid) (PLGA), resulting from cross-convergent synthesis using L-lactic acid and glycolic acid as the building blocks can store one bit per 60 Da of mass, which is a two-fold greater density than that of DNA. Theoretically, two grams of sequence-defined PLGA can store the entire information that is currently stored in data centers worldwide (~ 2500 exabytes).[20]

Despite step-economical synthesis of sequence-defined polymers using the cross-convergent method, realizing large-scale information storage in PLGA mandates the synthesis of a multitude of PLGA chains. For example, the encoding of 106 bit (1 Mbit) of information requires 15,625 sequence-defined PLGA chains

built with 64 monomers (64-mer). Here we demonstrate that this challenge could be overcome through automated and programmable synthesis of PLGA. Continuous flow chemistry to synthesize sequence-defined polymers could offer several advantages over the repetition of traditional batch processes such as integrating individual reactions in a single operation, reducing reaction time, easy scale-up synthesis, and massively parallel operation.[21–25] This process can handle the complexity involved in the generation of aperiodic sequences without resorting to the repetition of batch reactions by manually selecting the required building blocks.

3.3 Results and discussion

Synthesis of sequence-defined tetrads using continuous flow chemistry. The cross-convergent strategy for sequence-defined polymers consists of two distinct stages (Figure 3-1). In a divergent stage, two monomers, L-lactic acid (L) and glycolic acid (G) representing 1 and 0, respectively, were coupled to form four sequenced dimers (00, 01, 10, and 11) with orthogonal protecting groups; *tert*-butyldimethylsilyl (TBDMS) for the hydroxy group (O-terminus) and benzyl (Bz) for the carboxylic acid end group (C-terminus). Using a typical batch process, these dyads, that is, the dimers covering all possible permutations of two-bit signals, were cross-converged to generate 16 tetramers (tetrads) to encode all possible binary sequences via orthogonal deprotections and esterification. This divergent stage introduces complexity in synthesis as the encoding of tetrameric sequences requires cross-convergences between four dyads, thereby resulting in the exponential increase in the number of permutations of L and G. To synthesize all tetrads by conventional batch processes, 48 individual reactions with purifications over 16 days should be performed.

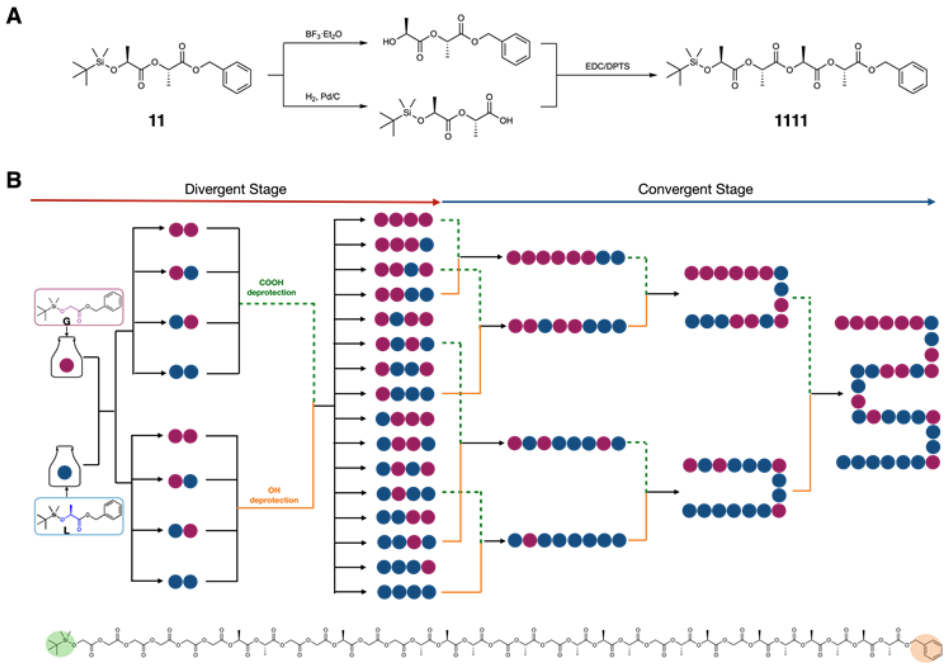
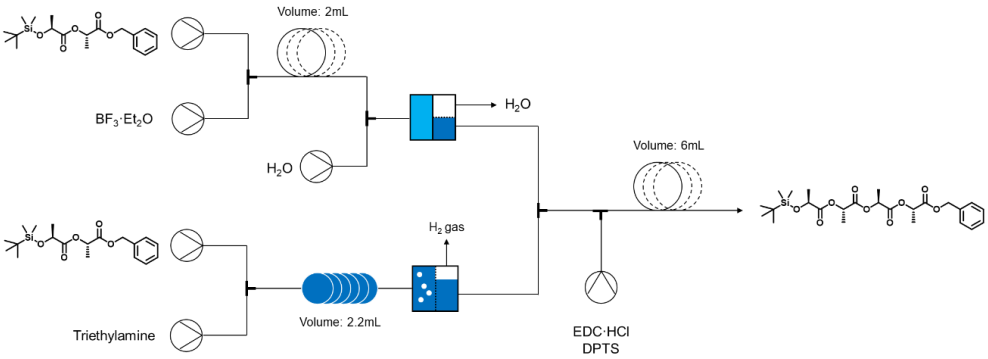


Figure 3-1. (A) Cross-convergent synthesis of a tetramer, LLLL. A dimer is converted to tetramer via orthogonal deprotection and esterification. (B) Synthetic strategy for sequence-defined polyester. In divergent stage, all possible tetramers are generated by cross-convergent synthesis of dimers. In convergent stage, sequence-defined polyester with 2^{n+2} repeating units is produced by n times successive cross-convergent synthesis of tetramers.



Scheme 3-1. Combined continuous flow setup for a tetrad, LLLL.

To convert the divergent stage to a single continuous flow, we devised a semi-automated process that can generate all sequence-defined tetrads in a single continuous flow setup (Scheme 3-1). The dyads of L and G having TBDMS and Bz protecting groups were synthesized by batch reactions on a large scale (>50 g). As a test reaction, we programmed the flow setup to synthesize a tetrad 1111 using a dyad 11 as a building block. The process started with the selective injection of 11 (100 mg, 0.8 M in CH_2Cl_2) out of four building block solutions to both deprotection lines at a 0.1 mL/min flow rate for 10 min through a Teflon capillary (diameter 0.5 mm), which was regulated by a programmable 6-way syringe connected to a computer. For the deprotection of the O-terminus of 11, the injected solution was mixed with trifluoroborane etherate ($BF_3{\cdot}Et_2O$, 4 M in CH_2Cl_2) through a T-mixer. The synchronization of flows in the T-mixer was achieved by adjusting the length of capillaries. This mixture was allowed to react in the reaction loop (the volume V_{rl} of 2 mL), followed by the passage through an in-line module for the extraction of the resulting solution with water. Similarly, for the deprotection of the C-terminus, the 0.8 M solution of 11 in tetrahydrofuran (THF) was injected to mix with the 2.4 M solution of triethylsilane (Et_3SiH) in THF, which was subsequently passed through a column of Pd/C equipped with a degasser to conduct hydrogenation. Finally, the deprotected dyads were synchronously combined and mixed with a solution of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) to perform esterification in the reaction loop (V_{rl} = 6 mL). Thereafter, the coupled tetrad 1111 was collected using a fraction collector, which was subsequently purified using a 10 min run on an automated column chromatography instrument using silica column. After purification, 83 mg of pure 1111 (60% yield) was obtained after a full flow of the dyad solutions to the coupled tetrad for 35 min. This modest coupling yield was attributed to the presence of impurities from the hydrogenation with

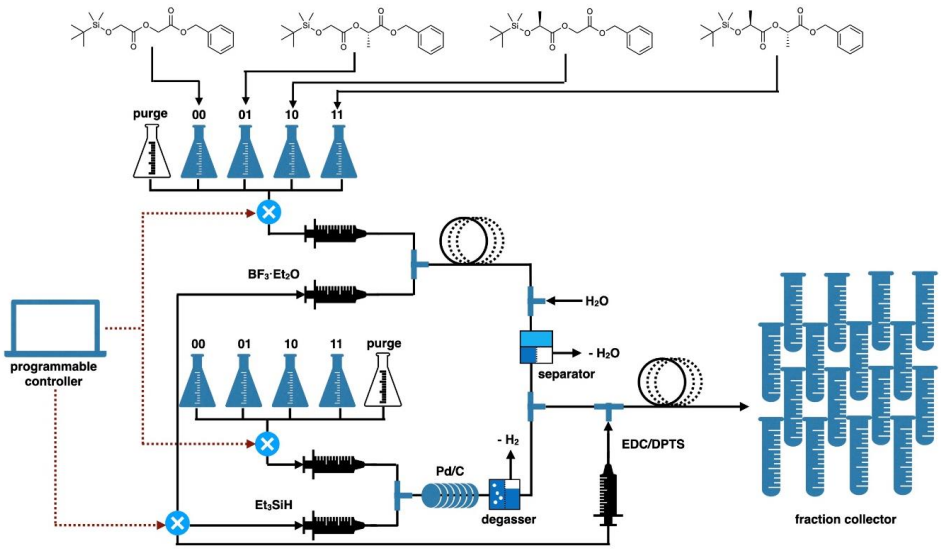


Figure 3-2. Schematic illustrations of the continuous flow process for the preparation of sequence-defined tetramers. Synchronous deprotection of selected dyads and subsequent coupling are performed. All permutations of tetrads are generated by a single continuous flow.

Table 3-1. The yield of all possible tetrads obtained by a single continuous flow.

LLLL	**LLLG**	**LLGG**	**LGGG**
1.2 g (59%)	1.11 g (58%)	1.12 g (61%)	1.08 g (56%)
LGLL	**LGLG**	**LGGL**	**LGGG**
1.07g (54%)	1.08g (56%)	1.02 g (53%)	1.03g (55%)
GLLL	**GLLG**	**GLGL**	**GLGG**
1.11g (56%)	1.04g (54%)	980 mg (51%)	940 mg (50%)
GGLL	**GGLG**	**GGGL**	**GGGG**
1 g (52%)	990 mg (53%)	940 mg (50%)	950 mg (52%)

Et_3SiH. However, upon addition of the off-line purification step after hydrogenation, the coupling yield from the continuous flow process was comparable to or higher than the yield of the corresponding batch reaction.

For serial synthesis of all possible tetrads, we performed 16 convergence steps to synthesize all tetrads in a gram scale in a single continuous flow with an independent feed of four dyads to the deprotection reactor lines through programmable syringes controlled by the software (Figure 3-2). Each convergence process consisted of 50 min reaction flow and 10 min purge with two pure solvents to clean up the lines. After the flow, the synthesized tetrads were individually collected in 16 test tubes in a fraction collector. The reaction time for the synthesis of all 16 tetramers in a single flow was 16 h. After off-line purification, the yields of the tetrads were in the range of 50 to 61% (0.94 ~ 1.20 g) (Table 3-1).

Encoding process to sequence-defined polymers via continuous flow synthesis. To demonstrate the potential of the continuous flow process for accelerated synthesis of a set of sequence-defined polymers, we encoded a low-resolution bitmap data (896 bits) converted from a photograph in a series of 64-mer PLGA chains (Figure 3-3). Each PLGA of the 14-chain library contained fragmented data (56 bits) with an 8-bit address at the O-terminus for the chain identification by sequencing. In a divergent stage, the reaction time for the synthesis of each tetrad was adjusted by setting the flow time of the required dyads for the coupling according to the relative occurrence of the tetrad in the sequences of a series of PLGA chains corresponding to the encoded information. After the divergent stage, the resulting library of tetrads was cross-converged to encode information in 64-mer PLGA through four successive couplings using the product of the previous convergence as the building blocks.[16] Each coupling process, required a flow time of 30 min, was followed by off-line purification using an automated silica column chromatography for up to 32-

mers and preparative size-exclusion chromatography (prep-SEC) for polymers with higher molecular weights. The total process time for the synthesis of 64-mer PLGA from tetrads was 8 h including off-line purifications after each convergent process. Finally, we repeated the cross-convergence to complete the synthesis of 14 PLGA chains

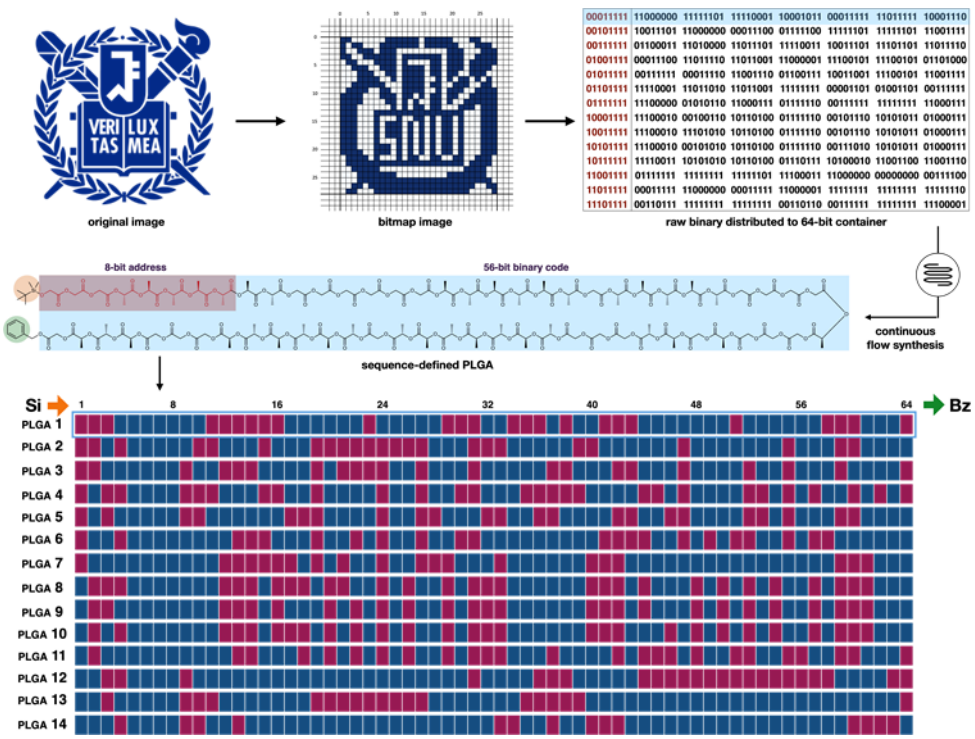


Figure 3-3. Process to encode a bitmap image in a multiple of sequence-defined PLGAs. Original image is converted to bitmap image and bitmap image is converted to binary code. Binary information is divided into several polymer chains with 64 repeating units by continuous flow synthesis. Sequence-defined PLGA is composed of 8 bits of address and 56 bits of fragmented data. Synthesized fourteen sequence-defined PLGAs are represented by bar code. Red and blue rectangle represent glycolic acid and lactic acid unit.

The resulting 64-mer PLGAs were unambiguously characterized by ^{1}H NMR, analytical SEC, and Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (Detailed characterization of PLGAs is described in Appendix). The mass analysis of the PLGA chains provides the number of L and G repeating units constituting the polymer chain since the total number of repeating units was fixed to 64 (Table 3-2). In addition, ^{1}H NMR analysis revealed the identity of the first repeating unit from the O-terminus of PLGA as the peak corresponding to the α-proton of the L-lactic acid unit in the vicinity of Si-atom typically appears at 4.39 or 4.43 ppm, and the glycolic acid unit appears at 4.32 – 4.36 ppm (Figure 3-4).

Table 3-2. Summarized table of the information of sequence-defined PLGA. N_{LA} indicates the number of the lactic acid units.

PLGA	Mass (Da)	N_{LA}
1	4504.29 Da	39
2	4518.48 Da	40
3	4546.61 Da	42
4	4462.59 Da	36
5	4532.63 Da	41
6	4546.61 Da	42
7	4574.73 Da	44
8	4448.55 Da	35
9	4490.54 Da	38
10	4461.98 Da	36
11	4504.38 Da	39
12	4518.76 Da	40
13	4588.41 Da	45
14	4659.00 Da	50

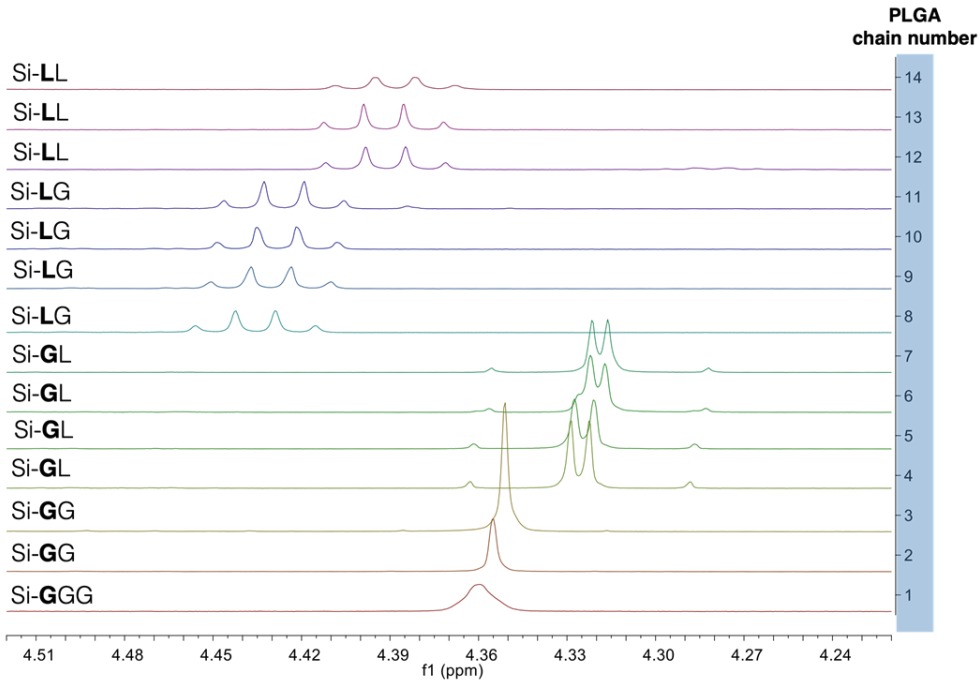
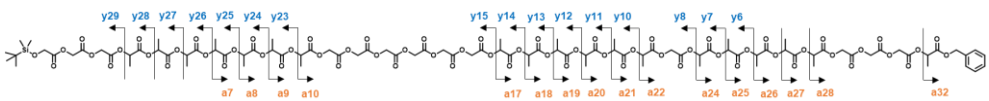


Figure 3-4. ^{1}H NMR spectra of sequence-defined PLGA. 4.3 – 4.5 ppm region indicating the first repeating unit from O-terminus.

Considering this information, the complete sequences of the PLGA chains could be decoded by tandem MS using MALDI-TOF MS. The fragmentation pattern under the condition used for the tandem mass sequencing revealed that the L was fragmented as a single repeating unit. In contrast, the G units were fragmented with an adjacent L because of unstable fragment resulting from 1,5-H rearrangement of ester bond in tandem mass condition (Figure 3-5). Reading two series of decreasing masses of the fragments (O to C terminus and C to O terminus) from the parent mass value, the entire sequence of the PLGA chain 5 was decoded without any error by a single tandem mass spectrum (Figure 3-6 and Table 3-3). Other PLGAs were successfully decoded, of which data are represented in Appendix (A.3.2).



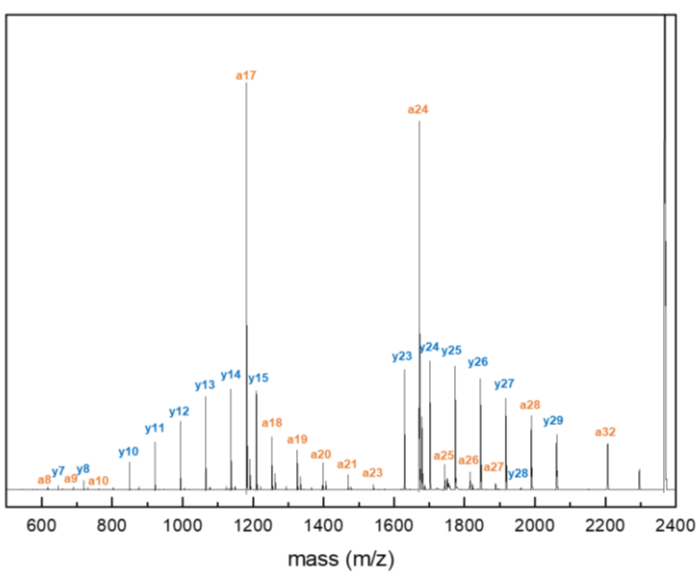


Figure 3-5. Fragment pattern of sequence-defined PLGA and its tandem mass spectrum. y*i* series are the fragments containing O-terminus. a*i* series are the fragments containing C-terminus.

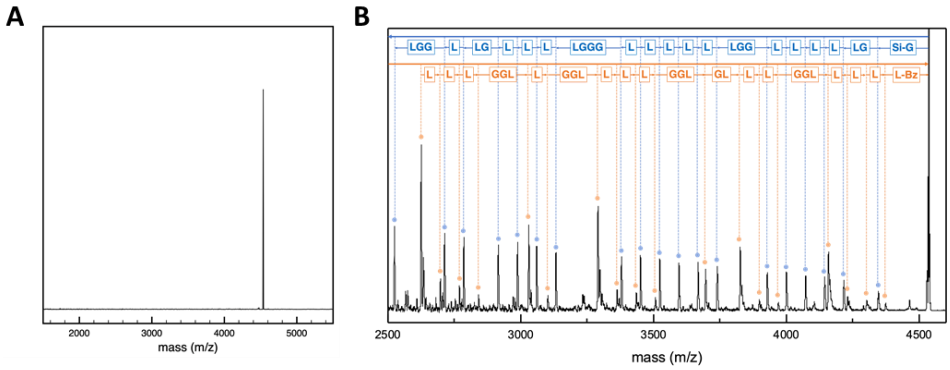


Figure 3-6. (A) MALDI-TOF mass spectrum of PLGA chain 5 (theoretical mass, 4532.13 Da; experimental mass, 4532.63 Da). (B) Tandem mass spectrum of PLGA chain 5 (parent ion, 4532.63 Da). The sequence-defined PLGA is decoded by reading two series of decreasing masses of fragments (lactic acid and glycolic acid units have 72 and 58 molecular weight). Blue and orange boxes represent the sequence from O terminus and C terminus. m/z, mass/charge ratio.

Table 3-3. Decoding table of PLGA chain 5.

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
$[M+Na]^+$	4532.13	4532.63			$[M+Na]^+$	4532.13	4532.63		
y63	4342.03	4345.15	187.48	Si-G	**a64**	4370.05	4373.49	159.14	Bz-L
y61	4212	4215.5	129.65	LG	**a63**	4298.03	4301.22	72.27	L
y60	4139.98	4143.4	72.1	L	**a62**	4226.01	4229.42	71.8	L
y59	4067.96	4071.45	71.95	L	**a61**	4153.99	4157.29	72.13	L
y58	3995.94	3998.8	72.65	L	**a58**	3965.95	3969.39	187.9	GGL
y57	3923.92	3927.56	71.24	L	**a57**	3893.93	3897.38	72.01	L
y54	3735.89	3738.75	188.81	LGG	**a56**	3821.91	3825.09	72.29	L
y53	3663.87	3666.79	71.96	L	**a54**	3691.89	3694.33	130.76	GL
y52	3591.85	3594.15	72.64	L	**a51**	3503.85	3506.21	188.12	GGL
y51	3519.83	3522.17	71.98	L	**a50**	3431.83	3433.7	72.51	L
y50	3447.8	3449.72	72.45	L	**a49**	3359.81	3361.97	71.73	L
y49	3375.78	3377.59	72.13	L	**a48**	3287.79	3289.47	72.5	L
y45	3129.75	3131.89	245.7	LGGG	**a45**	3099.76	3101.83	187.64	GGL
y44	3057.72	3058.8	73.09	L	**a44**	3027.74	3028.97	72.86	L
y43	2985.7	2986.71	72.09	L	**a41**	2839.7	2840.4	188.57	GGL
y42	2913.68	2914.65	72.06	L	**a40**	2767.68	2768.32	72.08	L
y40	2783.66	2784.35	130.3	LG	**a39**	2695.66	2696.25	72.07	L
y39	2711.63	2712.26	72.09	L	**a38**	2623.64	2624.22	72.03	L
y36	2523.6	2523.97	188.29	LGG	**a35**	2435.61	2434.88	189.34	GGL
y35	2451.58	2451.85	72.12	L	**a34**	2363.59	2363.82	71.06	L
y34	2379.56	2379.8	72.05	L	**a31**	2175.56	2174.7	189.12	GGL
y31	2191.53	2190.67	189.13	LGG	**a30**	2103.54	2102.67	72.03	L
y30	2119.51	2118.6	72.07	L	**a29**	2031.51	2030.67	72	L
y27	1931.47	1930.58	188.02	LGG	**a26**	1843.48	1842.6	188.07	GGL
y26	1859.45	1858.6	71.98	L	**a25**	1771.46	1770.61	71.99	L
y25	1787.43	1786.56	72.04	L	**a23**	1641.43	1640.66	129.95	GL
y24	1715.41	1714.6	71.96	L	**a22**	1569.41	1568.67	71.99	L
y21	1527.38	1526.61	187.99	LGG	**a21**	1497.39	1496.7	71.97	L
y20	1455.36	1454.72	71.89	L	**a20**	1425.37	1424.83	71.87	L
y17	1267.33	1266.87	187.85	LGG	**a16**	1179.33	1178.95	245.88	GGGL
y16	1195.31	1194.9	71.97	L	**a15**	1107.31	1107.08	71.87	L
y15	1123.28	1122.99	71.91	L	**a14**	1035.29	1035.1	71.98	L
y14	1051.26	1051.06	71.93	L	**a13**	963.27	963.19	71.91	L

Decoding process of sequence-defined polymers by tandem mass spectrometry. As indicated by the mass analysis of the PLGA library, the combination of two monomers at a fixed chain length (64 repeating units in this study) frequently introduces duplication of molecular weights of PLGAs, despite different encoded information. For example, the masses of chain 1 and 11 were identical at 4504.6 Da with the same composition of L and G units but different arrangement. The equimolar mixture of these PLGAs only showed a single peak in MALDI-TOF mass spectrum; thus, both the chains were indistinguishable in the mass spectrum. This requires each PLGA chain to be stored separately, which complicates the storage and decoding of the PLGA library. We performed tandem mass sequencing of a mixture of chain 1 and 11 (1:1 w/w) under the same condition as that for single-chain sequencing (Figure 3-7A). The resulting mass spectrum revealed two independent series of peaks, which were identified by reading the sequence of the first eight residues from the O-terminus, the 8-bit address implemented as a chain identifier in PLGA (Figure 3-7B). The assignment of these series of mass peaks allowed us to sequence chain 1 and 11 simultaneously from one tandem mass spectrum.

This simultaneous mass sequencing of multiple PLGA chains sharing molecular weights allowed bundling of the multiple chains, encoding different information, together as a piece of plastic (Figure 3-8A). The information stored in the bundle could be decoded by mass spectrometry without separating the mixture to individual chains. The MALDI-TOF mass spectrum of this plastic showed a set of 10 peaks, indicating the mass duplications between PLGA chains (Figure 3-8B).

This mixture of 14 chains was directly subjected to tandem mass sequencing by targeting each mass peak including four duplications of the masses of PLGA chains. For example, the tandem mass sequencing of 7th peak showing 4546 Da in MALDI-TOF spectrum could identify that the peak is corresponded to the mixture of PLGA

chain 3 and 6, and the stored information was could be decoded separately (Figure 3-8C). Consequently, the mass sequencing decoded the entire information distributed among the 14 PLGA chains. The other tandem mass sequencing results by targeting the rest of mass peaks are shown in Appendix (A.3.3).

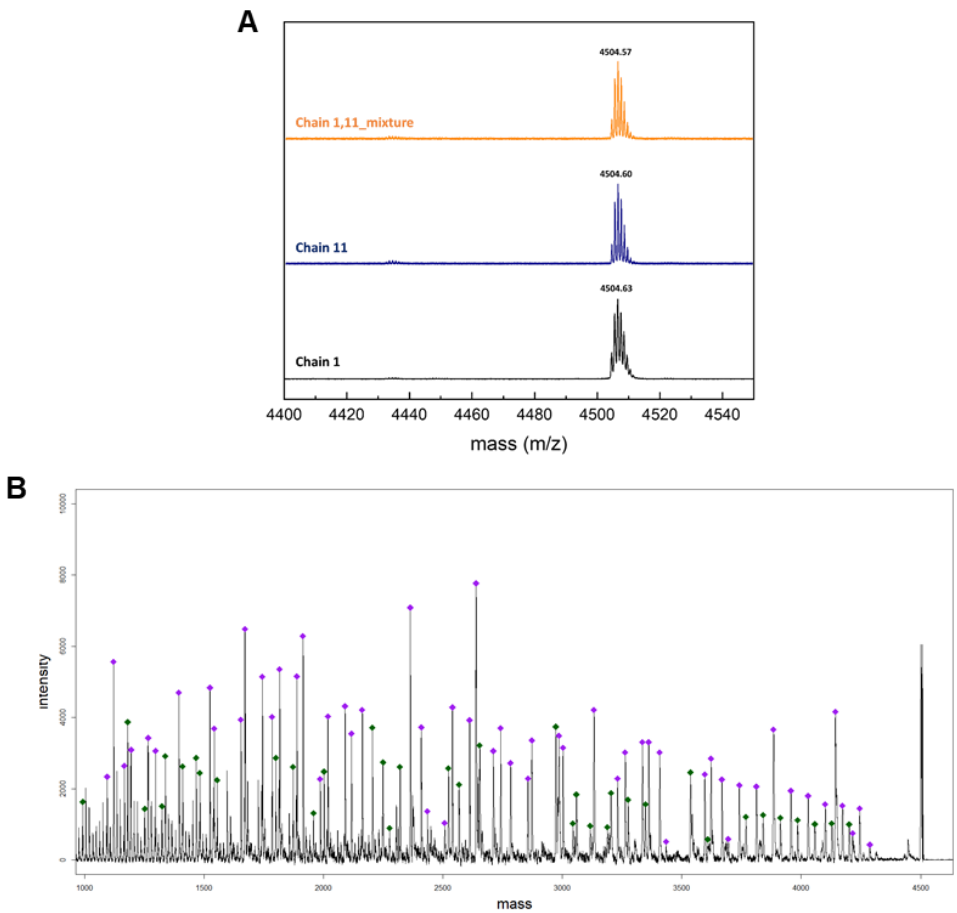


Figure 3-7. (A) Combined MALDI-TOF mass spectra of sequence defined PLGAs (B) Tandem mass spectrum of the sequence defined PLGAs mixture. Green diamond designates the fragements from PLGA chain 1. Violet diamond designates the fragments form PLGA chain 11.

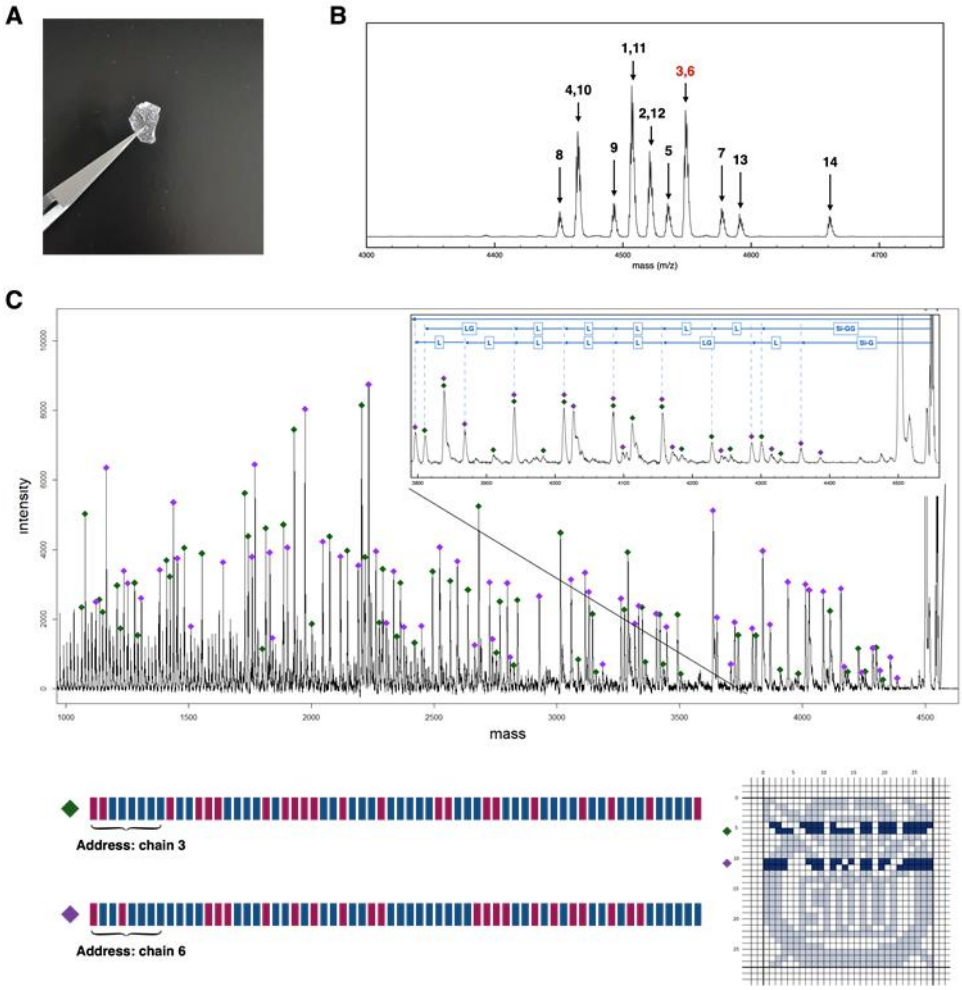


Figure 3-8. (A) Photograph of polymer resin mixed of fourteen sequence-defined PLGAs. (B) MALDI-TOF mass spectrum of the polymer resin. Each peak corresponds to the molecular weight of fourteen polymers. (C) Tandem mass spectrum of a parent ion (PLGA chain 3 and 6, 4546.86 Da). Both fragments generated from chain 3 (green diamond) and 6 (violet diamond) are shown in spectrum. Each stored information can be decoded by reading 8-bit address codes (Si-GGLLLLLL: chain 3, Si-GLLGLLLL: chain 6). Decoded information is represented by bar code. 56-bit fragmented code corresponds to the bit map image.

3.4 Conclusion

In conclusion, we devised accelerated and semiautomated flow synthesis of PLGAs. All permutations of tetrads could be generated in a single continuous flow synthesis. The sequence-defined tetrads were subjected to the flow synthesis for high molecular weight polymers. Consequently, sequence-defined PLGAs with 64 repeating units could be obtained in a fraction of the time required as compared to that of batch reactions. Furthermore, introducing 8-bit address code into the PLGAs enabled to decode the bundle of multitude PLGAs by direct tandem mass sequencing. Our demonstration suggested that the encoding-and-decoding of digital information distributed in a multitude of sequence-defined PLGAs could be streamlined via the efficient on-demand synthesis of polymer chains using continuous flow chemistry and the collective sequencing without handling individual polymer chains.[26] We envisage that the massively parallel synthesis of sequence-defined polymers through continuous flow chemistry should contribute to the molecular engineering of synthetic polymers for solving important technical and societal challenges such as the development of molecular media for information storage, degradable polymers, biomedical materials, and delivery of biomacromolecules.[27,28]

3.5 Experimental

Materials and Methods. All reagents and chemicals were purchased from commercial sources and used without further purification. Dichloromethane (DCM) and tetrahydrofuran (THF) were distilled over CaH_2 under N_2. Legato 101 syringe pump was purchased from Kd Scienctific. Cadent 3TM syringe pump was purchased from IMI Norgren. SEP-10 was purchased from Zaiput Flow Technologies. PFA

Tubing (1/16″ OD/0.02″ and 0.03″ ID) was purchased from Revodix. Omnifit EZ column was purchased from Diba Industries Inc. Gastorr AG-42-01 was purchased from GL Sciences. CF-2 Fraction Collector was purchased from Spectrum Chemical Mfg. Corp. ^{1}H and ^{13}C NMR spectra were recorded on a Varian INOVA 500 MHz NMR spectrometer using CDCl3 as solvent. Gel permeation chromatography (SEC) was performed on an Agilent 1260 Infinity equipped with a PL gel 5 μm mixed D column and differential refractive index detectors. THF was used as an eluent with a flow rate of 0.3 mL min^{-1} at 35 °C. A polystyrene standard kit (Agilent Technologies) was used for calibration. Automated column chromatography was performed on a Biotage Selekt flash chromatography purification system equipped with a Sfär silica column cartridge. *n*-hexane and ethyl acetate were used as eluent. Size exclusion chromatography (SEC) was performed on a Recycling Preparative HPLC (LC-9260 NEXT, Japan Analytical Industry) system equipped with JAIGEL-2.5HR/2HR columns and a differential refractometer. Chloroform (J.T.Baker) was used as an eluent with a flow rate of 10 mL min^{-1}. Before injection, the solution was filtered through a PTFE syringe filter (Whatman, 0.2 μm pore). The SEC was performed under a cycling mode until the coinciding peaks were separated. The desired fraction was collected using a fraction collector. Differential Scanning Calorimetry (DSC) was performed on a TA Instruments Q10 from –40 °C to 160 °C with a scan rate of 10 °C min^{-1} under N_2 atmosphere. The instrument was calibrated before measurements by measuring the melting temperature of indium (T_m = 429.75 K).

Synthesis of sequence-defined polymers using continuous flow synthesis. In divergent stage, 4 mmol of dyad (0.8 M, 5 mL) was injected to the deprotection of TBMDS or benzyl module for 50 min. Subsequently, flow lines were washed with pure solvents for 10 min. The fraction collector allowed to receive produced tetrad into separated glass tube. Sequence generation and washing cycle was repeated 16

times. Collected tetrads were purified with automated column chromatography as eluent of hexane/ethyl acetate mixture. In convergent stage, 3-way syringes were used for injecting sequence defined oLGA protected with TBDMS and benzyl groups (LA_xGA_m and LA_yGA_n, 0.8 ~ 0.1 M and 30 ~ 40 min according to the molecular weight). Crude product having twice repeating units ($LA_{x+y}GA_{m+n}$) was purified by column chromatography or preparative size-exclusion chromatography.

MADLI-TOF and Tandem mass sequencing of PcL. Molecular weights of sequence-defined PLGAs and their fragments were measured on a Bruker Ultraflex TOF/TOF mass spectrometer equipped with a smartbeam 2 at 2000 Hz (MALDI-MS) or 1000 Hz (MALDI-MS/MS). For MALDI-MS analysis, the instrument was operated in a positive reflector mode. Voltage for ion source and reflector is controlled depending on the molecular weight of a polymer. External calibration was based on peptide and protein (ProteoMass Peptide/Protein MADLI-MS Calibration Kit, Sigma). Tandem mass sequencing (MS/MS) was performed in positive reflector mode without gas option. The precursor ion was used as internal calibration. For MALDI and MS/MS analysis, 2-(4-Hydroxyphenylazo)benzoic acid (HABA) was used as a matrix and sodium fluoroacetate (NaTFA) was used as a cationizing agent. A polymer sample, matrix and cationizing agent were dissolved in THF at 5mg mL^{-1}, 30mg mL^{-1} and 2mg mL^{-1} respectively, and, these solutions were mixed in 1:1:1 to 1:5:1 ratio depending on the molecular weight of the analyte. 1 μL of the mixed solution was spotted on a MALDI plate, and dried in the air.

3.6 References

1. Church, G. M.; Gao, Y.; Kosuri, S. Next-Generation Digital Information Storage in DNA. *Science* **2012**, *337*, 1628.

2. Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; LeProust, E. M.; Sipos, B.; Birney, E. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **2013**, *494*, 77–80.
3. Ceze, L.; Nivala, J.; Strauss, K. Molecular digital data storage using DNA. *Nat. Rev. Genet.* **2019**, *20*, 456–466.
4. Organick, L.; Ang, S. D.; Chen. Y.-J.; Lopez, R.; Yekhanin, S.; Makarychev, K.; Racz, M. Z.; Kamath, G.; Gopalan, P.; Nguyen, B.; Takahashi, C. N.; Newman, S.; Paker, H.-Y.; Rashtchian, C.; Stewart, K.; Gupta, G.; Carlson, R.; Mulligan, J.; Carmean, D.; Seelig, G.; Ceze, L.; Strauss, K. Random access in large-scale DNA data storage. *Nat. Biotechnol.* **2018**, *36*, 242–248.
5. Yazdi, S. M. H. T.; Yuan, Y.; Ma, J.; Zhao, H.; Milenkovic, O. A Rewritable, Random-Access DNA-Based Storage System. *Sci. Rep.* **2015**, *5*, 14138.
6. Zhirnov, V.; Zadegan, R. M.; Sandhu, G. S.; Church, G. M.; Hughes, W. L. Nucleic acid memory. *Nat. Mater.* **2016**, *15*, 366–370.
7. Grassm, R. N.; Heckel, R.; Puddu, M.; Paunescu, D.; Stark, W. J. Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes. *Angew. Chen. Int. Ed.* **2015**, *54*, 2552–2555.
8. Erlich, Y.; Zielinski, D. DNA Fountain enables a robust and efficient storage architecture. *Science* **2017**, *355*, 950–954.
9. Kosuri, S.; Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **2014**, *11*, 499–507.
10. Lee, H. H.; Kalhor, R.; Goela, N.; Bolot, J.; Church, G. M. Terminator-free template-independent enzymatic DNA synthesis for digital information storage. *Nat. Commun.* **2019**, *10*, 2383.
11. Rutten, M. G. T. A.; Vaandrager, F. W.; Elemans, J. A. A. W.; Nolte, R. J. M. Encoding information into polymers. *Nat. Rev. Chem.* **2018**, *2*, 365–381.
12. Barnes, J. C.; Ehrlich, D. J. C.; Gao, A. X.; Leibfarth, F. A.; Jiang, Y.; Zhou, E.; Jamison, T. F.; Johnson, J. A. Iterative exponential growth of stereo- and sequence-controlled polymers. *Nat. Chem.* **2015**, *7*, 810–815.
13. Charles, L.; Lutz, J.-F. Design of Abiological Digital Poly(phosphodiester)s. *Acc.*

Chem. Res. **2021**, *54*, 1791–1800.

14. Martens, S.; Landuyt, A.; Espeel, P.; Decreese, B.; Dawyndt, P.; Du Prez, F. Multifunctional sequence-defined macromolecules for chemical data storage. *Nat. Commun.* **2018**, *9*, 4451.
15. Zydziak, N.; Konrad, W.; Feist, F.; Afonin, S.; Weidner, S.; Barner-Kowolik, C. Coding and decoding libraries of sequence-defined functional copolymers synthesized via photoligation. *Nat. Commun.* **2016**, *7*, 13672.
16. Lee, J. M.; Koo, M. B.; Lee, S. W.; Lee, H.; Kwon, J.; Shim, Y. H.; Kim, S. Y.; Kim, K. T. High-density information storage in an absolutely defined aperiodic sequence of monodisperse copolyester. *Nat. Commun.* **2020**, *11*, 56.
17. Al Ouahabi, A.; Kotera, M.; Charles, L.; Lutz, J.-F. Synthesis of monodisperse sequence-coded polymers with chain lengths above DP100. *ACS Macro Lett.* **2015**, *4*, 1077–1080.
18. Solleder, S. C.; Zengel, D.; Wetzel, K. S.; Meier, M. A. R. A scalable and high-yield strategy for the synthesis of sequence-defined macromolecules. *Angew. Chem. Int. Ed.* **2016**, *55*, 1204–1207.
19. Koo, M. B.; Lee, S. W.; Lee, J. M.; Kim, K. T. Iterative Convergent Synthesis of Large Cyclic Polymers and Block Copolymers with Discrete Molecular Weights. *J. Am. Chem. Soc.* **2020**, *142*, 14028–14032.
20. Mlitz, K. "Data center storage capacity worldwide from 2016 to 2021, by segment" *Statista* **2021**.
21. Breen, C. P.; Nambiar, A. M. K.; Jamison, T. F.; Jensen, K. F. Ready, Set, Flow! Automated Continuous Synthesis and Optimization. *Trends Chem.* **2021**, *3*, 373–386.
22. Liu, C.; Xie, J.; Wu, W.; Wang, M.; Chen, W.; Idres, S. B.; Rong, J.; Deng, L.-W.; Khan, S. A.; Wu, J. Automated synthesis of prexasertib and derivatives enabled by continuous-flow solid-phase synthesis. *Nat. Chem.* **2021**, *13*, 451–457.
23. Lin, B.; Hedrick, J. L.; Park, N. H.; Waymouth, R. M. Programmable High-Throughput Platform for the Rapid and Scalable Synthesis of Polyester and Polycarbonate Libraries. *J. Am. Chem. Soc.* **2019**, *141*, 8921–8927.

24. Leibfarth, F. A.; Johnson, J. A.; Jamison, T. F. Scalable synthesis of sequence-defined, unimolecular macromolecules by Flow-IEG. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 10617–10622.

25. Martens, S.; Van den Begin, J.; Madder, A.; Du Prez, F. E.; Espeel, P. Automated Synthesis of Monodisperse Oligomers, Featuring Sequence Control and Tailored Functionalization. *J. Am. Chem. Soc.* **2016**, *138*, 14182–14185.

26. Szweda, R.; Tschopp, M.; Felix, O.; Decher, G.; Lutz, J.-F. Sequences of Sequences: Spatial Organization of Coded Matter through Layer-by-Layer Assembly of Digital Polymers. *Angew. Chem. Int. Ed.* **2018**, *57*, 15817–15821.

27. Hoff, E. A.; De Hoe, G. X.; Mulvaney, C. M.; Hillmyer, M. A.; Alabi, C. A. Thiol−Ene Networks from Sequence-Defined Polyurethane Macromers. *J. Am. Chem. Soc.* **2020**, *142*, 6729–6736.

28. Zhu, J.-B.; Watson, E. M.; Tang, J.; Chen, E. Y.-X. A synthetic polymer system with repeatable chemical recyclability. *Science* **2018**, *360*, 398–403.

Chapter 4

Nondestructive Sequencing of Enantiopure Oligoesters by Nuclear Magnetic Resonance Spectroscopy

(A series of oMPs was synthesized and characterized by J.H.)

4.1 Abstract

Sequence-defined synthetic oligomers and polymers are promising molecular media for permanently storing digital information at densities higher than those of currently used media. However, the information decoding process relies on the degradative sequencing methods such as mass spectrometry, which consumes the information-storing polymers upon decoding. Here we demonstrate the nondestructive decoding of sequence-defined oligomers of enantiopure α-hydroxy acids, oligo(L-lactic-*co*-glycolic acid)s (oLGs) and oligo(L-mandelic-*co*-D-phenyllactic acid)s (oMPs) by ^{13}C nuclear magnetic resonance spectroscopy. We were able to nondestructively decode a bitmap image (192 bits) encoded using a library of 12 equimolar mixtures of an 8-bit-storing oLG and oMP, synthesized through semi-automated flow chemistry in less than 1% of the reaction time required for the repetition of conventional batch reactions. Our results highlight the potential of bundles of sequence-defined oligomers as efficient media for encoding and decoding large-scale information based on the automation of their synthesis and nondestructive sequencing processes.

4.2 Introduction

The storage of information produced by human activities is essential for civilization. The explosion in data production in recent decades demands a corresponding expansion in data storage capacity. However, conventional technologies based on magnetic, optical, and electronic media consume a significant amount of physical space and energy for storing and maintaining information.[1–3]

Long-term storage also requires the periodic refreshment of the data, given the deterioration of media over time.[4] Large-scale information storage in sequence-defined macromolecules has been evolved from curiosity-driven research to a promising alternative to existing information-storage technologies.[5–11] DNA and synthetic sequence-defined polymers can store digital information in their chemical structures using only a few atoms per bit, the unit of digital information.[12–16] Therefore, these macromolecular media can store large amounts of information in proportion to the number of chemically distinguishable repeating units composing polymer chains. In addition, the structural integrity of these information-storing macromolecules can be preserved for an extended period without requiring additional energy for extensive cooling or the periodic refreshment of the stored data.[17,18]

However, the realization of macromolecular media for information storage requires that several key challenges be overcome. In these media, information is encoded in the form of sequences of the monomers constituting the polymer chains through the repetitive coupling of the individual monomers in a stepwise manner.[19–21] Consequently, the encoding of information by chemical synthesis imposes significant cost and time constraints for large-scale information storage in macromolecular media. To overcome this challenge, the parallel synthesis of sequence-defined macromolecules by automated and continuous processes is necessary to accelerate the rate of chemical encoding.[22–26] The decoding of the information stored in sequence-defined macromolecules also presents a challenge. Current methods for the decoding of the information stored in sequence-defined macromolecules rely on destructive techniques such as tandem mass spectrometry, which involves the fragmentation of the parent molecules.[27–31] Consequently, these methods inevitably consume the information-storing polymers during each decoding

attempt, making large-scale or additional synthesis processes necessary for replenishing the polymers, especially when frequent decoding is required. Therefore, nondestructive methods for sequencing synthetic macromolecules must be developed for macromolecular media to become practically suitable for information storage.[32–37]

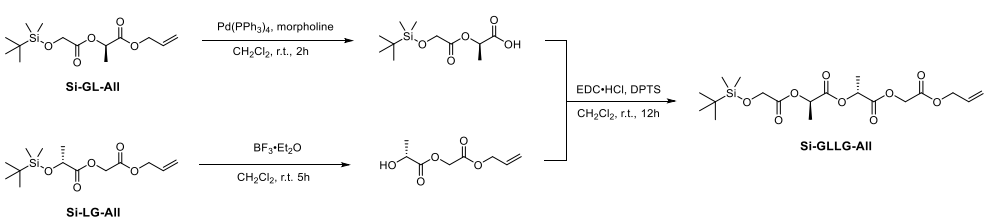
^{13}C NMR spectroscopy can detect the structural differences around the carbon atom of interest. Therefore, it is widely used for the analysis of microstructures (arrangements of enantiomeric repeating units along the polymer backbone) of stereoregular polymers such as poly(L/D-lactide)s.[38–40] However, the results of ^{13}C NMR spectroscopy only show the cumulative populations of the relative orientation of the enantiomeric repeating units constituting the polymer backbone;[41–43] this is especially true for high molecular-weight polylactides with a molecular weight distribution.

Here we report that the spectroscopic sequencing of sequence-defined oligoesters composed of enantiopure α-hydroxy acids. A library of sequence-defined oligomers of the enantiopure α-hydroxy acids, oligo(L-lactic-*co*-glycolic acid)s (oLGs) and oligo(L-mandelic-*co*-D-phenyllactic acid)s (oMPs) was constructed by semi-automated flow chemistry within less than 1% of the reaction time required to prepare the same set of oligomers by conventional batch reactions and the accompanying purification processes. The sequence of each oligoester could be unambiguously decoded from a single ^{13}C nuclear magnetic resonance (NMR) spectrum. In addition, we show that a maximum of 32 bits (4 bytes) of digital information can be stored in an NMR sample containing an equimolar mixture of oLG and oMP and that this information can be decoded by a single ^{13}C NMR measurement based on the nonoverlapping chemical shifts in the sequence-indicating peaks of oLG and oMP. Our results highlight the potential of bundles of

sequence-defined oligomers as efficient media for encoding and decoding large-scale information through the automation of their synthesis and nondestructive sequencing processes.

4.3 Results and discussion

Accelerated synthesis of sequence-defined octamers of α-hydroxy acids by flow chemistry. We employed the cross-convergent approach to synthesize the sequence-defined oligomers in a step-economical manner.[44–48] The cross-convergent approach involves the deconstruction of the target sequence into smaller segments built from building blocks composed of a minimum number of uniquely identifiable monomers. Therefore, the sequence-defined building blocks obtained from the permutation of the monomers covering all the possible sequences of a minimal number of repeating units are prerequisites for the cross-convergent approach (Scheme 4-1). The permutation in which L-lactic acid (L) represents 0 and glycolic acid (G) represents 1 yielded four dyads (LL, LG, GL, and GG), which covers all possible sequences of the dimers of L and G having protective groups for the hydroxyl and carboxyl end groups. The cross-convergence of these dyads would produce 16 tetramers of L and G (tetrads) covering all the possible sequences during the divergent stage of the synthesis, such that the number of possible products is maximized.



Scheme 4-1. Synthesis of sequence-defined oligoesters.

The cross-convergent synthesis of the sequence-defined oligomers and polymers relies on the repetition of a set of chemical reactions, wherein the number of required synthetic steps increases in proportion to the target molecular weight or number of sequence-defined products. To encode information in oligoesters at a rate higher than that of conventional batch processes, we used the semi-automated method to synthesize all the possible tetrads in a continuous flow process[26] (Figure 4-1). The controlled feeding of the desired dyads to the corresponding deprotection reactor line, one for the desilylation of the tert-butyldimethylsilyl (TBDMS) group with trifluoroborane etherate ($BF_3 \cdot Et_2O$) and the other for the allyl transfer reaction to morpholine catalyzed by tetrakis(triphenylphosphine)palladium(0) ($Pd(PPh_3)_4$), was achieved by using a set of the computer-controlled six-way valve systems. The hydroxyl product resulting from the desilylation process was purified with water while the carboxyl product from the deallylation process was purified with a 1M HCl aqueous solution in an in-line extractor. Finally, the two deprotected precursors were converged for esterification while injecting a coupling agent, 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC) and 4-(dimethylamino) pyridinium 4-toluenesulfonate (DPTS).

However, the coupling yield of this fully continuous setup was lower (35–40 % after purification) than that of the batch reaction (typically > 90%) to form an identical tetrad; this was owing to the residual byproducts produced during the deprotection steps, such as allyl morpholinium and PPh_3, which remained in the reaction mixture after the allyl transfer reaction. To counter the detrimental effects of these residual byproducts, we included an off-line purification step for their removal using an automated instrument for silica column chromatography. This step could be completed within 15 min. The reinjection of the purified dyads into the flow reactor for esterification improved the coupling yield to 90% or greater, which was

comparable to that of the batch reaction. The repetition of this semi-automated process allowed the synthesis of 16-tetrad library to be completed on the 0.5-g scale within 24 h.

The resulting tetrads of L and G were subsequently subjected to cross-convergence to form an octameric oMP having the desired sequence representing an 8-bit binary code. The target tetrads encoding 4-bit fragments of information (100 mg) were injected into flow reactor. This was followed by off-line purification via column chromatography. This flow process produced an octameric oLG encoding the target 8-bit information within 1 h; this period included the off-line purification step. The purified oLGs were fully characterized by ^{1}H and ^{13}C NMR spectroscopies, which confirmed their purity. NMR spectra of oLGs are shown in Appendix (A.4.1).

Based on the non-overlapping of the peaks of L-mandelic acid (M) and D-phenyllactic acid (P) units with those of the L and G units in ^{13}C NMR spectroscopy, we also synthesized a sequence-defined oligoester of M and P as an 8-bit-storing molecular medium using the flow chemistry setup described above. The injection of four dyads of M and P having the same protective groups as those of the dyads of L and G into the flow chemistry setup yielded 16 tetrads on the gram scale in 24 h. The purified tetrads, obtained in yields of 88–92%, were subsequently converged to the targeted octameric oMPs using the flow reactor employed for the synthesis of the oLGs.

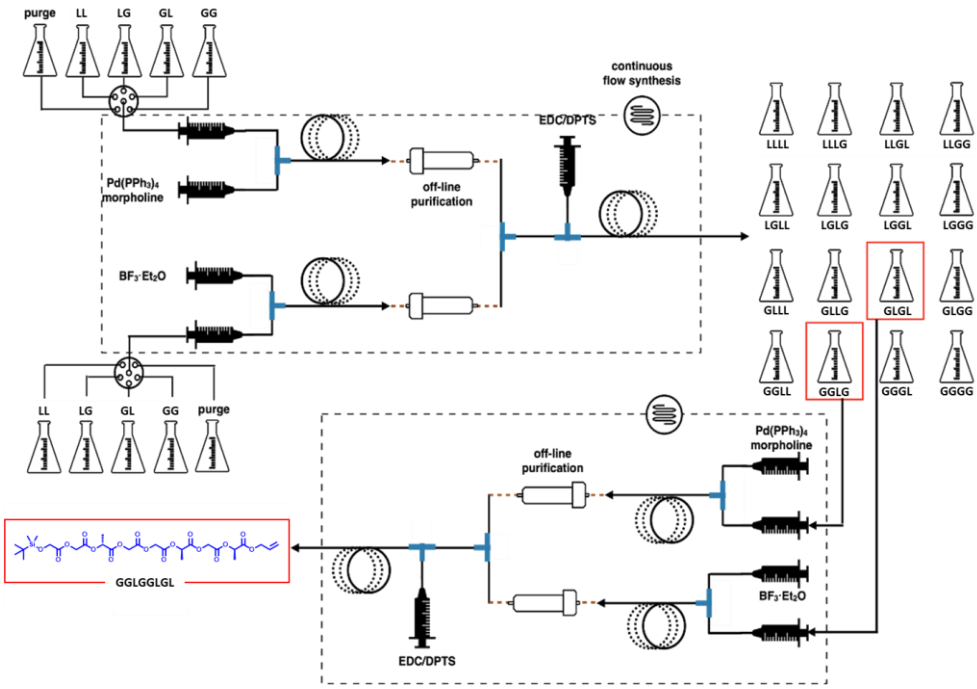


Figure 4-1. Flow chemistry for encoding 8-bit information in sequence-defined oLGs. A continuous flow synthesis of four dyads yielded 16-tetrad library. Subsequently, a sequence-defined octameric oLG, GGLGGLGL, could be obtained by cross-convergence of GGLG and GLGL using semi-automated flow system.

Nondestructive sequencing of oLGs and oMPs by NMR spectroscopy. The use of enantiomeric α-hydroxy acids as monomers for the cross-convergence to oLGs and oMPs render the resulting oligoesters to exhibit an absolutely defined stereochemical configuration. This simplified the analysis of the NMR signals, as it removed the complexities arising from the splitting of the peaks caused by the uncertainly of the chemistry and stereochemical configuration of the neighboring monomers.

We first examined the ^{13}C NMR spectra of the tetrads of L and G to obtain information for decoding the sequence of the constituting monomers of the oMP in relation to the neighboring monomer units. We determined that the peaks corresponding to the α-carbons of the aromatic substituents of L and G could be used for sequencing, as these peaks appeared as singlets at the specific chemical shifts corresponding to the relative position of the TBDMS protective group (Si-terminus) with respect to that of the carboxyl terminus having an allyl protective group (All-terminus). The positions of the monomers were numbered from A to D in the direction from the Si-terminus to the All-terminus. The peaks corresponding to the ipso-carbons of the four L units of the homotetrad LLLL from the Si-terminus to the All-terminus appeared in the order of L(D), L(C), L(B), and L(A) at the chemical shifts of 68.1, 68.7, 69.0, and 69.4 parts per million (ppm), respectively. Similarly, the ^{13}C NMR spectrum of the tetrad GGGG showed a set of α-carbon peaks of the G repeating units in the order of G(A), G(D), G(C), and G(B) at 161.3, 60.2, 60.7, and 61.1 ppm, respectively. With these assignments as references, all tetrads consisting of L and G could be sequenced by ^{13}C NMR (Figure 4-2) without using mass spectrometry. These slight variances in the chemical shifts of the peaks of the repeating units of the oligoesters arising from the chemical structures of the neighboring units were used to estimate the sequence of the repeating units.

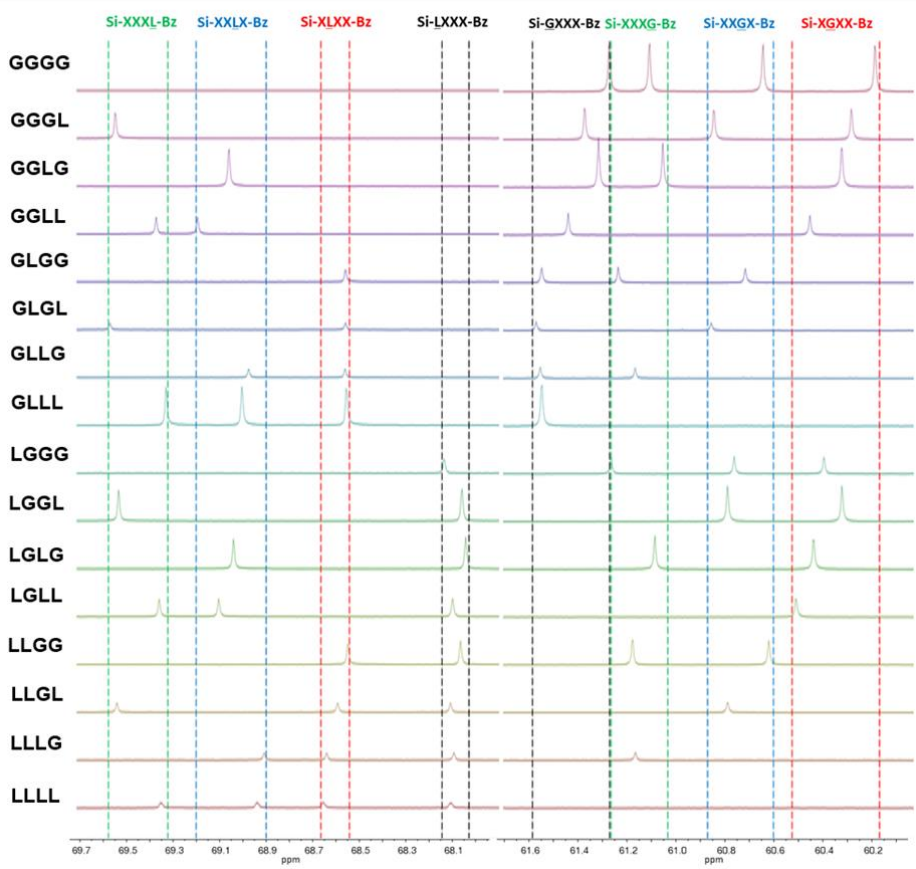


Figure 4-2. (A) ^{13}C NMR spectra of tetramers. In case of the constituent is lactic acid, first digit (Si-LXXX-Bz) is located in 68.18 – 68.07 ppm region; second digit (Si-XLXX-Bz) is located in 68.68 – 68.58 ppm region; third digit (Si-XXLX-Bz) is located in 69.22 – 68.93 ppm region; fourth digit (Si-XXXL-Bz) is located in 69.60 – 69.35 ppm region. In case of the constituent is glycolic acid, first digit (Si-GXXX-Bz) is located in 61.60 – 61.30 ppm region; second digit (Si-XGXX-Bz) is located in 61.29 – 61.08 ppm region; third digit (Si-XXGX-Bz) is located in 60.88 – 60.64 ppm region; fourth digit (Si-XXXG-Bz) is located in 60.54 – 60.21 ppm region.

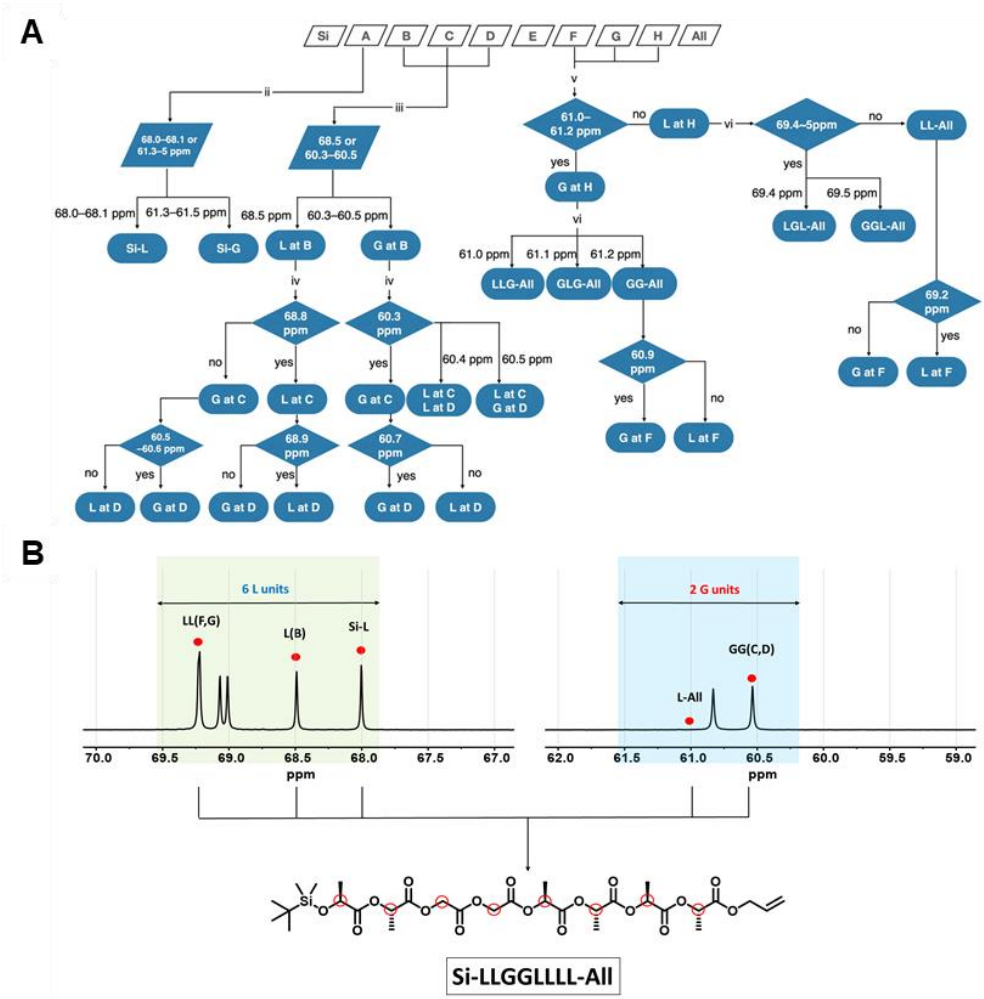


Figure 4-3. (A) Decoding diagrams of sequence-defined octameric oLGs. (B) Decoding of octameric oLGs based on ^{13}C NMR spectrum and deciphered chemical structure of oLG. Red circles indicate the chemical shift which should be checked for sequencing.

The sequencing of octameric oLGs by ^{13}C NMR spectroscopy was expectedly more complicated than the sequencing of tetrads because of the increased number of possible sequences corresponding to the eight peaks appearing within a narrow chemical shift range. Therefore, we established the following rules for decoding the sequence of oLGs based on the electronic shielding/deshielding effects caused by the neighboring monomers. These decoding rules are summarized in Figure 4-3A and the experimental section.

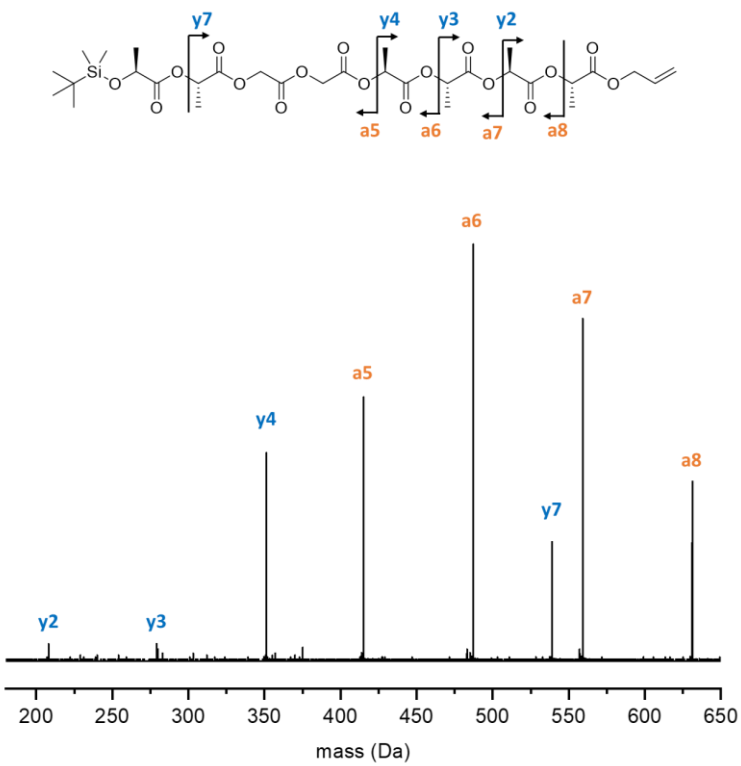


Figure 4-4. ESI-QTOF-MS/MS analysis of an octameric oLG, LLGGLLLL.

As a demonstration in Figure 4-3B, Si-LLGGLLLL-All was sequenced based on the decoding rules, as shown in Figure 4-5B. The number of peaks present in the 68.0–70.0 ppm range suggests that the oLG is composed of 6 L and 2 G units (Rule i). The presence of a peak at 68.0 ppm suggests that the first unit at the A position is L. (Rule iii) suggests that an L unit is present at position B, given the presence of a peak at 68.5 ppm. The absence of a peak at 68.8 ppm suggest that the third repeating unit is G. The appearance of a peak at 60.5 ppm is indicative of a fourth repeating unit of G. (Rule v) suggests that an L unit is present at position H, given the absence of a peak at 61.0–61.2 ppm region. The absence of a peak at 69.4–69.5 ppm range suggest that the seventh repeating unit is L. Finally, the appearance of a peak at 69.2 ppm is indicative of a L unit at position F. Therefore, the final sequence of the oLG was determined to be Si-LLGGLLLL-All, which is in keeping with the proposed structure of the oLG. Moreover, tandem mass sequencing performed using an electrospray ionization (ESI) mass spectrometer yielded a sequence identical to that obtained from the NMR sequencing of the oLG (Figure 4-4). Thus, this nondestructive sequencing method can be applied repeatedly without a loss of the oligoester, which remains intact in the solution. In addition, the solution can be stored in a conventional NMR tube for more than a year.

Similarly, the sequence of an oMP could be decoded by analyzing ^{13}C NMR spectrum to determine the chemical shifts of the peaks corresponding to the *ipso*-carbons of the M and P units composing the sequence-defined octamer. The positions of the monomers of the oMP were assigned from 1 to 8, starting from the Si-terminus to the All-terminus. In a similar manner to the case for the decoding of oLGs, we established the rules for decoding the sequence of oMPs based on the peak of the *ipso*-carbon. These decoding rules are summarized in the experimental section (Figure 4-5A).

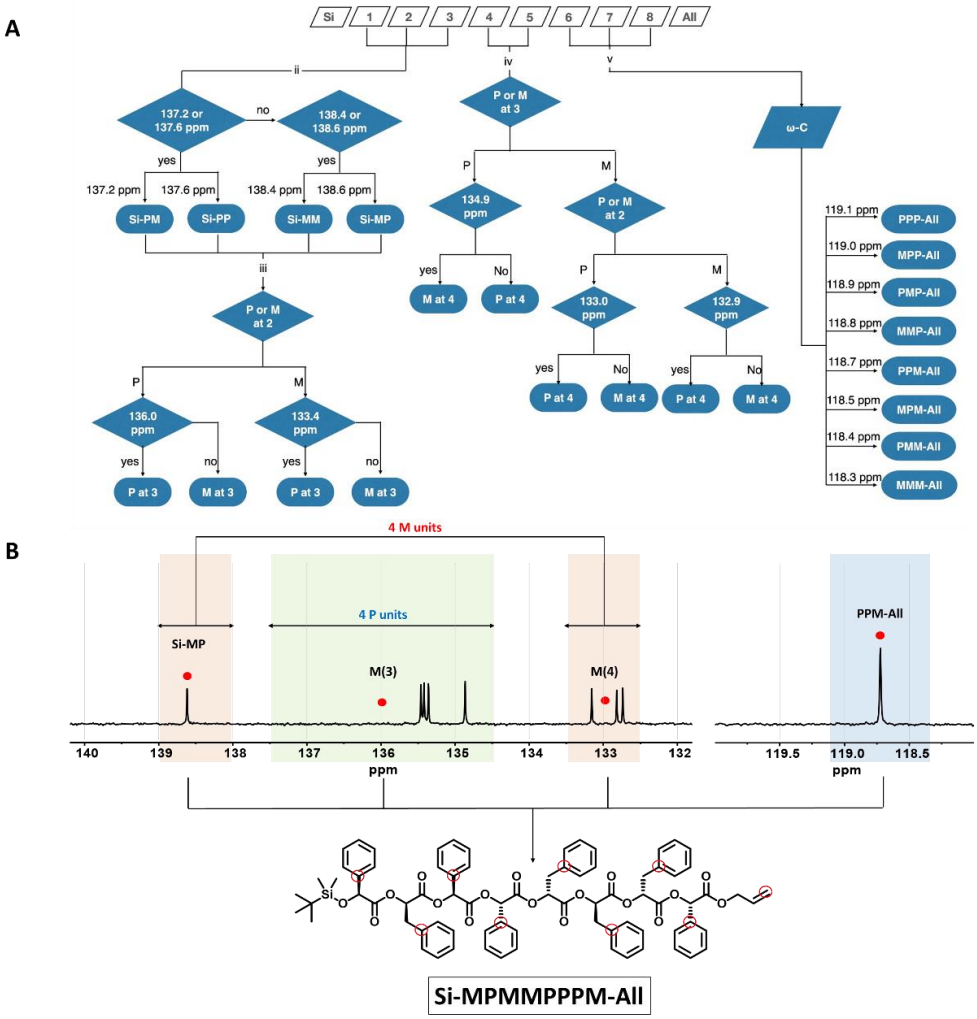


Figure 4-5. (A) Decoding diagrams of sequence-defined octameric oMPs. (B) Decoding of octameric oMPs based on ^{13}C NMR spectrum and deciphered chemical structure of oMP. Red circles indicate the chemical shift which should be checked for sequencing.

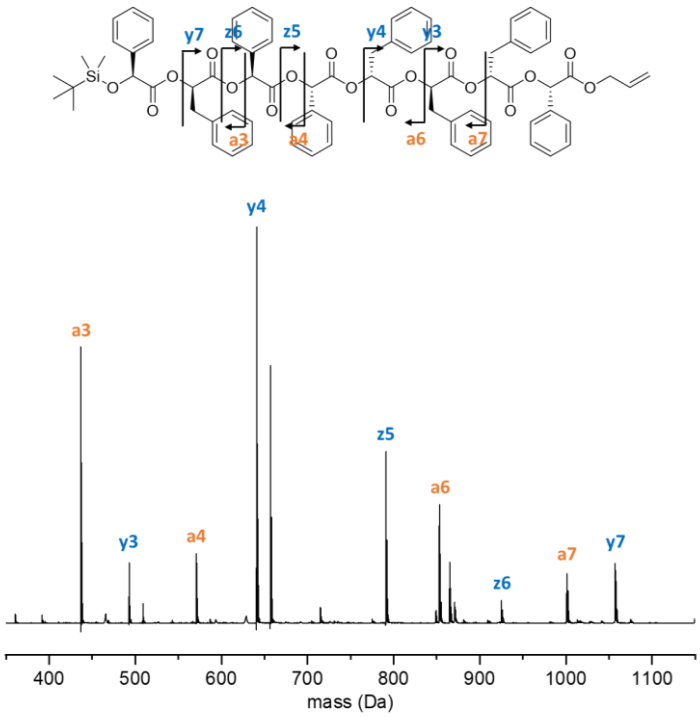


Figure 4-6. MALDI-TOF-MS/MS analysis of an octameric oMP, MPMMPPPM.

Based on these rules, we attempted the NMR sequencing of an oMP, Si-MPMMPPPM-All, as shown in Figure 4-5B. The ^{13}C NMR spectrum exhibited four peaks in the 134.5–137.5 ppm range, indicating that the oMP consists of four M and four P units as per (Rule i). The presence of the peak at 138.6 ppm suggests that the units at positions 1 and 2 are M and P, respectively. (Rule iii) suggests that an M unit is present at position 3, given the absence of a peak at 136.0 ppm. The three units at the All-terminus are P, P, and M units, which are present at positions 6, 7, and 8, respectively. This is based on the peak of ω-carbon of the allyl protective group at 118.7 ppm. Finally, given that no peak was present at 133.0 ppm, (Rule v) suggests that an M unit is at position 4 and a P unit at position 5. The results of the

nondestructive sequencing of the oMP using the above-described rules to decode its ^{13}C spectrum were verified by comparing them with the results of tandem mass sequencing performed using MALDI-TOF/TOF mass spectrometer (Figure 4-6).

Our rules for decoding the sequences of oligoesters based on their ^{13}C NMR spectra indicates that there is no duplication between the 256 possible ^{13}C NMR spectra of oLGs or the 256 possible spectra of oMPs. This one-to-one correspondence between oligoesters and their respective NMR spectra suggests that the sequences of these oligoesters can be determined by comparing the acquired spectrum with a library of the spectra of 8-bit-storing oligoesters. We envisage that this pre-synthesized oligoesters covering all the possible sequences could be used as a pool of 8-bit packets to compose and store large-size digital information that can be readily decoded by nondestructive NMR sequencing.

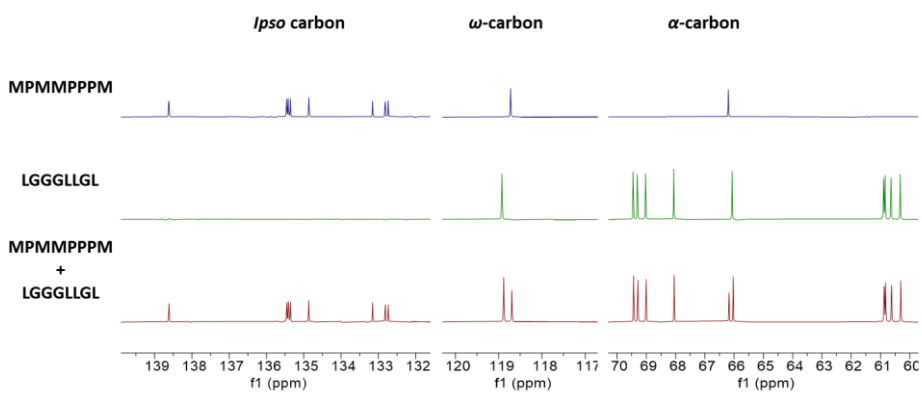


Figure 4-7. ^{13}C NMR spectra of octameric oMP, oLG, and mixture.

Encoding and decoding of digital information in oligoesters. We chose two pairs of enantiomeric α-hydroxy acids, L-mandelic acid/D-phenyllactic acid and L-lactic acid/glycolic acid, to compose oMPs and oLGs with the aim of confirming that there is no overlapping of the peaks of oligoesters having markedly different substituent chemistries. The NMR spectra of these oligoesters showed that their peaks of interest were present in different chemical-shift regions and did not overlap (Figure 4-7). Hence the sequences of an oMP and oLG could be decoded simultaneously from a single NMR spectrum of a mixture of the two oligoesters. Therefore, a mixture of an oMP and oLG can store 16 bits, which can be decoded based on a single ^{13}C NMR measurement of the mixture.

To demonstrate this idea, we attempted the accelerated encoding of information in sequence-defined oligoesters through flow chemistry to show that the nondestructive sequencing of oligoesters by NMR spectroscopy can be exploited for the archival storage and retrieval of digital information (Figure 4-8). A bitmap image (192 bit) was converted into a chemical sequence distributed into 12 sets of oMPs and oLGs. A set of an oMP and oLG, each storing 8 bits, was constructed by repeating the semi-automatic flow processes using tetrads of M and P or L and G as the precursors. The encoding of all 192 bits of information in the library of 12 oMPs and oLGs could be completed within 12 h by running two flow processes in parallel. The encoding time was significantly lower (~ 1%) than the time required to complete the synthesis of the same set of oligoesters by repeating the conventional batch reaction and purification processes.

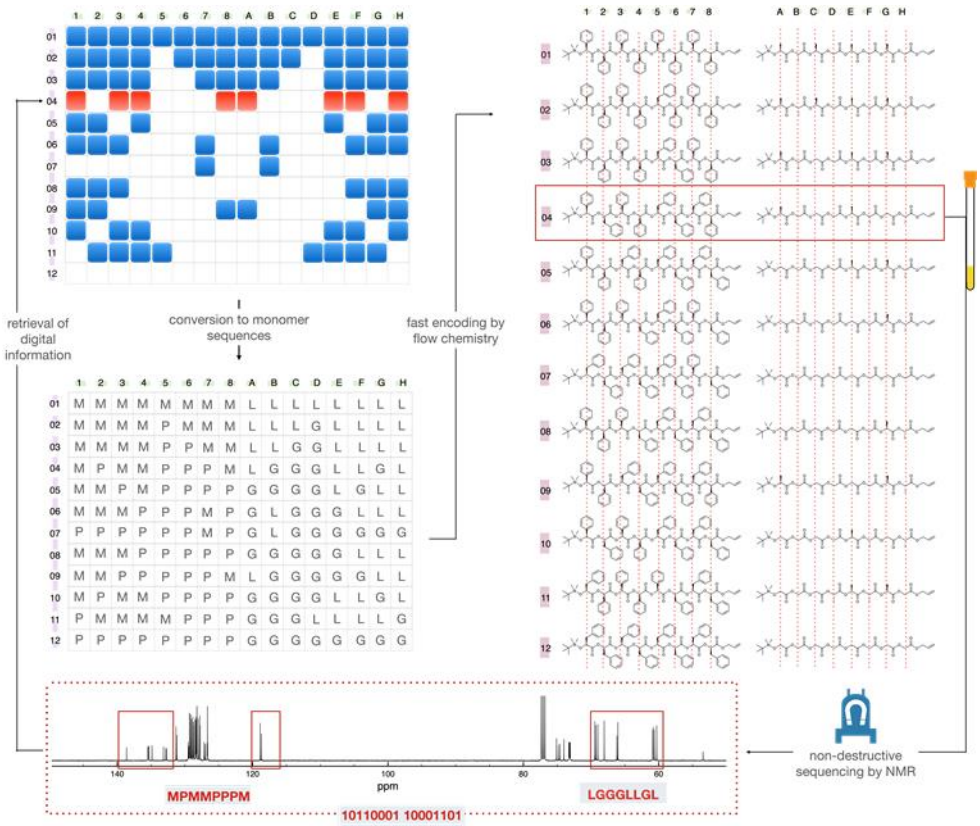


Figure 4-8. Encoding and decoding process of enantiopure oligoesters. Converted bitmap image was encoded into 12 sets of sequence-defined octameric oMP and oLG by semi-automated flow process. Nondestructive decoding of 04 mixture revealed absolute sequence of oMP and oLG, which could be retrieved to digital information (highlighted to red rectangles in bitmap image).

The synthesized oligoesters were grouped into 12 sets of equimolar mixtures of the oMP and oLG (1:2 w/w, 15 μmol), which were dissolved in 0.5 mL $CDCl_3$ and stored in conventional NMR tubes labeled 1–12, respectively. The assigned tube numbers corresponded to the externally given addresses for the 12 sets of 16 bits of information. Each tube was subjected to NMR spectrometer (Varian, 125 MHz for ^{13}C) to acquire the spectrum. The sequencing of the oMP and oLG was completed by using a single spectrum containing the *ipso*-carbon peaks of the oMPs and the α-carbon peaks of the oLG, which did not overlap. The 12 spectra of the mixtures of the oMP and oLG could be decoded completely. The detailed NMR sequencing results of oLGs are represented in Appendix (A.4.2). The stored information could be fully retrieved within 1 h by comparing the recorded spectra acquired by the minimun number of scans by NMR (32 scans for 110 sec) with the existing reference spectra of oMPs and oLGs (Figure 4-9). In addition, the decoding and retrieval of the stored image could be achieved without any reading errors even after 6 months using the same NMR samples. We also note that the storage capacity in the mixture of the oMP and oLG could be expanded by the correlation of two sequences. For example, 4-bit information can be encoded by the correlation of the position 1 of the oMP and the position A of the oLG, which makes the mixture of oMP and oLG to store 32 bits.

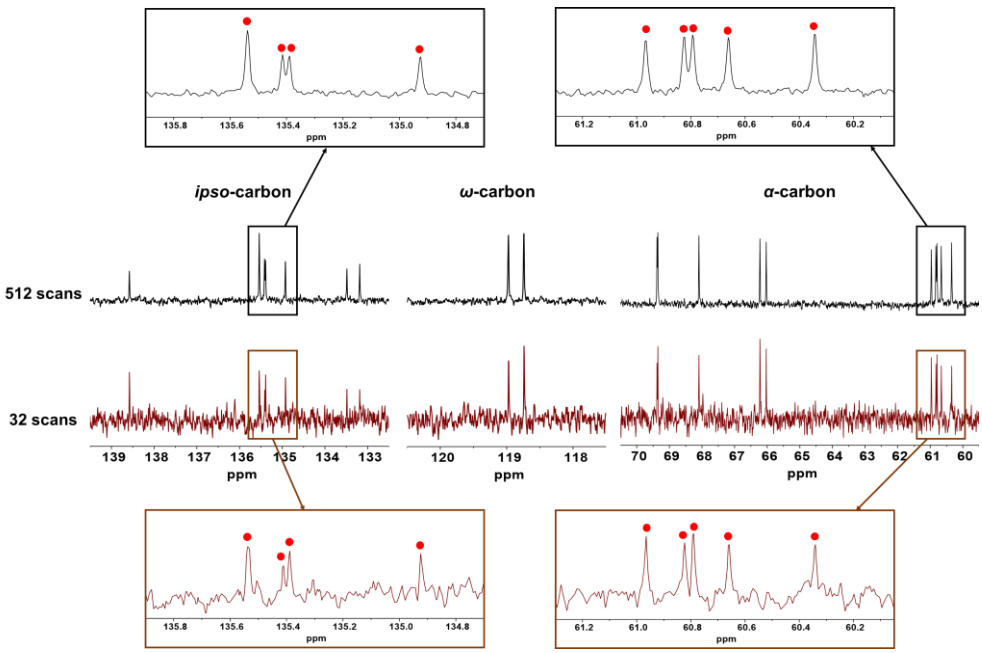


Figure 4-9. ^{13}C NMR spectra in a range of *ipso*-carbons (left), ω-carbon (center), and α-carbons (right) of the octameric mixture 09 with different number of scans. The peaks in spectrum with 32 scans (maroon) could be distinguishable and matched with reference spectrum (black).

4.4 Conclusion

In conclusion, we demonstrated the nondestructive sequencing of oligoesters composed of enantiopure α-hydroxy acids by NMR spectroscopy based on the sequence-specific ^{13}C NMR spectral peaks of absolutely configured oligoesters in the ^{13}C NMR spectra. The sequence-defined octameric oligoesters were synthesized at an accelerated rate by a step-economical cross-convergent synthesis based on semi-automated flow chemistry while using a feed of sequence-defined dyads and tetrads of M and P or L and G. The flow chemistry-based synthesis of the information-storing oligoesters accelerates the rate of encoding by a factor of 100 compared with the rate of synthesis for conventional batch processes. The synthesized sequence-defined octaesters, oMPs and oLGs, could both store 8-bit information. The ^{13}C NMR spectra of the oMPs and oLGs contained the peaks arising from the enantiopure monomers with respect to their relative positions between the Si- and All-protective groups. This, in turn, allowed for sequencing without the degradation of the information-storing molecules. We also demonstrated that the nondestructive decoding of the information-storing oligoesters can be combined with accelerated encoding through flow chemistry to allow for the permanent storage of digital information without requiring any additional energy or synthesis processes for maintaining the stored information. Thus, our results suggest that the bundles of sequence-defined oligomers can serve as efficient media for storing large-scale information based on their automated synthesis and nondestructive sequencing.

4.5 Experimental

Materials and Methods. All other reagents and chemicals were purchased from commercial sources and used without further purification. Dichloromethane (DCM), was distilled over CaH_2 under N_2.

Legato 101 syringe pump was purchased from Kd Scienctific. Cadent 3TM syringe pump was purchased from IMI Norgren. SEP-10 was purchased from Zaiput Flow Technologies. PFA Tubing (1/16″ OD/0.02″ and 0.03″ ID) was purchased from Revodix. Omnifit EZ column was purchased from Diba Industries Inc. Gastorr AG-42-01 was purchased from GL Sciences. CF-2 Fraction Collector was purchased from Spectrum Chemical Mfg. Corp. Automated column chromatography was performed on a Biotage Selekt flash chromatography purification system equipped with a Sfär silica column cartridge. Hexane and ethyl acetate were used as eluent. ^{1}H NMR and ^{13}C NMR spectra were recorded on a Varian INOVA 500 MHz NMR spectrometer in $CDCl_3$. ESI-MS analyses were performed on a SCIEX TripleTOF 5600. MALDI-TOF MS/MS analyses were performed on a Bruker Ultraflex TOF/TOF mass spectrometer. For MALDI-TOF MS/MS analysis, 2-(4-Hydroxyphenylazo)benzoic acid (HABA) used as a matrix and sodium trifluoroacetate (NaTFA) was used as a cationizing agent. The instrument was operated in positive reflector mode. oMPs were fragmented y*i* fragment containing the original omega group (allyl ester) and a new alkene terminus corresponding a*i* fragment having the original alpha group (TBDMS) and a new carboxylic acid, or z*i* fragment containing the original omega group and a new aldehyde terminus. oLGs were fragmented y*i* fragment corresponding a*i* fragment.

Continuous flow synthesis for sequence-defined oligoesters. A continuous flow system consists of synthetic step and cleaning step, which is operated via 6-way syringe pumps connected with programmable controller. 2 mmol of a dyad (1 M in

DCM) and BF_3•Et_2O (7 M in DCM) were injected at 0.1 mL/min and mixed through T-mixer. The mixture was allowed to react in the reaction loop (volume of 2 mL). Simultaneously, 2 mmol of a dyad (1 M in THF) and $Pd(PPh_3)_4$ and morpholine mixture (0.03 M and 1.05 M in THF) were injected at 0.1 mL/min and mixed through T-mixer. The mixture was allowed to react in the reaction loop (volume of 2 mL). Deprotected dyads, hydroxyl and carboxylic acid, were purified by automated flash column chromatography using HEX/EA and ether/MeOH eluents, respectively. Subsequently, the purified dyads (1 M in DCM, 0.1 mL/min) were reinjected and mixed with EDC•HCl and DPTS (0.7 M and 0.07 M in DCM, 0.2 mL/min). The mixture was allowed to react in reaction loop (volume of 6 mL). After synthetic step, DCM solvent (6 mL) was purged to flow reactor for cleaning reaction loop. Synthetic and cleaning cycle were repeated 16 times for generating every permutation of tetrads. Collected tetrads were purified with automated flash column chromatography (86–92% yield).

Synthesis of octameric oligoesters by semi-automated flow system. 200 mg of a tetrad (0.25–0.5 M in DCM) and BF_3·Et_2O (7 M in DCM) were injected at 0.1 mL/min and mixed through T-mixer. The mixture was allowed to react in the reaction loop (volume of 2 mL). Simultaneously, 200 mg of a tetrad (0.25–0.5 M in THF) and $Pd(PPh_3)_4$ and morpholine mixture (0.03 M and 1.05 M in THF) were injected at 0.1 mL/min and mixed through T-mixer. The mixture was allowed to react in the reaction loop (volume of 2 mL). Deprotected tetrads were purified by automated flash column chromatography. The purified tetrads (0.25–0.5 M in DCM, 0.1 mL/min) were reinjected and mixed with EDC·HCl and DPTS (0.7 M and 0.07 M in DCM, 0.3 mL/min). The mixture was allowed to react in reaction loop (volume of 9 mL). Collected octameric oligoester was purified with automated flash column chromatography (85–91% yield).

^{13}C NMR sequencing of octameric oMPs. (Rule i) The number of M units in the oMP can be determined by counting the number of peaks in the chemical-shift regions of 132.5–133.5 ppm and 138–139 ppm. The number of peaks in the range of 134.5–137.5 ppm range is identical to the number of P units (Figure S3); (Rule ii) The monomer units at positions 1 and 2 can be determined based on the peak of the *ipso*-carbon that is deshielded the most by a silicon atom present in its proximity (Figure S4). In a range of 138.4–138.6 ppm, the peak corresponding to the ipso-carbon of Si-M(1), which neighbors M(2), appears downfield to that of the ipso-carbon that neighbors P(2). Similarly, the peak related to the *ipso*-carbon of Si-P(1) appears in the range of 137.24–60 ppm owing to the adjacent monomer units; (Rule iii) The presence of a P unit at position 3 can be predicted based on the peak of the monomer unit at position 2 (Figure S5). The peak corresponding to P(3) appears at 133.4 ppm in the case of with M(2) and 135.9 ppm in the case of P(2); (Rule iv) The ω-carbon peak of the allyl protective group appears at the characteristic chemical shift in the range of 118.36–119.13 ppm because of the sequence of the monomer units at positions 6–8 (Figure S6); (Rule v) The monomer units on positions 4 and 5 can be predicted by an indirect method (Figure S7). With P at position 3, the presence of the peak at 134.9 ppm is indicative of M(4). In the case of M(3) and P(2), the peak at 133.0 ppm suggests that the P unit is at position 4. In the case of M(3) and M(2), the peak at 132.9 ppm suggests that the P unit is at position 4.

^{13}C NMR sequencing of octameric oLGs. A compound protected with TBDMS and benzyl groups was dissolved in dry DCM. The solution was cooled to 0°C on an ice bath, and BF_3•Et_2O was added dropwise. The reaction mixture was stirred at room temperature for 4 h. The reaction was monitored by TLC analysis. Upon completion of the reaction, the reaction was quenched with saturated $NaHCO_3$ followed by dilution with water. The organic layer was separated and washed with brine. The

combined organic layer was dried over $MgSO_4$, and the solvent was removed under reduced pressure. The crude product was purified by automated column chromatography.

General procedure for esterification reactions. (Rule i) The number of L units in the oLG can be determined by counting the number of peaks in the chemical-shift regions of 67.9–69.5 ppm. The number of peaks in the range of 60.2–61.5 ppm range is identical to the number of G units (Figures S9 and S10); (Rule ii) The monomer unit at position A can be determined based on the presence of the peak of the α-carbon at specific chemical shift (Figures S11 and S12). At 68.0–68.1 ppm, the peak corresponding to the α-carbon of L(A) appears and G(A) appears at 61.3−61.5 ppm; (Rule iii) The monomer unit at position B can also be revealed depending on the presence of the peak at specific regions (Figures S11 and S12). The peak corresponding to the α-carbon of L(B) appears at 68.5 ppm and that of G(B) appears at 60.3–5 ppm; (Rule iv) The monomer units at position C and D are predicted by an indirect method (Figure S13). When the L unit is at position B, the peak at 68.8 ppm suggests that the L unit is at position C. In the case of L(B) and G(C), the peak at 60.5–6 ppm suggests that the G unit is at position D. In the case of L(B) and L(C), the peak at 68.9 ppm suggests that the L unit is at position D. The α-carbon of G(B) appears at 60.3, 60.4 or 60.5 ppm as the case of neighboring G(C), (L(C) and L(D)) or (L(C) and G(D)), respectively. When the G units are at position B and C, the peak at 60.7 ppm suggests that the G unit is at position D; (Rule v) The monomer unit at position H is predicted by the presence of the α-carbon at specific chemical shift (Figure S14). The peak corresponding to the α-carbon of G(H) appears at 61.0–2 ppm; (Rule vi) the monomer units at position F and G are predicted by an indirect method (Figure S14). The α-carbon of G(H) appears at 61.0, 61.1 or 61.2 ppm as the case of neighboring (L(F) and L(G)), (G(F) and L(G)) or G(G), respectively. The

peak corresponding to the α-carbon of L(H) appears at 69.4 or 69.5 ppm as the case of neighboring (L(F) and G(G)) or (G(F) and G(G)), respectively. In the case of G(G) and G(H), the peak at 60.9 ppm suggests that the G unit is at position F. In the case of L(G) and L(H), the peak at 69.2 ppm suggests that the L unit is at position F.

4.6 References

1. Shulaker, M. M.; Hills, G.; Park, R. S.; Howe, R. T.; Saraswat, K.; Wong, H.-S. P.; Mitra, S. Three-dimensional integration of nanotechnologies for computing and data storage on a single chip. *Nature* **2017**, *547*, 74–78.
2. Lloyd, S. Ultimate physical limits to computation. *Nature* **2000**, *406*, 1047–1054.
3. Salahuddin, S.; Ni, K.; Datta, S. The era of hyper-scaling in electronics. *Nat. Electron.* **2018**, *1*, 442–450.
4. Wallace, C.; Pordesch, U.; Brandner, R. Long-term archive service requirements. *Request For Comments–RFC* **2007**, 4810.
5. Ceze, L.; Nivala, J.; Strauss, K. Molecular digital data storage using DNA. *Nat. Rev. Genet.* **2019**, *20*, 456–466.
6. Church, G. M.; Gao, Y.; Kosuri, S. Next-generation digital information storage in DNA. *Science* **2012**, *337*, 1628–1628.
7. Zhirnov, V.; Zadegan, R. M.; Sandhu, G. S.; Church, G. M.; Hughes, W. L. Nucleic acid memory. *Nat. Mater.* **2016**, *15*, 366–370.
8. Rutten, M. G.; Vaandrager, F. W.; Elemans, J. A.; Nolte, R. J. Encoding information into polymers. *Nat. Rev. Chem.* **2018**, *2*, 365–381.
9. Martens, S.; Landuyt, A.; Espeel, P.; Devreese, B.; Dawyndt, P.; Du Prez, F. Multifunctional sequence-defined macromolecules for chemical data storage. *Nat. Commun.* **2018**, *9*, 4451.
10. Charles, L.; Lutz, J.-F., Design of Abiological Digital Poly (phosphodiester)s. *Acc. Chem. Res.* **2021**, *54*, 1791–1800.
11. Laurent, E.; Amalian, J.-A.; Parmentier, M.; Oswald, L.; Al Ouahabi, A.; Dufour,

F.; Launay, K.; Clément, J.-L.; Gigmes, D.; Delsuc, M.-A. High-capacity digital polymers: storing images in single molecules. *Macromolecules* **2020**, *53*, 4022–4029.

12. Kosuri, S.; Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **2014**, *11*, 499–507.
13. Lee, H. H.; Kalhor, R.; Goela, N.; Bolot, J.; Church, G. M. Terminator-free template-independent enzymatic DNA synthesis for digital information storage. *Nat. Commun.* **2019**, *10*, 2383.
14. Al Ouahabi, A.; Charles, L.; Lutz, J.-F. Synthesis of non-natural sequence-encoded polymers using phosphoramidite chemistry. *J. Am. Chem. Soc.* **2015**, *137*, 5629–5635.
15. Lee, J. M.; Koo, M. B.; Lee, S. W.; Lee, H.; Kwon, J.; Shim, Y. H.; Kim, S. Y.; Kim, K. T. High-density information storage in an absolutely defined aperiodic sequence of monodisperse copolyester. *Nat. Commun.* **2020**, *11*, 56.
16. Dong, R.; Liu, R.; Gaffney, P. R.; Schaepertoens, M.; Marchetti, P.; Williams, C. M.; Chen, R.; Livingston, A. G. Sequence-defined multifunctional polyethers via liquid-phase synthesis with molecular sieving. *Nat. Chem.* **2019**, *11*, 136–145.
17. Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; LeProust, E. M.; Sipos, B.; Birney, E. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **2013**, *494*, 77–80.
18. Grass, R. N.; Heckel, R.; Puddu, M.; Paunescu, D.; Stark, W. J. Robust chemical preservation of digital information on DNA in silica with error-correcting codes. *Angew. Chem. Int. Ed.* **2015**, *54*, 2552–2555.
19. Solleder, S. C.; Zengel, D.; Wetzel, K. S.; Meier, M. A. A scalable and high-yield strategy for the synthesis of sequence-defined macromolecules. *Angew. Chem. Int. Ed.* **2016**, *55*, 1204–1207.
20. Al Ouahabi, A.; Kotera, M.; Charles, L.; Lutz, J.-F. Synthesis of monodisperse sequence-coded polymers with chain lengths above DP100. *ACS Macro Lett.* **2015**, *4*, 1077–1080.
21. Wetzel, K. S.; Frölich, M.; Solleder, S. C.; Nickisch, R.; Treu, P.; Meier, M. A.

Dual sequence definition increases the data storage capacity of sequence-defined macromolecules. *Commun. Chem.* **2020**, *3*, 63.

22. Breen, C. P.; Nambiar, A. M.; Jamison, T. F.; Jensen, K. F. Ready, set, flow! Automated continuous synthesis and optimization. *Trends Chem.* **2021**, *3*, 373–386.
23. Leibfarth, F. A.; Johnson, J. A.; Jamison, T. F. Scalable synthesis of sequence-defined, unimolecular macromolecules by Flow-IEG. *Proc. Natl. Acad. Sci.* **2015**, *112*, 10617–10622.
24. Martens, S.; Van den Begin, J.; Madder, A.; Du Prez, F. E.; Espeel, P. Automated synthesis of monodisperse oligomers, featuring sequence control and tailored functionalization. *J. Am. Chem. Soc.* **2016**, *138*, 14182–14185.
25. Van De Walle, M.; De Bruycker, K.; Junkers, T.; Blinco, J. P.; Barner-Kowollik, C. Scalable Synthesis of Sequence-Defined Oligomers via Photoflow Chemistry. *ChemPhotoChem* **2019**, *3*, 225–228.
26. Lee, J. M.; Kwon, J.; Lee, S. J.; Jang, H.; Kim, D.; Song, J.; Kim, K. T. Semiautomated synthesis of sequence-defined polymers for information storage. *Sci. Adv.* **2022**, *8*, eabl8614.
27. Soete, M.; Mertens, C.; Aksakal, R.; Badi, N.; Du Prez, F. Sequence-Encoded Macromolecules with Increased Data Storage Capacity through a Thiol-Epoxy Reaction. *ACS Macro Lett.* **2021**, *10*, 616–622.
28. Zydziak, N.; Konrad, W.; Feist, F.; Afonin, S.; Weidner, S.; Barner-Kowollik, C. Coding and decoding libraries of sequence-defined functional copolymers synthesized via photoligation. *Nat. Commun.* **2016**, *7*, 13672.
29. Launay, K.; Amalian, J. a.; Laurent, E.; Oswald, L.; Al Ouahabi, A.; Burel, A.; Dufour, F.; Carapito, C.; Clément, J. l.; Lutz, J.-F. Precise Alkoxyamine Design to Enable Automated Tandem Mass Spectrometry Sequencing of Digital Poly (phosphodiester)s. *Angew. Chem. Int. Ed.* **2021**, *133*, 930–939.
30. Huang, Z.; Shi, Q.; Guo, J.; Meng, F.; Zhang, Y.; Lu, Y.; Qian, Z.; Li, X.; Zhou, N.; Zhang, Z. Binary tree-inspired digital dendrimer. *Nat. Commun.* **2019**, *10*, 1918.

31. Song, B.; Lu, D.; Qin, A.; Tang, B. Z. Combining Hydroxyl-Yne and Thiol-Ene Click Reactions to Facilely Access Sequence-Defined Macromolecules for High-Density Data Storage. *J. Am. Chem. Soc.* **2022**, *144*, 1672–1680.
32. Mutlu, H.; Lutz, J.-F. Reading polymers: sequencing of natural and synthetic macromolecules. *Angew. Chem. Int. Ed.* **2014**, *53*, 13010–13019.
33. Reiner, J. E.; Kasianowicz, J. J.; Nablo, B. J.; Robertson, J. W. Theory for polymer analysis using nanopore-based single-molecule mass spectrometry. *Proc. Natl. Acad. Sci.* **2010**, *107*, 12080–12085.
34. Zhu, Z.; Cardin, C. J.; Gan, Y.; Colquhoun, H. M. Sequence-selective assembly of tweezer molecules on linear templates enables frameshift-reading of sequence information. *Nat. Chem.* **2010**, *2*, 653–660.
35. Zhu, Z.; Cardin, C. J.; Gan, Y.; Murray, C. A.; White, A. J.; Williams, D. J.; Colquhoun, H. M. Conformational modulation of sequence recognition in synthetic macromolecules. *J. Am. Chem. Soc.* **2011**, *133*, 19442–19447.
36. Cafferty, B. J.; Ten, A. S.; Fink, M. J.; Morey, S.; Preston, D. J.; Mrksich, M.; Whitesides, G. M. Storage of information using small organic molecules. *ACS Cent. Sci.* **2019**, *5*, 911–916.
37. Nagarkar, A. A.; Root, S. E.; Fink, M. J.; Ten, A. S.; Cafferty, B. J.; Richardson, D. S.; Mrksich, M.; Whitesides, G. M. Storing and Reading Information in Mixtures of Fluorescent Molecules. *ACS Cent. Sci.* **2021**, *7*, 1728–1735.
38. Zell, M. T.; Padden, B. E.; Paterick, A. J.; Thakur, K. A.; Kean, R. T.; Hillmyer, M. A.; Munson, E. J. Unambiguous determination of the 13C and 1H NMR stereosequence assignments of polylactide using high-resolution solution NMR spectroscopy. *Macromolecules* **2002**, *35*, 7700–7707.
39. Ovitt, T. M.; Coates, G. W. Stereochemistry of lactide polymerization with chiral catalysts: new opportunities for stereocontrol using polymer exchange mechanisms. *J. Am. Chem. Soc.* **2002**, *124*, 1316–1326.
40. Suganuma, K.; Horiuchi, K.; Matsuda, H.; Cheng, H.; Aoki, A.; Asakura, T. Stereoregularity of poly (lactic acid) and their model compounds as studied by NMR and quantum chemical calculations. *Macromolecules* **2011**, *44*, 9247–9253.

41. Stayshich, R. M.; Meyer, T. Y. New insights into poly (lactic-co-glycolic acid) microstructure: using repeating sequence copolymers to decipher complex NMR and thermal behavior. *J. Am. Chem. Soc.* **2010**, *132*, 10920–10934.
42. Bluhm, T. L.; Hamer, G. K.; Marchessault, R. H.; Fyfe, C. A.; Veregin, R. P. Isodimorphism in bacterial poly (β-hydroxybutyrate-co-β-hydroxyvalerate). *Macromolecules* **1986**, *19*, 2871–2876.
43. Cais, R.; O'Donnell, J.; Bovey, F. Copolymerization of styrene with sulfur dioxide. Determination of the monomer sequence distribution by carbon-13 NMR. *Macromolecules* **1977**, *10*, 254–260.
44. Barnes, J. C.; Ehrlich, D. J.; Gao, A. X.; Leibfarth, F. A.; Jiang, Y.; Zhou, E.; Jamison, T. F.; Johnson, J. A. Iterative exponential growth of stereo-and sequence-controlled polymers. *Nat. Chem.* **2015**, *7*, 810–815.
45. Koo, M. B.; Lee, S. W.; Lee, J. M.; Kim, K. T. Iterative Convergent Synthesis of Large Cyclic Polymers and Block Copolymers with Discrete Molecular Weights. *J. Am. Chem. Soc.* **2020**, *142*, 14028–14032.
46. Takizawa, K.; Tang, C.; Hawker, C. J. Molecularly defined caprolactone oligomers and polymers: synthesis and characterization. *J. Am. Chem. Soc.* **2008**, *130*, 1718–1726.
47. Franz, N.; Menin, L.; Klok, H. A. A Post-Modification Strategy for the Synthesis of Uniform, Hydrophilic/Hydrophobic Patterned α-Hydroxy Acid Oligomers. *Eur. J. Org. Chem.* **2009**, *2009*, 5390–5405.
48. van Genabeek, B.; de Waal, B. F.; Gosens, M. M.; Pitet, L. M.; Palmans, A. R.; Meijer, E. Synthesis and self-assembly of discrete dimethylsiloxane–lactic acid diblock co-oligomers: the dononacontamer and its shorter homologues. *J. Am. Chem. Soc.* **2016**, *138*, 4210–4218.

Chapter 5

Synthesis of Enantiomeric ω-Substituted Hydroxyalkanoates from Terminal Epoxides: Building Blocks for Sequence-defined Polyesters and Macromolecular Engineering

(K.D. and S.J. contributed to the synthesis of HAs)

5.1 Abstract

Polyesters with protein-like absolute atomic precision have unlimited potential as functional polymers and biomaterials. However, a library of enantiomeric monomers having a set of functional groups has to be constructed to generate polyesters in a sequence-defined manner. Herein, we report the synthesis of enantiomeric ω-substituted hydroxyalkanoates (HAs) from terminal epoxides and alkenes as starting materials. Our synthesis allows to synthesize a library of HAs having a well-defined atomic composition (carbon number), substituent chemistry, and stereochemical configuration, which serves as building blocks for discrete polyhydroxyalkanoates (PHAs) without molecular-weight distribution and sequence-defined PHAs. Moreover, high molecular weight sequence-regulated PHAs were obtained by the polycondensation of sequence-defined oligo(hydroxyalkanoate) (oHA) building blocks, which allows their crystallinity and thermal properties to be controlled by manipulating the stereoregularity and introducing medium-length chains into the oHA. Our results could afford a facile method to engineer PHAs and related polyesters with structural sophistication exhibited only by biopolymers, such as proteins and nucleic acids.

5.2 Introduction

Poly(hydroxyalkanoate) (PHA)[1–4] is a class of aliphatic polyesters produced naturally by microbes as an energy and carbon reserve. These bio-derived aliphatic polyesters have been attracting attention as degradable substitutes for hydrocarbon-based polymers that are used as plastic and packaging materials.[5–7] For example, poly(3-hydroxybutyrate) (P3HB) is a highly crystalline polymer (T_m = 173–180 ℃)

that can be hydrolyzed to the corresponding monomer via chemical or enzymatic catalysis.[8,9] However, the application of PHAs is hindered by their monotonous chemical structure arising from the limited number of monomers for biosynthesis.[10] To tailor the properties of PHAs for a desired application, a library of hydroxyalkanoates (HAs), containing diverse substituents, is required to be used as monomers for incorporation in PHAs in a well-defined manner.[10,11] Although fermentation via genetically engineered microbes allows the conversion of carbohydrates and lipids to substituted HAs, biosynthesis of PHAs with diverse side chains is challenging because of the limited number of engineered microbes available that can be employed for this purpose.[2,12]

The chemical synthesis of PHAs with desired functions and optimized properties using enantiomeric HAs with functional groups as valuable building blocks is attracting increasing attention. This endeavor, referred to as macromolecular engineering,[13] mandates the precise synthesis of polymers to tailor their structures and functions for the desired applications.[14–18] For macromolecular engineering, the controlled polymerization of functional monomers, along with organic synthesis, has been used as the essential method for optimizing chemical structures of the resulting polymers and copolymers.[19,20] The ring-opening polymerization of the lactones of HAs results in PHAs with a well-defined molecular weight and molecular-weight distribution.[21–23] In particular, Lewis-acid lanthanum catalysts with chiral ligands produced stereoregular PHAs with well-defined stereosequence, which have been demonstrated with a limited set of monomers.[23–25] A general method for the synthesis of enantiopure HAs with diverse substituents, a foundation of macromolecular engineering of PHAs, has not been established yet. Furthermore, the biochemical and chemical methods for synthesizing PHAs depend on the polymerization of a set of substituent-bearing monomers, which inevitably

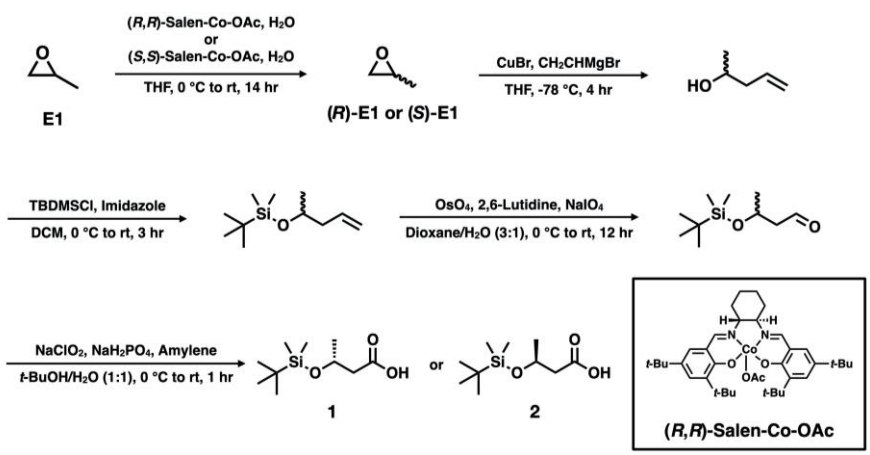
leads to the statistical distribution of the molecular weight, sequence of substituents, stereoregularity, and nature and location of the functional groups of the resulting polyesters.[13,26] We envisaged that controlling the chemical structures of PHAs with sophistication comparable to that of biological polymers, such as proteins, could be a powerful method for developing new biomaterials and polymeric materials.[26–28]

Here we report the synthesis of a library of HAs with the desired functional groups, stereochemical configurations, and atomic compositions using readily available terminal epoxides. The stereochemistry of the stereocenter and the number of carbon atoms of the monomer backbone can be determined. The resulting enantiopure HAs serve as building blocks for synthesizing discrete PHAs with absolutely-defined sequences and stereoregularities via iterative convergent method and cross-convergent method.[29] The sequence-defined oligo(hydroxyalkanoate) can also be used to synthesize sequence-regulated polyesters by condensation polymerization. These synthetic approaches allow us to control the structural parameters of the resulting polyesters, including their molecular weight, monomer-sequence, stereoregularity, and functionality.[13,28,30,31]

5.3 Results and discussion

Synthesis of enantiomeric HAs from terminal epoxides and alkenes. In this study, we define discrete PHAs as aliphatic polyesters composed of a predetermined number of HAs, bearing the desired substituents on ω-carbon. We aimed to establish a library of enantiopure HAs with the desired substituents that could be used for the convergent synthesis of sequence-defined PHAs. We began our synthesis by first determining the atomic composition (carbon number), substituent chemistry, and stereochemical configuration of HAs. Considering that various terminal epoxides are easily available, we devised a simple conversion of racemic mixtures of terminal

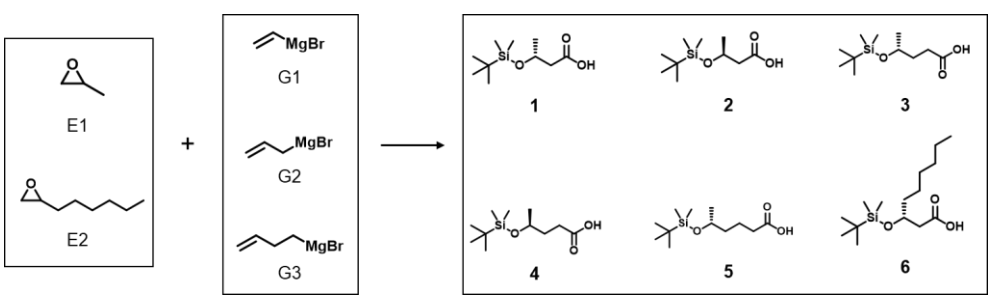
epoxides to enantiopure ω-substituted HAs, containing a desired functional group. We tested our strategy using commercially available racemic propylene oxide (E1) as a precursor (Scheme 5-1). The hydrolytic kinetic resolution (HKR) of racemic E1 was performed with an (*R,R*)-salen-Co complex[32–34] (Jacobson's catalyst) to selectively ring-open (*S*)-E1, and followed by the isolation of intact (*R*)-E1 by distillation (42.0% yield after isolation). Enantiomeric (*R*)-E1 was subjected to the regioselective ring-opening reaction[35] with vinylmagnesium bromide, which yielded (*R*)-4-penten-2-ol. After installing the protective *tert*-butyldimethylsilyl (TBDMS) group on the secondary alcohol, the resulting terminal alkene was oxidized to carboxylic acid via two-step procedures involving Lemieux–Johnson oxidation[36,37] of alkene to aldehyde with OsO_4 and subsequent Pinnick oxidation[38] with $NaClO_2$ and NaH_2PO_4 to afford **1**. The overall yield of **1** from (*R*)-E1 was 61.4%. Similarly, (*S*)-E1 could be isolated using an (*S,S*)-salen-Co catalyst for the selective ring-opening of the (*R*)-enantiomer, which was subsequently converted to **2** with a moderate yield (57.6%).



Scheme 5-1. Synthesis of **1** (from (*R,R*)-Salen-Co-OAc) and **2** (from (*S,S*)-Salen-Co-OAc) from racemic propylene oxide.

The number of carbon atoms, constituting the backbone of the HA, is an important structural factor that determines the conformation and solid-state properties of the resulting PHAs. The number of methylene units in HAs can be controlled by selecting a suitable Grignard reagent as a nucleophile for the ring-opening of a substituted epoxide (Table 5-1). We synthesized a series of HAs (3-hydroxybutanoate (**1**) to 5-hydroxyhexanoate (**5**) with moderate yields (56.5–64.6%) using the corresponding 1-alkenyl magnesium bromide (G1–G3) as a nucleophile with propylene oxide (E1) in the regioselective epoxide-opening step. Our procedure could be extended to HAs containing hydrocarbon chains as substituents. For HAs with a medium-chain-length (MCL) hydrocarbon, 1,2-epoxyoctane was resolved using HKR to afford (*R*)-E2. Ring-opening with G1 yielded the ring-opened secondary alcohol, which was oxidized to carboxylic acid **6**. Our procedure allowed the synthesis of various enantiomeric HAs using terminal epoxides as precursors. Detailed characterization of HAs is described in Appendix (A.5.1).

Table 5-1. Molecular structures of enantiomeric HAs from a combination of terminal epoxides and Grignard reagents



Iterative convergence of enantiomeric HAs. The library of enantiomeric HAs synthesized in this study can be employed as building blocks for constructing discrete PHAs via iterative convergent synthesis. This method allows a rapid increase in the molecular weight of discrete PHAs through a minimum number of synthetic steps.[28,29,31,39–42]

To test the scalability of the molecular weight of discrete PHAs, we used the HAs as building blocks. The carboxylic acid termini of HAs were converted to benzyl ester as a protective group for convergent synthesis. The convergent step was composed of selective deprotections of TBDMS with $BF_3 \cdot Et_2O$ and benzyl ester by hydrogenation, and followed by esterification using *N*,*N'*-diisopropylcarbodiimide (DIC); 4-(dimethylamino)pyridinium 4-toluenesulfonate (DPTS) was used as a coupling agent (Figure 5-1A). For example, five iterations of convergent growth from dimer of hydroxypentanoate (**3**) afforded discrete PHAs with 64 repeating units. A twice the number of repeating units compared to that of the reactant after each convergent step could be observed by gel permeation chromatography (GPC) analysis (Figure 5-1B). However, MALDI-TOF analysis showed that terminal repeating units were deleted during the esterification of the deprotected precursors (Figure 5-1C). This deletion was caused by the intramolecular lactonization of terminal units of the OH-deprotected precursor, which also occurred in the hydroxyhexanoate (**5**). Similar deletion errors were reported from the convergent synthesis of discrete poly(3-hydroxybutanote) $(\mathbf{1})_n$, due to the serial β-elimination of COOH-deprotected precursors under Steglich reaction conditions.[39]

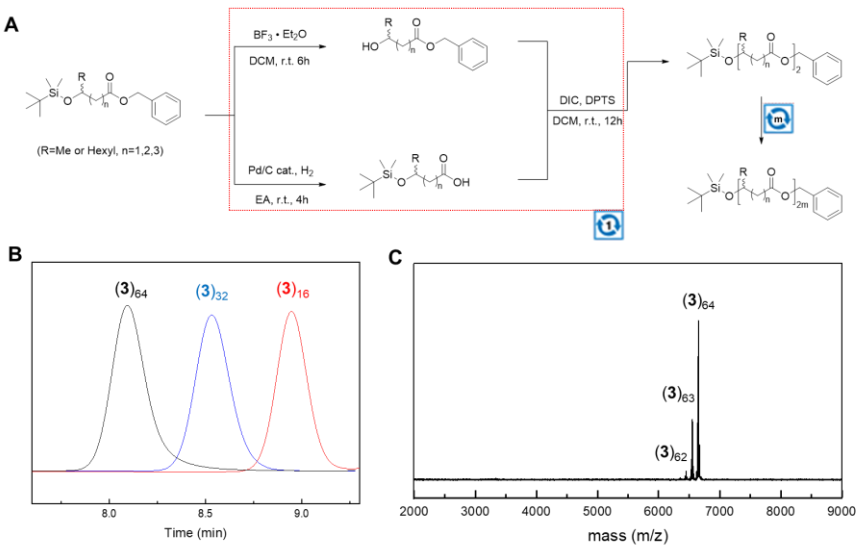


Figure 5-1. (A) Scheme for iterative convergent synthesis of HAs. (B) GPC traces of discrete $(\mathbf{3})_{16}$, $(\mathbf{3})_{32}$, and $(\mathbf{3})_{64}$. (C) MALDI-TOF analysis of $(\mathbf{3})_{64}$. Deletion of the monomers resulting from intramolecular cyclization was observed.

To prevent the unwanted deletion of the terminal repeating unit, we synthesized **1·3** by coupling **1** with **3** as a permanent protective unit for the COOH-terminus of 1 and the OH-terminus of **3**, which was utilized as a monomer for convergent growth (Figure 5-2A). The convergence of **1·3** to a discrete alternating copolymer of **1** and **3** yielded $(\mathbf{1}\cdot\mathbf{3})_{16}$ with an isotactic arrangement of stereocenters. The structures of the resultant products $(\mathbf{1}\cdot\mathbf{3})_{16}$, without any unreacted precursors and deletion of repeating units, were characterized using ^{1}H and ^{13}C NMR, MALDI-TOF mass spectrometry (Figure 5-2B), and GPC analysis (Figure 5-2C). Additionally, syndiotactic $(\mathbf{1}\cdot\mathbf{L})_{32}$ was also synthesized by coupling **1** with L-lactic acid (L). The attempted synthesis of discrete polymers of **1·D** by coupling **1** and D-lactic acid (D) with an isotactic microstructure was unsuccessful because of the limited solubility of oligomeric $(\mathbf{1}\cdot\mathbf{D})_{n}$, owing to its high crystallinity.[43]

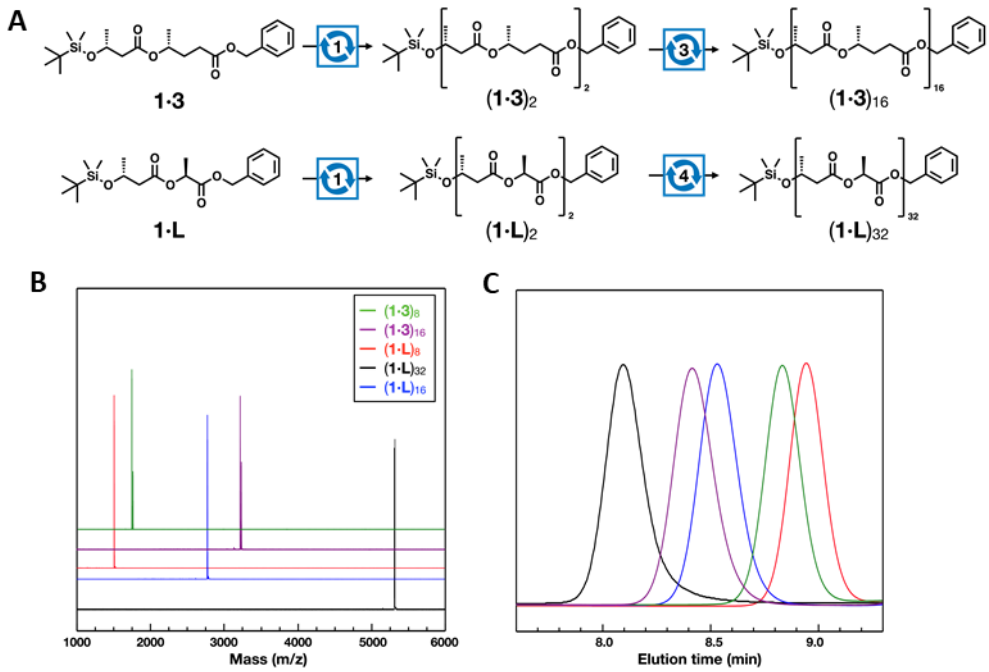


Figure 5-2. (A) Synthetic route of (**1**·**3**)$_{16}$ and (**1**·**L**)$_{32}$ via iterative convergent method. (B) MALDI-TOF mass spectra of (**1**·**3**)$_{8}$, (**1**·**3**)$_{16}$, (**1**·**L**)$_{8}$, (**1**·**L**)$_{16}$, and (**1**·**L**)$_{32}$. (C) GPC traces of (**1**·**3**)$_{8}$ (green), (**1**·**3**)$_{16}$ (purple), (**1**·**L**)$_{8}$ (red), (**1**·**L**)$_{16}$ (blue), and (**1**·**L**)$_{32}$ (black).

Synthesis of high molecular weight sequence-regulated PHAs. Polymerization of sequence-defined oligomers is a facile strategy for synthesizing high molecular weight sequence-regulated polymers. It is known that the sequence of repeating segments affects the thermal[43], physical properties[44], and biodegradability[44,45] of sequence-regulated polymers. We demonstrated the effects of the appearance of a MCL hydrocarbon substituent in the repeating sequence and tacticity on physical properties of high molecular weight sequence-regulated PHAs.

To investigate the effect of complicated tacticity on polymers, we synthesized a series of stereoregular PHAs using diastereomeric building blocks composed of enantiomeric **1**, **2**, **3**, and **4**. Four tetrameric building blocks were built from the dimers (X·Y·X·Y, X=**1** or **2**, Y=**3** or **4**) having two stereocenters (Figure 5-3A). The iterative convergence of these building blocks resulted in the synthesis of discrete $(\mathbf{1}\cdot\mathbf{3}\cdot\mathbf{1}\cdot\mathbf{3})_8$ and $(\mathbf{1}\cdot\mathbf{4}\cdot\mathbf{2}\cdot\mathbf{3})_8$ having two stereoregular repeating units alternating each other. Sequence-defined PHAs without molecular weight distribution have not been synthesized by conventional polymerization or biosynthesis using microbes. Subsequently, a high molecular weight PHAs was obtained via the step-growth polymerization of a deprotected segment having hydroxyl and carboxylic acid termini with a DIC/DPTS coupling agent (Figure 5-3B). The resulting sequence-regulated polymers composed of repeating sequences of the sequenced segments were polydisperse in their molecular weights ($Đ$ = 1.6–2.5), but the tacticity of the segment was maintained.

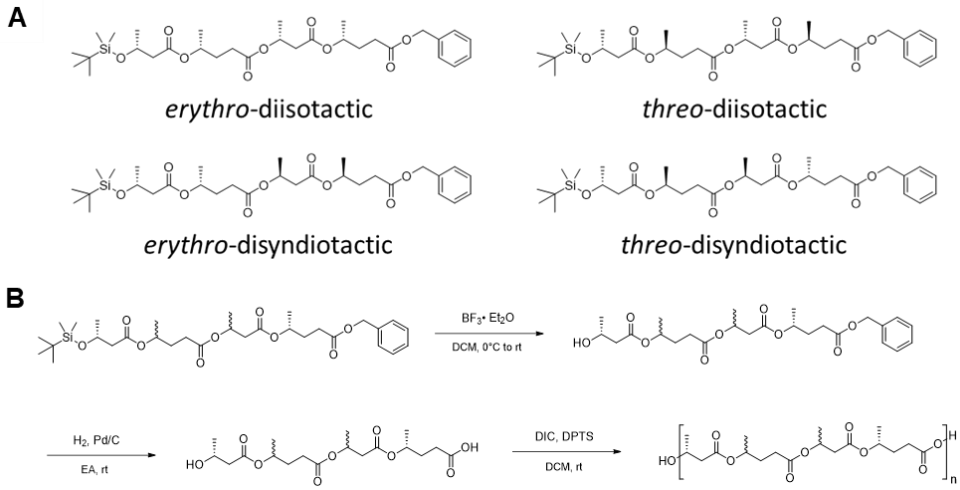


Figure 5-3. (A) Chemical structures of four tetrameric building blocks. (B) Synthetic procedure for stereo-regulated polymers with molecular weight distribution by polycondensation of stereospecific tetramer.

The screening of their structural, optical, and physical characteristics by ^{13}C NMR, circular dichroism spectroscopy, and differential scanning calorimetry of a series of stereoregular PHAs revealed the differences between *erythro*-diisotactic and *threo*-disyndiotactc PHAs without molecular weight distribution (Figure 5-4A). *Erythro*-diisotactic polymer was semi-crystalline (Figure 5-4C), in contrast, *threo*-disyndiotactic polymer was amorphous (Figure 5-4D). These differences in properties ,such as optical and thermal properties, arising from the difference in stereochemical regularity persisted when the PHAs became polydisperse in molecular weight, which were prepared by the condensation polymerization of tetrameric building blocks (Figure 5-4).

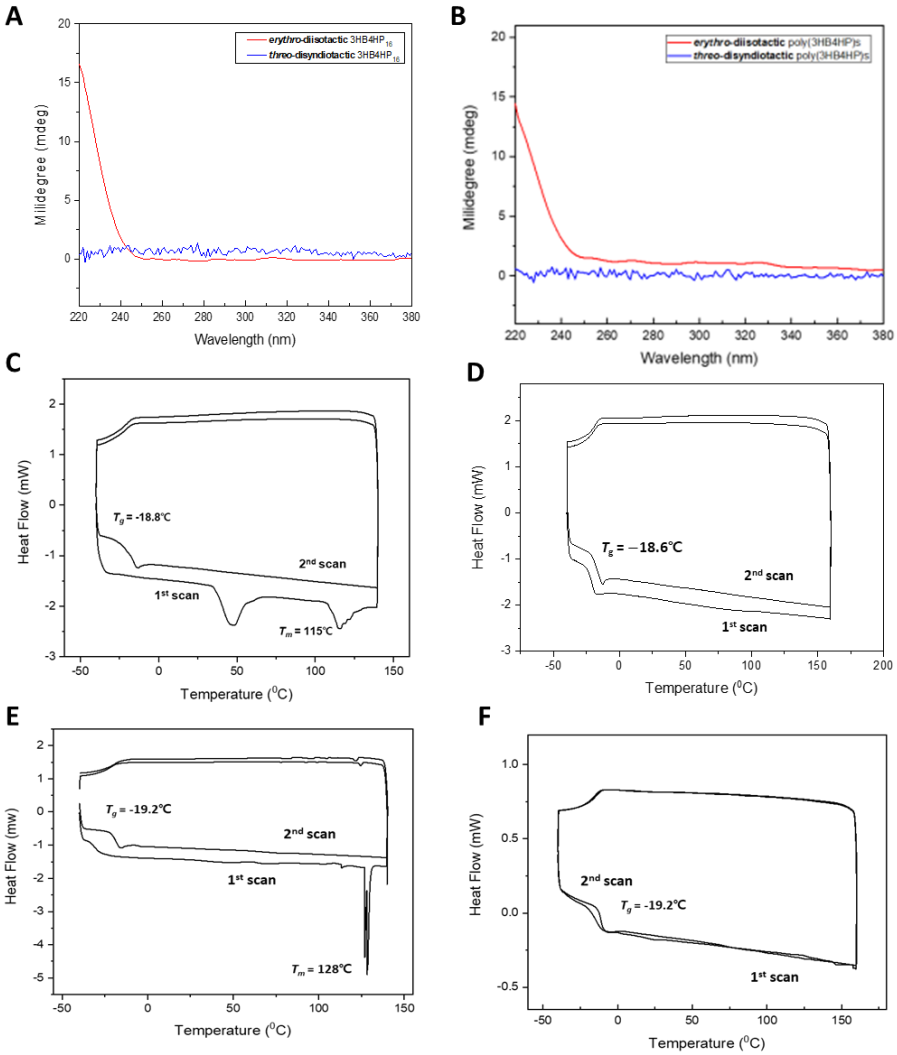


Figure 5-4. (A) Circular dichroism spectra of *erythro*-diisotactic (**1**·**3**·**1**·**3**)$_8$ and *threo*-disyndiotactic (**1**·**4**·**2**·**3**)$_8$ synthesized by cross-convergent method. (B) Circular dichroism spectra of *erythro*-diisotactic poly(**1**·**3**·**1**·**3**)s and *threo*-disyndiotactic poly(**1**·**4**·**2**·**3**)s synthesized by polycondensation. DSC curves of (C) *erythro*-diisotactic (**1**·**3**·**1**·**3**)$_8$, (D) *threo*-disyndiotactic (**1**·**4**·**2**·**3**)$_8$, (E) *erythro*-diisotactic poly(**1**·**3**·**1**·**3**)s, (F) *threo*-disyndiotactic poly(**1**·**4**·**2**·**3**)s. *erythro*-diisotactic polymers are semi-crystalline, but *threo*-disyndiotactic polymers are amorphous.

Next, we chose the diastereomeric dimers synthesized by coupling **1** with L- or D-lactic acid (**1·L** and **1·D**, respectively). Additionally, the hexyl-substituted dimers (**6·L** and **6·D**) were selected to introduce an MCL-hydrocarbon substituent in the repeating oligomers (Figure 5-5A). Subsequently, sequenced octameric segments (1·X·Y·X·1·X·Y·X, X=**L** or **D**, Y=**1** or **6**) were synthesized by the convergent method using diastereomeric dimers. Several sequence-regulated PHAs were synthesized by step-growth polymerization used above experiment. The resulting sequence-regulated polymers remained tacticity and sequence of the building block. For example, the deprotections of silyl and benzyl ester of a syndiotactic octamer containing one hexyl side chain, **1·L·1·L·1·L·6·L**, and the subsequent esterification of this segment afforded syndiotactic poly(**1·L·1·L·1·L·6·L**) (M_n = 110 kDa, $Đ$ =1.8) having one hexyl group per eight monomeric units.

The resulting structures of these polymers could be confirmed using ^{1}H NMR, ^{13}C NMR, and GPC analyses. As shown in the ^{13}C NMR spectra (Figure 5-5C), the methine groups of **6** and **L** of syndiotactic poly(**1·L·1·L·1·L·6·L**) were observed at 71.56 and 68.87 ppm, respectively, whereas those of isotactic poly(**1·D·1·D·1·D·6·D**) (M_n = 25 kDa, $Đ$ =3.1) shifted to 71.34 and 68.99 ppm, respectively. Furthermore, the number of the hexyl group in the repeating sequence was verified by comparing the integral of terminal protons in the hexyl group and methine protons in the lactic acid in ^{1}H NMR spectra (Figure 5-5D). The thermal properties of these sequence-regulated PHLs were investigated using DSC analysis (Figure 5-5E). The syndiotactic and MCL hydrocarbon substituted polymers poly(**1·L·1·L·1·L·6·L**) demonstrated a glass transition temperature of 1.7 °C, which is lower than that of poly(**1·L·1·L·1·L·1·L**) without a hexyl substituent in the repeating sequence (T_g = 9.7 °C). The isotactic poly(**1·D·1·D·1·D·6·D**) were crystalline, and only showed a melting temperature at 145 °C, while the

poly(**1·D·6·D·1·D·6·D**)s (M_n = 25 kDa, $Đ$ =3.8) containing two hexyl side chains in the repeating sequence indicated a lower melting temperature (T_m = 137 °C). These results suggest that the crystallinity and thermal properties could be controlled by defining the repeated sequence of monomers.

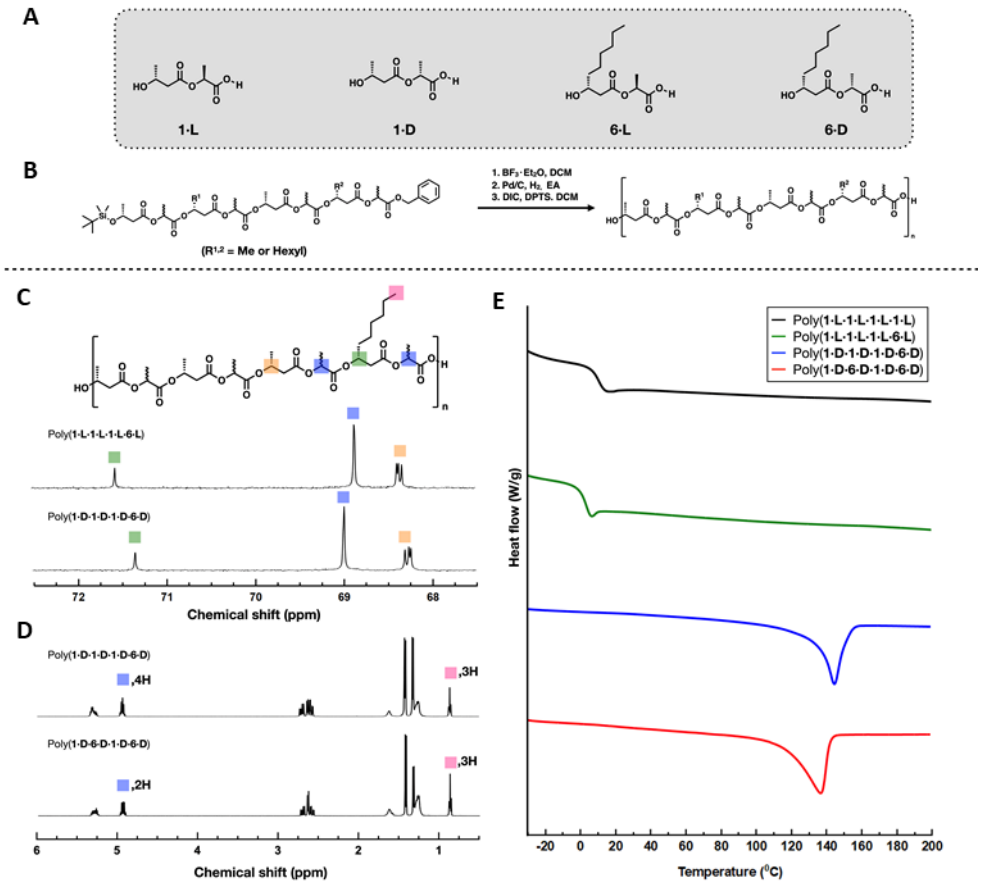


Figure 5-5. (A) Chemical structures of diastereomeric dimers: **1·L**, **1·D**, **6·L**, and **6·D**. (B) Condensation polymerization of sequenced segments for sequence-regulated PHLs. (C) ^{13}C NMR spectra ($CDCl_3$) of poly(**1·L·1·L·1·L·6·L**) and poly(**1·D·1·D·1·D·6·D**) at methine region. (D) ^{1}H NMR spectra ($CDCl_3$) of poly(**1·D·1·D·1·D·6·D**) and poly(**1·D·6·D·1·D·6·D**). The peaks are indicated by the colored box corresponding to its position. (E) DSC curves of sequence-regulated PHLs: poly(**1·L·1·L·1·L·1·L**), poly(**1·L·1·L·1·L·6·L**), poly(**1·D·1·D·1·D·6·D**), and poly(**1·D·1·D·1·D·6·D**).

5.4 Conclusion

In this study, a new synthetic strategy for enantiopure HAs from terminal epoxides was developed to obtain PHA with absolute atomic precision. A combination of HKR of terminal epoxides and subsequent ring-opening with alkenyl Grignard reagents yielded enantiopure HAs with desired substituents. The library of enantiomeric HAs serves as building blocks to synthesize polyesters without molecular weight distribution and absolutely defined sequence of repeating units via iterative convergent and cross-convergent synthesis. This study showed that the structural parameters of the aliphatic polyesters could be precisely controlled, which could serve a synthetic platform for macromolecular engineering with precision comparable to that exhibited by biopolymers such as proteins. We envisage that sequence-defined and multifunctional polyesters could help solve the structure–function relationship of polymers, and could be utilized as precision nanostructures, biomedical polymers, and nanoreactors. Furthermore, high molecular weight sequence-controlled polyesters could be obtained from the polymerization of sequence-defined and stereoregular oligomers. We contemplate that these macromolecular engineering methods may help in devising a suitable solution for replacing commercial plastics with degradable polymers.

5.5 Experimental

Materials and Methods. All other reagents and chemicals were purchased from commercial sources and used without further purification. Dichloromethane (DCM) and tetrahydrofuran (THF) were distilled over CaH_2 and sodium/benzophenone, respectively. ^{1}H NMR and ^{13}C NMR spectra were recorded by Varian INOVA 500

MHz NMR spectrometer using $CDCl_3$ as solvent. Gel permeation chromatography (GPC) was performed on an Agilent 1260 Infinity equipped with a PL gel 5 μm mixed D column and differential refractive index detectors. THF was used as an eluent with a flow rate of 0.3 mL min^{-1} at 35 °C. A polystyrene standard kit (Agilent Technologies) was used for calibration. Mass spectra of PHAs were measured on a Bruker Ultraflex III TOF-TOF mass spectrometer equipped with a Nd:YAG laser (355 nm). The instrument was operated in linear and refractive positive ion modes. Trans-2-[3-(4-tert-Butylphenyl)-2-methyl-2-propenylidene]malononitrile (DCTB) were used as a matrix. Sodium trifluoroacetate (NaTFA) was used as cationizing agent. Differential Scanning Calorimetry (DSC) was performed on a TA Instruments DSC 25 from –50 °C to 210 °C with a scan rate of 10 °C min^{-1} under N_2 atmosphere for polycondensed PHAs samples and from –130 °C to 100 °C with a scan rate of 10 °C min^{-1} under N_2 atmosphere for PCLs samples. Calibration was performed by measuring the melting temperature of indium (T_m = 429.75 K).

General procedure for hydrolytic kinetic resolution of terminal epoxides. All enantioenriched epoxides were prepared by the literature procedure earlier. Racemic epoxide was added to (*R,R*)-Salen-Co(II)-OAc (0.5–0.8 mol%). To this mixture, H_2O (0.55 eq.) was added over 30 min under ice-bath. The mixture was stirred at ambient temperature for 14–72 hrs. The enantioenriched epoxides were obtained by distillation under reduced pressure.

General procedure for ring-opening of enantioenriched epoxides. CuBr (5 mol%) was dispersed in dry THF (0.3 M) under N2 gas and cooled to -78℃ in an acetone/CO_2(s) bath. Grignard reagent (1 M in THF, 3.0 eq.) was added to the suspension and stirred for 30 minutes. Enantioenreiched epoxide was added slowly and stirred for 4 hours. After completion, the mixture was quenched with saturated NH_4Cl (aq), diluted with water and extracted with Et_2O. Combined organic layers

were washed with brine and dried over $MgSO_4$. The crude product was purified by distillation under reduced pressure.

General procedure for TBDMS silyl ether protection of enols. The enol was dissolved in dry DCM (0.4 M). To this solution, imidazole (1.2 eq.) and *tert*-butyldimethylsilyl chloride (1.1 eq.) was added sequentially under ice-bath. The reaction had proceeded at ambient temperature for 3 hours. The reaction mixture was diluted with water and extracted with CH_2Cl_2. Combined organic layers were washed with ice-cold 5% HCl(aq), washed with brine, and dried over $MgSO_4$. Residual solvents are removed under reduced pressure.

General procedure for Lemieux-Johnson oxidation of terminal olefins. The TBDMS-protected terminal olefin was dissolved in Dioxane/H_2O (3:1) (0.3 M). To this solution, 2,6-lutidine (2.0 eq.), 4% OsO_4(aq) (0.5 mol%) and sodium periodate (2.5 eq.) was added sequentially under ice-bath. The reaction had proceeded at ambient temperature for 12 hours. The reaction mixture was diluted with water and extracted with CH_2Cl_2. Combined organic layers were washed with brine, dried over $MgSO_4$ and concentrated under reduced pressure. The crude product was purified by flash column chromatography using hexane/EtOAc as an eluent.

General procedure for Pinnick oxidation of aldehydes. The TBDMS-protected aldehyde and 2-methyl-2-butene (10 eq.) were dissolved in *t*-BuOH/H_2O (1:1) (0.3 M). To this solution, an aqueous solution of $NaClO_2$ (2.5 eq.) and NaH2PO4 was transferred under an ice bath. The reaction had proceeded at ambient temperature for an hour. The reaction mixture was concentrated under reduced pressure. Then the concentrate was extracted with CH_2Cl_2, washed with brine, and 1M NaOH(aq). The organic layer was discarded. The aqueous layer was then acidified to pH 3 using 1M $KHSO_4$(aq), then was extracted with Et_2O. and dried over $MgSO_4$. Residual solvents are removed under reduced pressure.

General procedure for debenzylation giving carboxylic acid end functional group. The benzyl-protected compound was dissolved in EtOAc (0.2 M) and palladium on activated carbon (10 wt%) was added to the solution. After bubbled with nitrogen gas for a few minutes, Hydrogen gas-charged balloon was connected to the reaction vessel and stirred at room temperature. After completion, the palladium-suspended product was filtered through celite. The solvent in the filtrate was removed under reduced pressure, giving debenzylated product.

General procedure for deallylation giving carboxylic acid end functional group. The allyl-protected compound was dissolved in dry THF (0.1 M) under N_2. $Pd(PPh_3)_4$ (3.0 mol%) and morpholine (1.05 eq.) were added to the solution. The reaction was proceeded at room temperature. After completion, the reaction mixture was diluted with DCM and washed with 1N HCl(aq) and brine. The organic layer was dried over $MgSO_4$ and concentrated under reduced pressure, giving deallylated product.

General procedure for esterification. The hydroxy group and carboxylic acid group containing compounds were dissolved in CH_2Cl_2 (0.4 M) under N_2 and cooled to 0℃ in an ice bath. To the solution, DPTS (0.15 eq.) and DIC (2.0 eq.) were added slowly. The reaction mixture was stirred at ambient temperature for 3 hours. The reaction mixture was washed with water and brine. The concentrate was diluted with Et_2O and filtered to remove residual 1,3-diisopropylurea. The filtrate was concentrated under reduced pressure and purified by flash column chromatography using hexane/EtOAc as an eluent.

Synthesis of sequence-regulated PHAs by polycondensation. Building unit for sequence and tacticity defined PHAs having TBDMS and benzyl terminal ends were deprotected following the general procedure for desilylation and debenzylation. Subsequently, deprotected oligo(hydroxyalkanoate)s (400 mg) were dissolved in

CH_2Cl_2 (0.5 M). DIC (0.75 g, 1.5 eq.) and DPTS (0.15 eq.) were added under ice-bath, and the reaction mixture was stirred at room temperature for 4h. The product was obtained after precipitation twice in cold methanol.

5.6 References

1. Choi, S. Y.; Cho, I. J.; Lee, Y.; Kim, Y.-J.; Kim, K.-J.; Lee, S. Y. Microbial Polyhydroxyalkanoates and Nonnatural Polyesters. *Adv. Mater.* **2020**, *32*, 1907138.
2. Chen, G.-Q.; Patel, M. K. Plastics derived from biological sources: present and future: a technical and environmental review. *Chem. Rev.* **2012**, *112*, 2082–2099.
3. Muhammadi, S.; Afzal, M.; Hameed, S. Bacterial polyhydroxyalkanoates-eco-friendly next generation plastic: Production, biocompatibility, biodegradation, physical properties and applications. *Green Chem. Lett. Rev.* **2015**, *8*, 56–77.
4. Anjum, A.; Zuber, M.; Zia, K. M.; Noreen, A.; Anjum, M. N.; Tabasum, S. Microbial production of polyhydroxyalkanoates (PHAs) and its copolymers: A review of recent advancements. *Int. J. Biol. Macromol.* **2016**, *89*, 161–174.
5. Albertsson, A.-C.; Varma, I. K. Aliphatic polyesters: Synthesis, properties and applications. *Degradable aliphatic polyesters* **2002**, 1–40.
6. Chen, G.-Q. A microbial polyhydroxyalkanoates (PHA) based bio-and materials industry. *Chem. Soc. Rev.* **2009**, *38*, 2434–2446.
7. Hillmyer, M. A.; Tolman, W. B. Aliphatic Polyester Block Polymers: Renewable, Degradable, and Sustainable. *Acc. Chem. Res.* **2014**, *47*, 2390–2396.
8. Saito, T.; Saegusa, H.; Miyata, Y.; Fukui, T. Intracellular degradation of poly(3-hydroxybutyrate) granules of Zoogloea ramigera I-16-M. FEMS *Microbiol. Lett.* **1992**, *103*, 333–338.
9. Ong, S. Y.; Chee, J. Y.; Sudesh, K. Degradation of polyhydroxyalkanoate (PHA): a review. **2017**, *10*, 211–225.
10. Wang, Y.; Yin, J.; Chen, G.-Q. Polyhydroxyalkanoates, challenges and opportunities. *Curr. Opin. Biotechnol.* **2014**, *30*, 59–65.

11. Możejko-Ciesielska, J.; Kiewisz, R. Bacterial polyhydroxyalkanoates: Still fabulous? *Microbiol. Res.* **2019**, *192*, 271–282.
12. Hartmann, R.; Hany, R.; Geiger, T.; Egli, T.; Witholt, B.; Zinn, M. Tailored Biosynthesis of Olefinic Medium-Chain-Length Poly [(R)-3-hydroxyalkanoates] in Pseudomonas p utida GPo1 with Improved Thermal Properties. *Macromolecules* **2004**, *37*, 6780–6785.
13. Aksakal, R.; Mertens, C.; Soete, M.; Badi, N.; Du Prez, F. Applications of Discrete Synthetic Macromolecules in Life and Materials Science: Recent and Future Trends. *Adv. Sci.* **2021**, *8*, 2004038.
14. Solleder, S. C.; Meier, M. A. Sequence control in polymer chemistry through the Passerini three-component reaction. *Angew. Chem. Int. Ed.* **2014**, *53*, 711–714.
15. Al Ouahabi, A.; Charles, L.; Lutz, J.-F. Synthesis of Non-Natural Sequence-Encoded Polymers Using Phosphoramidite Chemistry. *J. Am. Chem. Soc.* **2015**, *137*, 5629–5635.
16. Roy, R. K.; Meszynska, A.; Laure, C.; Charles, L.; Verchin, C.; Lutz, J.-F. Design and synthesis of digitally encoded polymers that can be decoded and erased. *Nat. commun.* **2015**, *6*, 7237.
17. van Genabeek, B.; de Waal, B. F. M.; Ligt, B.; Palmans, A. R. A.; Meijer, E. W. Dispersity under Scrutiny: Phase Behavior Differences between Disperse and Discrete Low Molecular Weight Block Co-Oligomers. *ACS Macro Lett.* **2017**, *6*, 674–678.
18. Nguyen, H. V.-T.; Jiang, Y.; Mohapatra, S.; Wang, W.; Barnes, J. C.; Oldenhuis, N. J.; Chen, K. K.; Axelrod, S.; Huang, Z.; Chen, Q.; Golder, M. R.; Young, K.; Suvlu, D.; Shen, Y.; Willard, A. P.; Hore, M. J. A.; Gómez-Bombarelli, R.; Johnson, J. A. Bottlebrush polymers with flexible enantiomeric side chains display differential biological properties. *Nat. Chem.* **2022**, *14*, 85–93.
19. Matyjaszewski, K. Architecturally Complex Polymers with Controlled Heterogeneity. *Science* **2011**, *333*, 1104–1105.
20. Perrier, S. 50th Anniversary Perspective: RAFT Polymerization—A User Guide. *Macromolecules* **2017**, *50*, 7433–7447.

21. Rieth, L. R.; Moore, D. R.; Lobkovsky, E. B.; Coates, G. W. Single-site β-diiminate zinc catalysts for the ring-opening polymerization of β-butyrolactone and β-valerolactone to poly (3-hydroxyalkanoates). *J. Am. Chem. Soc.* **2002**, *124*, 15239–15248.
22. Tang, X.; Chen, E. Y.-X. Chemical synthesis of perfectly isotactic and high melting bacterial poly (3-hydroxybutyrate) from bio-sourced racemic cyclic diolide. *Nat. Commun.* **2018**, *9*, 2345.
23. Tang, X.; Westlie, A. H.; Watson, E. M.; Chen, E. Y.-X. Stereosequenced crystalline polyhydroxyalkanoates from diastereomeric monomer mixtures. *Science* **2019**, *366*, 754–758.
24. Ajellal, N. Bouyahyi, M.; Amgoune, A.; Thomas, C. M.; Bondon, A.; Pillin, I.; Grohens, Y.; Carpentier, J.-F. Syndiotactic-enriched poly (3-hydroxybutyrate) s via stereoselective ring-opening polymerization of racemic β-butyrolactone with discrete yttrium catalysts. *Macromolecules* **2009**, *42*, 987–993.
25. Tang, X.; Chen, E. Y.-X. Chemical synthesis of perfectly isotactic and high melting bacterial poly(3-hydroxybutyrate) from bio-sourced racemic cyclic diolide. *Nat. Commun.* **2018**, *9*, 2345.
26. Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-Controlled Polymers. *Science* **2013**, *341*, 1238149.
27. van Genabeek, B.; Lamers, B. A. G.; Hawker, C. J.; Meijer, E. W.; Gutekunst, W. R.; Schmidt, B. V. K. J. Properties and applications of precision oligomer materials; where organic and polymer chemistry join forces. *J. Polym. Sci.* **2021**, *59*, 373–403.
28. Takizawa, K.; Tang, C.; Hawker, C. J. Molecularly defined caprolactone oligomers and polymers: synthesis and characterization. *J. Am. Chem. Soc.* **2008**, *130*, 1718–1726.
29. Lee, J. M.; Koo, M. B.; Lee, S. W.; Lee, H.; Kwon, J.; Shim, Y. H.; Kim, S. Y.; Kim, K. T. High-density information storage in an absolutely defined aperiodic sequence of monodisperse copolyester. *Nat. Commun.* **2020**, *11*, 56.
30. Koo, M. B.; Lee, S. W.; Lee, J. M.; Kim, K. T. Iterative Convergent Synthesis of

Large Cyclic Polymers and Block Copolymers with Discrete Molecular Weights. *J. Am. Chem. Soc.* **2020**, *142*, 14028–14032.

31. Barnes, J. C.; Ehrlich, D. J. C.; Gao, A. X.; Leibfarth, F. A.; Jiang, Y.; Zhou, E.; Jamison, T. F.; Johnson, J. A. Iterative exponential growth of stereo-and sequence-controlled polymers. *Nat. Chem.* **2015**, *7*, 810–815.
32. Nielsen, L. P.; Stevenson, C. P.; Blackmond, D. G.; Jacobsen, E. N. Mechanistic investigation leads to a synthetic improvement in the hydrolytic kinetic resolution of terminal epoxides. *J. Am. Chem. Soc.* **2004**, *126*, 1360–1362.
33. Schaus, S. E. Brandes, B. D.; Larrow, J. F.; Tokunaga, M.; Hansen, K. B.; Gould, A. E.; Furrow, M. E.; Jacobsen, E. N. Highly selective hydrolytic kinetic resolution of terminal epoxides catalyzed by chiral (salen) CoIII complexes. Practical synthesis of enantioenriched terminal epoxides and 1, 2-diols. *J. Am. Chem. Soc.* **2002**, *124*, 1307–1315.
34. Tokunaga, M.; Larrow, J. F.; Kakiuchi, F.; Jacobsen, E. N. Asymmetric catalysis with water: efficient kinetic resolution of terminal epoxides by means of catalytic hydrolysis. *Science* **1997**, *277*, 936–938.
35. Bonini, C.; Chiummiento, L.; Lopardo, M. T.; Pullez, M.; Colobert, F.; Solladie, G. A general protocol for the regio high yielding opening of different glycidol derivatives. *Tetrahedron Lett.* **2003**, *44*, 2695–2697.
36. Pappo, R., Allen Jr, D. S.; Lemieux, R. U.; Johnson, W. S. Osmium tetroxide-catalyzed periodate oxidation of olefinic bonds. *J. Org. Chem.* **1956**, *21*, 478–479.
37. Yu, W.; Mei, Y.; Kang, Y.; Hua, Z.; Jin, Z. Improved procedure for the oxidative cleavage of olefins by OsO4−NaIO4. *Org. Lett.* **2004**, *6*, 3217–3219.
38. Bal, B. S.; Childers Jr, W. E.; Pinnick, H. W. Oxidation of α, β-un saturated aldehydes. *Tetrahedron* **1981**, *37*, 2091–2096.
39. Lengweiler, U. D.; Fritz, M. G.; Seebach, D. Synthese monodisperser linearer und cyclischer Oligomere der (R)-3-Hydroxybuttersaure mit bis zu 128 Einheiten. *Helv. Chim. Acta* **1996**, *79*, 670–701.
40. Brooke, G.; MacBride, J. H.; Mohammed, S.; Whiting, M. Versatile syntheses of oligomers related to nylon 6, nylon 4 6 and nylon 6 6. *Polymer* **2000**, *41*, 6457–

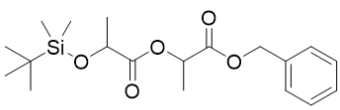
6471.

41. Williams, J. B.; Chapman, T. M.; Hercules, D. M. Synthesis of discrete mass poly (butylene glutarate) oligomers. *Macromolecules* **2003**, *36*, 3898–3908.
42. Li, Z. Ren, X.; Sun, P.; Ding, H.; Li, S.; Zhao, Y.; Zhang, K. Protecting-Group-Free Iterative Exponential Growth Method for Synthesizing Sequence-Defined Polymers. *ACS Macro Lett.* **2021**, *10*, 223–230.
43. Tabata, Y.; Abe, H. Synthesis and Properties of Alternating Copolymers of 3-Hydroxybutyrate and Lactate Units with Different Stereocompositions. *Macromolecules* **2014**, *47*, 7354–7361.
44. Li, J.; Rothstein, S. N.; Little, S. R.; Edenborn, H. M.; Meyer, T. Y. The effect of monomer order on the hydrolysis of biodegradable poly(lactic-co-glycolic acid) repeating sequence copolymers. *J. Am. Chem. Soc.* **2012**, *134*, 16352–16359.
45. Swisher, J. H.; Nowalk, J. A.; Meyer, T. Y. Property impact of common linker segments in sequence-controlled polyesters. *Polym. Chem.* **2019**, *10*, 244–252.

Appendix

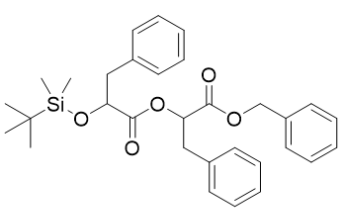
Chapter 2: High-Density Information Storage in an Absolutely Defined Aperiodic Sequence of Monodisperse Copolyester

A.2.1 Characterization of PcLs



LL

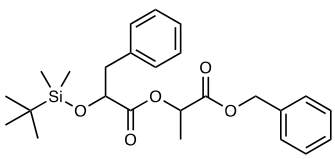
LL. Colorless liquid (25 g, 91%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.37-7.27 (m, 5H, Ph-**H**), 5.18 (m, 3H, COO-C**H**(CH_3)C=O and Ph-C**H$_2$**-O), 4.39 (q, *J* = 6.8Hz, 1H, SiO-C**H**(CH_3)C=O), 1.53 (d, *J* = 7.1 Hz, 3H, C**H$_3$**-CH), 1.43 (d, *J* = 6.8 Hz, 1H, C**H$_3$**-CH), 0.88 (s, 9H, (C**H$_3$**)$_3$C-Si), 0.10 (dd, *J* = 12.2, 5.7 Hz, 6H, (CH_3)$_3$C-Si(C**H$_3$**)$_2$-O) ppm; ^{13}C NMR (400MHz, $CDCl_3$): δ 173.2 (SiO-CH(CH_3)**C**=O), 170.1 (COO-CH(CH_3)**C**=O), 135.3 (**C$_5$**H$_5$-CH$_2$-COO), 128.5 (**C$_5$**H$_5$-CH$_2$-COO), 128.3 (**C$_5$**H$_5$-CH$_2$-COO), 128.1 (**C$_5$**H$_5$-CH$_2$-COO), 68.7 (CH_3-**C**H-SiO), 68.0 (CH_3-**C**H-COO), 66.8 (C$_5$H$_5$-**C**H$_2$-COO), 25.7 ((CH_3)$_3$**C**-Si(CH_3)$_2$-O), 21.1 (($\mathbf{C}H_3$)$_3$C-Si(CH_3)$_2$-O), 18.2 (**C**H$_3$-CH-SiO), 16.8 (**C**H$_3$-CH-COO), -4.9 ((CH_3)$_3$C-Si(**C**H$_3$)$_2$-O), -5.3 ((CH_3)$_3$C-Si(**C**H$_3$)$_2$-O) ppm; MS (ESI/Q-TOF): m/z (isotopic maximum value) calcd for $C_{19}H_{30}O_5Si+Na^+$ [**M**+Na]$^+$: 389; found 390.



PP

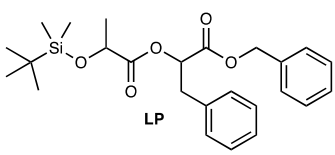
PP. colorless oil (16.07 g, 72%); ^{1}H NMR (500MHz, $CDCl_3$): 7.37-7.01 (m, 15H, Ph-**H**), 5.32 (m, 1H, COO-C**H**(CH_3)C=O), 5.14 (m, 2H, Ph-C**H$_2$**-O), 4.30 (ddd, *J* = 29.2, 9.2, 3.4 Hz, 1H, SiO-C**H**(CH_2Ph)C=O), 2.90 (m, 4H, Ph-C**H$_2$**-CH), 0.71 (d, *J* = 91. Hz, 9H,

($C\mathbf{H_3}$)$_3$C-Si), -0.29 (dd, J = 51.5, 23.9 Hz, 6H, (CH_3)$_3$C-Si($C\mathbf{H_3}$)$_2$-O) ppm; ^{13}C NMR (400MHz, $CDCl_3$): δ 172.4 (SiO-CH(CH_2Ph)**C**=O), 169.0 (COO-CH(CH_2Ph)**C**=O), 137.7 (**C**$_5$H$_5$-CH$_2$-COO), 135.8 (**C**$_5$H$_5$-CH$_2$-CH-SiO), 135.4 (**C**$_5$H$_5$-CH$_2$-CH-COO), 130.0 (**C**$_5$H$_5$-CH$_2$-CH-SiO, **C**$_5$H$_5$-CH$_2$-CH-COO and **C**$_5$H$_5$-CH$_2$-COO), 128.7 (**C**$_5$H$_5$-CH$_2$-CH-SiO, **C**$_5$H$_5$-CH$_2$-CH-COO and **C**$_5$H$_5$-CH$_2$-COO), 127.2 (**C**$_5$H$_5$-CH$_2$-CH-SiO, **C**$_5$H$_5$-CH$_2$-CH-COO and **C**$_5$H$_5$-CH$_2$-COO), 73.5 (Ph-CH$_2$-**C**H-SiO), 73.2 (Ph-CH$_2$-**C**H-COO), 67.2 (C$_5$H$_5$-**C**H$_2$-COO), 41.6 (Ph-**C**H$_2$-CH-SiO), 37.4 (Ph-**C**H$_2$-CH-COO), 25.8 (($\mathbf{C}H_3$)$_3$C-Si(CH_3)$_2$-O), 18.3 ((CH_3)$_3$**C**-Si($\mathbf{C}H_3$)$_2$-O), -5.2 ((CH_3)$_3$C-Si($\mathbf{C}H_3$)$_2$-O), -5.7 ((CH_3)$_3$C-Si($\mathbf{C}H_3$)$_2$-O) ppm; MS (ESI/Q-TOF): m/z calcd for $C_{31}H_{38}O_5Si+Na^+$ [**M**+Na]$^+$: 541; found 542.



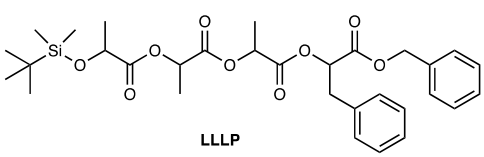
PL

PL. Colorless liquid (20.21 g, 74%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.43-6.91 (m, 10H, Ph-**H**), 5.15 (m, 3H, COO-C**H**(CH_3)C=O and Ph-C**H**$_2$-O), 4.36 (td, J = 9.5, 3.5 Hz, 1H, SiO-C**H**(CH_2Ph)C=O), 3.17-2.76 (m, 2H, Ph-C**H**$_2$-CH), 1.49 (m, 3H, C**H**$_3$-CH), 0.75 (m, 9H, ($C\mathbf{H_3}$)$_3$C-Si), -0.21 (m, 6H, (CH_3)$_3$C-Si($C\mathbf{H_3}$)$_2$-O) ppm; MS (ESI/Q-TOF): m/z calcd for $C_{25}H_{34}O_5Si+Na^+$ [**M**+Na]$^+$: 465; found 466.



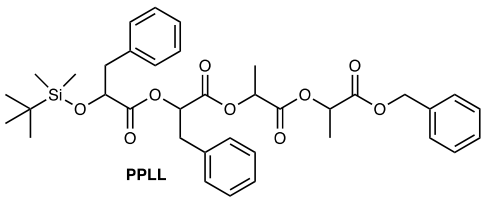
LP

LP. Colorless liquid (2.73g, 71%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.46-6.96 (m, 10H, Ph-**H**), 5.35 (m, 1H, COO-C**H**(CH_2Ph)C=O), 5.15 (m, 2H, Ph-C**H**$_2$-O), 4.34 (m, 1H, SiO-C**H**(CH_3)C=O), 3.27-3.08 (m, 2H, Ph-C**H**$_2$-CH), 1.31 (dd, J = 47.3, 6.8 Hz, 3H, C**H**$_3$-CH), 0.88 (dd, J = 6.9, 4.9 Hz, 9H, ($C\mathbf{H_3}$)$_3$C-Si), 0.02 (dd, J = 15.6, 13.5 Hz, 6H, (CH_3)$_3$C-Si($C\mathbf{H_3}$)$_2$-O) ppm; MS (ESI/Q-TOF): m/z calcd for $C_{25}H_{34}O_5Si+Na^+$ [**M**+Na]$^+$: 465; found 466.

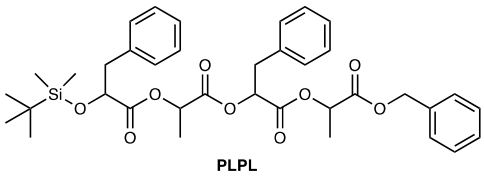


LLLP

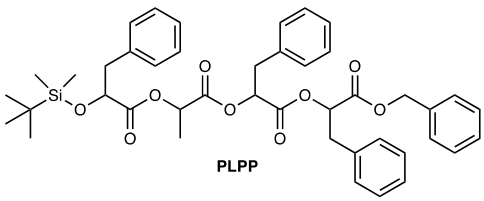
LLLP. Colorless oil (2.78 g, 93%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.37-7.10 (m, 10H, Ph-**H**), 5.34 (dd, *J* = 6.9, 5.3 Hz, 1H, COO-C**H**(CH_2Ph)C=O), 5.25-5.03 (m, 4H, COO-C**H**(CH_3)C=O and Ph-C**H$_2$**-O), 4.41 (q, *J* = 6.7 Hz, 1H, SiO-C**H**(CH_3)C=O), 3.16 (m, 2H, Ph-C**H$_2$**-CH), 1.57-1.32 (m, 9H, C**H$_3$**-CH), 0.92 (s, 9H, (C**H$_3$**)$_3$C-Si), 0.10 (dd, *J* = 9.9, 3.4 Hz, 6H, (CH_3)$_3$C-Si(C**H$_3$**)$_2$-O) ppm; MS (ESI/Q-TOF): m/z (isotopic maximum value) calcd for $C_{31}H_{42}O_9Si+Na^+$ [**M**+Na]$^+$: 610; found 610.



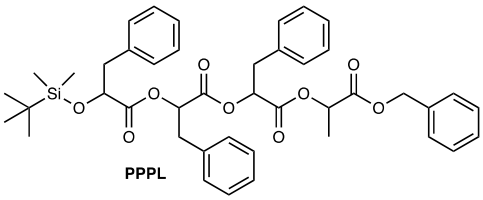
PPLL. Colorless oil (5.19 g, 97%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-7.00 (m, 15H, Ph-**H**), 5.36 (m, 1H, COO-C**H**(CH_2Ph)C=O), 5.30-5.12 (m, 4H, COO-C**H**(CH_3)C=O and Ph-C**H$_2$**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.45-2.55 (m, 4H, Ph-C**H$_2$**-CH), 1.59-1.42 (m, 6H, C**H$_3$**-CH), 0.74 (m, 9H, (C**H$_3$**)$_3$C-Si), -0.26 (dd, 6H, (CH_3)$_3$C-Si(C**H$_3$**)$_2$-O) ppm; MS (ESI/Q-TOF): m/z calcd for $C_{37}H_{46}O_9Si+Na^+$ [**M**+Na]$^+$: 686; found 686.



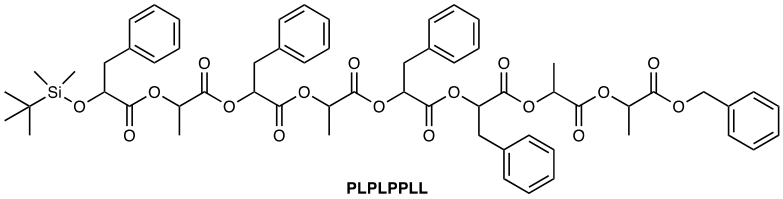
PLPL. Colorless oil (12.51 g, 92%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-7.20 (m, 15H, Ph-**H**), 5.36 (m, 1H, COO-C**H**(CH_2Ph)C=O), 5.25-5.12 (m, 4H, COO-C**H**(CH_3)C=O and Ph-C**H$_2$**-O), 4.35 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.35-2.75 (m, 4H, Ph-C**H$_2$**-CH), 1.57-1.34 (m, 6H, C**H$_3$**-CH), 0.77 (m, 9H, (C**H$_3$**)$_3$C-Si), -0.20 (m, 6H, (CH_3)$_3$C-Si(C**H$_3$**)$_2$-O) ppm; MS (ESI/Q-TOF): m/z calcd for $C_{37}H_{46}O_9Si+Na^+$ [**M**+Na]$^+$: 686; found 686.



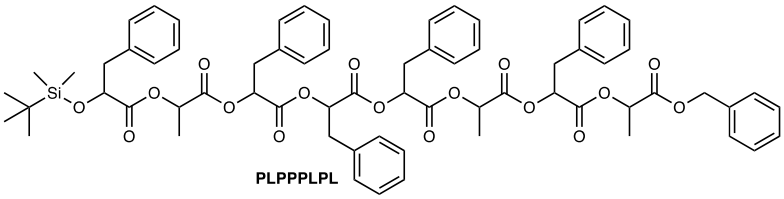
PLPP. Colorless oil (6.52 g, 92%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.37-6.90 (m, 20H, Ph-**H**), 5.42-5.02 (m, 5H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H$_2$**-O), 4.33 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.25-2.70 (m, 6H, Ph-C**H$_2$**-CH), 1.46-1.28 (m, 3H, C**H$_3$**-CH), 0.74 (dd, J = 6.1, 3.4 Hz, 9H, (C**H$_3$**)$_3$C-Si), -0.25 (m, 6H, (CH_3)$_3$C-Si(C**H$_3$**)$_2$-O) ppm; MS (ESI/Q-TOF): m/z calcd for $C_{43}H_{50}O_9Si+Na^+$ [**M**+Na]$^+$: 762; found 762.



PPPL. Colorless oil (2.34 g, 95%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.50-6.95 (m, 20H, Ph-**H**), 5.45-5.12 (m, 5H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H$_2$**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.60 (m, 6H, Ph-C**H$_2$**-CH), 1.57-1.44 (m, 3H, C**H$_3$**-CH), 0.72 (m, 9H, (C**H$_3$**)$_3$C-Si), -0.30 (m, 6H, (CH_3)$_3$C-Si(C**H$_3$**)$_2$-O) ppm; MS (ESI/Q-TOF): m/z calcd for $C_{43}H_{50}O_9Si+Na^+$ [**M**+Na]$^+$: 762; found 762.

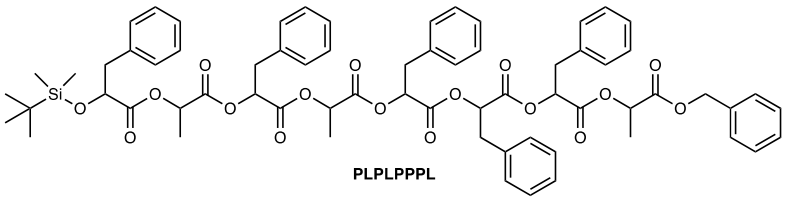


PLPLPPLL (**S**)**.** Viscous oil (2.1 g, 87%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-6.80 (m, 25H, Ph-**H**), 5.40-5.02 (m, 9H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H$_2$**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.67 (m, 8H, Ph-C**H$_2$**-CH), 1.57-1.17 (m, 12H, C**H$_3$**-CH), 0.75 (m, 9H, (C**H$_3$**)$_3$C-Si), -0.24 (m, 6H, (CH_3)$_3$C-Si(C**H$_3$**)$_2$-O) ppm; M_n and $Đ$ (GPC): 1210 Da and 1.03; MS (MALDI-TOF): m/z (isotopic maximum value) calcd for $C_{61}H_{70}O_{17}Si+Na^+$ [**M**+Na]$^+$: 1125; found 1126.

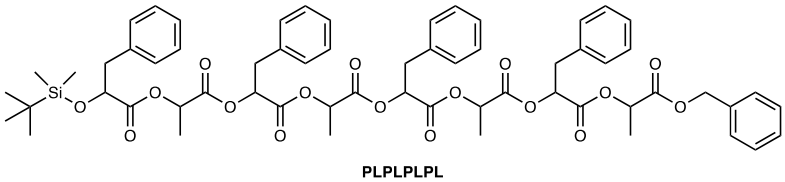


PLPPPLPL (**E**)**.** Viscous oil (8.37 g,90%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-6.65 (m, 30H, Ph-**H**), 5.43-5.02 (m, 9H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-

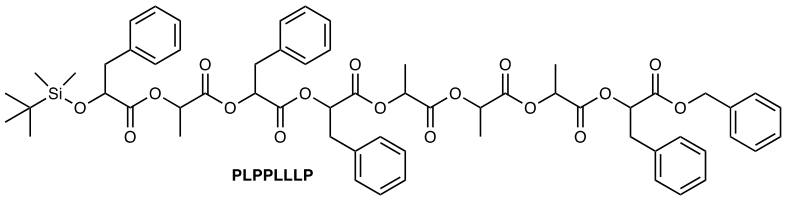
C**H₂**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.67 (m, 10H, Ph-C**H₂**-CH), 1.57-1.20 (m, 9H, C**H₃**-CH), 0.75 (m, 9H, (C**H₃**)$_3$C-Si), -0.24 (m, 6H, $(CH_3)_3$C-Si(C**H₃**)$_2$-O) ppm; M_n and Đ (GPC): 1320 Da and 1.03; MS (MALDI-TOF): m/z calcd for $C_{67}H_{74}O_{17}Si+Na^+$ [**M**+Na]$^+$: 1201; found 1202.



PLPLPPPL (**Q**)**.** Viscous oil (2.23 g, 92%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-6.65 (m, 30H, Ph-**H**), 5.43-5.02 (m, 9H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H₂**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.67 (m, 10H, Ph-C**H₂**-CH), 1.57-1.18 (m, 9H, C**H₃**-CH), 0.75 (m, 9H, (C**H₃**)$_3$C-Si), -0.24 (m, 6H, $(CH_3)_3$C-Si(C**H₃**)$_2$-O) ppm; M_n and Đ (GPC): 1280 Da and 1.03; MS (MALDI-TOF): m/z calcd for $C_{67}H_{74}O_{17}Si+Na^+$ [**M**+Na]$^+$: 1201; found 1202.

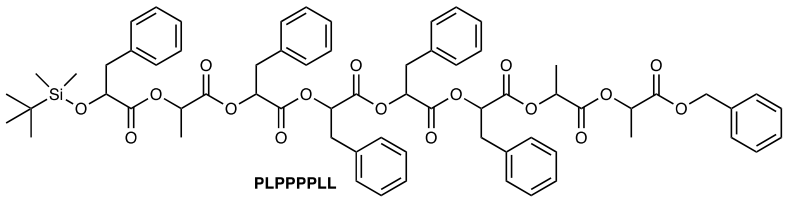


PLPLPLPL (**U**)**.** Viscous oil (2.35 g, 92%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-7.15 (m, 25H, Ph-**H**), 5.40-5.02 (m, 9H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H₂**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.67 (m, 8H, Ph-C**H₂**-CH), 1.57-1.20 (m, 12H, C**H₃**-CH), 0.75 (m, 9H, (C**H₃**)$_3$C-Si), -0.24 (m, 6H, $(CH_3)_3$C-Si(C**H₃**)$_2$-O) ppm; M_n and Đ (GPC): 1240 Da and 1.03; MS (MALDI-TOF): m/z calcd for $C_{61}H_{70}O_{17}Si+Na^+$ [**M**+Na]$^+$: 1125; found 1126.

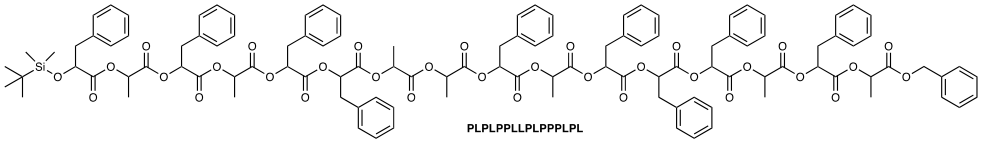


PLPPLLLP (**N**)**.** Viscous oil (1.75 g, 89%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-6.80 (m, 25H, Ph-**H**), 5.40-5.02 (m, 9H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-

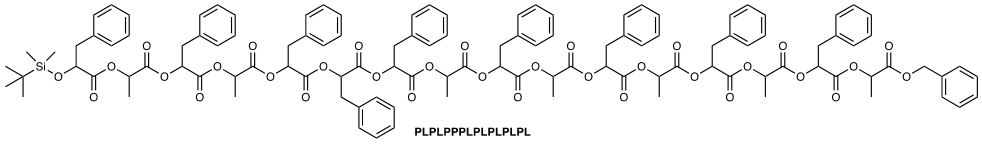
C**H₂**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.67 (m, 8H, Ph-C**H₂**-CH), 1.57-1.20 (m, 12H, C**H₃**-CH), 0.75 (m, 9H, (C**H₃**)$_3$C-Si), -0.24 (m, 6H, (CH_3)$_3$C-Si(C**H₃**)$_2$-O) ppm; M_n and Đ (GPC): 1210 Da and 1.03; MS (MALDI-TOF): m/z calcd for $C_{61}H_{70}O_{17}Si+Na^+$ [**M**+Na]$^+$: 1125; found 1126.



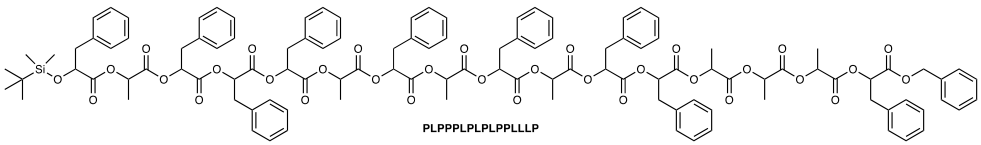
PLPPPPLL (**C**)**.** Viscous oil (1.80 g, 87%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-6.65 (m, 30H, Ph-**H**), 5.43-5.02 (m, 9H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H₂**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.67 (m, 10H, Ph-C**H₂**-CH), 1.57-1.18 (m, 9H, C**H₃**-CH), 0.75 (m, 9H, (C**H₃**)$_3$C-Si), -0.24 (m, 6H, (CH_3)$_3$C-Si(C**H₃**)$_2$-O) ppm; M_n and Đ (GPC): 1260 Da and 1.03; MS (MALDI-TOF): m/z calcd for $C_{67}H_{74}O_{17}Si+Na^+$ [**M**+Na]$^+$: 1201; found 1202.



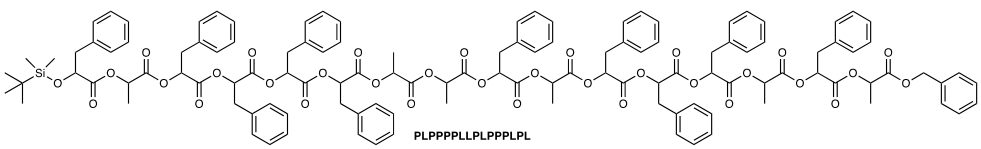
PLPLPPLLPLPPPLPL (**SE**)**.** A white solid (2.68 g, 83%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-6.65 (m, 50H, Ph-**H**), 5.43-5.02 (m, 17H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H₂**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.67 (m, 18H, Ph-C**H₂**-CH), 1.54-1.14 (m, 21H, C**H₃**-CH), 0.75 (m, 9H, (C**H₃**)$_3$C-Si), -0.24 (m, 6H, (CH_3)$_3$C-Si(C**H₃**)$_2$-O) ppm; M_n and Đ (GPC): 2430 Da and 1.01; MS (MALDI-TOF): m/z calcd for $C_{115}H_{122}O_{33}Si+Na^+$ [**M**+Na]$^+$: 2083; found 2083.



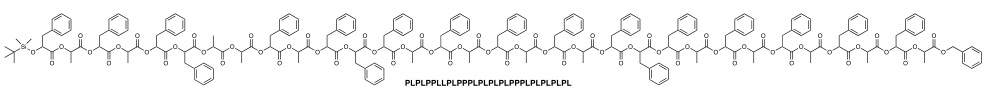
PLPLPPPLPLPLPLPL (**QU**). A white solid (2.83 g, 83%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-6.65 (m, 50H, Ph-**H**), 5.43-5.02 (m, 17H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H₂**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.67 (m, 18H, Ph-C**H₂**-CH), 1.54-1.14 (m, 21H, C**H₃**-CH), 0.75 (m, 9H, (C**H₃**)$_3$C-Si), -0.24 (m, 6H, (CH_3)$_3$C-Si(C**H₃**)$_2$-O) ppm; M_n and Đ (GPC): 2400 Da and 1.01; MS (MALDI-TOF): m/z calcd for $C_{115}H_{122}O_{33}Si+Na^+$ [**M**+Na]$^+$: 2083; found 2083.



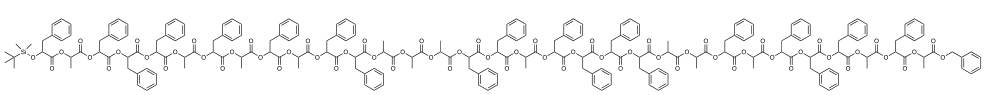
PLPPPLPLPLPPLLLP (**EN**). A white solid (2.02 g, 74%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-6.65 (m, 50H, Ph-**H**), 5.43-5.02 (m, 17H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H$_2$**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.67 (m, 18H, Ph-C**H$_2$**-CH), 1.54-1.14 (m, 21H, C**H$_3$**-CH), 0.75 (m, 9H, (C**H$_3$**)$_3$C-Si), -0.24 (m, 6H, (CH_3)$_3$C-Si(C**H$_3$**)$_2$-O) ppm; M_n and $Đ$ (GPC): 2390 Da and 1.01; MS (MALDI-TOF): m/z calcd for $C_{115}H_{122}O_{33}Si+Na^+$ [**M**+Na]$^+$: 2083; found 2083.



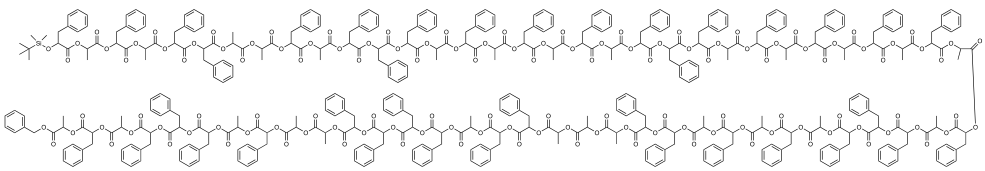
PLPPPPLLPLPPPLPL (**CE**). A white solid (1.47 g, 81%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-6.65 (m, 55H, Ph-**H**), 5.43-5.02 (m, 17H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H$_2$**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.71 (m, 20H, Ph-C**H$_2$**-CH), 1.54-1.14 (m, 18H, C**H$_3$**-CH), 0.75 (m, 9H, (C**H$_3$**)$_3$C-Si), -0.24 (m, 6H, (CH_3)$_3$C-Si(C**H$_3$**)$_2$-O) ppm; M_n and $Đ$ (GPC): 2440 Da and 1.01; MS (MALDI-TOF): m/z calcd for $C_{121}H_{126}O_{33}Si+Na^+$ [**M**+Na]$^+$: 2159; found 2159.



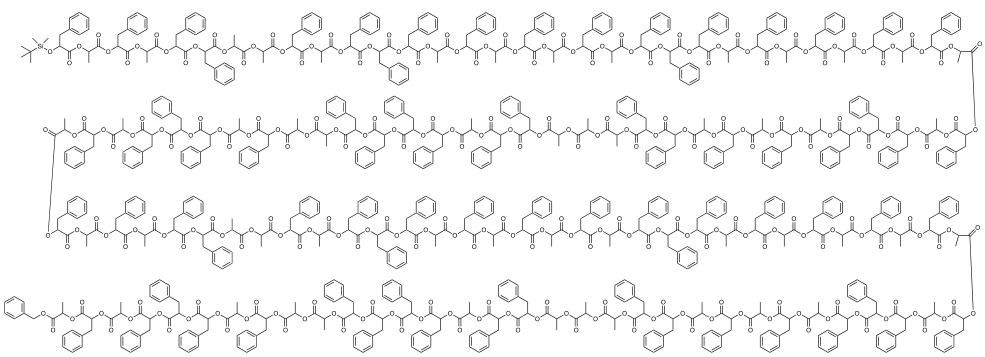
PLPLPPLLPLPPPLPLPLPLPPPLPLPLPLPL (**SEQU**). A white solid (2.95 g, 71%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-6.65 (m, 95H, Ph-**H**), 5.43-5.02 (m, 33H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H$_2$**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.67 (m, 36H, Ph-C**H$_2$**-CH), 1.54-1.14 (m, 42H, C**H$_3$**-CH), 0.75 (m, 9H, (C**H$_3$**)$_3$C-Si), -0.24 (m, 6H, (CH_3)$_3$C-Si(C**H$_3$**)$_2$-O) ppm; M_n and $Đ$ (GPC): 4120 Da and 1.02; MS (MALDI-TOF): m/z calcd for $C_{217}H_{222}O_{65}Si+Na^+$ [**M**+Na]$^+$: 3920; found 3920.



PLPPPLPLPLPPLLLPPLPPPPLLPLPPPLPL (**ENCE**). A white solid (2.06 g, 88%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-6.65 (m, 100H, Ph-**H**), 5.43-5.02 (m, 33H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H$_2$**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.67 (m, 38H, Ph-C**H$_2$**-CH), 1.54-1.14 (m, 39H, C**H$_3$**-CH), 0.75 (m, 9H, (C**H$_3$**)$_3$C-Si), -0.24 (m, 6H, (CH_3)$_3$C-Si(C**H$_3$**)$_2$-O) ppm; M_n and Đ (GPC): 4090 Da and 1.02; MS (MALDI-TOF): m/z calcd for $C_{223}H_{226}O_{65}Si+Na^+$ [**M**+Na]$^+$: 3996; found 3996.



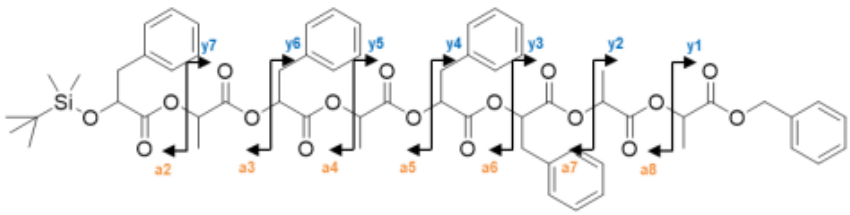
PLPLPPLLPLPPPLPLPLPLPPPLPLPLPLPLPLPPPLPLPLPPLLLPPLPPPPLLPLPPPLPL (**SEQUENCE**). A white solid (2.49 g, 71%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-6.65 (m, 190H, Ph-**H**), 5.43-5.02 (m, 65H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H$_2$**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.67 (m, 74H, Ph-C**H$_2$**-CH), 1.54-1.14 (m, 81H, C**H$_3$**-CH), 0.75 (m, 9H, (C**H$_3$**)$_3$C-Si), -0.24 (m, 6H, (CH_3)$_3$C-Si(C**H$_3$**)$_2$-O) ppm; M_n and Đ (GPC): 8170 Da and 1.03; MS (MALDI-TOF): m/z calcd for $C_{427}H_{426}O_{129}Si+Na^+$ [**M**+Na]$^+$: 7673; found 7674.

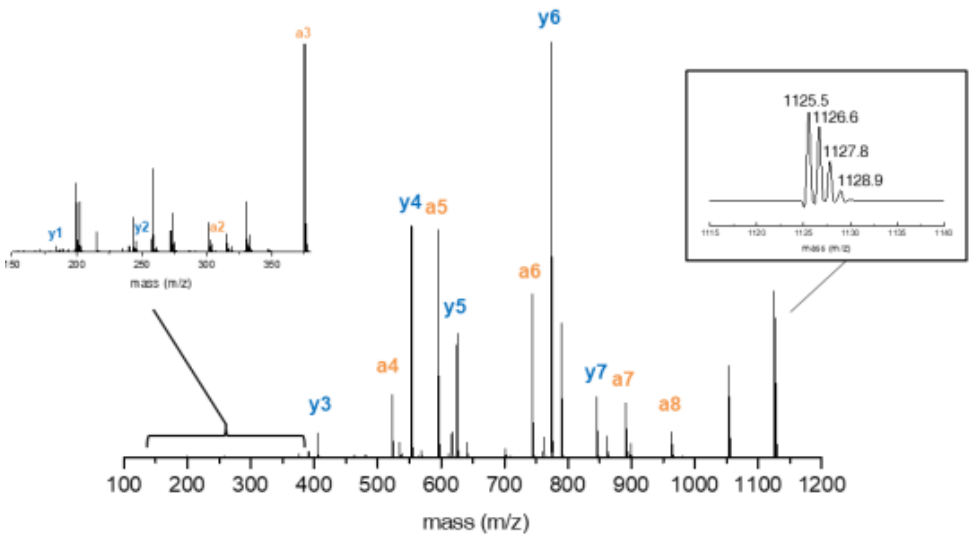


PLPLPPLLPLPPPLPLPLPLPPPLPLPLPLPLPLPPPLPLPLPPLLLPPLPPPPLLPLPPPLPLPLPPLLPLPPPLPLPLPPPLPLPLPLPLPPPLPLPLPPLLLPPLPPPPLLPLPPPLPL (**SEQUENCESEQUENCE**). A white solid (0.63 g, 70%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.40-6.65 (m, 375H, Ph-**H**), 5.43-5.02 (m, 129H, COO-C**H**(CH_2Ph)C=O, COO-C**H**(CH_3)C=O and Ph-C**H$_2$**-O), 4.32 (m, 1H, SiO-C**H**(CH_2Ph)C=O), 3.40-2.67 (m, 148H, Ph-C**H$_2$**-CH), 1.54-1.14 (m, 162H, C**H$_3$**-CH), 0.75 (m, 9H, (C**H$_3$**)$_3$C-Si), -0.24 (m, 6H, (CH_3)$_3$C-Si(C**H$_3$**)$_2$-O) ppm; M_n and Đ (GPC): 16230 Da and 1.03; MS (MALDI-TOF): molecular m/z calcd for $C_{841}H_{830}O_{257}Si+Na^+$ [**M**+Na]$^+$: 15101; found 15105.

A.2.2 MALDI-TOF/TOF tandem mass sequencing results of PcLs

MALDI-TOF/TOF tandem mass analysis and decoding table of the 8-bit PcL (S).

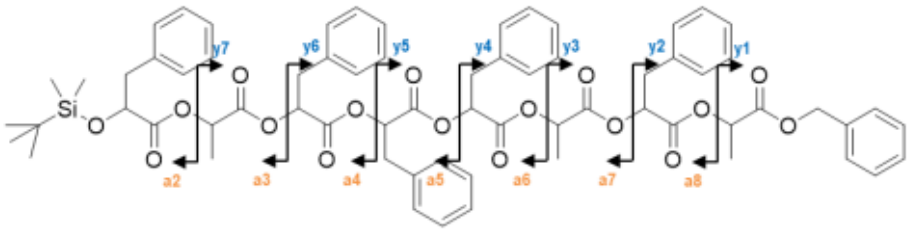


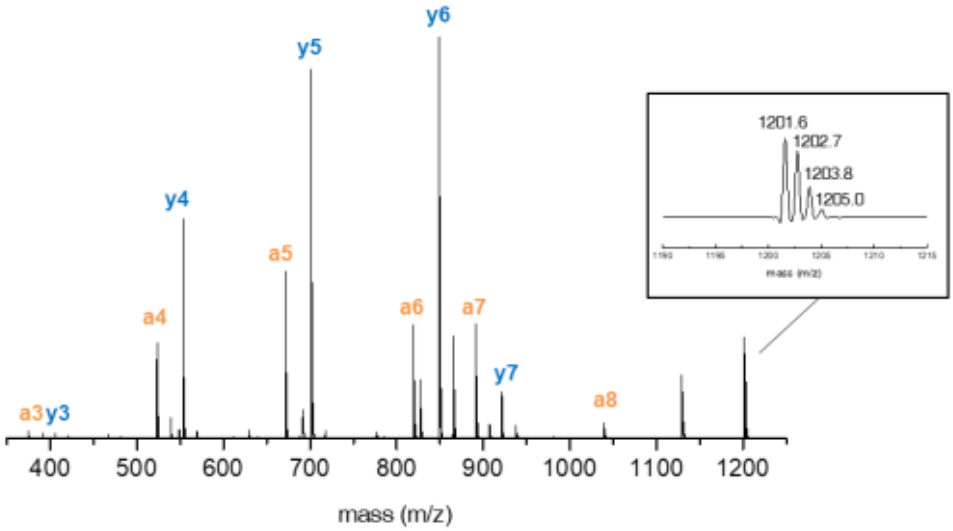


S (PLPLPPLL)

Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	1125.4	1125.5			$[M+Na]^+$	1125.4	1125.5		
y7	845.3	845.5	280.0	Si-P	a8	963.4	963.5	162.0	Bz-L
y6	773.3	773.5	72.0	L	a7	891.3	891.5	72.0	L
y5	625.2	625.5	148.0	P	a6	743.3	743.5	148.0	P
y4	553.2	553.4	72.1	L	a5	595.2	595.5	148.0	P
y3	405.1	405.3	148.1	P	a4	523.2	523.5	72.0	L
y2	257.1	257.3	148.0	P	a3	375.2	375.4	148.1	P
y1	185.1	185.2	72.1	L	a2	303.1	303.4	72.0	L

MALDI-TOF/TOF tandem mass analysis and decoding table of the 8-bit PcL (E).

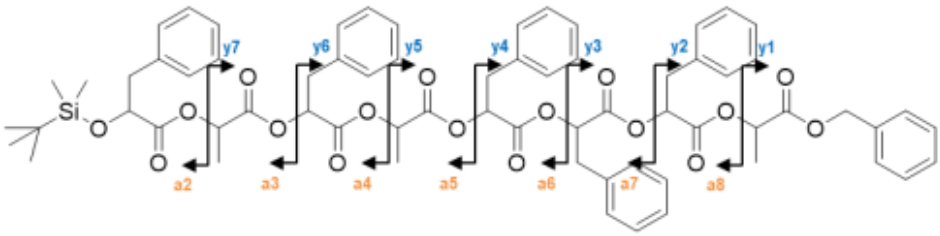


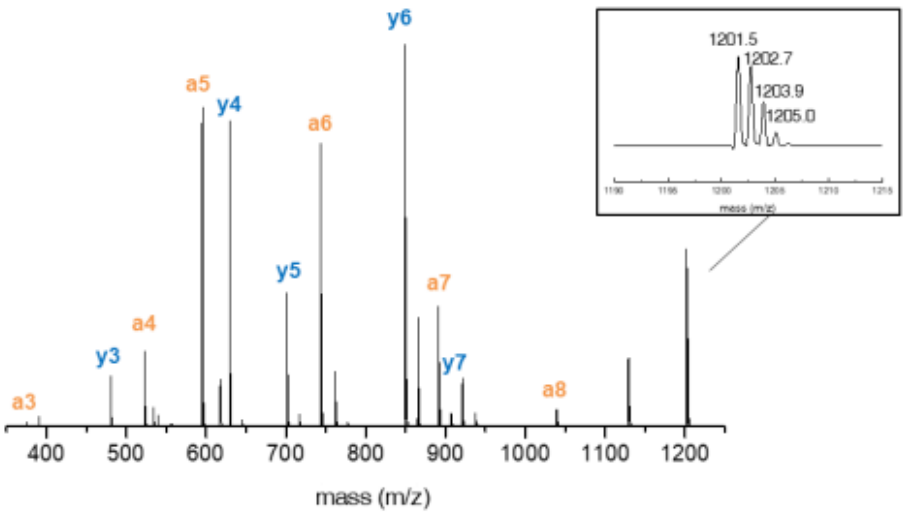


E (PLPPPLPL)

Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	1201.5	1201.5			$[M+Na]^+$	1201.5	1201.5		
y7	921.3	921.4	280.1	Si-P	a8	1039.4	1039.5	162.0	Bz-L
y6	849.3	849.4	72.0	L	a7	891.3	891.4	148.1	P
y5	701.2	701.4	148.0	P	a6	819.3	819.5	71.9	L
y4	553.2	553.4	148.0	P	a5	672.3	671.4	148.1	P
y3	405.1	405.3	148.1	P	a4	523.2	523.4	148.0	P
y2	333.1	333.3	72.0	L	a3	375.2	375.3	148.1	P
y1	185.1	185.2	148.1	P	a2	303.1	303.3	72.0	L

MALDI-TOF/TOF tandem mass analysis and decoding table of the 8-bit PcL (Q).

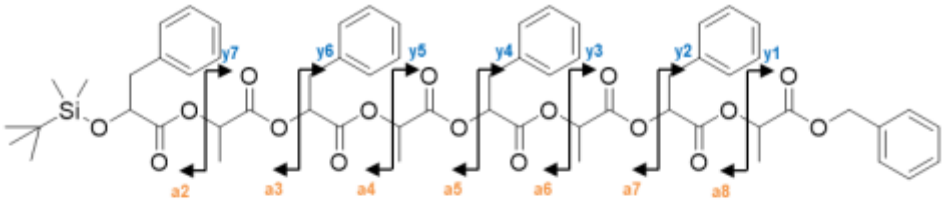


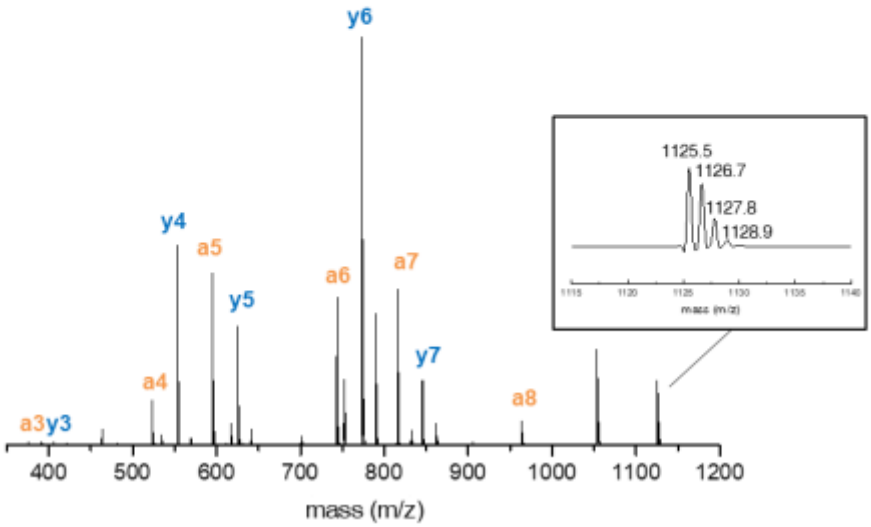


Q (PLPLPPPL)

Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	1201.5	1201.5			$[M+Na]^+$	1201.5	1201.5		
y7	921.3	921.4	280.1	Si-P	a8	1039.4	1039.6	161.9	Bz-L
y6	849.3	849.4	72.0	L	a7	891.3	891.4	148.2	P
y5	701.2	701.4	148.0	P	a6	743.3	743.4	148.0	P
y4	629.2	629.3	72.1	L	a5	595.2	595.4	148.0	P
y3	481.2	481.3	148.0	P	a4	523.2	523.4	72.0	L
y2	333.1	333.2	148.1	P	a3	375.2	375.3	148.1	P
y1	185.1	185.2	148.0	P	a2	303.1	303.3	72.0	L

MALDI-TOF/TOF tandem mass analysis and decoding table of the 8-bit PcL (U).

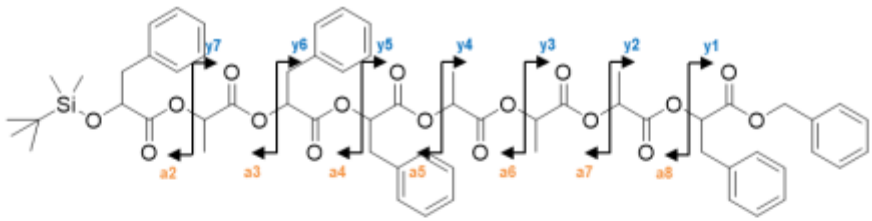


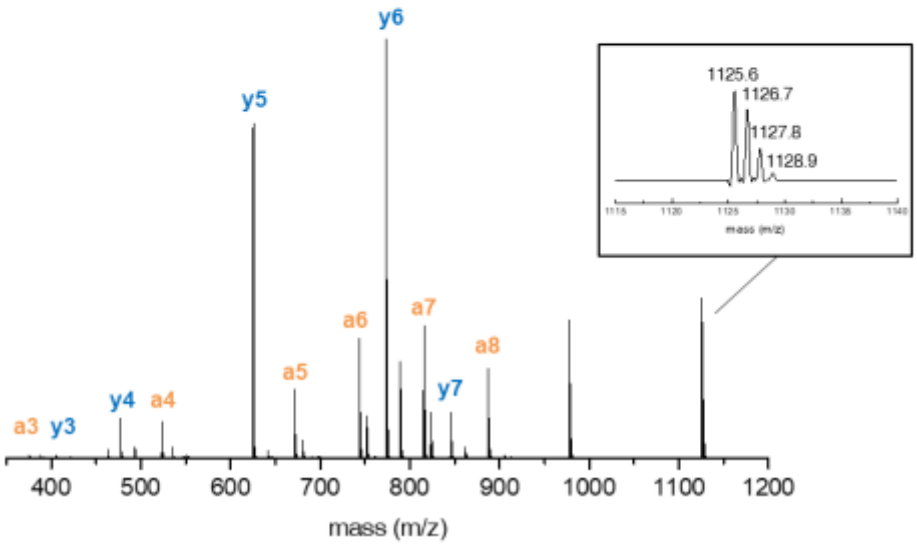


U (PLPLPLPL)

Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	1125.4	1125.5			$[M+Na]^+$	1125.4	1125.5		
y7	845.3	845.5	280.0	Si-P	a8	963.4	963.5	162.0	Bz-L
y6	773.3	773.5	72.0	L	a7	815.3	815.4	148.1	P
y5	625.2	625.4	148.1	P	a6	743.3	743.5	71.9	L
y4	553.2	553.4	72.0	L	a5	595.2	595.5	148	P
y3	405.1	405.4	148.0	P	a4	523.2	523.5	72.0	L
y2	333.1	333.4	72.0	L	a3	375.2	375.4	148.1	P
y1	185.1	185.2	148.2	P	a2	303.1	303.3	72.1	L

MALDI-TOF/TOF tandem mass analysis and decoding table of the 8-bit PcL (N).

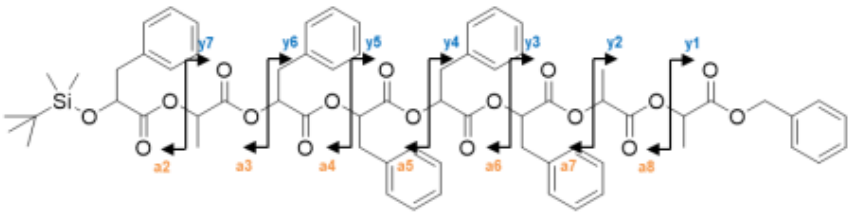


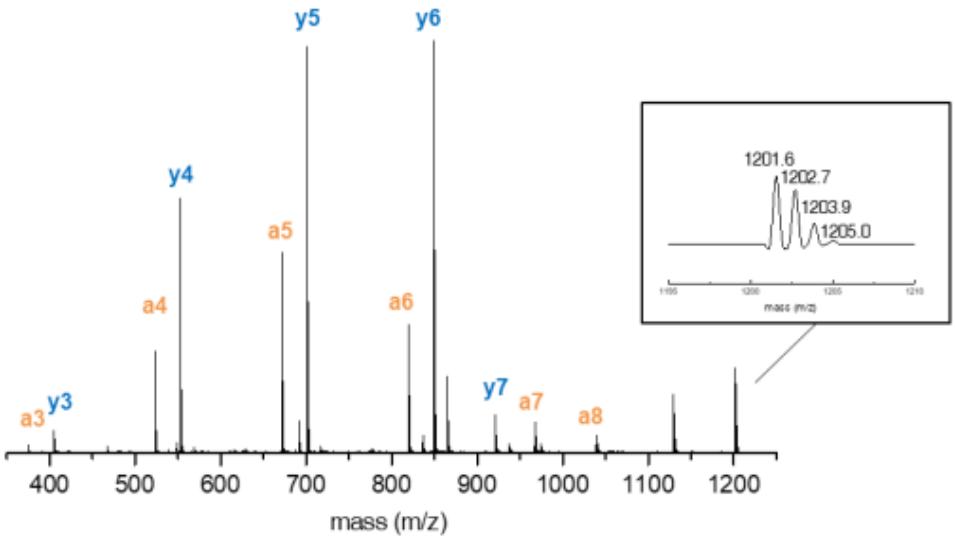


N (PLPPLLLP)

Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	1125.4	1125.6			$[M+Na]^+$	1125.4	1125.6		
y7	845.3	845.5	280.1	Si-P	a8	887.3	887.5	238.1	Bz-P
y6	773.3	773.6	71.9	L	a7	815.3	815.6	71.9	L
y5	625.2	625.5	148.1	P	a6	743.3	743.6	72.0	L
y4	477.2	477.4	148.1	P	a5	671.3	671.6	72.0	L
y3	405.1	405.4	72.0	L	a4	523.2	523.5	148.1	P
y2	333.1	333.3	72.1	L	a3	375.2	375.4	148.1	P
y1	261.1	261.3	72.0	L	a2	303.1	303.3	72.1	L

MALDI-TOF/TOF tandem mass analysis and decoding table of the 8-bit PcL (C).

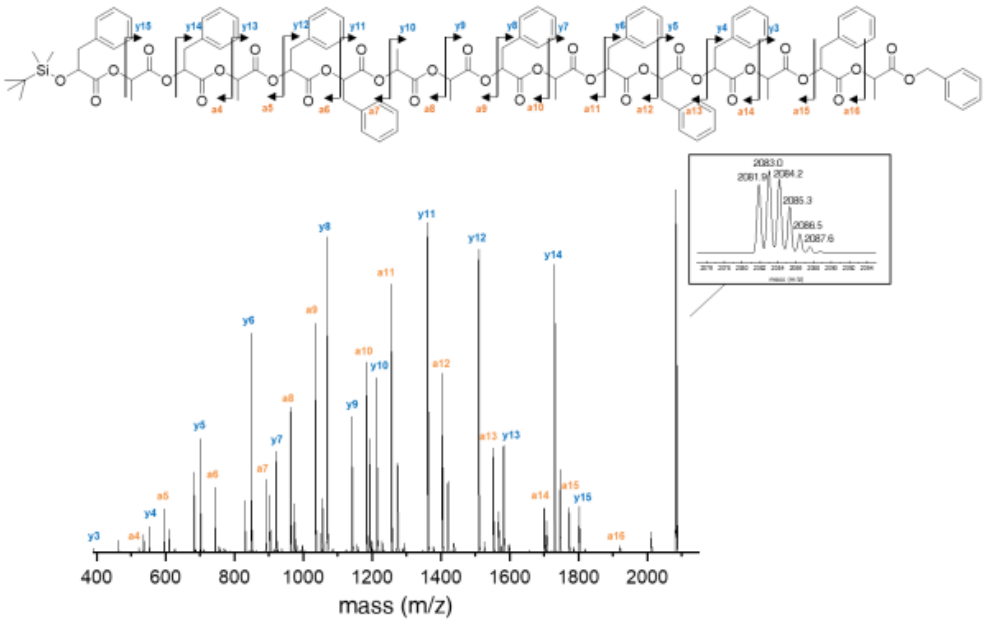




C (PLPPPPLL)

Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	1201.5	1201.6			$[M+Na]^+$	1201.5	1201.6		
y7	921.3	921.5	280.1	Si-P	a8	1039.4	1039.5	162.1	Bz-L
y6	849.3	849.5	72.0	L	a7	967.4	967.5	72.0	L
y5	701.2	701.4	148.1	P	a6	819.3	819.5	148.0	P
y4	553.2	553.4	148.0	P	a5	671.3	671.4	148.1	P
y3	405.1	405.3	148.1	P	a4	523.2	523.4	148.0	P
y2	257.1	257.3	148.0	P	a3	375.2	375.3	148.1	P
y1	185.1	185.2	72.1	L	a2	303.1	303.3	72.0	L

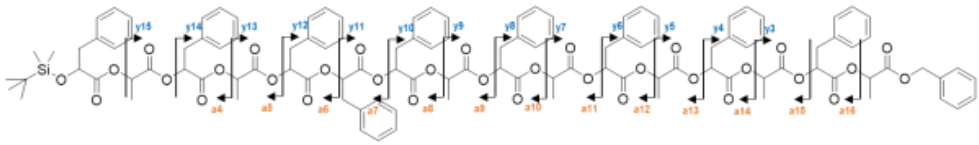
MALDI-TOF/TOF tandem mass analysis and decoding table of the 16-bit PcL (SE).

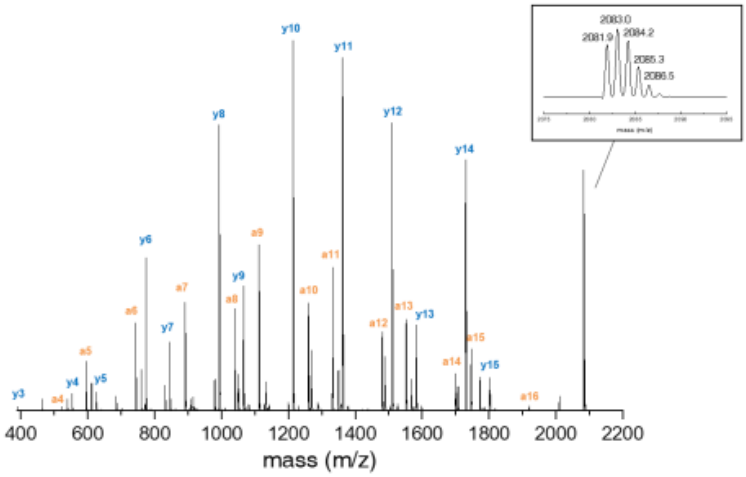


SE (PLPLPPLLPLPPPLPL)

Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	2081.8	2083.0			$[M+Na]^+$	2081.8	2083.0		
y15	1801.6	1802.7	280.3	Si-P	a16	1919.7	1920.8	162.2	Bz-L
y14	1729.6	1730.7	72.0	L	a15	1771.6	1772.7	148.1	P
y13	1581.5	1582.7	148.0	P	a14	1699.6	1700.7	72.0	L
y12	1509.5	1510.7	72.0	L	a13	1551.6	1552.7	148.0	P
y11	1361.5	1361.7	149.0	P	a12	1403.5	1403.7	149.0	P
y10	1213.4	1213.6	148.1	P	a11	1255.5	1255.7	148.0	P
y9	1141.4	1141.6	72.0	L	a10	1183.4	1183.7	72.0	L
y8	1069.4	1069.6	72.0	L	a9	1035.4	1035.7	148.0	P
y7	921.3	921.5	148.1	P	a8	963.4	963.6	72.1	L
y6	849.3	849.5	72.0	L	a7	891.3	891.6	72.0	L
y5	701.2	701.4	148.1	P	a6	743.3	743.5	148.1	P
y4	553.2	553.4	148.0	P	a5	595.2	595.5	148.0	P
y3	405.1	405.3	148.1	P	a4	523.2	523.4	72.1	L

MALDI-TOF/TOF tandem mass analysis and decoding table of the 16-bit PcL (QU).

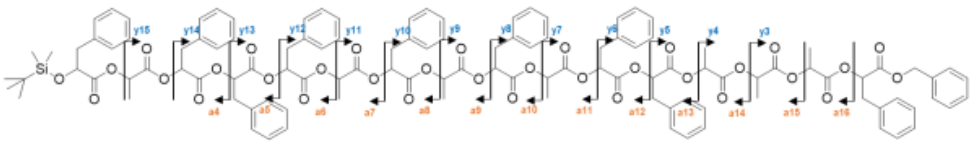


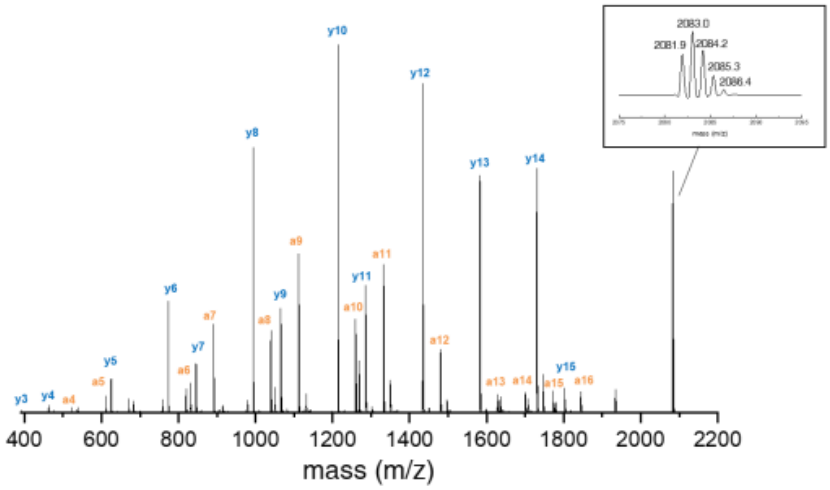


QU (PLPLPPPLPLPLPLPL)

Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	2081.8	2083.1			$[M+Na]^+$	2081.8	2083.1		
y15	1801.6	1802.6	280.5	Si-P	a16	1919.7	1920.7	162.4	Bz-L
y14	1729.6	1730.6	72.0	L	a15	1771.6	1772.6	148.1	P
y13	1581.5	1582.6	148.0	P	a14	1699.6	1700.6	72.0	L
y12	1509.5	1510.6	72.0	L	a13	1551.6	1552.6	148.0	P
y11	1361.5	1361.6	149.0	P	a12	1479.5	1480.6	72.0	L
y10	1213.4	1213.6	148.0	P	a11	1331.5	1331.6	149.0	P
y9	1065.4	1065.6	148.0	P	a10	1259.5	1259.7	71.9	L
y8	993.3	993.5	72.1	L	a9	1111.4	1111.6	148.1	P
y7	845.3	845.5	148.0	P	a8	1039.4	1039.6	72.0	L
y6	773.3	773.5	72.0	L	a7	891.3	891.5	148.1	P
y5	625.2	625.4	148.1	P	a6	743.3	743.5	148.0	P
y4	553.2	553.4	72.0	L	a5	595.2	595.4	148.1	P
y3	405.1	405.4	148.0	P	a4	523.2	523.4	72.0	L

MALDI-TOF/TOF tandem mass analysis and decoding table of the 16-bit PcL (EN).

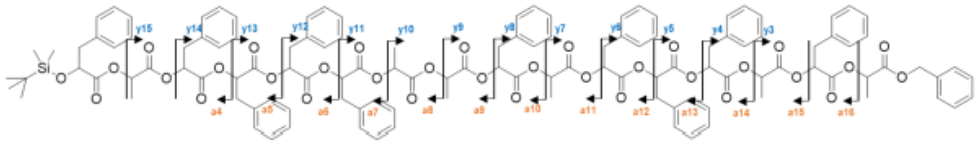


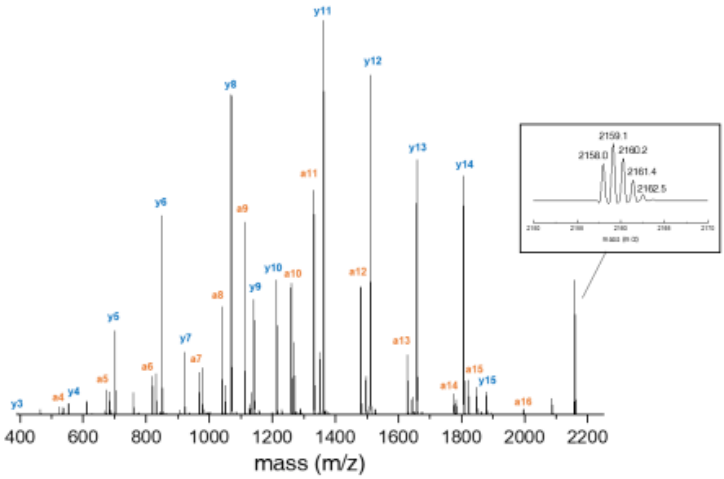


EN (PLPPPLPLPLPPLLLP)

Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	2081.8	2083.0			$[M+Na]^+$	2081.8	2083.0		
y15	1801.6	1802.8	280.2	Si-P	a16	1843.7	1844.8	238.2	Bz-P
y14	1729.6	1730.8	72.0	L	a15	1771.6	1772.8	72.0	L
y13	1581.5	1582.8	148.0	P	a14	1699.6	1700.8	72.0	L
y12	1433.5	1433.8	149.0	P	a13	1627.6	1628.8	72.0	L
y11	1285.4	1285.7	148.1	P	a12	1479.5	1480.8	148.0	P
y10	1213.4	1213.7	72.0	L	a11	1331.5	1331.8	149.0	P
y9	1065.4	1065.7	148.0	P	a10	1259.5	1259.8	72.0	L
y8	993.3	993.6	72.1	L	a9	1111.4	1111.7	148.1	P
y7	845.3	845.6	148.0	P	a8	1039.4	1039.7	72.0	L
y6	773.3	773.6	72.0	L	a7	891.3	891.6	148.1	P
y5	625.2	625.5	148.1	P	a6	819.3	819.6	72.0	L
y4	477.2	477.3	148.2	P	a5	671.3	671.5	148.1	P
y3	405.1	405.3	72.0	L	a4	523.2	523.5	148.0	P

MALDI-TOF/TOF tandem mass analysis and decoding table of the 16-bit PcL (CE).

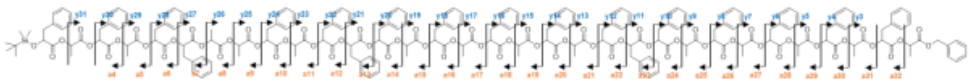


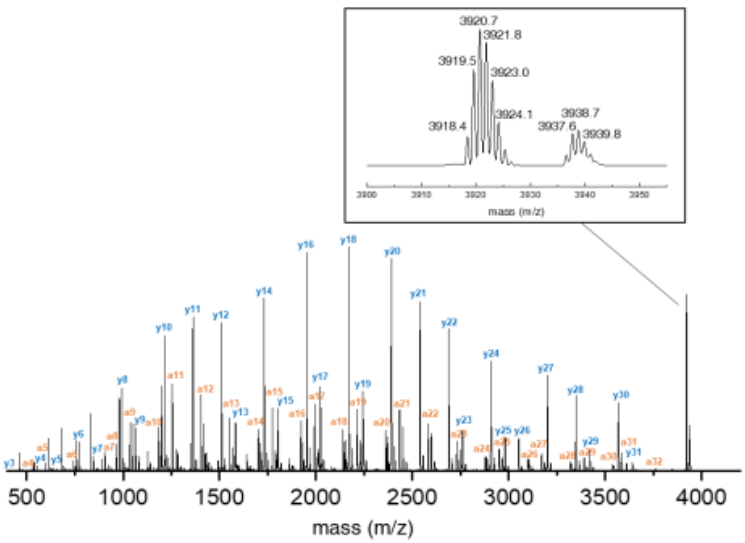


CE (PLPPPPLLPLPPPLPL)

Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	2157.8	2159.1			$[M+Na]^+$	2157.8	2159.1		
y15	1877.6	1878.8	280.3	Si-P	a16	1995.7	1996.9	162.2	Bz-L
y14	1805.6	1806.8	72.0	L	a15	1847.7	1848.8	148.1	P
y13	1657.6	1658.8	148.0	P	a14	1775.6	1776.8	72.0	L
y12	1509.5	1510.8	148.0	P	a13	1627.6	1628.8	148.0	P
y11	1361.5	1361.8	149.0	P	a12	1479.5	1479.8	149.0	P
y10	1213.4	1213.7	148.1	P	a11	1331.5	1331.8	148.0	P
y9	1141.4	1141.7	72.0	L	a10	1259.5	1259.8	72.0	L
y8	1069.4	1069.7	72.0	L	a9	1111.4	1111.7	148.1	P
y7	921.3	921.6	148.1	P	a8	1039.4	1039.7	72.0	L
y6	849.3	849.6	72.0	L	a7	967.4	967.7	72.0	L
y5	701.2	701.5	148.1	P	a6	819.3	819.6	148.1	P
y4	553.2	553.4	148.1	P	a5	671.3	671.5	148.1	P
y3	405.1	405.2	148.2	P	a4	523.2	523.5	148.0	P

MALDI-TOF/TOF tandem mass analysis and decoding table of the 32-bit PcL (SEQU).

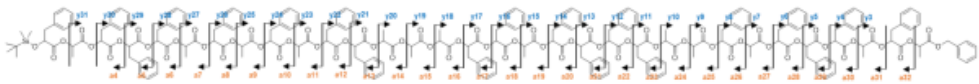


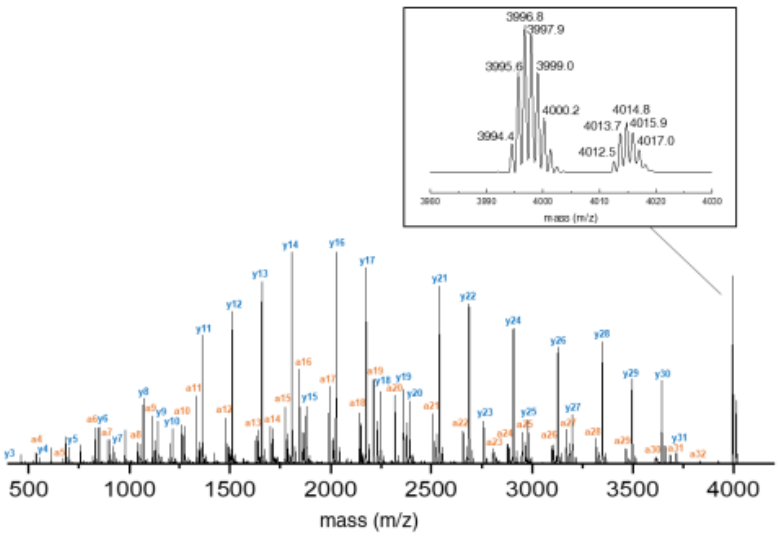


SEQU (PLPLPPLLPLPPPLPLPLPLPPPLPLPLPLPL)

Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	3918.4	3920.7			$[M+Na]^+$	3918.4	3920.7		
y31	3638.2	3640.6	280.1	Si-P	a32	3756.3	3758.7	162.0	Bz-L
y30	3566.2	3568.6	72.0	L	a31	3608.2	3610.7	148.0	P
y29	3418.2	3420.5	148.1	P	a30	3536.2	3538.6	72.1	L
y28	3346.1	3348.5	72.0	L	a29	3388.2	3390.6	148.0	P
y27	3198.1	3200.4	148.1	P	a28	3316.2	3318.6	72.0	L
y26	3050.0	3052.5	147.9	P	a27	3168.1	3170.4	148.2	P
y25	2978.0	2980.5	72.0	L	a26	3096.1	3098.6	71.8	L
y24	2906.0	2907.5	73.0	L	a25	2948.0	2949.5	149.1	P
y23	2757.9	2759.4	148.1	P	a24	2876.0	2877.5	72.0	L
y22	2685.9	2687.5	71.9	L	a23	2728.0	2729.5	148.0	P
y21	2537.9	2539.5	148.0	P	a22	2579.9	2581.5	148.0	P
y20	2389.8	2391.5	148.0	P	a21	2431.8	2433.5	148.0	P
y19	2241.8	2243.4	148.1	P	a20	2359.8	2361.5	72.0	L
y18	2169.7	2171.5	71.9	L	a19	2211.8	2213.5	148.0	P
y17	2021.7	2023.4	148.1	P	a18	2139.8	2141.5	72.0	L
y16	1949.7	1951.4	72.0	L	a17	1991.7	1993.5	148.0	P
y15	1801.6	1803.4	148.0	P	a16	1919.7	1921.5	72.0	L
y14	1729.6	1731.4	72.0	L	a15	1771.6	1773.4	148.1	P
y13	1581.5	1583.3	148.1	P	a14	1699.6	1701.4	72.0	L
y12	1509.5	1510.3	73.0	L	a13	1551.6	1553.3	148.1	P
y11	1361.5	1362.2	148.1	P	a12	1403.5	1404.3	149.0	P
y10	1213.4	1214.2	148.0	P	a11	1255.4	1256.2	148.1	P
y9	1065.4	1066	148.2	P	a10	1183.4	1184.2	72.0	L
y8	993.3	994.0	72.0	L	a9	1035.4	1036.1	148.1	P
y7	845.3	846.0	148.0	P	a8	963.4	964.1	72.0	L
y6	773.3	774.0	72.0	L	a7	891.3	892.0	72.1	L
y5	625.2	625.8	148.2	P	a6	743.3	743.9	148.1	P
y4	553.2	553.8	72.0	L	a5	595.2	595.8	148.1	P
y3	405.1	405.6	148.2	P	a4	523.2	523.8	72.0	L

MALDI-TOF/TOF tandem mass analysis and decoding table of the 32-bit PcL (ENCE).



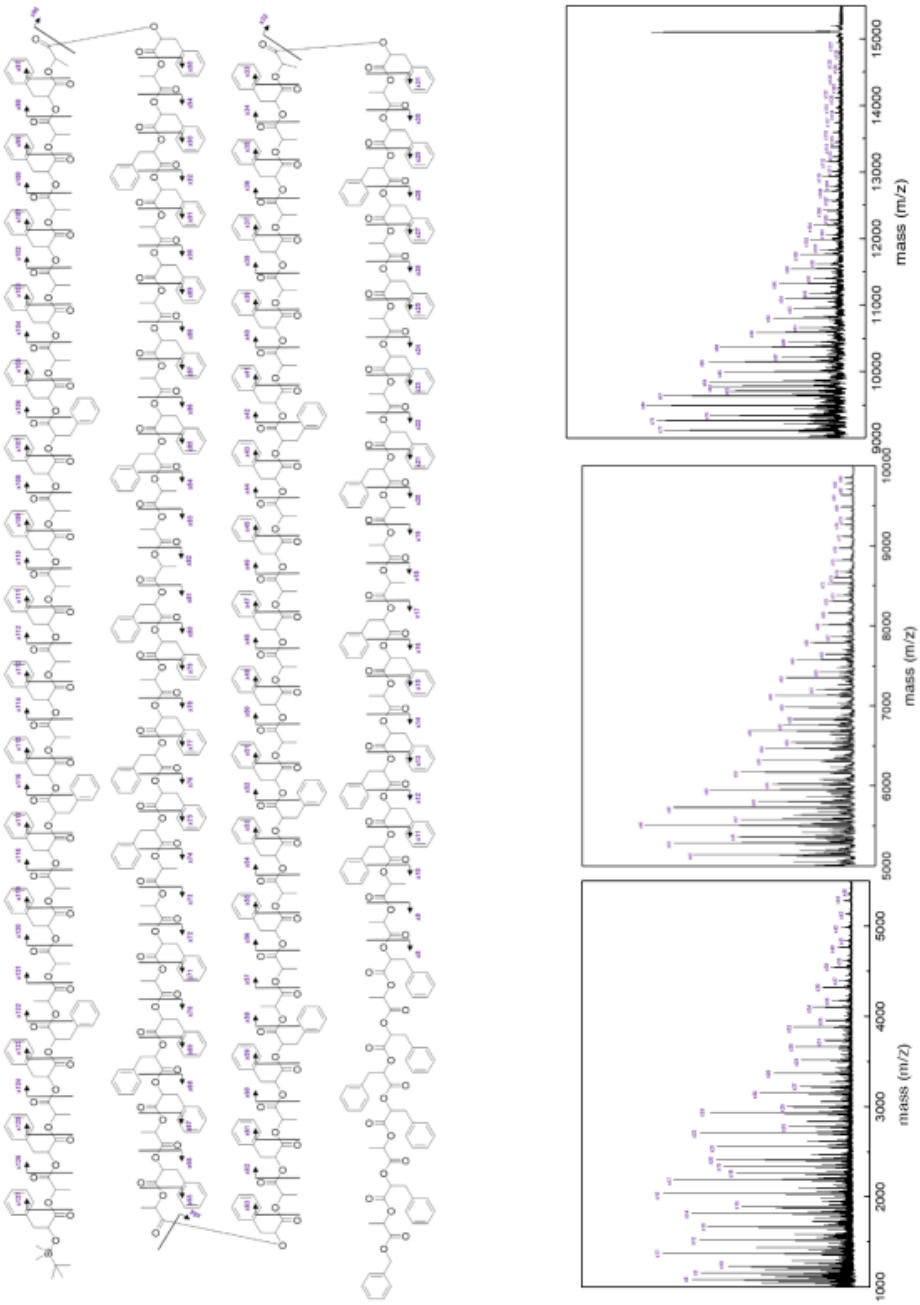


ENCE (PLPPPLPLPLPPLLLPPLPPPPLLPLPPPLPL)

Si → Bz	Calculated m/z	Found m/z	Difference	Sequence	Bz → Si	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	3994.4	3996.8			$[M+Na]^+$	3994.4	3996.8		
y31	3714.3	3716.7	280.1	Si-P	a32	3832.3	3834.4	162.4	Bz-L
y30	3642.2	3644.6	72.1	L	a31	3684.3	3686.5	147.9	P
y29	3494.2	3496.6	148.0	P	a30	3612.3	3614.6	71.9	L
y28	3346.1	3348.5	148.1	P	a29	3464.2	3466.4	148.2	P
y27	3198.1	3200.2	148.3	P	a28	3316.2	3318.6	147.8	P
y26	3126.1	3128.4	71.8	L	a27	3168.1	3170.5	148.1	P
y25	2978.0	2980.5	147.9	P	a26	3096.1	3097.5	73.0	L
y24	2906.0	2908.4	72.1	L	a25	2948.0	2950.4	147.1	P
y23	2757.9	2759.4	149.0	P	a24	2876.0	2877.5	72.9	L
y22	2685.9	2687.5	71.9	L	a23	2804.0	2805.6	71.9	L
y21	2537.9	2539.5	148.0	P	a22	2655.9	2657.5	148.1	P
y20	2389.8	2391.4	148.1	P	a21	2507.9	2509.5	148.0	P
y19	2317.8	2319.4	72.0	L	a20	2359.8	2361.5	148.0	P
y18	2245.8	2247.4	72.0	L	a19	2211.8	2213.5	148.0	P
y17	2173.8	2175.4	72.0	L	a18	2139.8	2141.4	72.1	L
y16	2025.7	2027.4	148.0	P	a17	1991.7	1993.4	148.0	P
y15	1877.6	1879.4	148.0	P	a16	1843.6	1845.4	148.0	P
y14	1805.6	1807.3	72.1	L	a15	1771.6	1773.4	72.0	L
y13	1657.6	1659.3	148.0	P	a14	1699.6	1701.4	72.0	L
y12	1509.5	1510.3	149.0	P	a13	1627.6	1629.3	72.1	L
y11	1361.5	1362.2	148.1	P	a12	1479.5	1481.3	148.0	P
y10	1213.4	1214.2	148.0	P	a11	1331.5	1332.3	149.0	P
y9	1141.4	1142.1	72.1	L	a10	1259.5	1260.3	72.0	L
y8	1069.4	1070.1	72.0	L	a9	1111.4	1112.2	148.1	P
y7	921.3	922.0	148.1	P	a8	1039.4	1040.1	72.1	L
y6	849.3	849.9	72.1	L	a7	891.3	892.0	148.1	P
y5	701.2	701.8	148.1	P	a6	819.3	820.0	72.0	L
y4	553.2	553.7	148.1	P	a5	671.3	671.8	148.2	P
y3	405.1	405.5	148.2	P	a4	523.2	523.7	148.1	P

A.2.3 Degradative sequencing result of 128-bit PcL

MALDI-TOF analysis and decoding table of the hydrolyzed 128-bit PcL.

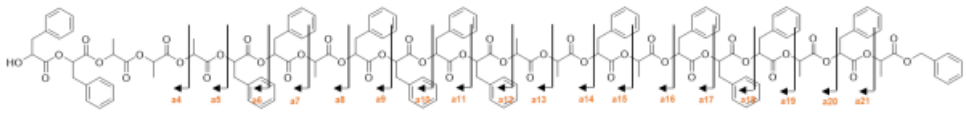


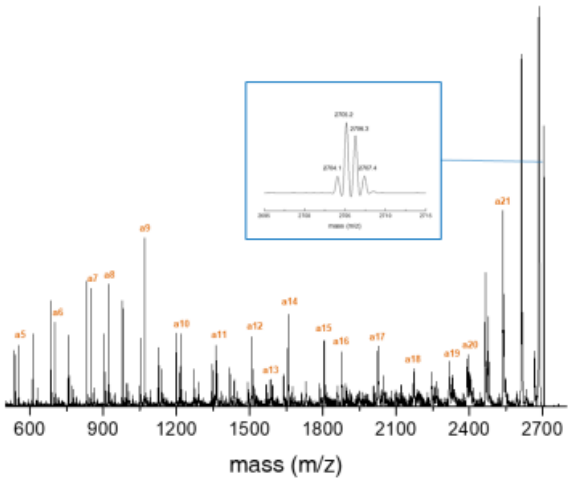
Si → Bz	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	15101	15104		
x127	14838	14841	263	Si-P
x126	14766	14769	72	L
x125	14618	14621	148	P
x124	14546	14550	71	L
x123	14398	14402	148	P
x122	14250	14254	148	P
x121	14178	14181	73	L
x120	14106	14109	72	L
x119	13957	13962	147	P
x118	13885	13888	74	L
x117	13737	13739	149	P
x116	13589	13591	148	P
x115	13441	13442	149	P
x114	13369	13371	71	L
x113	13221	13223	148	P
x112	13149	13150	73	L
x111	13000	13003	147	P
x110	12928	12931	72	L
x109	12780	12782	149	P
x108	12708	12710	72	L
x107	12560	12562	148	P
x106	12412	12414	148	P
x105	12264	12266	148	P
x104	12192	12194	72	L
x103	12043	12045	149	P
x102	11971	11973	72	L
x101	11823	11825	148	P
x100	11751	11753	72	L
x99	11603	11605	148	P
x98	11531	11532	73	L
x97	11383	11383	149	P
x96	11311	11313	70	L
x95	11163	11164	149	P
x94	11090	11092	72	L
x93	10942	10942	150	P
x92	10794	10795	147	P

x91	10646	10647	148	P
x90	10574	10575	72	L
x89	10426	10428	147	P
x88	10354	10354	74	L
x87	10206	10206	148	P
x86	10133	10134	72	L
x85	9985	9986	148	P
x84	9837	9837	149	P
x83	9765	9766	71	L
x82	9693	9694	72	L
x81	9621	9622	72	L
x80	9473	9473	149	P
x79	9325	9324	149	P
x78	9253	9252	72	L
x77	9104	9104	148	P
x76	8956	8956	148	P
x75	8808	8808	148	P
x74	8660	8660	148	P
x73	8588	8588	72	L
x72	8516	8516	72	L
x71	8368	8367	149	P
x70	8296	8296	71	L
x69	8147	8147	149	P
x68	7999	7999	148	P
x67	7851	7850	149	P
x66	7779	7779	71	L
x65	7631	7630	149	P
x64	7559	7558	72	L
x63	7411	7410	148	P
x62	7339	7338	72	L
x61	7190	7190	148	P
x60	7118	7118	72	L
x59	6970	6969	149	P
x58	6822	6821	148	P
x57	6750	6749	72	L
x56	6678	6677	72	L
x55	6530	6529	148	P
x54	6458	6456	73	L
x53	6310	6308	148	P

x52	6161	6160	148	P
x51	6013	6012	148	P
x50	5941	5941	71	L
x49	5793	5792	149	P
x48	5721	5721	71	L
x47	5573	5572	149	P
x46	5501	5500	72	L
x45	5353	5352	148	P
x44	5280	5281	71	L
x43	5132	5132	149	P
x42	4984	4983	149	P
x41	4836	4836	147	P
x40	4764	4764	72	L
x39	4616	4615	149	P
x38	4544	4544	71	L
x37	4396	4395	149	P
x36	4323	4322	73	L
x35	4175	4174	148	P
x34	4103	4102	72	L
x33	3955	3954	148	P
x32	3883	3882	72	L
x31	3735	3734	148	P
x30	3663	3662	72	L
x29	3515	3514	148	P
x28	3366	3366	148	P
x27	3218	3218	148	P
x26	3146	3146	72	L
x25	2998	2998	148	P
x24	2926	2925	73	L
x23	2778	2777	148	P
x22	2706	2705	72	L
x21	2558	2557	148	P
x20	2409	2409	148	P
x19	2337	2337	72	L
x18	2265	2265	72	L
x17	2193	2193	72	L
x16	2045	2045	148	P
x15	1897	1897	148	P
x14	1825	1825	72	L
x13	1677	1677	148	P

Tandem mass sequencing of the hydrolyzed fragment, x22 (HO-PPLLLPPLPPPP LLPLPPPLPL-Bz)



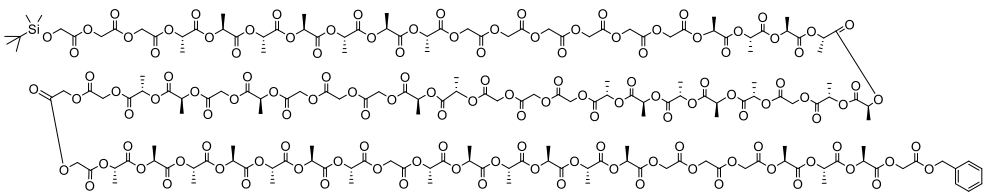


Bz → OH	Calculated m/z	Found m/z	Difference	Sequence
$[M+Na]^+$	2703.9	2705.2		
a21	2541.8	2543.0	162.2	Bz-L
a20	2393.8	2395.1	147.9	P
a19	2321.8	2323.0	72.1	L
a18	2173.7	2174.9	148.1	P
a17	2025.7	2026.3	148.6	P
a16	1877.6	1877.3	149.0	P
a15	1805.6	1805.5	71.8	L
a14	1657.5	1657.4	148.1	P
a13	1585.5	1585.4	72.0	L
a12	1513.5	1513.4	72.0	L
a11	1365.4	1365.4	148.0	P
a10	1217.4	1217.4	148.0	P
a9	1069.3	1069.4	148.0	P
a8	921.3	921.4	148.0	P
a7	849.3	849.4	72.0	L
a6	701.2	701.3	148.1	P
a5	553.2	553.3	148.0	P
a4	481.1	481.3	72.0	L

Chapter 3: Semiautomated Synthesis of Sequence-defined Polymers for Information Storage

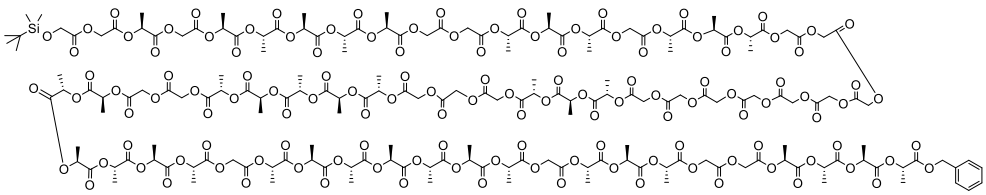
A.3.1 Characterization of PLGAs

Chain 1



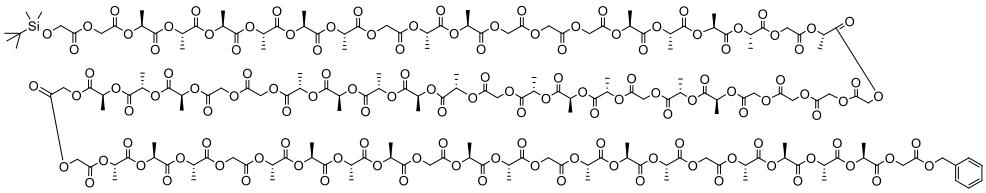
GGGLLLLLLLGGGGGGLLLLLLGLLLLLGGGLLGGGLGLLGGGLLLLLLLGLLLLLLGGGLLLG. A white solid (154 mg, 44%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.33 (d, *J* = 2.9 Hz, 5H), 5.27 – 5.07 (m, 41H), 4.92 – 4.51 (m, 48H), 4.34 (s, 2H), 1.63 – 1.50 (m, 117H), 0.89 (s, 9H), 0.09 (s, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 171.20, 169.71, 169.65, 169.58, 169.52, 169.43, 167.01, 166.63, 166.54, 166.51, 166.48, 166.45, 128.79, 128.61, 69.45, 69.33, 69.24, 69.13, 69.07, 67.44, 61.51, 61.23, 61.05, 60.94, 60.87, 60.78, 60.41, 25.85, 18.51, 16.81, 16.77, -5.36; M_n and $Đ$ (SEC): 7460 Da and 1.02; MS (MALDI-TOF): m/z calcd for $C_{180}H_{228}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4504.10; found 4504.29.

Chain 2



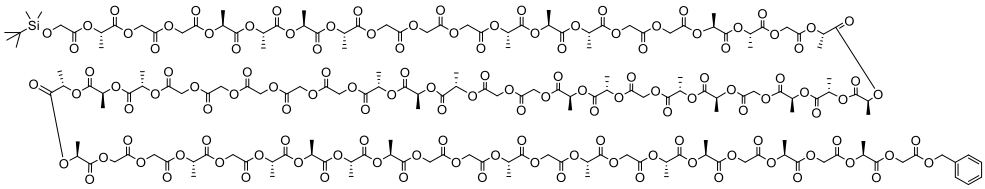
GGLGLLLLLGGLLLGLLLGGGGGGGGLLLGGGLLLLLGGLLLLLLLGLLLLLLLLGLLLGGLLLL. A white solid (145 mg, 41%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 – 7.29 (m, 5H), 5.29 – 5.09 (m, 42H), 4.91 – 4.57 (m, 46H), 4.36 (s, 2H), 1.61 – 1.55 (m, 114H), 1.52 (dd, *J* = 7.1, 3.0 Hz, 6H), 0.91 (s, *J* = 2.8 Hz, 9H), 0.11 (d, *J* = 2.9 Hz, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 171.23, 170.02, 169.73, 169.71, 169.66, 169.64, 169.59, 169.54, 167.02, 166.64, 166.58, 166.55, 166.53, 166.49, 166.46, 135.18, 128.75, 128.66, 128.38, 69.43, 69.33, 69.23, 69.18, 69.14, 69.08, 67.35, 61.51, 61.07, 61.05, 61.00, 60.95, 60.92, 60.87, 60.79, 60.52, 25.86, 18.52, 16.92, 16.89, 16.83, 16.81, 16.78, 16.74, 16.73, -5.33, -5.36; M_n and $Đ$ (SEC): 7370 Da and 1.02; MS (MALDI-TOF): m/z calcd for $C_{181}H_{230}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4518.11; found 4518.48.

Chain 3



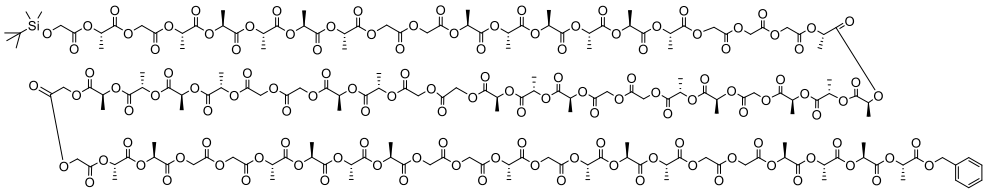
GGLLLLLLGLLGGGLLLLGLGGGGLLGLLLGLLLLLGGLLLGGLLLGL LLLGLLGLLLGLLLLG. A white solid (1157 mg, 46%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 – 7.29 (m, 5H), 5.31 – 5.10 (m, 43H), 4.93 – 4.55 (m, 44H), 4.35 (s, 2H), 1.57 (dt, J = 8.0, 2.9 Hz, 123H), 0.92 – 0.88 (m, 9H), 0.10 (s, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 171.21, 169.72, 169.70, 169.64, 169.58, 169.56, 169.53, 169.50, 169.44, 167.07, 166.99, 166.63, 166.57, 166.55, 166.51, 166.48, 166.46, 134.96, 128.79, 128.61, 77.41, 77.16, 76.91, 69.43, 69.33, 69.28, 69.23, 69.20, 69.14, 69.08, 67.44, 61.51, 61.22, 61.06, 61.00, 60.92, 60.89, 60.83, 60.79, 60.53, 25.86, 18.51, 16.84, 16.81, 16.79, 16.76, 16.75, -5.34, -5.37; M_n and $Đ$ (SEC): 7470 Da and 1.02; MS (MALDI-TOF): m/z calcd for $C_{183}H_{234}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4546.14; found 4546.61.

Chain 4



GLGGLLLLGGGLLLGGLLGLLLLGLLGLLGGLLLGGGGGLLLLGGLGL LLLGGLGLGLLGLGLG. A white solid (144 mg, 42%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41 – 7.27 (m, 5H), 5.29 – 5.13 (m, 37H), 4.91 – 4.57 (m, 56H), 4.33 (d, J = 3.1 Hz, 2H), 1.61 – 1.55 (m, 105H), 0.92 – 0.89 (m, 9H), 0.10 (s, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 171.27, 169.94, 169.70, 169.64, 169.63, 169.59, 169.56, 169.53, 169.50, 169.45, 167.00, 166.70, 166.62, 166.59, 166.58, 166.55, 166.51, 166.48, 166.46, 166.44, 134.98, 128.81, 128.61, 69.46, 69.42, 69.36, 69.30, 69.28, 69.25, 69.15, 69.09, 68.61, 67.46, 61.62, 61.28, 61.06, 61.00, 60.98, 60.93, 60.88, 60.80, 60.73, 25.86, 18.51, 16.99, 16.88, 16.84, 16.80, 16.79, 16.76, 16.75, -5.32, -5.37; M_n and $Đ$ (SEC): 7020 Da and 1.03; MS (MALDI-TOF): m/z calcd for $C_{177}H_{222}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4462.05; found 4462.59.

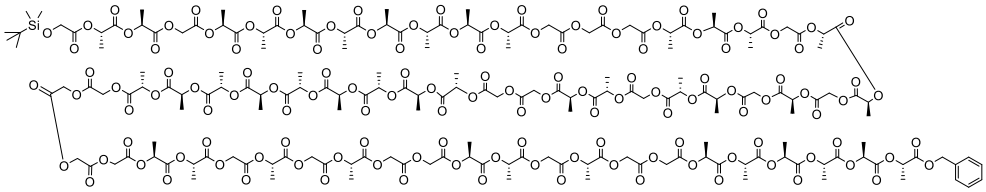
Chain 5



GLGLLLLLGGLLLLLLGGGLLLLGLLGGLLLGGLLGGLLLLGGLLGGL LLLGGLGLLLGGLLLL. A white solid (150 mg, 39%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 – 7.28 (m, 5H), 5.29 – 5.11 (m, 43H), 4.92 – 4.57 (m, 44H), 4.32 (d, J = 3.3 Hz, 2H),

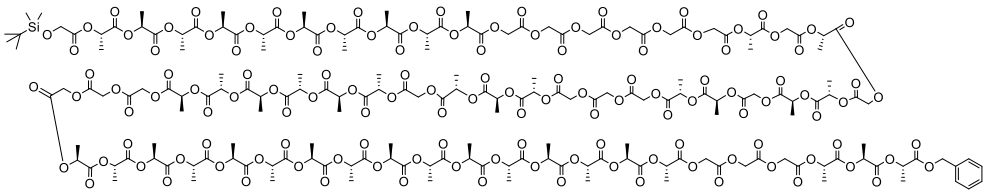
1.58 (d, J = 6.9 Hz, 117H), 1.52 (dd, J = 7.1, 2.8 Hz, 6H), 0.94 – 0.88 (m, 9H), 0.10 (s, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 171.27, 170.01, 169.71, 169.69, 169.66, 169.64, 169.58, 169.57, 169.54, 169.50, 169.45, 166.75, 166.64, 166.62, 166.58, 166.55, 166.52, 166.48, 135.21, 128.76, 128.67, 128.38, 69.44, 69.36, 69.31, 69.25, 69.15, 69.09, 68.58, 67.36, 61.62, 61.06, 61.01, 60.93, 60.86, 60.80, 25.86, 18.51, 16.99, 16.89, 16.83, 16.82, 16.77, 16.74, -5.31, -5.37; M_n and $Đ$ (SEC): 7270 Da and 1.02; MS (MALDI-TOF): m/z calcd for $C_{182}H_{232}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4532.13; found 4532.63.

Chain 6



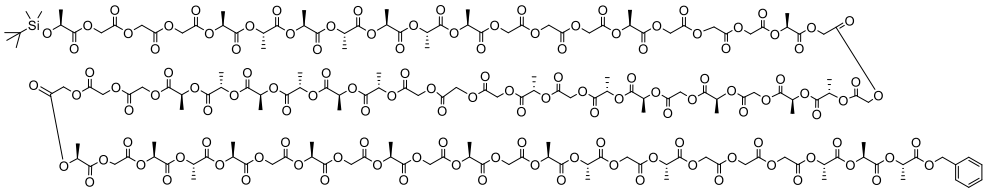
GLLGLLLLLLLLGGGLLLGLLGLGLLGLLGGLLLLLLLLLGGGGLLGLGLGGLLGLGGLLLLLL. A white solid (147 mg, 42%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.39 – 7.29 (m, 5H), 5.29 – 5.10 (m, 44H), 4.91 – 4.56 (m, 42H), 4.32 (d, J = 2.4 Hz, 2H), 1.57 (dq, J = 6.9, 2.4 Hz, 120H), 1.51 (dd, J = 7.1, 2.1 Hz, 6H), 0.92 – 0.90 (m, 9H), 0.10 (d, J = 1.1 Hz, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 171.34, 170.02, 169.71, 169.62, 169.59, 169.50, 169.44, 166.63, 166.57, 166.46, 135.21, 128.76, 128.66, 128.38, 77.41, 77.16, 76.91, 69.45, 69.42, 69.35, 69.32, 69.29, 69.25, 69.15, 69.09, 68.97, 68.61, 67.35, 61.62, 61.07, 61.02, 60.98, 60.94, 60.88, 60.85, 60.80, 25.86, 18.51, 16.89, 16.83, 16.79, 16.74, 16.71, -5.30, -5.37; M_n and $Đ$ (SEC): 7320 Da and 1.03; MS (MALDI-TOF): m/z calcd for $C_{183}H_{234}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4546.14; found 4546.61.

Chain 7



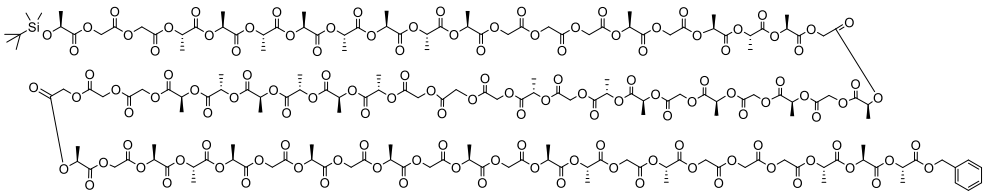
GLLLLLLLLLLGGGGGGLGLGLLGLLGGGLLLGLLLLLLGGGLLLLLLLLLLLLLLLLLGGGLLL. A white solid (142 mg, 42%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.39 – 7.29 (m, 5H), 5.29 – 5.10 (m, 46H), 4.91 – 4.57 (m, 38H), 4.32 (d, J = 2.6 Hz, 2H), 1.60 – 1.56 (m, 126H), 1.52 (dd, J = 7.1, 5.1 Hz, 6H), 0.92 – 0.89 (m, 9H), 0.10 (d, J = 1.2 Hz, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 171.33, 170.08, 170.01, 169.83, 169.72, 169.65, 169.63, 169.58, 169.53, 169.50, 169.44, 166.63, 166.56, 166.54, 166.48, 135.27, 128.76, 128.66, 128.38, 77.41, 77.16, 76.91, 69.45, 69.43, 69.39, 69.34, 69.30, 69.27, 69.24, 69.14, 69.09, 69.03, 68.60, 67.36, 61.60, 61.06, 60.98, 60.95, 60.93, 60.87, 60.79, 31.06, 25.86, 18.51, 16.92, 16.89, 16.85, 16.82, 16.78, 16.74, 16.70, -5.30, -5.38; M_n and $Đ$ (SEC): 7550 Da and 1.02; MS (MALDI-TOF): m/z calcd for $C_{185}H_{238}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4574.17; found 4574.73.

Chain 8



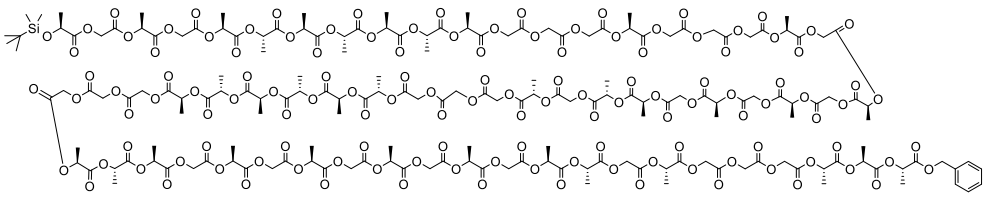
LGGGLLLLLLLGGGLGGGLGGLLGLGLLGLGGGLLLLLLGGGLGLLLG LGLGLGLLGLGGGLLL. A white solid (139 mg, 38%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 – 7.29 (m, 5H), 5.30 – 5.09 (m, 36H), 4.91 – 4.58 (m, 58H), 4.44 (q, *J* = 6.8 Hz, 1H), 1.61 – 1.55 (m, 96H), 1.52 (dd, *J* = 7.1, 4.9 Hz, 6H), 1.45 (d, *J* = 6.8 Hz, 3H), 0.90 (s, 9H), 0.09 (d, J = 10.6 Hz, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 173.51, 170.01, 169.71, 169.65, 169.63, 169.62, 169.58, 169.53, 169.50, 169.46, 169.44, 169.40, 169.39, 166.99, 166.62, 166.59, 166.58, 166.55, 166.53, 166.50, 166.49, 166.45, 135.20, 128.75, 128.66, 128.37, 69.43, 69.39, 69.37, 69.35, 69.29, 69.27, 69.23, 69.14, 69.08, 68.19, 67.35, 61.05, 60.97, 60.90, 60.87, 60.86, 60.79, 60.45, 25.82, 21.45, 18.41, 16.87, 16.84, 16.82, 16.81, 16.78, 16.77, 16.73, 16.70, -4.84, -5.23; M_n and *Đ* (SEC): 6980 Da and 1.02; MS (MALDI-TOF): m/z calcd for $C_{176}H_{220}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4448.03; found 4448.55.

Chain 9



LGGLLLLLLLLGGGLGLLLGLGLGLGLLGLGGGLLLLLLGGGLGLLLG LGLGLGLLGLGGGLLL. A white solid (144 mg, 40%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.39 – 7.29 (m, 5H), 5.28 – 5.10 (m, 39H), 4.91 – 4.59 (m, 52H), 4.43 (q, *J* = 6.8 Hz, 1H), 1.62 – 1.55 (m, 105H), 1.52 (dd, *J* = 7.1, 4.8 Hz, 6H), 1.45 (d, *J* = 6.8 Hz, 3H), 0.89 (s, 9H), 0.08 (d, *J* = 10.4 Hz, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 173.48, 170.00, 169.70, 169.68, 169.64, 169.62, 169.57, 169.56, 169.49, 169.43, 167.00, 166.64, 166.61, 166.56, 166.55, 166.54, 166.51, 166.48, 166.44, 135.19, 128.74, 128.65, 128.36, 69.44, 69.41, 69.36, 69.34, 69.28, 69.26, 69.22, 69.12, 69.07, 68.18, 67.34, 61.04, 60.96, 60.91, 60.88, 60.86, 60.83, 60.78, 60.45, 25.81, 21.44, 18.40, 16.87, 16.82, 16.81, 16.80, 16.78, 16.76, 16.72, 16.69, -4.85, -5.23; M_n and *Đ* (SEC): 7110 Da and 1.03; MS (MALDI-TOF): m/z calcd for $C_{179}H_{226}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4490.08; found 4490.54.

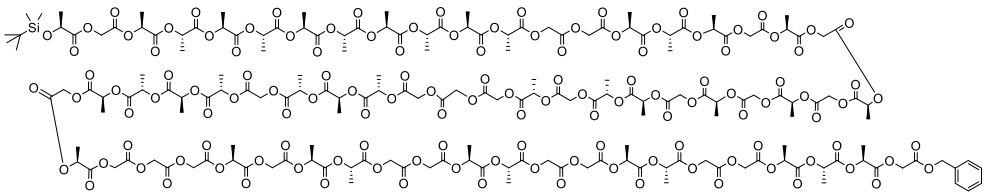
Chain 10



LGLGLLLLLLLGGGLGGGLGLGLGLGLLGLGGGLLLLLLGGGLLLGLG LGLGLGLLGLGGGLLL. A white solid (158 mg, 42%); ^{1}H NMR (500 MHz, $CDCl_3$) δ

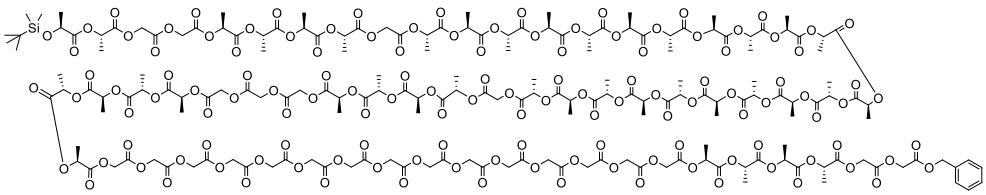
7.39 – 7.29 (m, 5H), 5.28 – 5.10 (m, 37H), 4.91 – 4.59 (m, 56H), 4.43 (q, J = 6.8 Hz, 1H), 1.61 – 1.55 (m, 99H), 1.52 (dd, J = 7.1, 4.9 Hz, 6H), 1.45 (d, J = 6.8 Hz, 3H), 0.89 (s, 9H), 0.09 (d, J = 10.3 Hz, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 173.53, 170.01, 169.72, 169.70, 169.65, 169.63, 169.59, 169.50, 169.44, 167.00, 166.66, 166.62, 166.56, 166.54, 166.52, 166.49, 166.45, 135.19, 128.75, 128.66, 128.37, 69.45, 69.42, 69.37, 69.35, 69.33, 69.29, 69.27, 69.24, 69.14, 69.13, 69.08, 68.18, 67.35, 61.05, 60.96, 60.92, 60.89, 60.87, 60.84, 60.79, 60.58, 25.83, 21.48, 18.41, 16.94, 16.88, 16.83, 16.82, 16.79, 16.77, 16.73, 16.69, -4.81, -5.19; M_n and $Đ$ (SEC): 6970 Da and 1.03; MS (MALDI-TOF): m/z calcd for $C_{177}H_{222}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4462.05; found 4461.98.

Chain 11



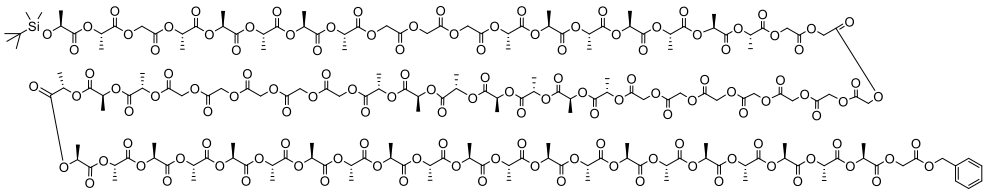
LGLLLLLLLLLLGGLLLGLGLGLGLGLLGLGGGLLLGLLLLGLGGGLG LLGGLLGGLLGGLLLG. A white solid (135 mg, 47%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.39 – 7.31 (m, 5H), 5.27 – 5.12 (m, 40H), 4.91 – 4.57 (m, 50H), 4.43 (q, J = 6.8 Hz, 1H), 1.61 – 1.54 (m, 114H), 1.45 (d, J = 6.8 Hz, 3H), 0.89 (s, 9H), 0.08 (d, J = 10.4 Hz, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 173.53, 169.76, 169.73, 169.71, 169.65, 169.64, 169.60, 169.58, 169.55, 169.54, 169.51, 169.49, 169.45, 167.06, 167.00, 166.64, 166.63, 166.58, 166.57, 166.56, 166.49, 166.46, 134.96, 128.79, 128.62, 69.43, 69.38, 69.36, 69.33, 69.29, 69.24, 69.19, 69.16, 69.13, 69.09, 68.18, 67.45, 61.23, 61.05, 61.00, 60.96, 60.92, 60.89, 60.84, 60.81, 60.59, 25.83, 21.49, 18.41, 16.88, 16.83, 16.79, 16.78, 16.77, 16.74, -4.81, -5.19; M_n and $Đ$ (SEC): 7010 Da and 1.02; MS (MALDI-TOF): m/z calcd for $C_{180}H_{228}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4504.10; found 4504.38.

Chain 12



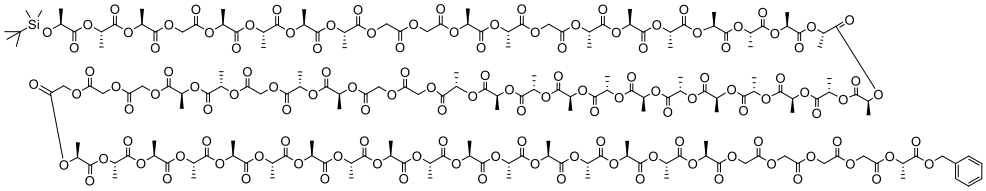
LLGGLLLLGLLLLLLLLLLLLLLLLLLLLLGLLLLGGGLLLLLGGGGGG GGGGGGGGGLLLLGG. A white solid (154 mg, 37%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.39 – 7.31 (m, 5H), 5.26 – 5.12 (m, 41H), 4.90 – 4.58 (m, 48H), 4.39 (q, J = 6.8 Hz, 1H), 1.60 – 1.54 (m, 117H), 1.44 (d, J = 6.8 Hz, 3H), 0.90 – 0.88 (m, 9H), 0.09 (d, J = 12.2 Hz, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 173.57, 169.98, 169.71, 169.65, 169.63, 169.58, 169.55, 169.52, 166.93, 166.70, 166.66, 166.63, 166.58, 166.57, 166.54, 166.52, 166.48, 166.47, 134.96, 128.80, 128.78, 128.57, 69.44, 69.40, 69.33, 69.23, 69.13, 69.08, 69.07, 68.64, 68.15, 67.48, 61.30, 61.06, 60.95, 60.87, 60.83, 60.79, 60.67, 25.81, 21.31, 18.40, 16.94, 16.82, 16.80, 16.77, -4.81, -5.19; M_n and $Đ$ (SEC): 7710 Da and 1.02; MS (MALDI-TOF): m/z calcd for $C_{181}H_{230}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4518.11; found 4518.76.

Chain 13



LLGLLLLLGGGLLLLLLLGGGGGGGGLLLLLLLGGGGGLLLLLLLLLLLLLLLLLLLLLG. A white solid (169 mg, 49%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.39 – 7.30 (m, 5H), 5.25 – 5.11 (m, 46H), 4.90 – 4.56 (m, 38H), 4.39 (q, *J* = 6.8 Hz, 1H), 1.60 – 1.55 (m, 132H), 1.44 (d, *J* = 6.8 Hz, 3H), 0.89 (s, 9H), 0.09 (d, *J* = 12.0 Hz, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 173.58, 170.06, 169.72, 169.69, 169.65, 169.63, 169.59, 169.54, 167.00, 166.77, 166.56, 166.55, 166.52, 166.49, 166.46, 166.45, 134.95, 128.79, 128.61, 69.45, 69.29, 69.25, 69.21, 69.14, 69.08, 68.62, 68.15, 67.45, 61.22, 61.07, 60.95, 60.93, 60.87, 60.79, 25.81, 21.33, 18.41, 16.96, 16.82, 16.81, 16.77, 16.73, -4.80, -5.19; M_n and $Đ$ (SEC): 7680 Da and 1.02; MS (MALDI-TOF): m/z calcd for $C_{186}H_{240}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4588.19; found 4588.41.

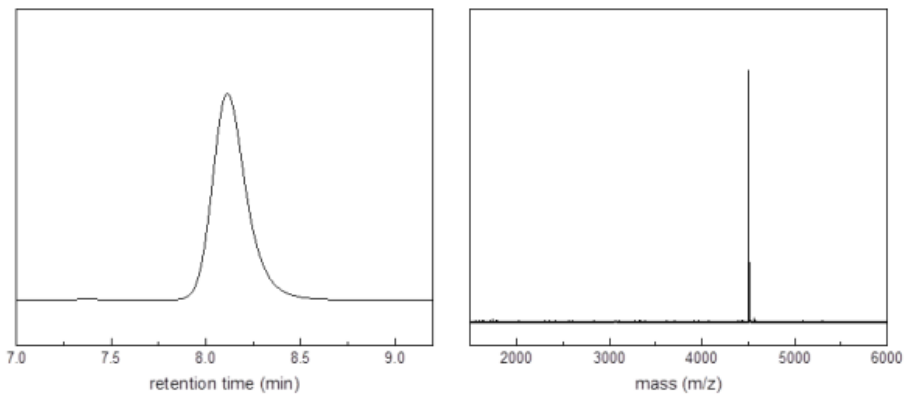
Chain 14

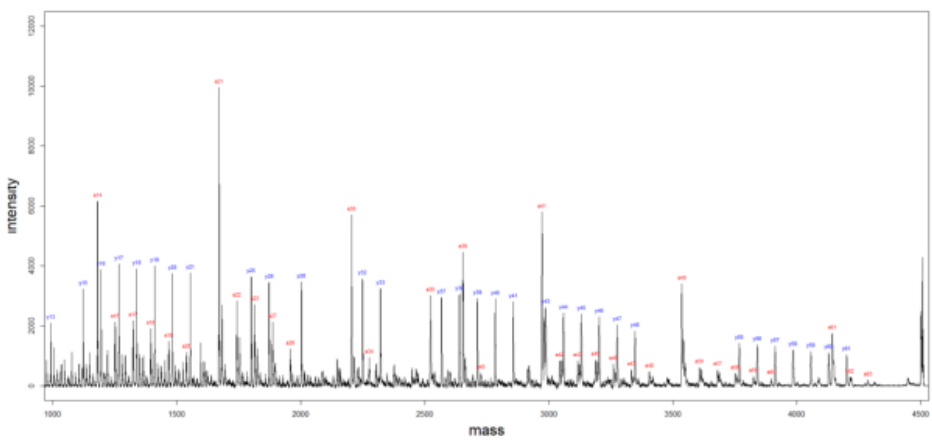


LLLGLLLLGGLLGLLLLLLLLLLLLLLLLLLLLGGLLGLLGGGLLLLLLLLLLLLLLLLLLLLGGGGL. A white solid (310 mg, 45%); ^{1}H NMR (500 MHz, cdcl3) δ 7.39 – 7.30 (m, 5H), 5.26 – 5.10 (m, 51H), 4.91 – 4.58 (m, 28H), 4.39 (q, *J* = 6.7 Hz, 1H), 1.62 – 1.54 (m, 144H), 1.52 (d, *J* = 7.1 Hz, 3H), 1.44 (d, *J* = 6.8 Hz, 3H), 0.89 (s, 9H), 0.09 (d, *J* = 11.8 Hz, 6H); ^{13}C NMR (500 MHz, $CDCl_3$) δ 173.66, 170.05, 169.85, 169.78, 169.72, 169.65, 169.61, 169.58, 169.56, 169.53, 169.49, 169.46, 166.68, 166.62, 166.58, 166.54, 166.52, 166.48, 166.47, 166.46, 135.20, 128.77, 128.65, 128.31, 77.41, 77.16, 76.91, 69.72, 69.43, 69.33, 69.30, 69.23, 69.13, 69.07, 68.88, 68.66, 68.12, 67.41, 61.10, 61.05, 60.99, 60.93, 60.87, 60.78, 25.82, 21.34, 18.41, 16.92, 16.87, 16.84, 16.81, 16.77, 16.73, -4.79, -5.18; M_n and $Đ$ (SEC): 7830 Da and 1.03; MS (MALDI-TOF): m/z calcd for $C_{191}H_{250}O_{129}Si+Na^+$ [**M**+Na]$^+$: 4658.27; found 4659.00.

A.3.2 MALDI-TOF/TOF tandem mass sequencing results of PLGAs

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 1.

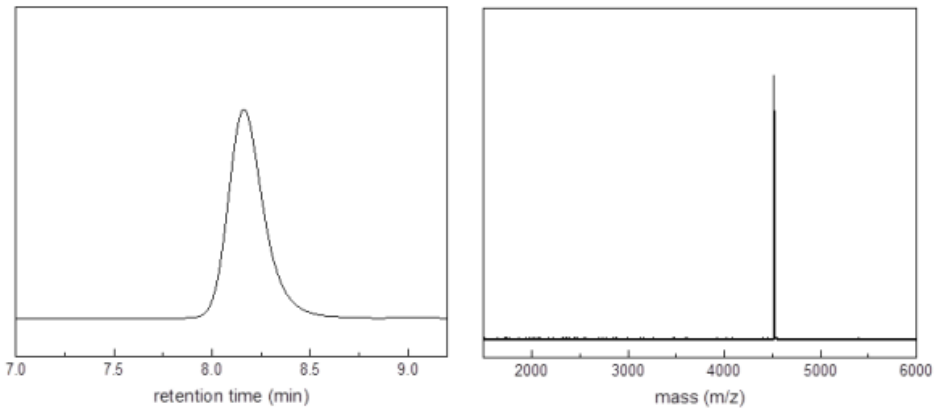


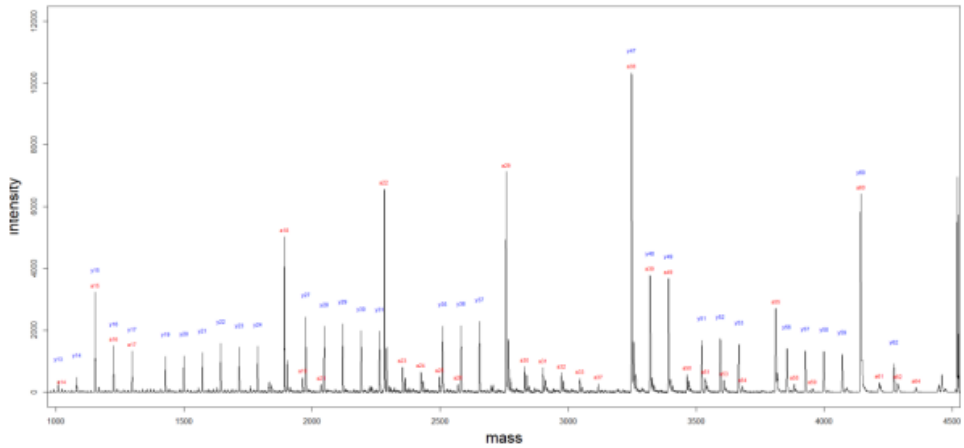


Decoding table of PLGA chain 1

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
$[M+Na]^+$	4504.1	4504.29			$[M+Na]^+$	4504.1	4504.29		
y61	4197.99	4200.07	304.22	Si-GGG	**a63**	4284.01	4287.16	217.13	Bz-GL
y60	4125.97	4128.28	71.79	L	**a62**	4211.99	4214.24	72.92	L
y59	4053.95	4056.46	71.82	L	**a61**	4139.97	4143.21	71.03	L
y58	3981.93	3984.99	71.47	L	**a60**	3893.93	3896.87	246.34	GGGL
y57	3909.9	3912.42	72.57	L	**a59**	3821.91	3824.34	72.53	L
y56	3837.88	3840.16	72.26	L	**a58**	3749.89	3751.73	72.61	L
y55	3765.86	3768.16	72	L	**a57**	3677.87	3680.81	70.92	L
y48	3345.81	3347.66	420.5	LGGGG GG	**a50**	3605.85	3607.61	73.2	L
y47	3273.79	3275.08	72.58	L	**a49**	3533.83	3536.55	71.06	L
y46	3201.77	3202.87	72.21	L	**a48**	3403.8	3405.27	131.28	GL
y45	3129.75	3131.12	71.75	L	**a47**	3331.78	3333.93	71.34	L
y44	3057.72	3059.47	71.65	L	**a46**	3259.76	3260.96	72.97	L
y43	2985.7	2986.54	72.93	L	**a45**	3187.74	3188.91	72.05	L
y41	2855.68	2856.19	130.35	LG	**a43**	3115.72	3116.87	72.04	L
y40	2783.66	2784.03	72.16	L	**a42**	3043.7	3044.42	72.45	L
y39	2711.63	2711.86	72.17	L	**a41**	2971.67	2972.43	71.99	L
y38	2639.61	2639.74	72.12	L	**a40**	2725.64	2725.88	246.55	GGGL
y37	2567.59	2567.69	72.05	L	**a39**	2653.62	2653.83	72.05	L
y33	2321.55	2321.46	246.23	LGGG	**a35**	2523.59	2523.63	130.2	GL
y32	2249.53	2249.38	72.08	L	**a34**	2277.55	2276.38	247.25	GGGL
y28	2003.5	2002.32	247.06	LGGG	**a30**	2205.53	2205.38	71	L
y26	1873.47	1872.36	129.96	LG	**a28**	1959.49	1958.26	247.12	GGGL
y25	1801.45	1800.3	72.06	L	**a27**	1887.47	1886.35	71.91	L
y21	1555.41	1554.46	245.84	LGGG	**a23**	1815.45	1814.34	72.01	L
y20	1483.39	1482.52	71.94	L	**a22**	1743.43	1742.38	71.96	L
y19	1411.37	1410.55	71.97	L	**a21**	1671.41	1670.44	71.94	L
y18	1339.35	1338.59	71.96	L	**a20**	1541.38	1540.39	130.05	GL
y17	1267.33	1266.66	71.93	L	**a19**	1469.36	1468.44	71.95	L
y16	1195.31	1194.73	71.93	L	**a18**	1397.34	1396.55	71.89	L
y15	1123.28	1122.81	71.92	L	**a17**	1325.32	1324.59	71.96	L
y13	993.26	992.93	129.88	LG	**a15**	1253.3	1252.63	71.96	L
y12	921.24	921	71.93	L	**a14**	1181.28	1180.77	71.86	L

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 2.

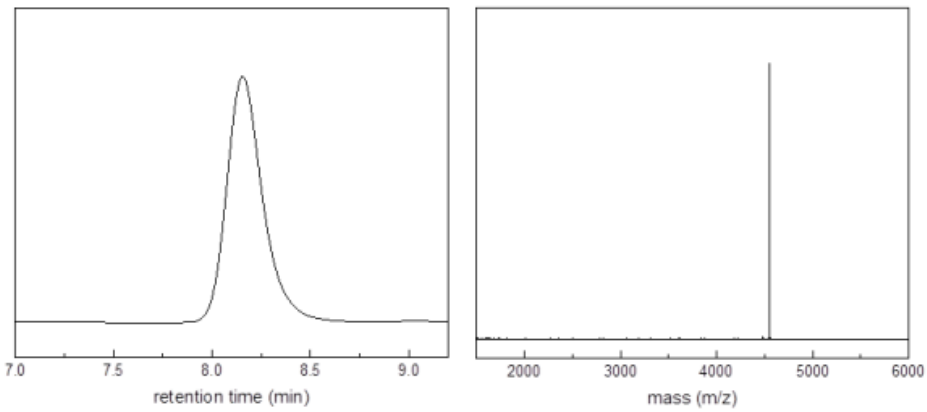


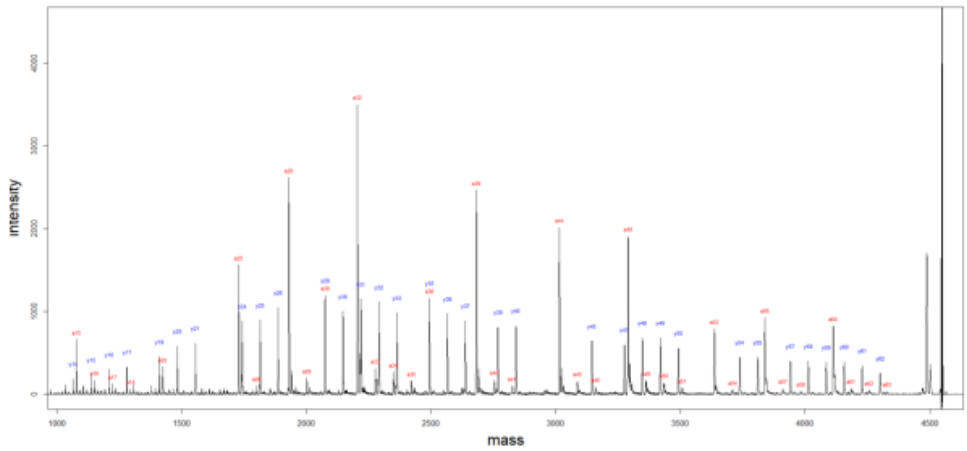


Decoding table of PLGA chain 2

Si to Bz	Calc.	Found	Difference	Sequence	Bz to Si	Calc.	Found	Difference	Sequence
[M+Na]$^+$	4518.11	4518.48			**[M+Na]$^+$**	4518.11	4518.48		
y62	4270.01	4272.52	245.96	Si-GG	**a64**	4356.03	4358.4	160.08	Bz-L
y60	4139.98	4142.6	129.92	LG	**a62**	4284.01	4286.64	71.76	L
y59	4067.96	4069.66	72.94	L	**a61**	4211.99	4214.94	71.7	L
y58	3995.94	3998.58	71.08	L	**a60**	4139.97	4142.6	72.34	L
y57	3923.92	3925.82	72.76	L	**a59**	3951.94	3953.69	188.91	GGL
y56	3851.9	3853.83	71.99	L	**a58**	3879.92	3882.89	70.8	L
y53	3663.87	3666.86	186.97	LGG	**a55**	3807.9	3810.7	72.19	L
y52	3591.85	3593.92	72.94	L	**a54**	3677.87	3681.09	129.61	GL
y51	3519.83	3521.74	72.18	L	**a53**	3605.85	3608.01	73.08	L
y49	3389.8	3391.89	129.85	LG	**a51**	3533.83	3536.9	71.11	L
y48	3317.78	3319.91	71.98	L	**a50**	3461.81	3464.76	72.14	L
y47	3245.76	3247.89	72.02	L	**a49**	3389.79	3391.89	72.87	L
y37	2651.69	2653.84	594.05	LGGGG GGGGG	**a39**	3317.76	3319.91	71.98	L
y36	2579.66	2581.88	71.96	L	**a38**	3245.74	3247.89	72.02	L
y35	2507.64	2509.82	72.06	L	**a37**	3115.72	3117.79	130.1	GL
y31	2261.61	2263.62	246.2	LGGG	**a33**	3043.7	3045.88	71.91	L
y30	2189.58	2191.68	71.94	L	**a32**	2971.67	2973.97	71.91	L
y29	2117.56	2119.56	72.12	L	**a31**	2899.65	2901.65	72.32	L
y28	2045.54	2047.5	72.06	L	**a30**	2827.63	2829.84	71.81	L
y27	1973.52	1975.52	71.98	L	**a29**	2755.61	2757.78	72.06	L
y24	1785.49	1787.55	187.97	LGG	**a26**	2567.58	2569.87	187.91	GGL
y23	1713.47	1715.59	71.96	L	**a25**	2495.56	2497.99	71.88	L
y22	1641.45	1643.59	72	L	**a24**	2423.54	2425.54	72.45	L
y21	1569.43	1571.58	72.01	L	**a23**	2351.52	2353.5	72.04	L
y20	1497.41	1499.53	72.05	L	**a22**	2279.49	2281.58	71.92	L
y19	1425.38	1427.47	72.06	L	**a21**	2033.46	2035.38	246.2	GGGL
y17	1295.36	1297.48	129.99	LG	**a19**	1961.44	1963.45	71.93	L
y16	1223.34	1225.47	72.01	L	**a18**	1889.41	1891.5	71.95	L
y15	1151.32	1153.4	72.07	L	**a17**	1295.34	1297.48	594.02	GGGGG GGGGL
y14	1079.29	1081.33	72.07	L	**a16**	1223.32	1225.47	72.01	L
y13	1007.27	1009.2	72.13	L	**a15**	1151.3	1153.4	72.07	L
y12	935.25	937.04	72.16	L	**a14**	1021.28	1023.19	130.21	GL

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 3.

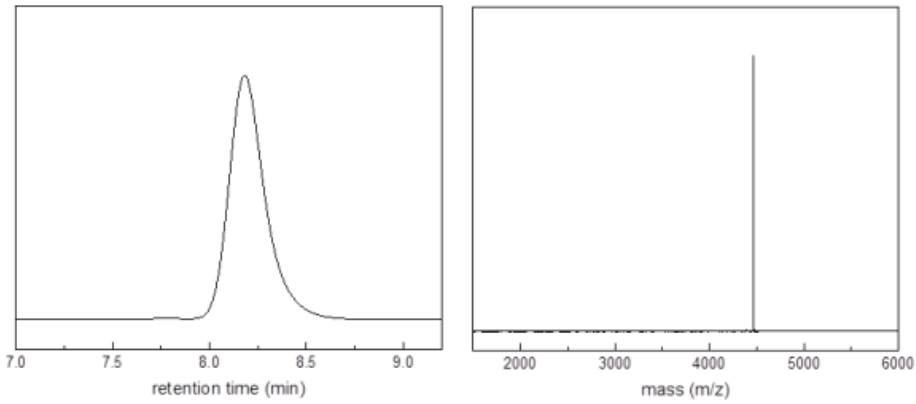


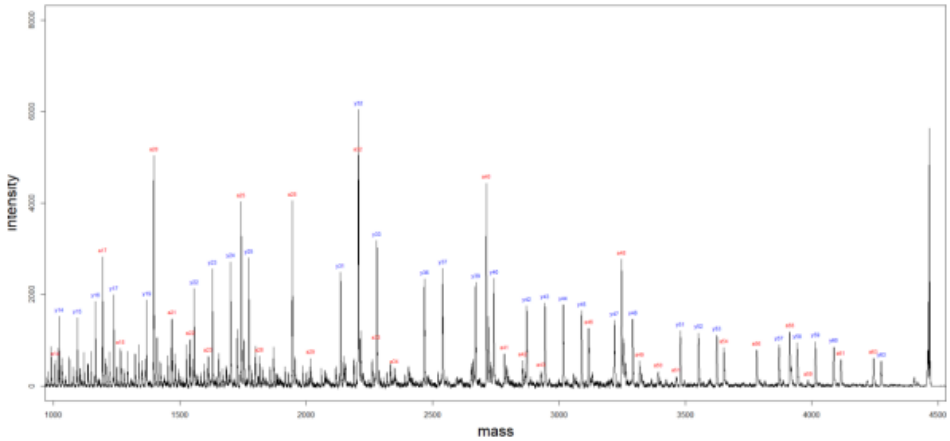


Decoding table of PLGA chain 3

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
$[M+Na]^+$	4546.14	4546.61			$[M+Na]^+$	4546.14	4546.61		
y62	4298.04	4300.94	245.67	Si-GG	**a63**	4326.06	4328.89	217.72	Bz-GL
y61	4226.02	4229.03	71.91	L	**a62**	4254.04	4256.68	72.21	L
y60	4154	4156.99	72.04	L	**a61**	4182.02	4184.41	72.27	L
y59	4081.98	4084.92	72.07	L	**a60**	4110	4113.34	71.07	L
y58	4009.96	4013.36	71.56	L	**a58**	3979.97	3983.81	129.53	GL
y57	3937.94	3941.56	71.8	L	**a57**	3907.95	3910.64	73.17	L
y55	3807.91	3810.67	130.89	LG	**a56**	3835.93	3839.16	71.48	L
y54	3735.89	3738.6	72.07	L	**a54**	3705.9	3709.49	129.67	GL
y50	3489.85	3492.59	246.01	LGGG	**a53**	3633.88	3636.92	72.57	L
y49	3417.83	3420	72.59	L	**a51**	3503.85	3508.11	128.81	GL
y48	3345.81	3347.88	72.12	L	**a50**	3431.83	3434.39	73.72	L
y47	3273.79	3275.46	72.42	L	**a49**	3359.81	3362.43	71.96	L
y45	3143.76	3145.12	130.34	LG	**a48**	3287.79	3290.16	72.27	L
y40	2839.72	2840.48	304.64	LGGGG	**a46**	3157.76	3159.81	130.35	GL
y39	2767.7	2768.41	72.07	L	**a45**	3085.74	3087.84	71.97	L
y37	2637.67	2638.16	130.25	LG	**a44**	3013.72	3014.79	73.05	L
y36	2565.65	2566.02	72.14	L	**a41**	2825.69	2826.3	188.49	GGL
y35	2493.63	2493.88	72.14	L	**a40**	2753.67	2754.36	71.94	L
y33	2363.6	2363.78	130.1	LG	**a39**	2681.65	2682.21	72.15	L
y32	2291.58	2291.66	72.12	L	**a36**	2493.61	2493.88	188.33	GGL
y31	2219.56	2218.65	73.01	L	**a35**	2421.59	2421.84	72.04	L
y30	2147.54	2146.66	71.99	L	**a34**	2349.57	2349.82	72.02	L
y29	2075.52	2075.58	71.08	L	**a33**	2277.55	2277.81	72.01	L
y26	1887.48	1886.65	188.93	LGG	**a32**	2205.53	2205.67	72.14	L
y25	1815.46	1814.66	71.99	L	**a30**	2075.5	2075.58	130.09	GL
y24	1743.44	1742.65	72.01	L	**a29**	2003.48	2002.59	72.99	L
y21	1555.41	1554.72	187.93	LGG	**a28**	1931.46	1930.65	71.94	L
y20	1483.39	1482.8	71.92	L	**a26**	1801.44	1800.51	130.14	GL
y19	1411.37	1410.84	71.96	L	**a25**	1729.41	1728.64	71.87	L
y17	1281.34	1280.84	130	LG	**a20**	1425.37	1424.92	303.72	GGGGL
y16	1209.32	1208.91	71.93	L	**a18**	1295.34	1294.79	130.13	GL
y15	1137.3	1136.96	71.95	L	**a17**	1223.32	1222.94	71.85	L
y14	1065.28	1065.02	71.94	L	**a16**	1151.3	1151.17	71.77	L
y12	935.25	935.38	129.64	LG	**a15**	1079.28	1079.23	71.94	L

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 4.

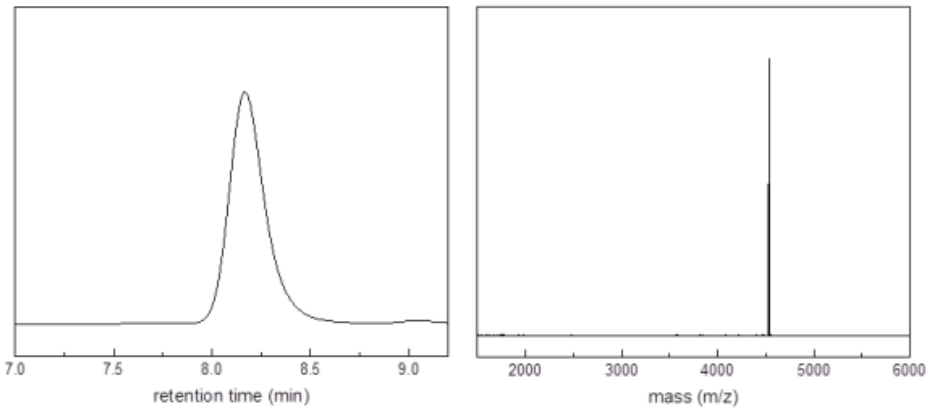


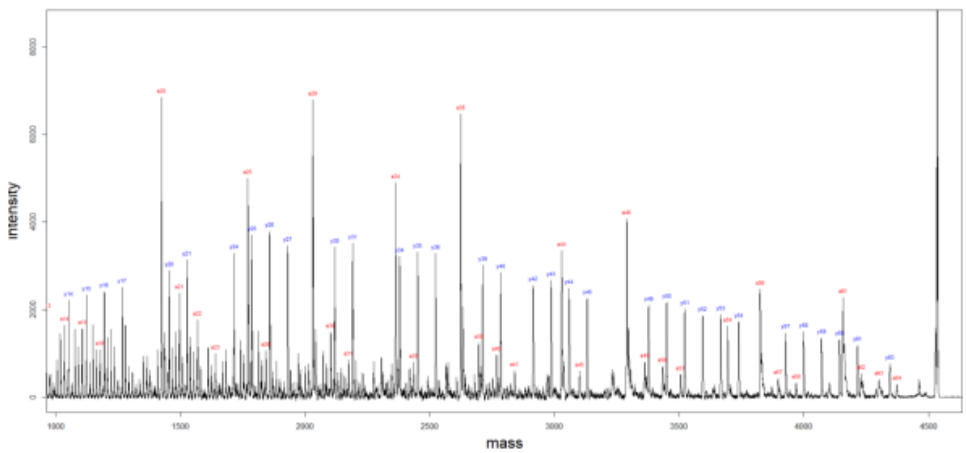


Decoding table of PLGA chain 4

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
[M+Na]$^+$	4462.05	4462.59			**[M+Na]$^+$**	4462.05	4462.59		
y63	4271.95	4274.96	187.63	Si-G	**a63**	4241.97	4244.07	218.52	Bz-GL
y60	4083.92	4087.27	187.69	LG	**a61**	4111.94	4114.9	129.17	GL
y59	4011.9	4014.33	72.94	L	**a59**	3981.91	3985.72	129.18	GL
y58	3939.88	3942.88	71.45	L	**a58**	3909.89	3913.24	72.48	L
y57	3867.86	3870.01	72.87	L	**a56**	3779.87	3782.24	131	GL
y53	3621.82	3624.03	245.98	LGGG	**a54**	3649.84	3652.97	129.27	GL
y52	3549.8	3551.82	72.21	L	**a51**	3461.81	3463.77	189.2	GGL
y51	3477.78	3480.84	70.98	L	**a50**	3389.79	3392.51	71.26	L
y48	3289.75	3291.56	189.28	LGG	**a49**	3317.76	3319.42	73.09	L
y47	3217.72	3219.22	72.34	L	**a48**	3245.74	3247.39	72.03	L
y45	3087.7	3088.97	130.25	LG	**a46**	3115.72	3117.21	130.18	GL
y44	3015.68	3016.78	72.19	L	**a43**	2927.68	2928.68	188.53	GGL
y43	2943.66	2944.61	72.17	L	**a42**	2855.66	2856.55	72.13	L
y42	2871.64	2872.42	72.19	L	**a41**	2783.64	2784.37	72.18	L
y40	2741.61	2742.32	130.1	LG	**a40**	2711.62	2712.33	72.04	L
y39	2669.59	2670.24	72.08	L	**a34**	2349.57	2349.73	362.6	GGGGGL
y37	2539.56	2540.05	130.19	LG	**a33**	2277.55	2276.91	72.82	L
y36	2467.54	2468	72.05	L	**a32**	2205.53	2204.8	72.11	L
y33	2279.51	2278.77	189.23	LGG	**a29**	2017.5	2016.63	188.17	GGL
y32	2207.49	2206.76	72.01	L	**a28**	1945.48	1944.73	71.9	L
y31	2135.47	2134.69	72.07	L	**a26**	1815.45	1814.67	130.06	GL
y25	1773.42	1772.65	362.04	LGGGGG	**a25**	1743.43	1742.74	71.93	L
y24	1701.4	1700.63	72.02	L	**a23**	1613.4	1612.66	130.08	GL
y23	1629.37	1628.7	71.93	L	**a22**	1541.38	1540.82	71.84	L
y22	1557.35	1556.68	72.02	L	**a21**	1469.36	1468.79	72.03	L
y19	1369.32	1368.86	187.82	LGG	**a20**	1397.34	1396.86	71.93	L
y17	1239.29	1238.85	130.01	LG	**a18**	1267.31	1266.87	129.99	GL
y16	1167.27	1166.94	71.91	L	**a17**	1195.29	1195	71.87	L
y15	1095.25	1094.99	71.95	L	**a14**	1007.26	1007.07	187.93	GGL
y14	1023.23	1023.02	71.97	L	**a13**	935.24	935.17	71.9	L

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 5.

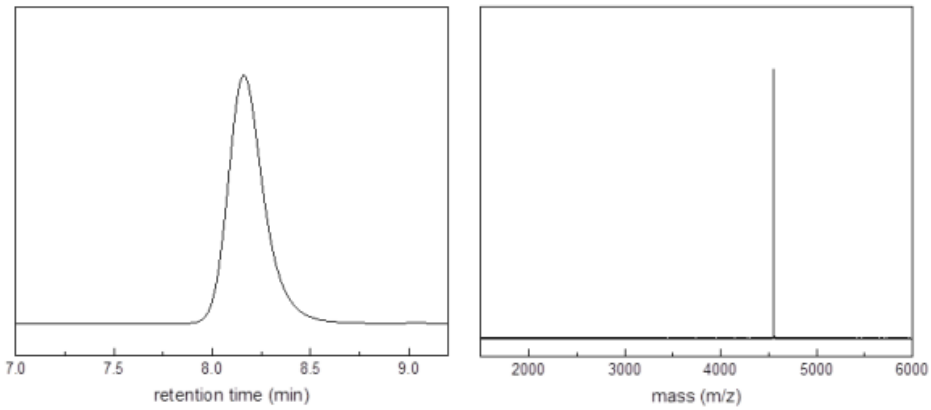


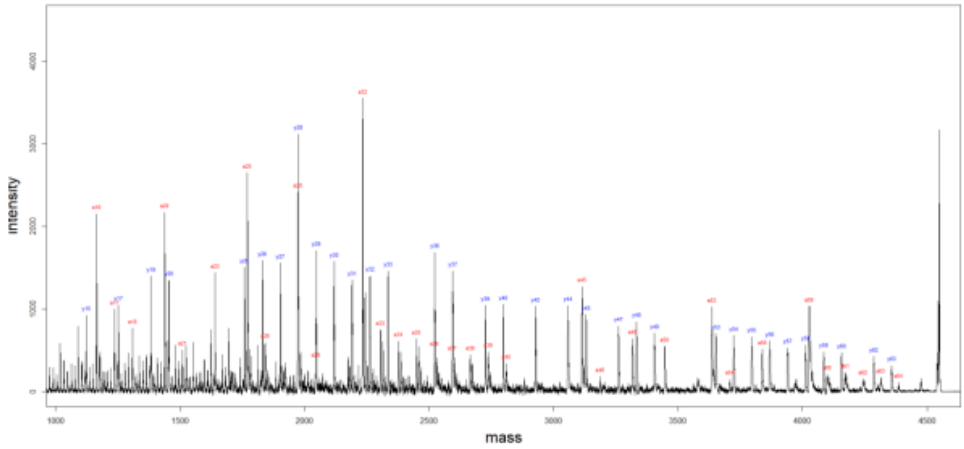


Decoding table of PLGA chain 5

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
[M+Na]$^+$	4532.13	4532.63			**[M+Na]$^+$**	4532.13	4532.63		
y63	4342.03	4345.15	187.48	Si-G	**a64**	4370.05	4373.49	159.14	Bz-L
y61	4212	4215.5	129.65	LG	**a63**	4298.03	4301.22	72.27	L
y60	4139.98	4143.4	72.1	L	**a62**	4226.01	4229.42	71.8	L
y59	4067.96	4071.45	71.95	L	**a61**	4153.99	4157.29	72.13	L
y58	3995.94	3998.8	72.65	L	**a58**	3965.95	3969.39	187.9	GGL
y57	3923.92	3927.56	71.24	L	**a57**	3893.93	3897.38	72.01	L
y54	3735.89	3738.75	188.81	LGG	**a56**	3821.91	3825.09	72.29	L
y53	3663.87	3666.79	71.96	L	**a54**	3691.89	3694.33	130.76	GL
y52	3591.85	3594.15	72.64	L	**a51**	3503.85	3506.21	188.12	GGL
y51	3519.83	3522.17	71.98	L	**a50**	3431.83	3433.7	72.51	L
y50	3447.8	3449.72	72.45	L	**a49**	3359.81	3361.97	71.73	L
y49	3375.78	3377.59	72.13	L	**a48**	3287.79	3289.47	72.5	L
y45	3129.75	3131.89	245.7	LGGG	**a45**	3099.76	3101.83	187.64	GGL
y44	3057.72	3058.8	73.09	L	**a44**	3027.74	3028.97	72.86	L
y43	2985.7	2986.71	72.09	L	**a41**	2839.7	2840.4	188.57	GGL
y42	2913.68	2914.65	72.06	L	**a40**	2767.68	2768.32	72.08	L
y40	2783.66	2784.35	130.3	LG	**a39**	2695.66	2696.25	72.07	L
y39	2711.63	2712.26	72.09	L	**a38**	2623.64	2624.22	72.03	L
y36	2523.6	2523.97	188.29	LGG	**a35**	2435.61	2434.88	189.34	GGL
y35	2451.58	2451.85	72.12	L	**a34**	2363.59	2363.82	71.06	L
y34	2379.56	2379.8	72.05	L	**a31**	2175.56	2174.7	189.12	GGL
y31	2191.53	2190.67	189.13	LGG	**a30**	2103.54	2102.67	72.03	L
y30	2119.51	2118.6	72.07	L	**a29**	2031.51	2030.67	72	L
y27	1931.47	1930.58	188.02	LGG	**a26**	1843.48	1842.6	188.07	GGL
y26	1859.45	1858.6	71.98	L	**a25**	1771.46	1770.61	71.99	L
y25	1787.43	1786.56	72.04	L	**a23**	1641.43	1640.66	129.95	GL
y24	1715.41	1714.6	71.96	L	**a22**	1569.41	1568.67	71.99	L
y21	1527.38	1526.61	187.99	LGG	**a21**	1497.39	1496.7	71.97	L
y20	1455.36	1454.72	71.89	L	**a20**	1425.37	1424.83	71.87	L
y17	1267.33	1266.87	187.85	LGG	**a16**	1179.33	1178.95	245.88	GGGL
y16	1195.31	1194.9	71.97	L	**a15**	1107.31	1107.08	71.87	L
y15	1123.28	1122.99	71.91	L	**a14**	1035.29	1035.1	71.98	L
y14	1051.26	1051.06	71.93	L	**a13**	963.27	963.19	71.91	L

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 6.

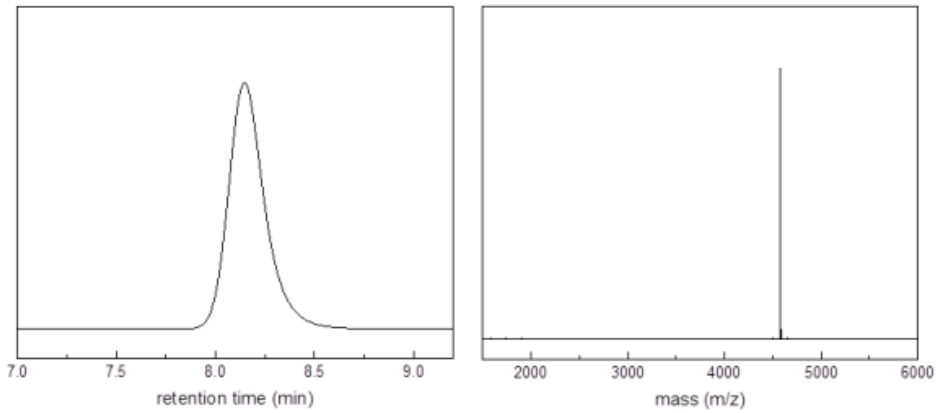


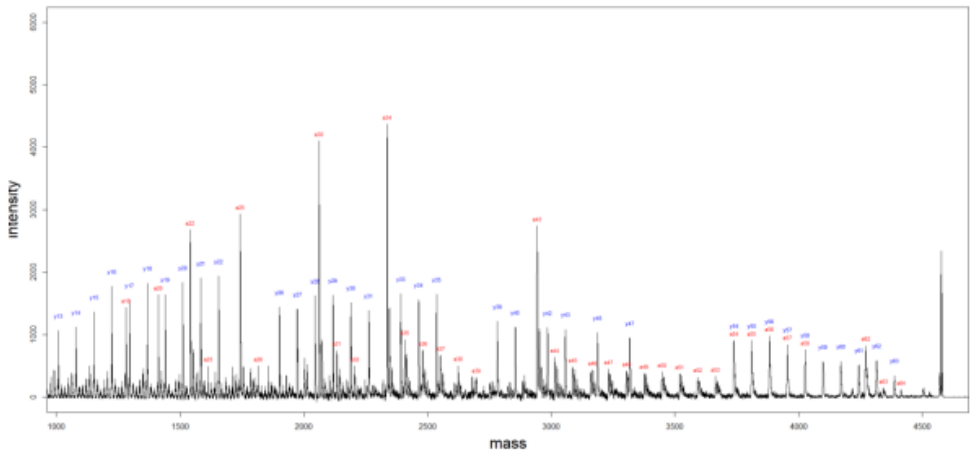


Decoding table of PLGA chain 6

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
[M+Na]$^+$	4546.14	4546.61			**[M+Na]$^+$**	4546.14	4546.61		
y63	4356.05	4358.85	187.76	Si-G	**a64**	4384.07	4385.62	160.99	Bz-L
y62	4284.03	4286.68	72.17	L	**a63**	4312.04	4314.5	71.12	L
y60	4154	4156.85	129.83	LG	**a62**	4240.02	4243.37	71.13	L
y59	4081.98	4085.39	71.46	L	**a61**	4168	4171.21	72.16	L
y58	4009.96	4012.7	72.69	L	**a60**	4095.98	4099.37	71.84	L
y57	3937.94	3941.25	71.45	L	**a59**	4023.96	4026.83	72.54	L
y56	3865.92	3868.56	72.69	L	**a56**	3835.93	3838.44	188.39	GGL
y55	3793.89	3796.2	72.36	L	**a54**	3705.9	3708.02	130.42	GL
y54	3721.87	3723.78	72.42	L	**a53**	3633.88	3636.04	71.98	L
y53	3649.85	3651.89	71.89	L	**a50**	3445.85	3447.72	188.32	GGL
y49	3403.81	3406.14	245.75	LGGG	**a48**	3315.82	3317.25	130.47	GL
y48	3331.79	3333.39	72.75	L	**a46**	3185.79	3186.49	130.76	GL
y47	3259.77	3261.92	71.47	L	**a45**	3113.77	3114.88	71.61	L
y45	3129.75	3130.64	131.28	LG	**a40**	2809.73	2810.03	304.85	GGGGL
y44	3057.72	3058.72	71.92	L	**a39**	2737.71	2738.1	71.93	L
y42	2927.7	2928.19	130.53	LG	**a38**	2665.69	2665.9	72.2	L
y40	2797.67	2798.24	129.95	LG	**a37**	2593.67	2592.94	72.96	L
y39	2725.65	2725.97	72.27	L	**a36**	2521.65	2520.73	72.21	L
y37	2595.62	2594.87	131.1	LG	**a35**	2449.63	2448.74	71.99	L
y36	2523.6	2522.77	72.1	L	**a34**	2377.6	2377.6	71.14	L
y33	2335.57	2335.54	187.23	LGG	**a33**	2305.58	2305.65	71.95	L
y32	2263.55	2263.57	71.97	L	**a32**	2233.56	2233.61	72.04	L
y31	2191.53	2191.49	72.08	L	**a29**	2045.53	2044.51	189.1	GGL
y30	2119.51	2118.49	73	L	**a28**	1973.51	1972.48	72.03	L
y29	2047.49	2046.39	72.1	L	**a26**	1843.48	1842.47	130.01	GL
y28	1975.46	1974.38	72.01	L	**a25**	1771.46	1770.49	71.98	L
y27	1903.44	1902.41	71.97	L	**a23**	1641.43	1640.55	129.94	GL
y26	1831.42	1830.43	71.98	L	**a21**	1511.41	1510.57	129.98	GL
y25	1759.4	1758.43	72	L	**a20**	1439.39	1438.64	71.93	L
y20	1455.36	1454.59	303.84	LGGGG	**a18**	1309.36	1308.73	129.91	GL
y19	1383.34	1382.63	71.96	L	**a17**	1237.34	1236.85	71.88	L
y17	1253.31	1252.78	129.85	LG	**a16**	1165.32	1164.91	71.94	L
y15	1123.28	1122.86	129.92	LG	**a12**	919.28	919.12	245.79	GGGL

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 7.

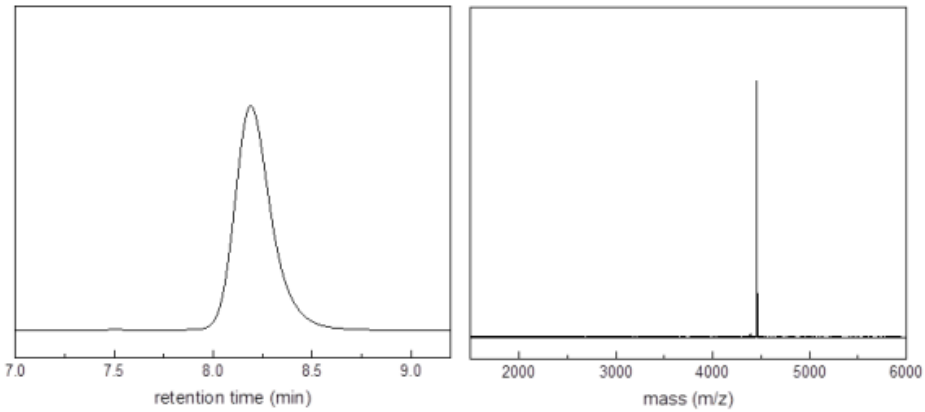


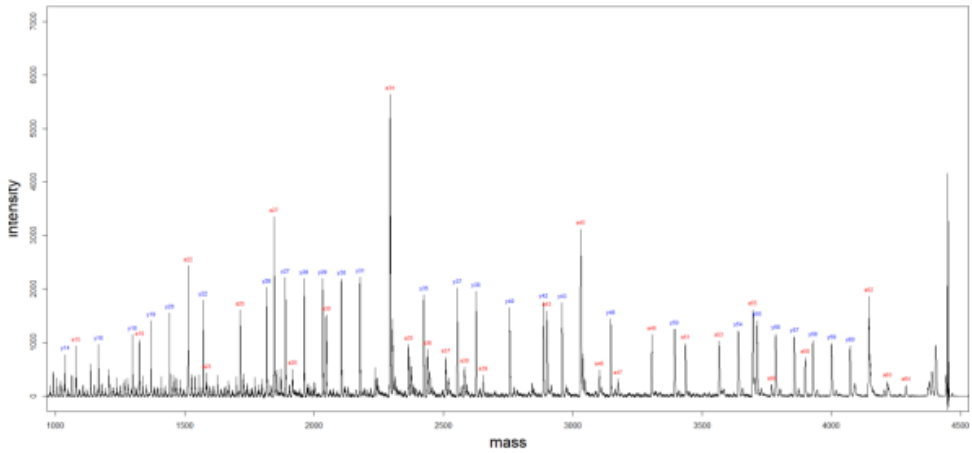


Decoding table of PLGA chain 7

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
[M+Na]$^+$	4574.17	4574.73			**[M+Na]$^+$**	4574.17	4574.73		
y63	4384.08	4386.86	187.87	Si-G	**a64**	4412.1	4413.38	161.35	Bz-L
y62	4312.06	4314.85	72.01	L	**a63**	4340.08	4342.06	71.32	L
y61	4240.04	4242.96	71.89	L	**a62**	4268.05	4271.02	71.04	L
y60	4168.02	4170.93	72.03	L	**a58**	4022.02	4025.89	245.13	GGGL
y59	4095.99	4098.01	72.92	L	**a57**	3950	3954.22	71.67	L
y58	4023.97	4025.89	72.12	L	**a56**	3877.97	3882.09	72.13	L
y57	3951.95	3954.22	71.67	L	**a55**	3805.95	3809.53	72.56	L
y56	3879.93	3882.09	72.13	L	**a54**	3733.93	3737.91	71.62	L
y55	3807.91	3809.53	72.56	L	**a53**	3661.91	3663.58	74.33	L
y54	3735.89	3737.91	71.62	L	**a52**	3589.89	3592.81	70.77	L
y47	3315.83	3317.77	420.14	LGGGG GG	**a51**	3517.87	3519.82	72.99	L
y45	3185.81	3186.85	130.92	LG	**a50**	3445.85	3447.21	72.61	L
y43	3055.78	3056.67	130.18	LG	**a49**	3373.83	3375.15	72.06	L
y42	2983.76	2984.59	72.08	L	**a48**	3301.81	3304.12	71.03	L
y40	2853.73	2854.26	130.33	LG	**a47**	3229.78	3231.32	72.8	L
y39	2781.71	2782.34	71.92	L	**a46**	3157.76	3158.84	72.48	L
y35	2535.67	2536.07	246.27	LGGG	**a45**	3085.74	3086.77	72.07	L
y34	2463.65	2464.02	72.05	L	**a44**	3013.72	3014.54	72.23	L
y33	2391.63	2391.95	72.07	L	**a43**	2941.7	2942.62	71.92	L
y31	2261.61	2261.87	130.08	LG	**a39**	2695.66	2696.67	245.95	GGGL
y30	2189.59	2189.78	72.09	L	**a38**	2623.64	2623.99	72.68	L
y29	2117.56	2116.84	72.94	L	**a37**	2551.62	2552.11	71.88	L
y28	2045.54	2045.76	71.08	L	**a36**	2479.6	2479.94	72.17	L
y27	1973.52	1972.83	72.93	L	**a35**	2407.58	2408.02	71.92	L
y26	1901.5	1900.78	72.05	L	**a34**	2335.56	2335.93	72.09	L
y22	1655.46	1654.82	245.96	LGGG	**a32**	2205.53	2204.86	131.07	GL
y21	1583.44	1582.77	72.05	L	**a31**	2133.51	2132.76	72.1	L
y20	1511.42	1510.87	71.9	L	**a30**	2061.49	2060.79	71.97	L
y19	1439.4	1438.85	72.02	L	**a26**	1815.45	1814.7	246.09	GGGL
y18	1367.38	1366.97	71.88	L	**a25**	1743.43	1742.76	71.94	L
y17	1295.36	1294.96	72.01	L	**a23**	1613.4	1612.77	129.99	GL
y16	1223.34	1223.08	71.88	L	**a22**	1541.38	1540.83	71.94	L
y15	1151.32	1151.1	71.98	L	**a20**	1411.36	1410.97	129.86	GL
y14	1079.29	1079.13	71.97	L	**a18**	1281.33	1281.02	129.95	GL

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 8.

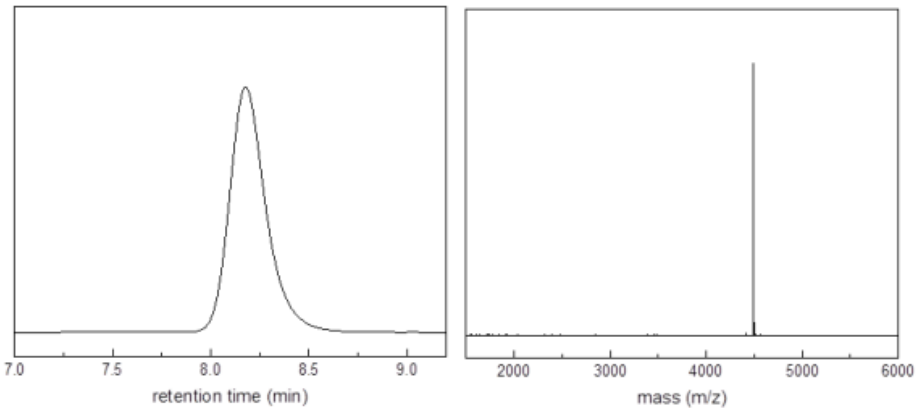


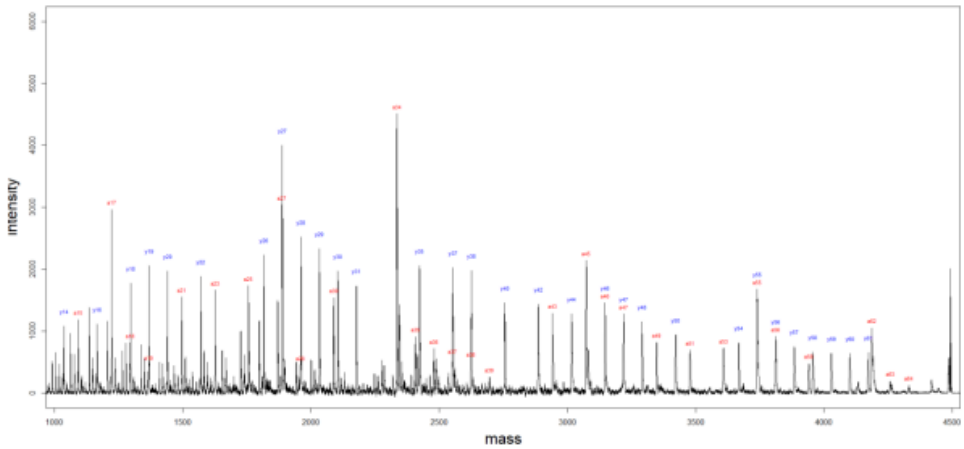


Decoding table of PLGA chain 8

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
$[M+Na]^+$	4448.03	4448.55			$[M+Na]^+$	4448.03	4448.55		
y60	4069.91	4073.25	375.3	Si-LGGG	**a64**	4285.96	4288.81	159.74	Bz-L
y59	3997.88	4001.21	72.04	L	**a63**	4213.94	4216.94	71.87	L
y58	3925.86	3929.19	72.02	L	**a62**	4141.91	4145.11	71.83	L
y57	3853.84	3857.3	71.89	L	**a58**	3895.88	3898.32	246.79	GGGL
y56	3781.82	3784.38	72.92	L	**a56**	3765.85	3769.09	129.23	GL
y55	3709.8	3712.74	71.64	L	**a55**	3693.83	3696.34	72.75	L
y54	3637.78	3640.04	72.7	L	**a53**	3563.8	3565.96	130.38	GL
y50	3391.74	3393.81	246.23	LGGG	**a51**	3433.78	3435.72	130.24	GL
y46	3145.7	3147.15	246.66	LGGG	**a49**	3303.75	3305.54	130.18	GL
y43	2957.67	2958.75	188.4	LGG	**a47**	3173.72	3175.52	130.02	GL
y42	2885.65	2886.58	72.17	L	**a46**	3101.7	3102.94	72.58	L
y40	2755.62	2756.35	130.23	LG	**a45**	3029.68	3030.88	72.06	L
y38	2625.6	2626.2	130.15	LG	**a43**	2899.65	2900.78	130.1	GL
y37	2553.58	2554.06	72.14	L	**a39**	2653.62	2653.37	247.41	GGGL
y35	2423.55	2423.9	130.16	LG	**a38**	2581.59	2582	71.37	L
y31	2177.51	2177.72	246.18	LGGG	**a37**	2509.57	2509.97	72.03	L
y30	2105.49	2105.72	72	L	**a36**	2437.55	2437.97	72	L
y29	2033.47	2032.65	73.07	L	**a35**	2365.53	2365.9	72.07	L
y28	1961.45	1960.67	71.98	L	**a34**	2293.51	2293.84	72.06	L
y27	1889.43	1888.63	72.04	L	**a30**	2047.47	2046.67	247.17	GGGL
y26	1817.41	1816.65	71.98	L	**a28**	1917.45	1916.74	129.93	GL
y22	1571.37	1570.7	245.95	LGGG	**a27**	1845.43	1844.65	72.09	L
y20	1441.34	1440.73	129.97	LG	**a25**	1715.4	1714.67	129.98	GL
y19	1369.32	1368.8	71.93	L	**a23**	1585.37	1584.71	129.96	GL
y18	1297.3	1296.92	71.88	L	**a22**	1513.35	1512.82	71.89	L
y16	1167.27	1166.91	130.01	LG	**a19**	1325.32	1324.81	188.01	GGL
y14	1037.25	1037.15	129.76	LG	**a15**	1079.28	1079.07	245.74	GGGL

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 9.

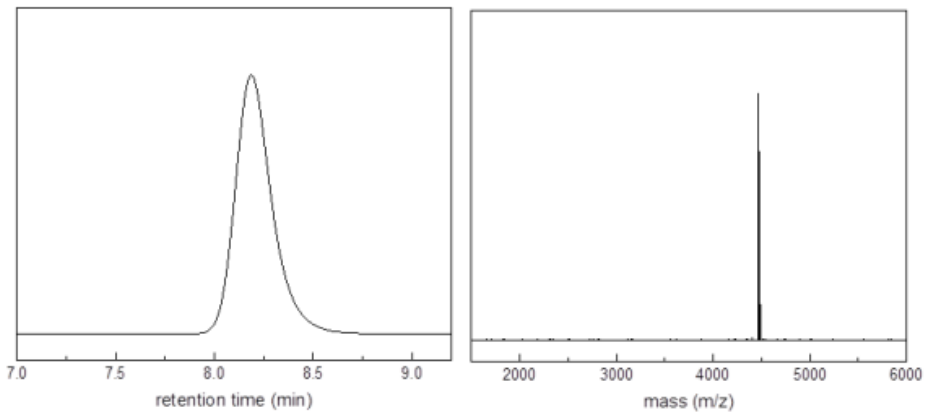


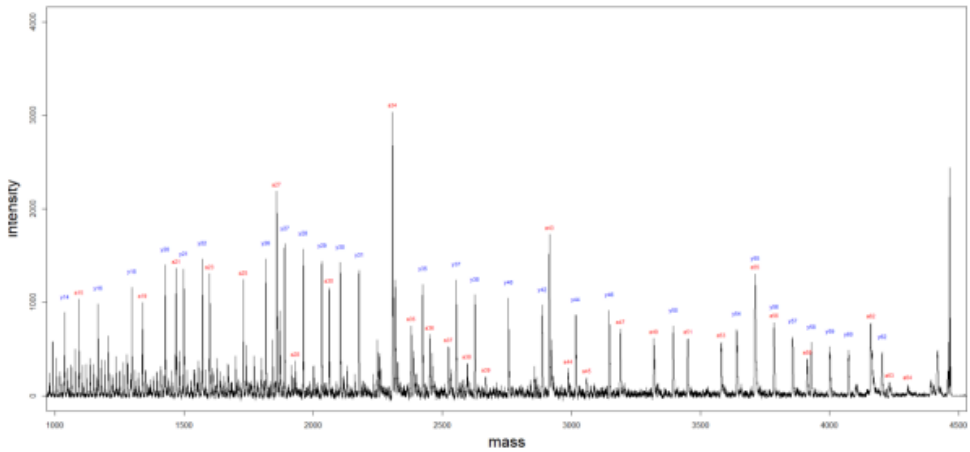


Decoding table of PLGA chain 9

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
$[M+Na]^+$	4490.08	4490.54			$[M+Na]^+$	4490.08	4490.54		
y61	4169.96	4173.22	317.32	Si-LGG	**a64**	4328	4330.65	159.89	Bz-L
y60	4097.94	4101.12	72.1	L	**a63**	4255.98	4258	72.65	L
y59	4025.92	4029.01	72.11	L	**a62**	4183.96	4187.16	70.84	L
y58	3953.89	3956.14	72.87	L	**a58**	3937.92	3940.43	246.73	GGGL
y57	3881.87	3883.83	72.31	L	**a56**	3807.9	3812.1	128.33	GL
y56	3809.85	3812.1	71.73	L	**a55**	3735.88	3740.02	72.08	L
y55	3737.83	3740.02	72.08	L	**a53**	3605.85	3608.03	131.99	GL
y54	3665.81	3668.38	71.64	L	**a51**	3475.82	3478.05	129.98	GL
y50	3419.77	3422.2	246.18	LGGG	**a49**	3345.8	3348.3	129.75	GL
y48	3289.75	3290.96	131.24	LG	**a47**	3215.77	3218.87	129.43	GL
y47	3217.73	3218.87	72.09	L	**a46**	3143.75	3146.86	72.01	L
y46	3145.7	3146.86	72.01	L	**a45**	3071.73	3072.82	74.04	L
y44	3015.68	3016.53	130.33	LG	**a43**	2941.7	2942.52	130.3	GL
y42	2885.65	2886.19	130.34	LG	**a39**	2695.66	2695.05	247.47	GGGL
y40	2755.62	2756.06	130.13	LG	**a38**	2623.64	2623.02	72.03	L
y38	2625.6	2625.87	130.19	LG	**a37**	2551.62	2550.81	72.21	L
y37	2553.58	2552.79	73.08	L	**a36**	2479.6	2479.81	71	L
y35	2423.55	2423.7	129.09	LG	**a35**	2407.58	2407.63	72.18	L
y31	2177.51	2176.52	247.18	LGGG	**a34**	2335.56	2335.68	71.95	L
y30	2105.49	2104.5	72.02	L	**a30**	2089.52	2088.5	247.18	GGGL
y29	2033.47	2032.45	72.05	L	**a28**	1959.49	1958.36	130.14	GL
y28	1961.45	1960.44	72.01	L	**a27**	1887.47	1886.5	71.86	L
y27	1889.43	1888.42	72.02	L	**a25**	1757.45	1756.47	130.03	GL
y26	1817.41	1816.42	72	L	**a23**	1627.42	1626.51	129.96	GL
y22	1571.37	1570.56	245.86	LGGG	**a21**	1497.39	1496.61	129.9	GL
y20	1441.34	1440.59	129.97	LG	**a19**	1367.37	1366.62	129.99	GL
y19	1369.32	1368.68	71.91	L	**a18**	1295.34	1294.87	71.75	L
y18	1297.3	1296.73	71.95	L	**a17**	1223.32	1222.86	72.01	L
y16	1167.27	1166.82	129.91	LG	**a15**	1093.3	1092.97	129.89	GL
y14	1037.25	1036.93	129.89	LG	**a11**	847.26	847.09	245.88	GGGL

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 10.

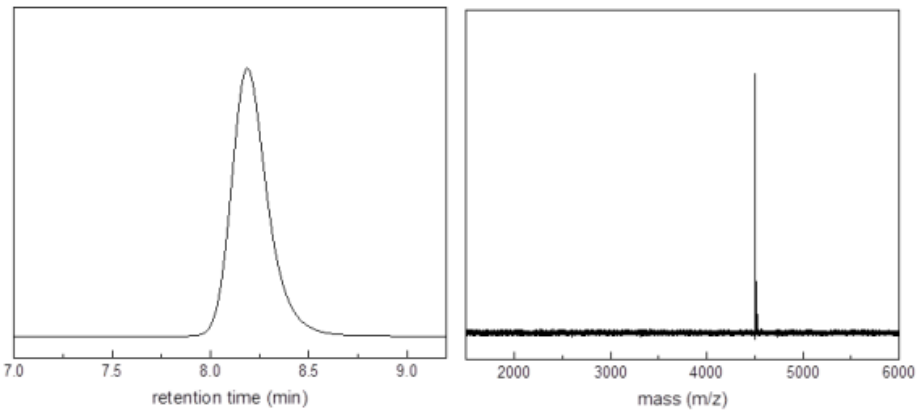


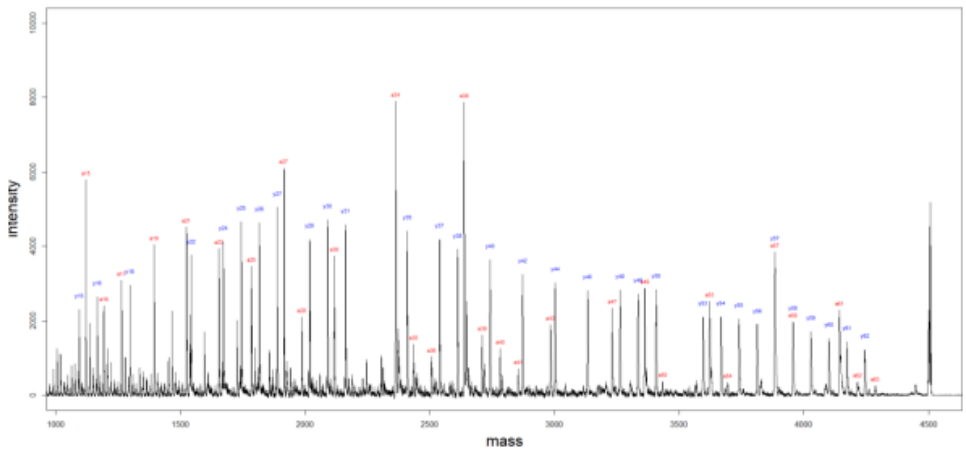


Decoding table of PLGA chain 10

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
[M+Na]$^+$	4462.05	4461.98			**[M+Na]$^+$**	4462.05	4461.98		
y62	4199.93	4203.69	258.29	Si-LG	**a64**	4299.97	4303.05	158.93	Bz-L
y60	4069.91	4072.75	130.94	LG	**a63**	4227.95	4232.48	70.57	L
y59	3997.88	4001.14	71.61	L	**a62**	4155.93	4159.56	72.92	L
y58	3925.86	3929.29	71.85	L	**a58**	3909.89	3913.03	246.53	GGGL
y57	3853.84	3856.3	72.99	L	**a56**	3779.87	3784.48	128.55	GL
y56	3781.82	3784.48	71.82	L	**a55**	3707.84	3711.58	72.9	L
y55	3709.8	3711.58	72.9	L	**a53**	3577.82	3581.17	130.41	GL
y54	3637.78	3640.06	71.52	L	**a51**	3447.79	3450.96	130.21	GL
y50	3391.74	3394.3	245.76	LGGG	**a49**	3317.76	3319.46	131.5	GL
y46	3145.7	3147.24	247.06	LGGG	**a47**	3187.74	3189.38	130.08	GL
y44	3015.68	3016.74	130.5	LG	**a45**	3057.71	3059.09	130.29	GL
y42	2885.65	2886.61	130.13	LG	**a44**	2985.69	2987.05	72.04	L
y40	2755.62	2756.53	130.08	LG	**a43**	2913.67	2914.74	72.31	L
y38	2625.6	2626.31	130.22	LG	**a39**	2667.63	2668.18	246.56	GGGL
y37	2553.58	2554.17	72.14	L	**a38**	2595.61	2596.04	72.14	L
y35	2423.55	2423.96	130.21	LG	**a37**	2523.59	2524.08	71.96	L
y31	2177.51	2177.85	246.11	LGGG	**a36**	2451.57	2451.13	72.95	L
y30	2105.49	2104.75	73.1	L	**a35**	2379.55	2378.93	72.2	L
y29	2033.47	2033.71	71.04	L	**a34**	2307.53	2307.95	70.98	L
y28	1961.45	1960.77	72.94	L	**a30**	2061.49	2061.8	246.15	GGGL
y27	1889.43	1889.77	71	L	**a28**	1931.46	1930.77	131.03	GL
y26	1817.41	1816.7	73.07	L	**a27**	1859.44	1858.75	72.02	L
y22	1571.37	1570.77	245.93	LGGG	**a25**	1729.41	1728.75	130	GL
y21	1499.35	1498.81	71.96	L	**a23**	1599.39	1598.81	129.94	GL
y20	1427.33	1426.89	71.92	L	**a21**	1469.36	1468.92	129.89	GL
y18	1297.3	1296.91	129.98	LG	**a19**	1339.33	1338.86	130.06	GL
y16	1167.27	1167.1	129.81	LG	**a15**	1093.3	1093.14	245.72	GGGL
y14	1037.25	1037.17	129.93	LG	**a11**	847.26	847.46	245.68	GGGL

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 11.

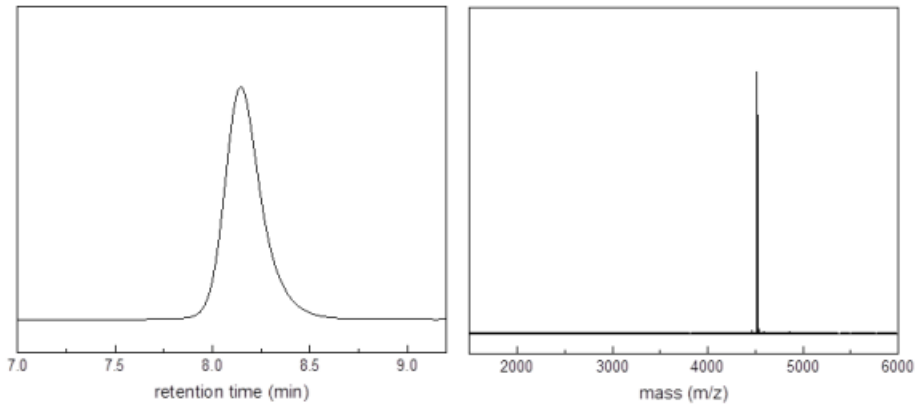


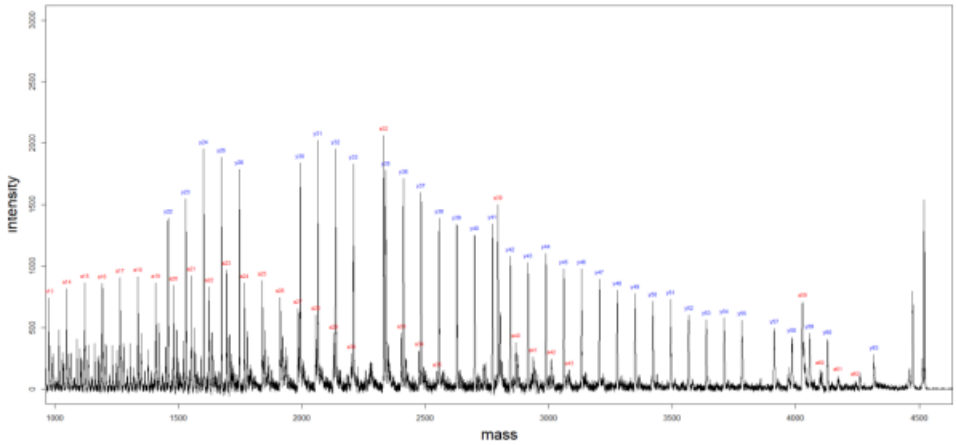


Decoding table of PLGA chain 11

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
$[M+Na]^+$	4504.1	4504.38			$[M+Na]^+$	4504.1	4504.38		
y62	4241.98	4244.33	260.05	Si-L	**a63**	4284.01	4286.92	217.46	Bz-GL
y61	4169.96	4172.36	71.97	L	**a62**	4211.99	4215.32	71.6	L
y60	4097.94	4100.88	71.48	L	**a61**	4139.97	4143.4	71.92	L
y59	4025.92	4028.42	72.46	L	**a58**	3951.94	3957.19	186.21	GGL
y58	3953.89	3957.19	71.23	L	**a57**	3879.92	3883.59	73.6	L
y57	3881.87	3883.59	73.6	L	**a54**	3691.89	3694.94	188.65	GGL
y56	3809.85	3812.87	70.72	L	**a53**	3619.86	3622.83	72.11	L
y55	3737.83	3740.28	72.59	L	**a50**	3431.83	3434.84	187.99	GGL
y54	3665.81	3668.31	71.97	L	**a49**	3359.81	3362.07	72.77	L
y53	3593.79	3595.98	72.33	L	**a47**	3229.78	3231.44	130.63	GL
y50	3405.76	3407.76	188.22	LGG	**a43**	2983.75	2984.77	246.67	GGGL
y49	3333.74	3335.6	72.16	L	**a41**	2853.72	2854.75	130.02	GL
y48	3261.72	3263.39	72.21	L	**a40**	2781.7	2782.47	72.28	L
y46	3131.69	3133.12	130.27	LG	**a39**	2709.68	2710.25	72.22	L
y44	3001.66	3002.88	130.24	LG	**a38**	2637.66	2638.22	72.03	L
y42	2871.64	2872.5	130.38	LG	**a36**	2507.63	2506.95	131.27	GL
y40	2741.61	2742.28	130.22	LG	**a35**	2435.61	2435.88	71.07	L
y38	2611.58	2612.11	130.17	LG	**a34**	2363.59	2363.87	72.01	L
y37	2539.56	2539.92	72.19	L	**a30**	2117.55	2116.69	247.18	GGGL
y35	2409.53	2409.8	130.12	LG	**a28**	1987.52	1986.63	130.06	GL
y31	2163.5	2162.68	247.12	LGGG	**a27**	1915.5	1914.68	71.95	L
y30	2091.48	2090.6	72.08	L	**a25**	1785.48	1784.63	130.05	GL
y29	2019.45	2018.58	72.02	L	**a23**	1655.45	1654.69	129.94	GL
y27	1889.43	1888.63	129.95	LG	**a21**	1525.42	1524.73	129.96	GL
y26	1817.41	1816.57	72.06	L	**a19**	1395.4	1394.82	129.91	GL
y25	1745.39	1744.62	71.95	L	**a17**	1265.37	1264.88	129.94	GL
y24	1673.36	1672.58	72.04	L	**a16**	1193.35	1193.05	71.83	L
y22	1543.34	1542.64	129.94	LG	**a15**	1121.33	1121.06	71.99	L
y18	1297.3	1296.91	245.73	LGGG	**a12**	933.3	933.21	187.85	GGL
y16	1167.27	1166.96	129.95	LG	**a11**	861.27	861.26	71.95	L
y15	1095.25	1095.03	71.93	L	**a10**	789.25	789.39	71.87	L

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 12.

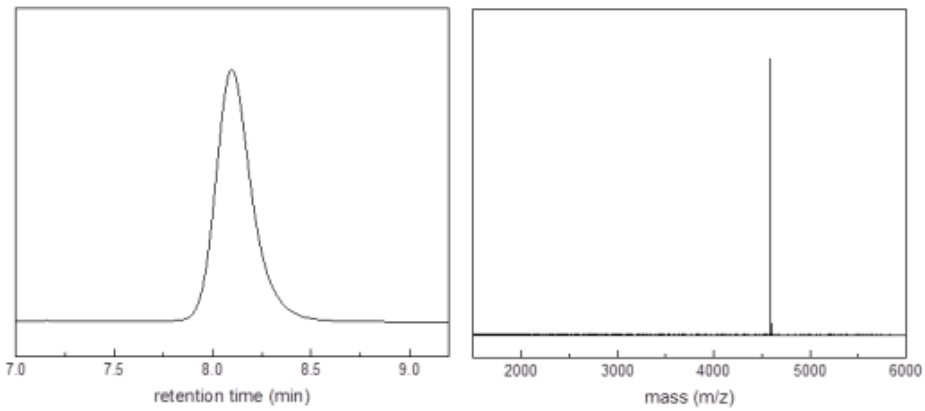


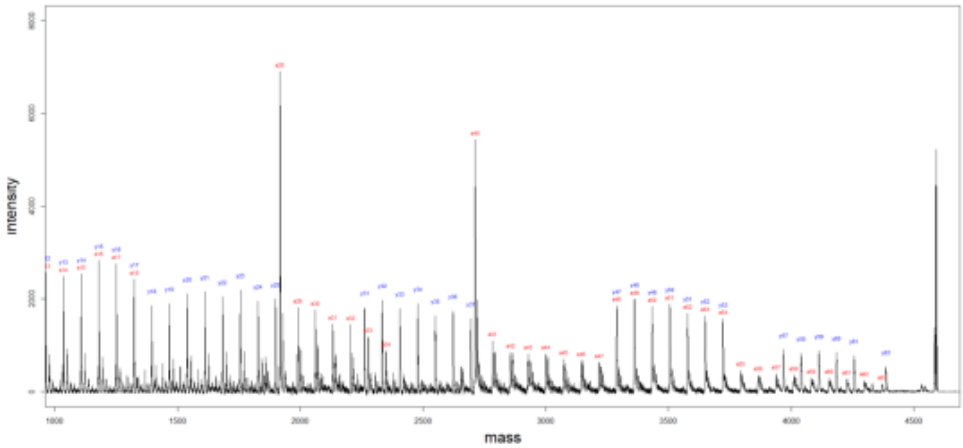


Decoding table of PLGA chain 12

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
$[M+Na]^+$	4518.11	4518.76			$[M+Na]^+$	4518.11	4518.76		
y63	4314	4317.33	201.43	Si-L	**a62**	4240.02	4240.79	277.97	Bz-GGL
y60	4125.97	4128.04	189.29	LGG	**a61**	4168	4170.52	70.27	L
y59	4053.95	4056.39	71.65	L	**a60**	4095.98	4098.52	72	L
y58	3981.93	3984.23	72.16	L	**a59**	4023.96	4027.73	70.79	L
y57	3909.9	3913.29	70.94	L	**a43**	3081.86	3083.27	944.46	GGGGG GGGGG GGGGGL
y55	3779.88	3782.7	130.59	LG	**a42**	3009.84	3010.87	72.4	L
y54	3707.86	3710.07	72.63	L	**a41**	2937.81	2937.98	72.89	L
y53	3635.84	3637.81	72.26	L	**a40**	2865.79	2866.61	71.37	L
y52	3563.82	3567.18	70.63	L	**a39**	2793.77	2794.32	72.29	L
y51	3491.79	3493.78	73.4	L	**a35**	2547.73	2547.81	246.51	GGGL
y50	3419.77	3421.48	72.3	L	**a34**	2475.71	2475.65	72.16	L
y49	3347.75	3349.25	72.23	L	**a33**	2403.69	2403.68	71.97	L
y48	3275.73	3277.44	71.81	L	**a32**	2331.67	2331.67	72.01	L
y47	3203.71	3204.89	72.55	L	**a30**	2201.64	2200.71	130.96	GL
y46	3131.69	3132.89	72	L	**a29**	2129.62	2129.46	71.25	L
y45	3059.67	3060.85	72.04	L	**a28**	2057.6	2057.51	71.95	L
y44	2987.65	2988.8	72.05	L	**a27**	1985.58	1984.55	72.96	L
y43	2915.63	2916.41	72.39	L	**a26**	1913.56	1912.48	72.07	L
y42	2843.6	2843.95	72.46	L	**a25**	1841.54	1840.57	71.91	L
y41	2771.58	2772.1	71.85	L	**a24**	1769.52	1768.51	72.06	L
y40	2699.56	2700.04	72.06	L	**a23**	1697.5	1696.55	71.96	L
y39	2627.54	2627.81	72.23	L	**a22**	1625.48	1624.58	71.97	L
y38	2555.52	2555.74	72.07	L	**a21**	1553.45	1552.64	71.94	L
y37	2483.5	2482.66	73.08	L	**a20**	1481.43	1480.81	71.83	L
y36	2411.48	2411.7	70.96	L	**a19**	1409.41	1408.69	72.12	L
y35	2339.46	2339.63	72.07	L	**a18**	1337.39	1336.79	71.9	L
y33	2209.43	2208.43	131.2	LG	**a17**	1265.37	1264.87	71.92	L
y32	2137.41	2136.35	72.08	L	**a16**	1193.35	1192.91	71.96	L
y31	2065.39	2064.34	72.01	L	**a15**	1121.33	1120.93	71.98	L
y30	1993.37	1992.33	72.01	L	**a14**	1049.31	1048.97	71.96	L
y26	1747.33	1746.39	245.94	LGGG	**a13**	977.29	977.1	71.87	L
y25	1675.31	1674.37	72.02	L	**a12**	905.26	905.32	71.78	L
y24	1603.29	1602.43	71.94	L	**a11**	833.24	833.25	72.07	L
y23	1531.27	1530.46	71.97	L					

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 13.

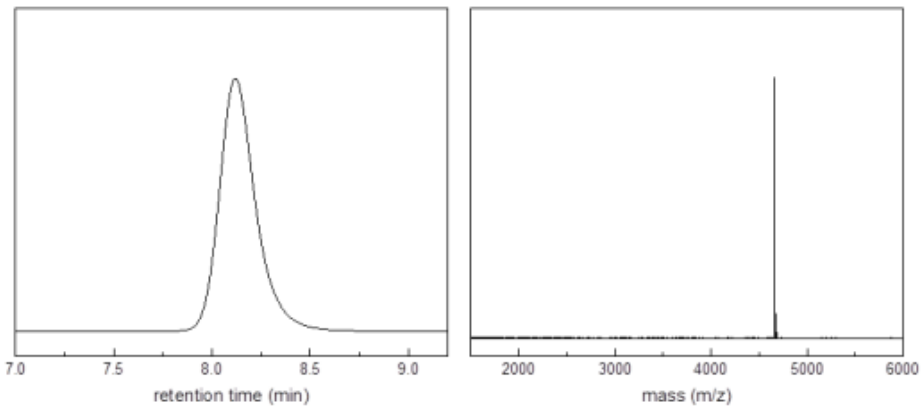


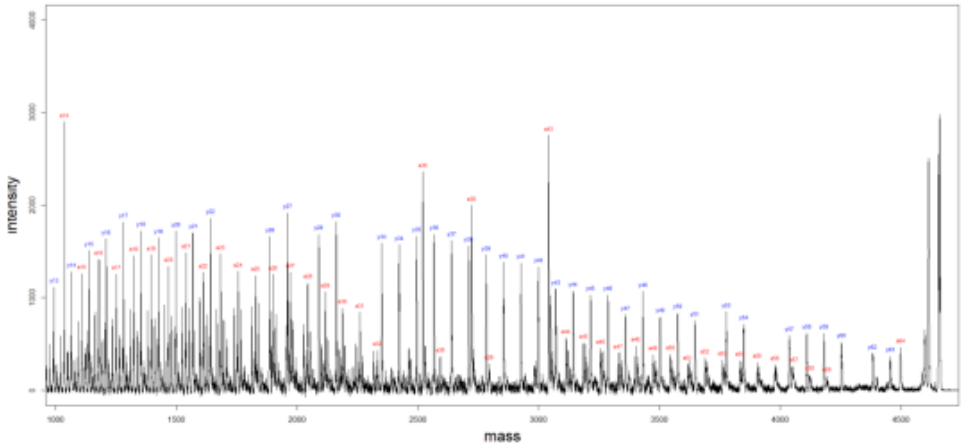


Decoding table of PLGA chain 13

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
[M+Na]⁺	4588.19	4588.41			**[M+Na]⁺**	4588.19	4588.41		
y63	4384.08	4386.77	201.64	Si-L	**a63**	4368.11	4369.88	218.53	Bz-GL
y61	4254.05	4255.89	130.88	LG	**a62**	4296.09	4298.05	71.83	L
y60	4182.03	4184.89	71	L	**a61**	4224.06	4226.27	71.78	L
y59	4110.01	4112.97	71.92	L	**a60**	4152.04	4154.8	71.47	L
y58	4037.99	4040.02	72.95	L	**a59**	4080.02	4081.74	73.06	L
y57	3965.97	3968.14	71.88	L	**a58**	4008	4010.1	71.64	L
y53	3719.93	3722.71	245.43	LGGG	**a57**	3935.98	3938.4	71.7	L
y52	3647.91	3649.87	72.84	L	**a56**	3863.96	3865.99	72.41	L
y51	3575.89	3577.78	72.09	L	**a55**	3791.94	3794.85	71.14	L
y50	3503.87	3505.61	72.17	L	**a54**	3719.92	3722.71	72.14	L
y49	3431.85	3433.57	72.04	L	**a53**	3647.9	3649.87	72.84	L
y48	3359.82	3361.45	72.12	L	**a52**	3575.87	3577.78	72.09	L
y47	3287.8	3289.02	72.43	L	**a51**	3503.85	3505.61	72.17	L
y37	2693.73	2694.55	594.47	LGGGG GGGGG	**a50**	3431.83	3433.57	72.04	L
y36	2621.71	2622.42	72.13	L	**a49**	3359.81	3361.45	72.12	L
y35	2549.69	2550.37	72.05	L	**a48**	3287.79	3289.02	72.43	L
y34	2477.67	2478.32	72.05	L	**a47**	3215.77	3217.36	71.66	L
y33	2405.65	2406.24	72.08	L	**a46**	3143.75	3144.99	72.37	L
y32	2333.63	2334.26	71.98	L	**a45**	3071.73	3072.96	72.03	L
y31	2261.61	2262.23	72.03	L	**a44**	2999.71	3000.78	72.18	L
y25	1899.56	1899.11	363.12	LGGGGG	**a43**	2927.68	2928.84	71.94	L
y24	1827.54	1827.11	72	L	**a42**	2855.66	2856.6	72.24	L
y23	1755.52	1755.13	71.98	L	**a41**	2783.64	2784.53	72.07	L
y22	1683.49	1683.11	72.02	L	**a40**	2711.62	2712.52	72.01	L
y21	1611.47	1611.1	72.01	L	**a34**	2349.57	2349.27	363.25	GGGGGL
y20	1539.45	1539.14	71.96	L	**a33**	2277.55	2277.12	72.15	L
y19	1467.43	1467.21	71.93	L	**a32**	2205.53	2205.03	72.09	L
y18	1395.41	1395.22	71.99	L	**a31**	2133.51	2133.13	71.9	L
y17	1323.39	1323.24	71.98	L	**a30**	2061.49	2061.1	72.03	L
y16	1251.37	1251.27	71.97	L	**a29**	1989.47	1989.01	72.09	L
y15	1179.35	1179.35	71.92	L	**a28**	1917.45	1917.12	71.89	L
y14	1107.33	1107.41	71.94	L	**a18**	1323.38	1323.24	593.88	GGGGG GGGGL
y13	1035.3	1035.47	71.94	L	**a17**	1251.35	1251.27	71.97	L
y12	963.28	963.55	71.92	L	**a16**	1179.33	1179.35	71.92	L

GPC, MALDI-TOF, and MS/MS spectra of PLGA chain 14.



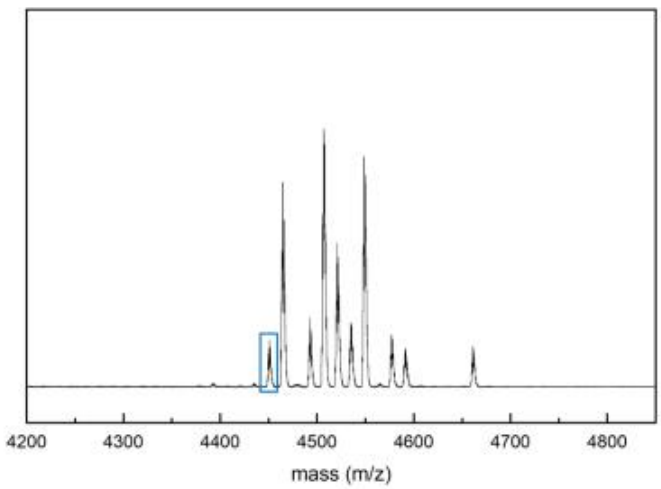


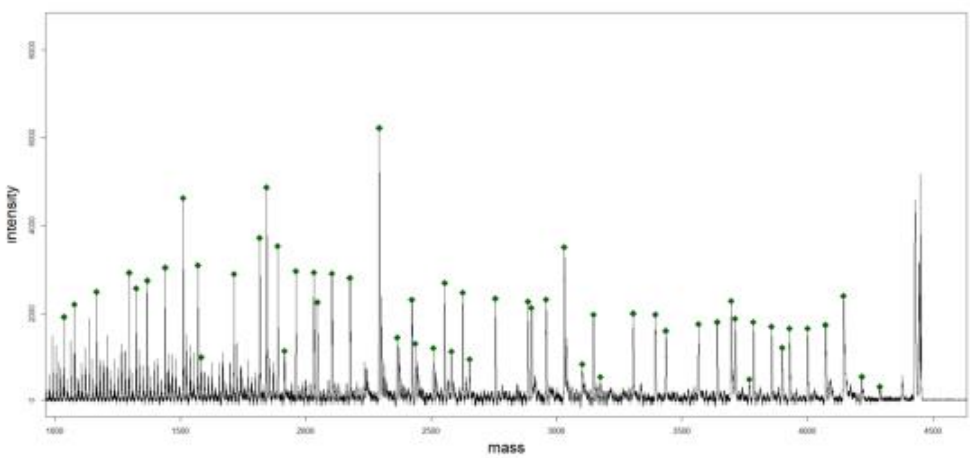
Decoding table of PLGA chain 14

Si to Bz	Calc.	Found	Different	Sequence	Bz to Si	Calc.	Found	Different	Sequence
$[M+Na]^+$	4658.27	4659			$[M+Na]^+$	4658.27	4659		
y63	4454.16	4455.97	203.03	Si-L	**a64**	4496.19	4498.25	160.75	Bz-L
y62	4382.14	4382.69	73.28	L	**a59**	4192.15	4193.55	304.7	GGGGL
y60	4252.11	4253.72	128.97	LG	**a58**	4120.13	4122.21	71.34	L
y59	4180.09	4181.25	72.47	L	**a57**	4048.1	4049.84	72.37	L
y58	4108.07	4110.01	71.24	L	**a56**	3976.08	3978.9	70.94	L
y57	4036.05	4037.67	72.34	L	**a55**	3904.06	3905.19	73.71	L
y54	3848.01	3850.14	187.53	LGG	**a54**	3832.04	3833.21	71.98	L
y53	3775.99	3776.93	73.21	L	**a53**	3760.02	3761.33	71.88	L
y51	3645.97	3646.85	130.08	LG	**a52**	3688	3689.09	72.24	L
y50	3573.94	3575.19	71.66	L	**a51**	3615.98	3616.92	72.17	L
y49	3501.92	3503.8	71.39	L	**a50**	3543.96	3546.06	70.86	L
y48	3429.9	3430.92	72.88	L	**a49**	3471.94	3473.28	72.78	L
y47	3357.88	3358.86	72.06	L	**a48**	3399.91	3400.74	72.54	L
y46	3285.86	3286.42	72.44	L	**a47**	3327.89	3328.54	72.2	L
y45	3213.84	3214.98	71.44	L	**a46**	3255.87	3256.69	71.85	L
y44	3141.82	3142.83	72.15	L	**a45**	3183.85	3185.28	71.41	L
y43	3069.8	3070.81	72.02	L	**a44**	3111.83	3112.66	72.62	L
y42	2997.78	2998.47	72.34	L	**a43**	3039.81	3040.77	71.89	L
y41	2925.76	2926.76	71.71	L	**a39**	2793.77	2794.67	246.1	GGGL
y40	2853.73	2854.73	72.03	L	**a38**	2721.75	2722.78	71.89	L
y39	2781.71	2782.73	72	L	**a36**	2591.72	2592.54	130.24	GL
y38	2709.69	2710.78	71.95	L	**a35**	2519.7	2520.76	71.78	L
y37	2637.67	2638.8	71.98	L	**a32**	2331.67	2331.88	188.88	GGL
y36	2565.65	2566.74	72.06	L	**a31**	2259.65	2259.76	72.12	L
y35	2493.63	2494.67	72.07	L	**a30**	2187.63	2187.65	72.11	L
y34	2421.61	2421.69	72.98	L	**a29**	2115.61	2116.72	70.93	L
y33	2349.59	2350.58	71.11	L	**a28**	2043.59	2043.76	72.96	L
y30	2161.55	2161.64	188.94	LGG	**a27**	1971.57	1971.63	72.13	L
y29	2089.53	2089.57	72.07	L	**a26**	1899.54	1899.64	71.99	L
y27	1959.51	1959.57	130	LG	**a25**	1827.52	1827.65	71.99	L
y26	1887.49	1887.53	72.04	L	**a24**	1755.5	1755.63	72.02	L
y22	1641.45	1641.51	246.02	LGGG	**a23**	1683.48	1683.56	72.07	L
y21	1569.43	1569.53	71.98	L	**a22**	1611.46	1611.55	72.01	L
y20	1497.41	1497.46	72.07	L	**a21**	1539.44	1539.54	72.01	L
y19	1425.38	1425.49	71.97	L	**a20**	1467.42	1467.56	71.98	L
y18	1353.36	1353.47	72.02	L	**a19**	1395.4	1395.59	71.97	L
y17	1281.34	1281.54	71.93	L	**a18**	1323.38	1323.52	72.07	L

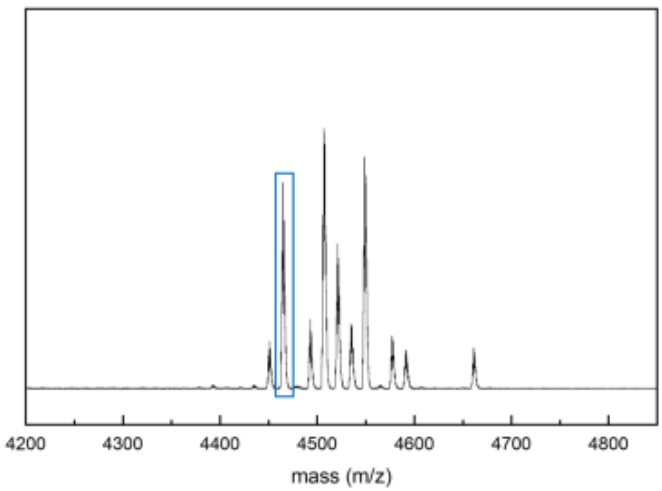
A.3.3 MALDI-TOF/TOF tandem mass sequencing of the mixture of the fourteen sequence-defined PLGAs

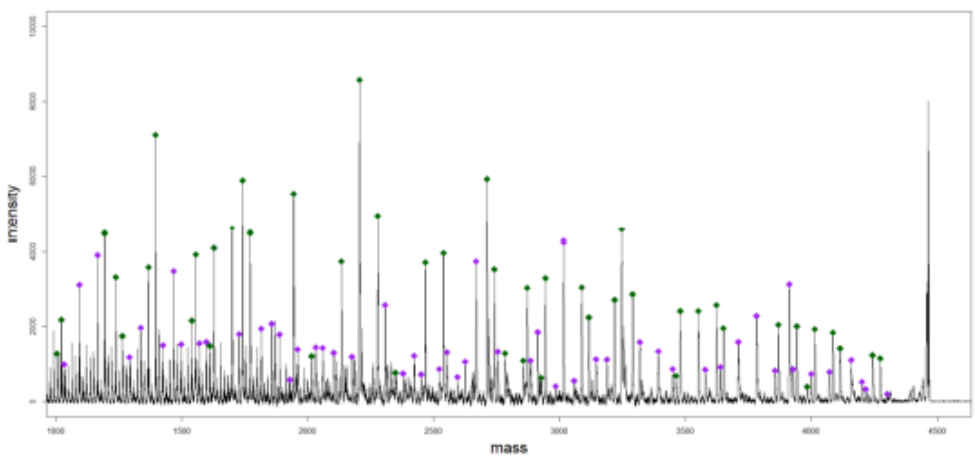
MALDI-TOF mass spectrum of mixed polymer resin and tandem mass spectrum of PLGA chain 8 (parent ion: 4448.80 Da).



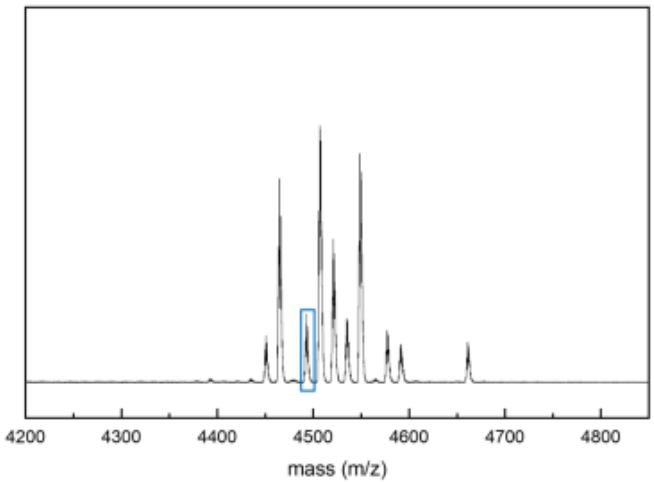


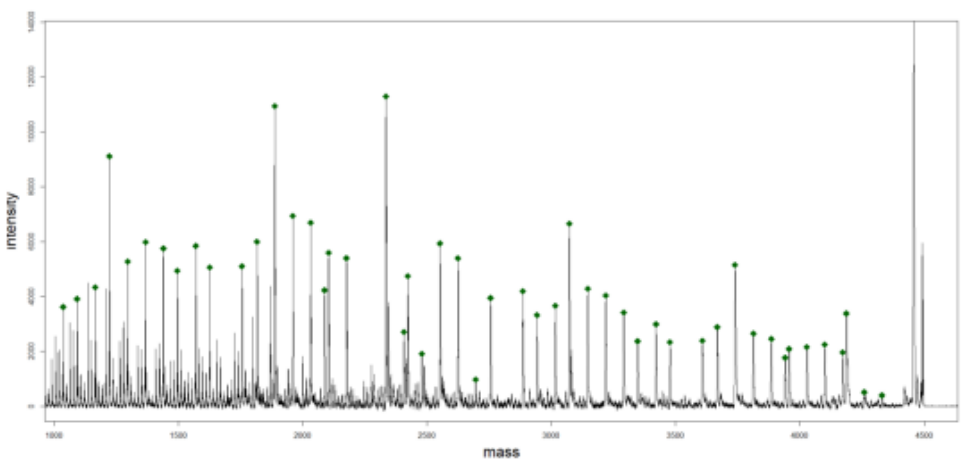
MALDI-TOF mass spectrum of mixed polymer resin and tandem mass spectrum of PLGA chain 4 and 10 (parent ion: 4462.84 Da).



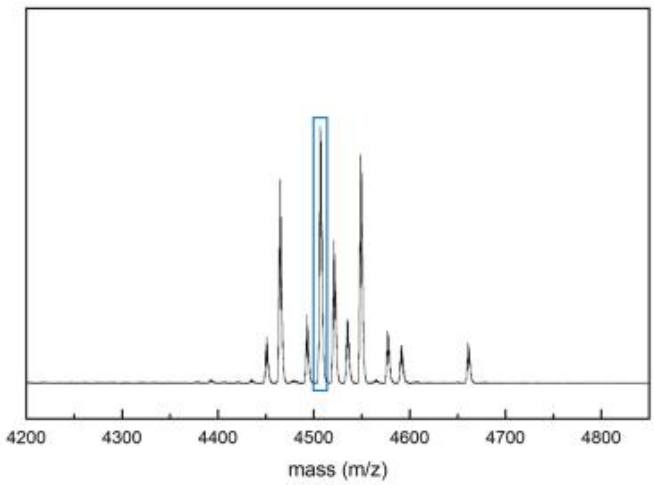


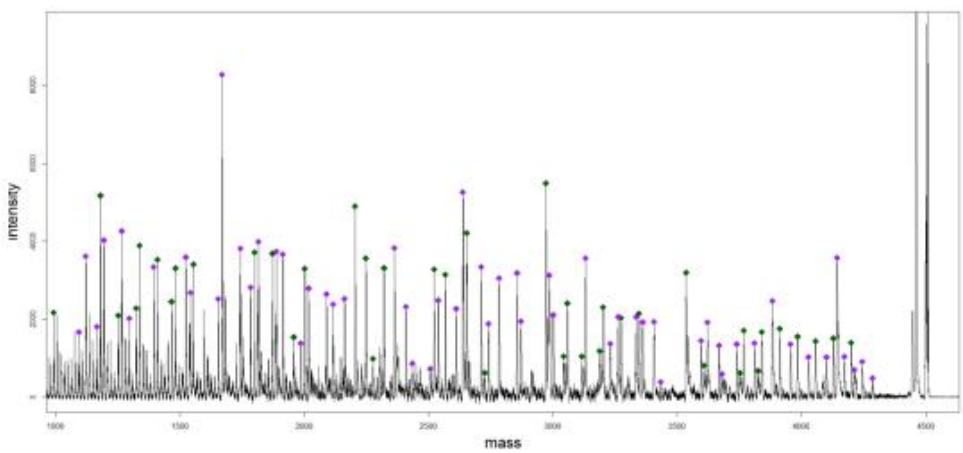
MALDI-TOF mass spectrum of mixed polymer resin and tandem mass spectrum of PLGA chain 9 (parent ion: 4490.79 Da).



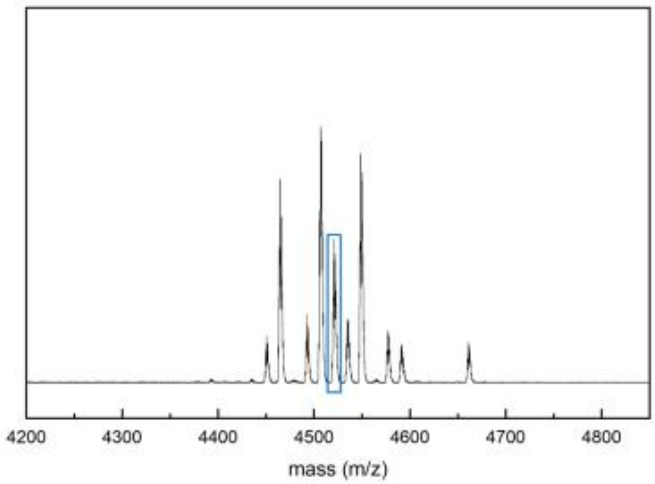


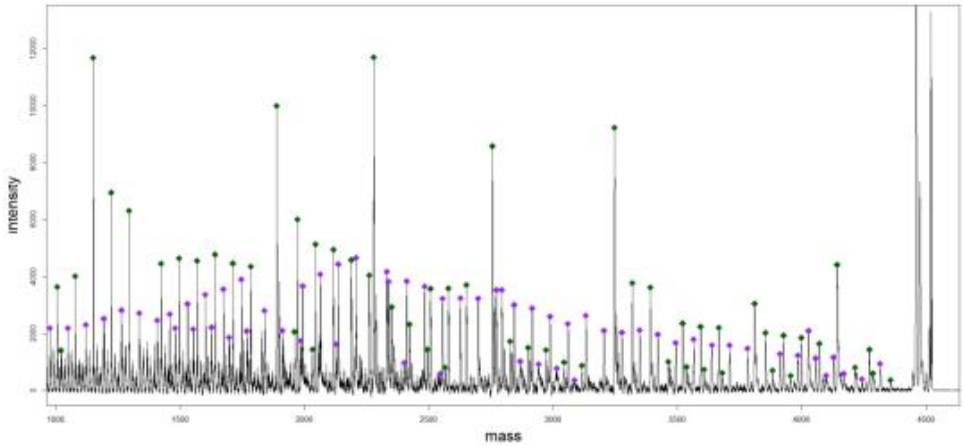
MALDI-TOF mass spectrum of mixed polymer resin and tandem mass spectrum of PLGA chain 1 and 11 (parent ion: 4504.980 Da).



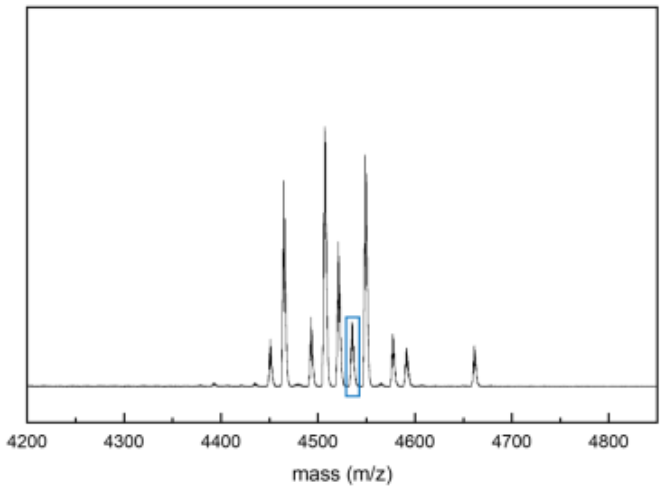


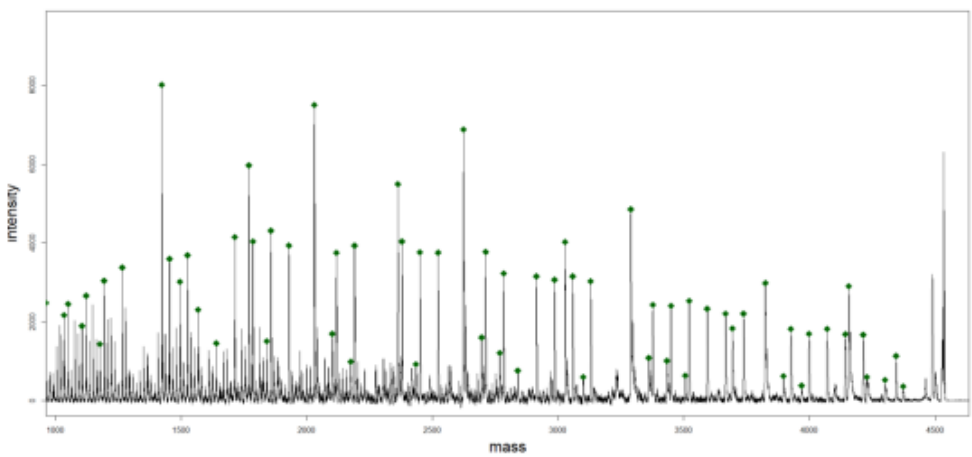
MALDI-TOF mass spectrum of mixed polymer resin and tandem mass spectrum of PLGA chain 2 and 12 (parent ion: 4518.83 Da).



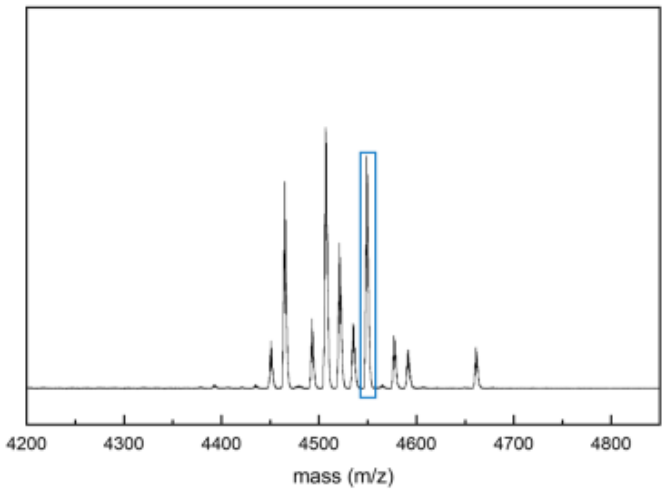


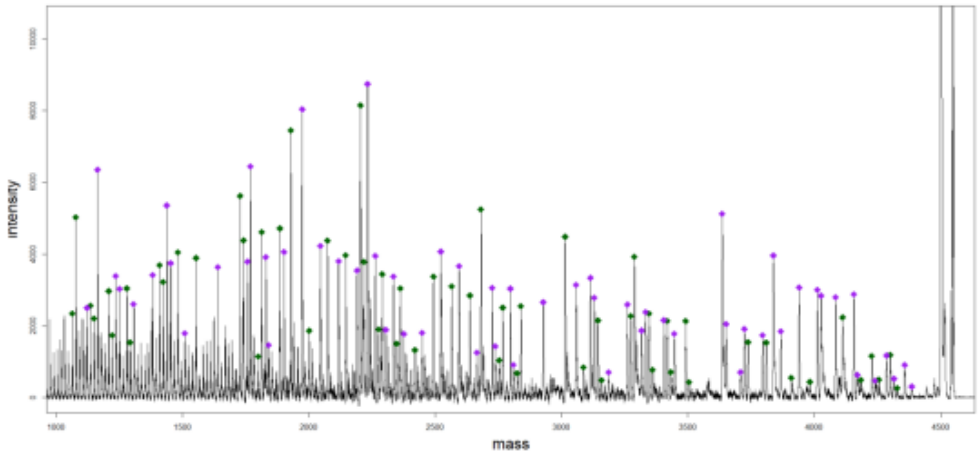
MALDI-TOF mass spectrum of mixed polymer resin and tandem mass spectrum of PLGA chain 5 (parent ion: 4532.97 Da).



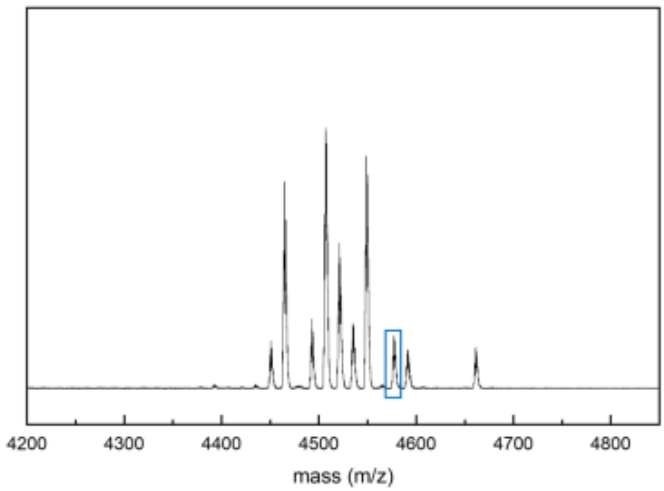


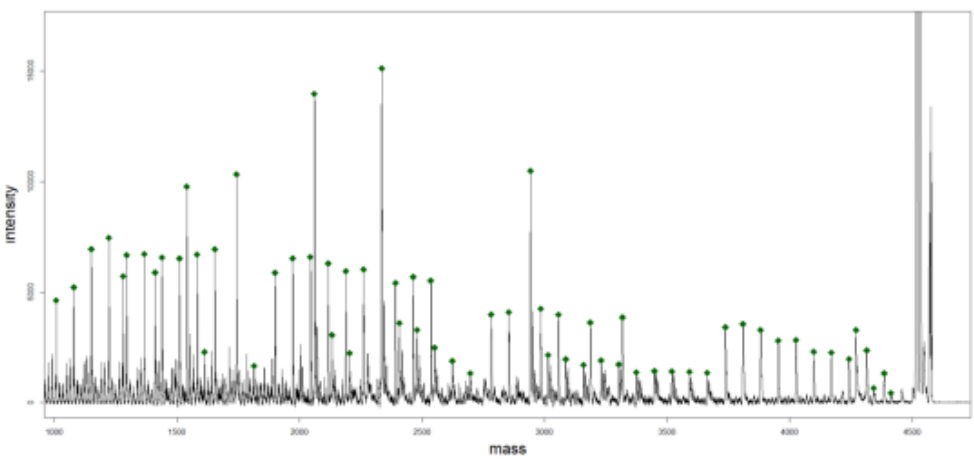
MALDI-TOF mass spectrum of mixed polymer resin and tandem mass spectrum of PLGA chain 3 and 6 (parent ion: 4546.86 Da).



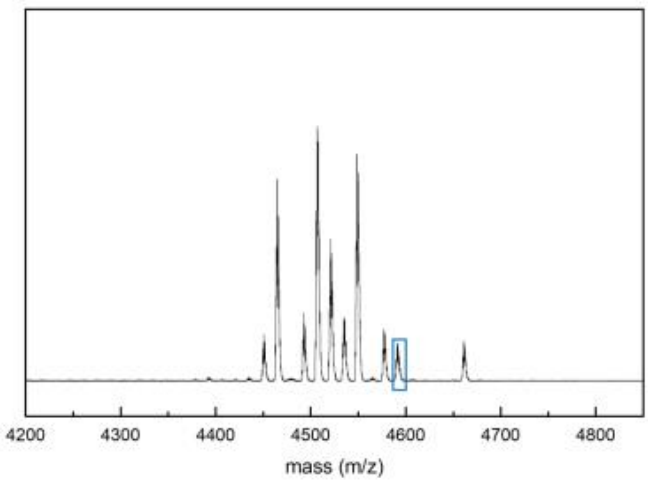


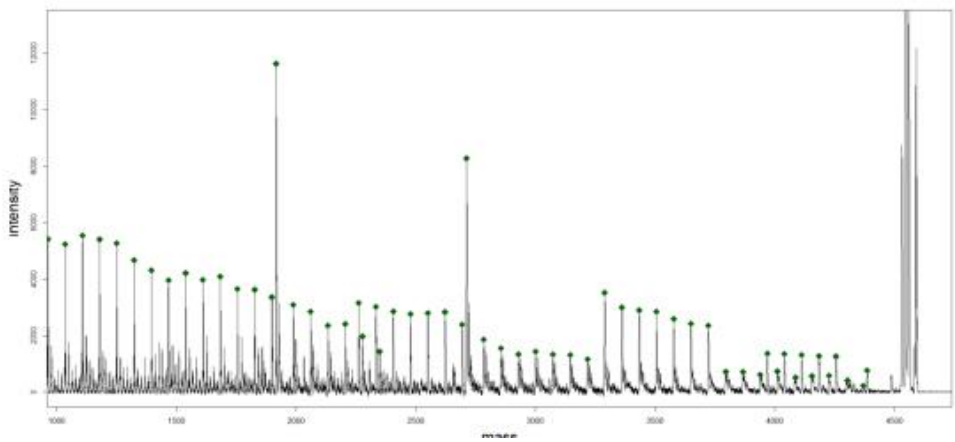
MALDI-TOF mass spectrum of mixed polymer resin and tandem mass spectrum of PLGA chain 7 (parent ion: 4575.07 Da).



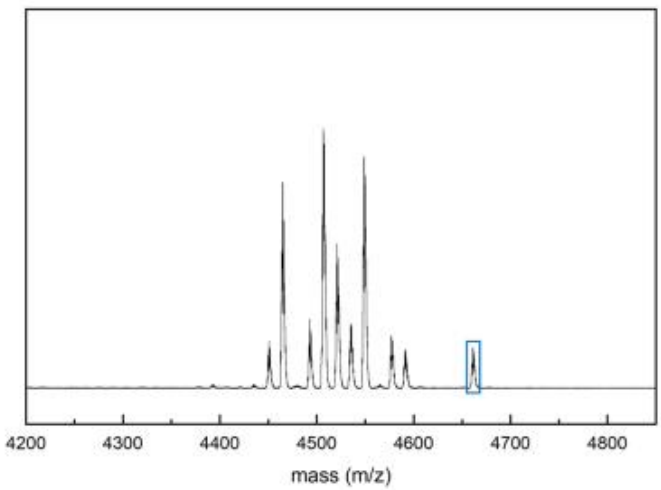


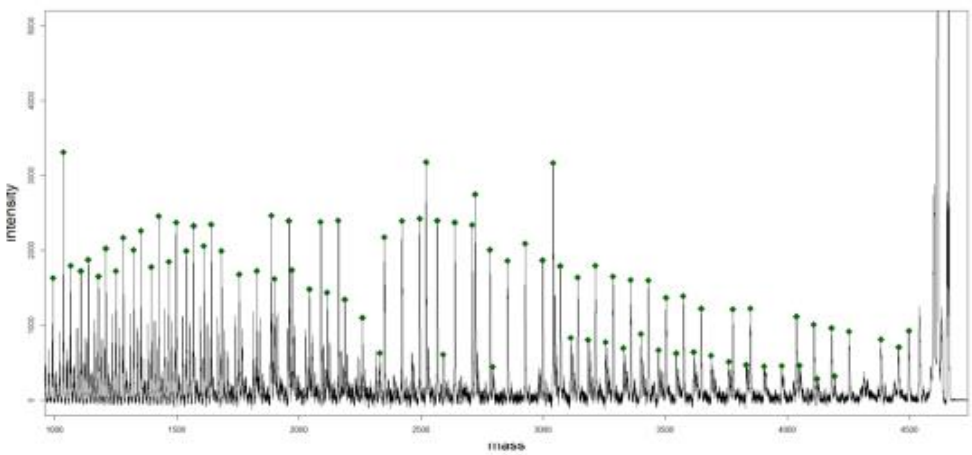
MALDI-TOF mass spectrum of mixed polymer resin and tandem mass spectrum of PLGA chain 13 (parent ion: 4589.21 Da).





MALDI-TOF mass spectrum of mixed polymer resin and tandem mass spectrum of PLGA chain 14 (parent ion: 4659.40 Da).

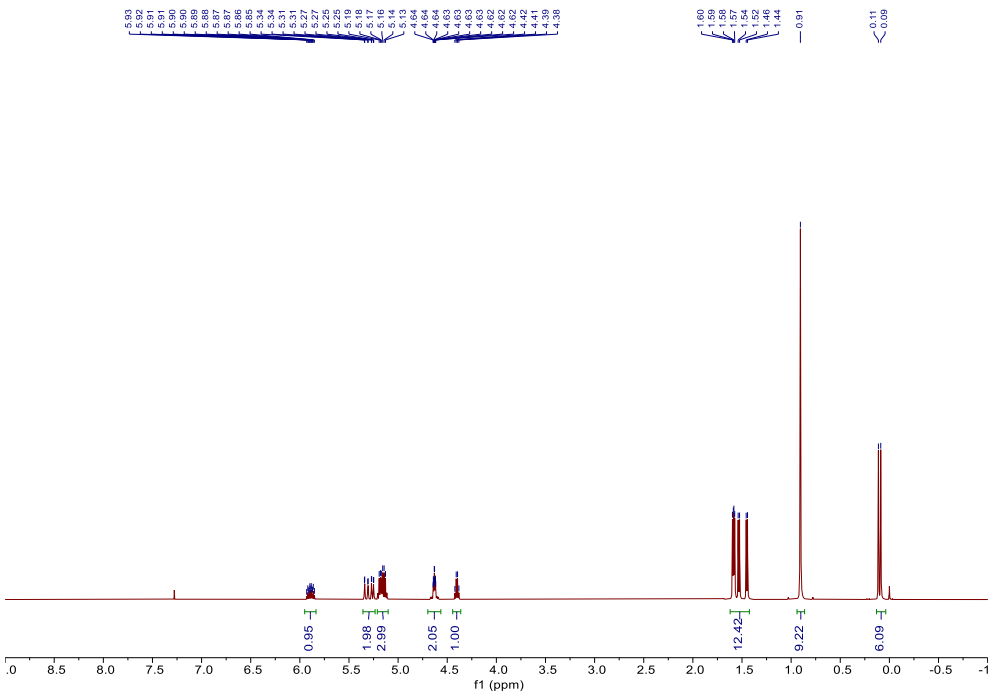


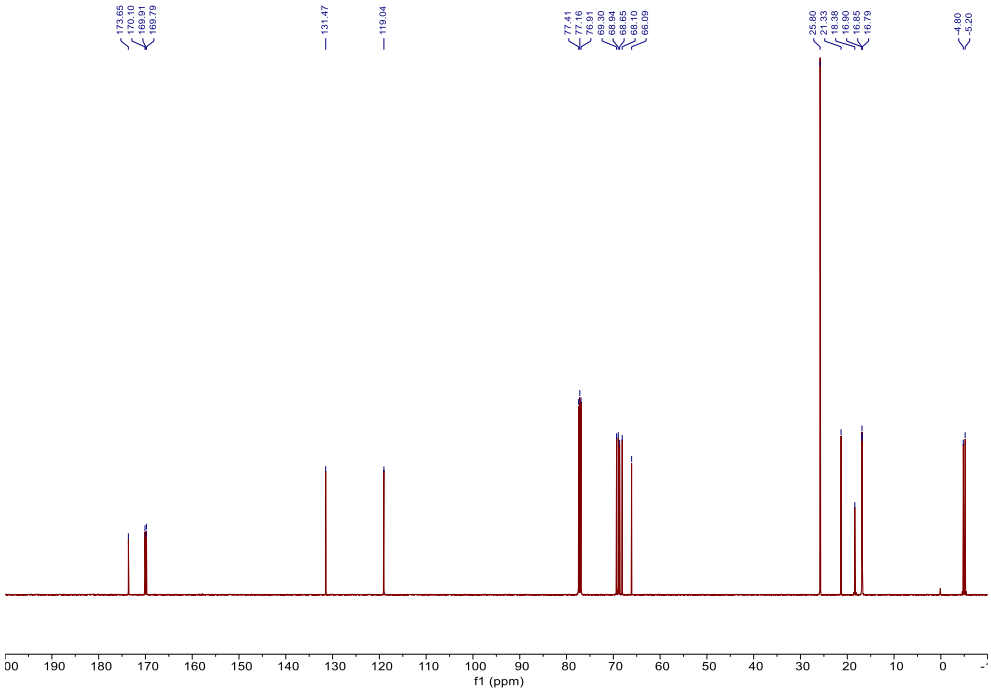


Chapter 4: Nondestructive Sequencing of Enantiopure Oligoesters by Nuclear Magnetic Resonance Spectroscopy

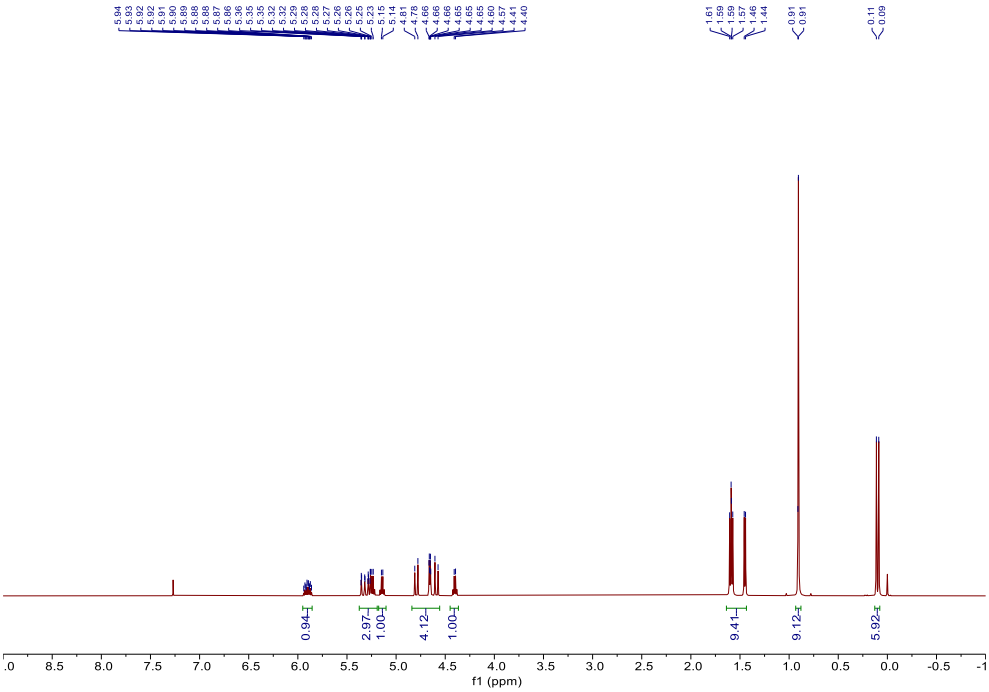
A.4.1 ^{1}H and ^{13}C NMR spectra of sequence-defined enantiopure oLGs

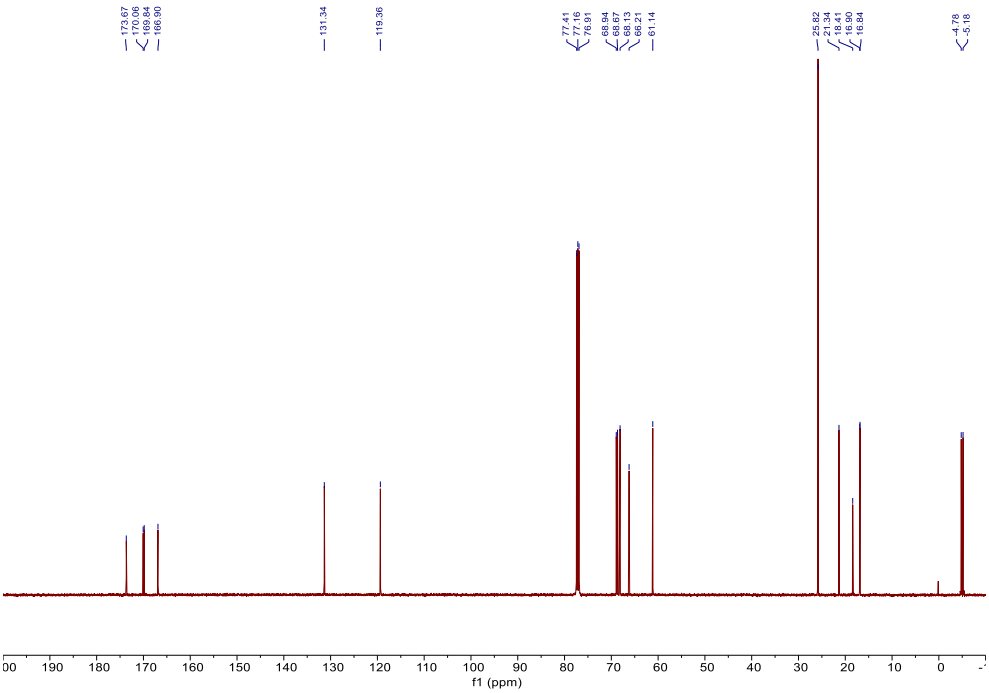
^{1}H and ^{13}C NMR spectra of **LLLL**.



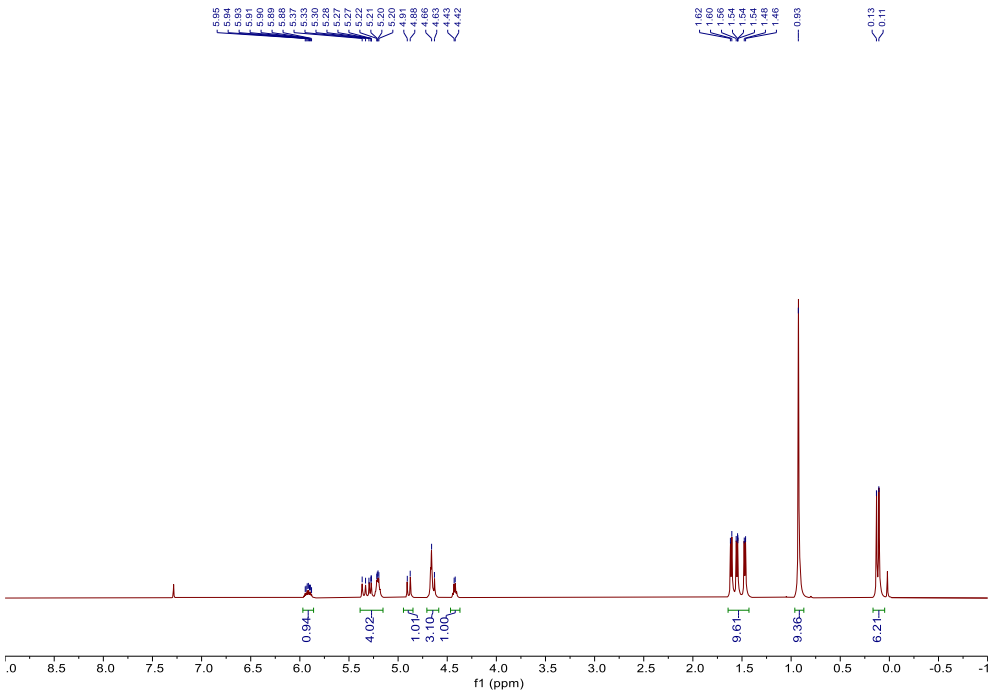

173.65
170.10
169.91
169.79
131.47
119.04
77.41
77.16
76.91
69.30
68.94
68.65
68.10
66.09
25.80
21.33
18.38
16.90
16.65
16.79
-4.80
-5.20
00 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

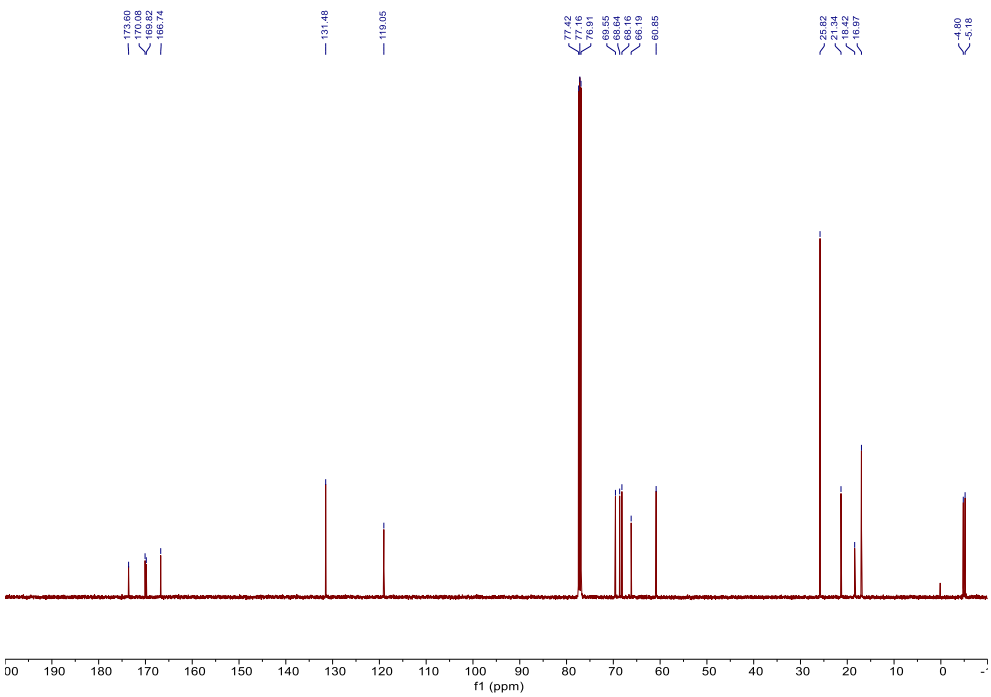
^{1}H and ^{13}C NMR spectra of **LLLG**.



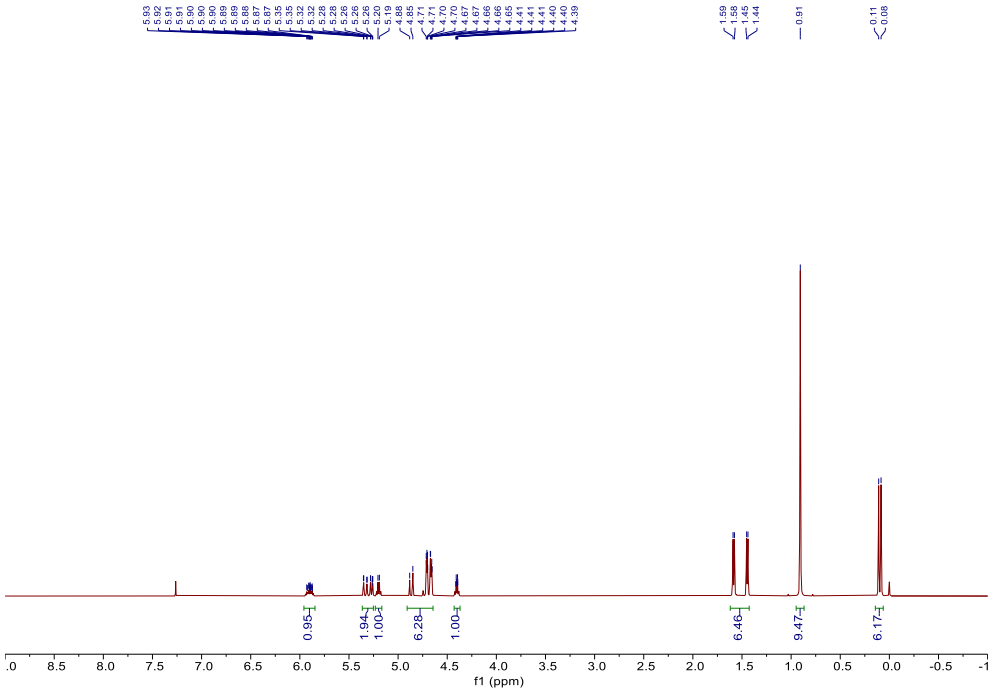


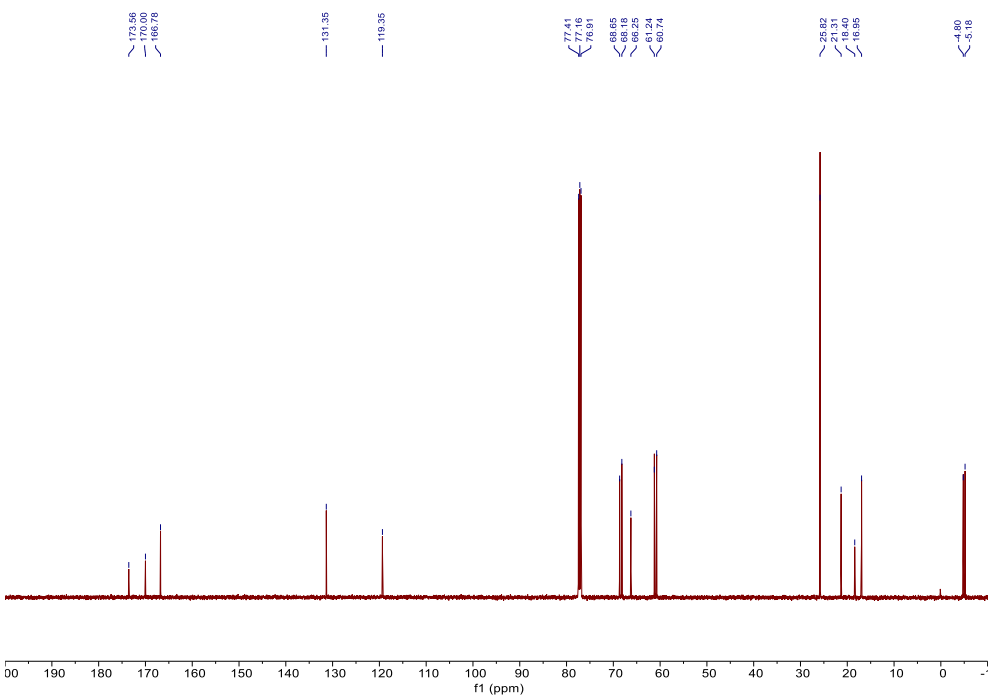
^{1}H and ^{13}C NMR spectra of **LLGL**.



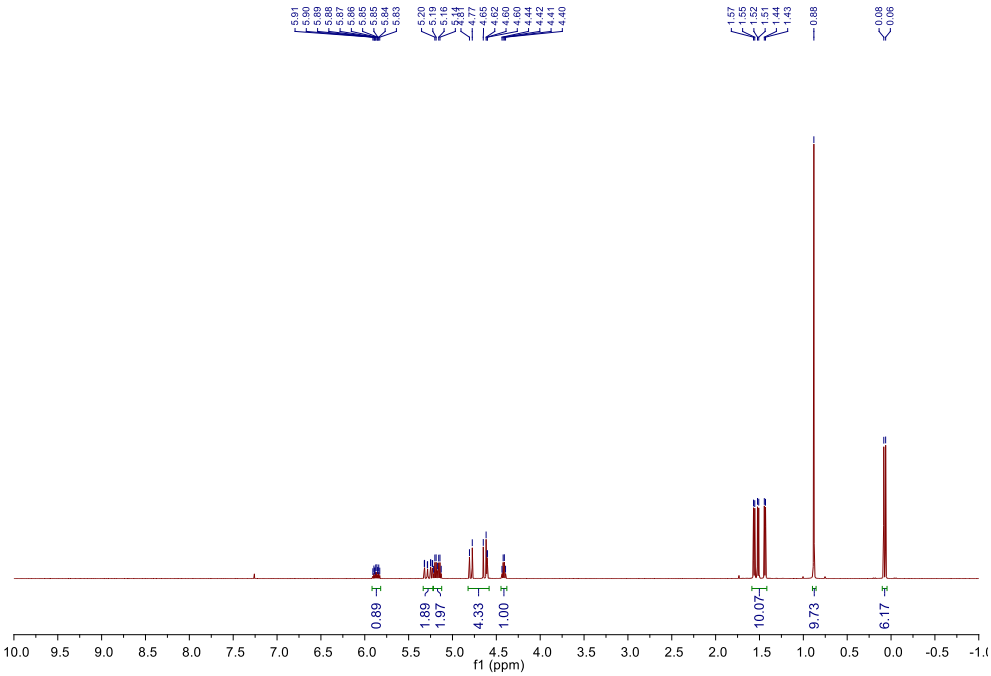


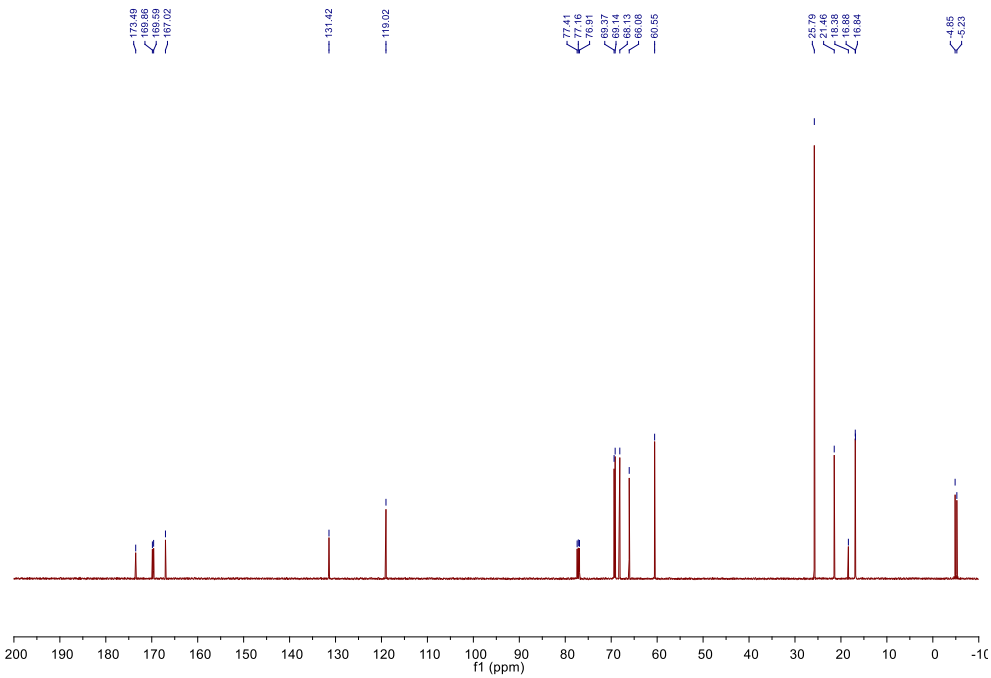
^{1}H and ^{13}C NMR spectra of **LLGG**.



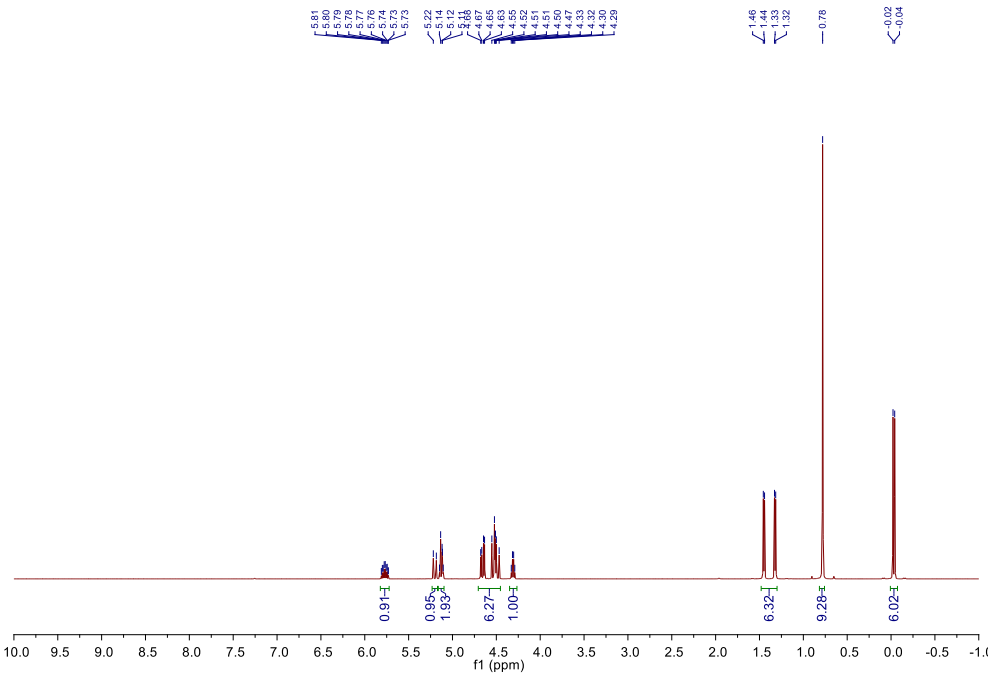


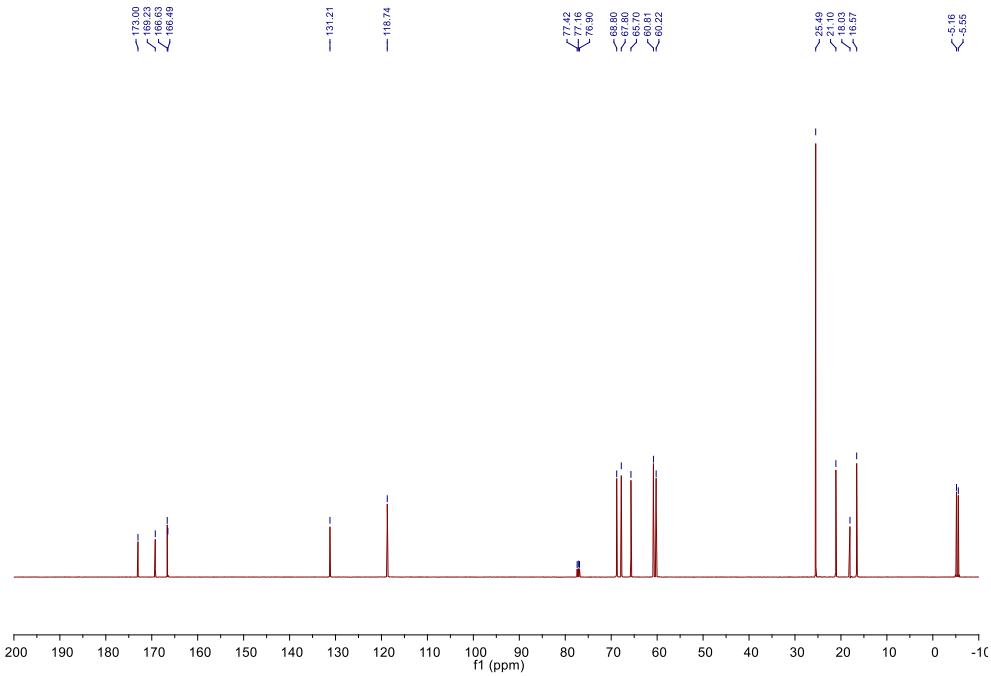
^{1}H and ^{13}C NMR spectra of **LGLL**.



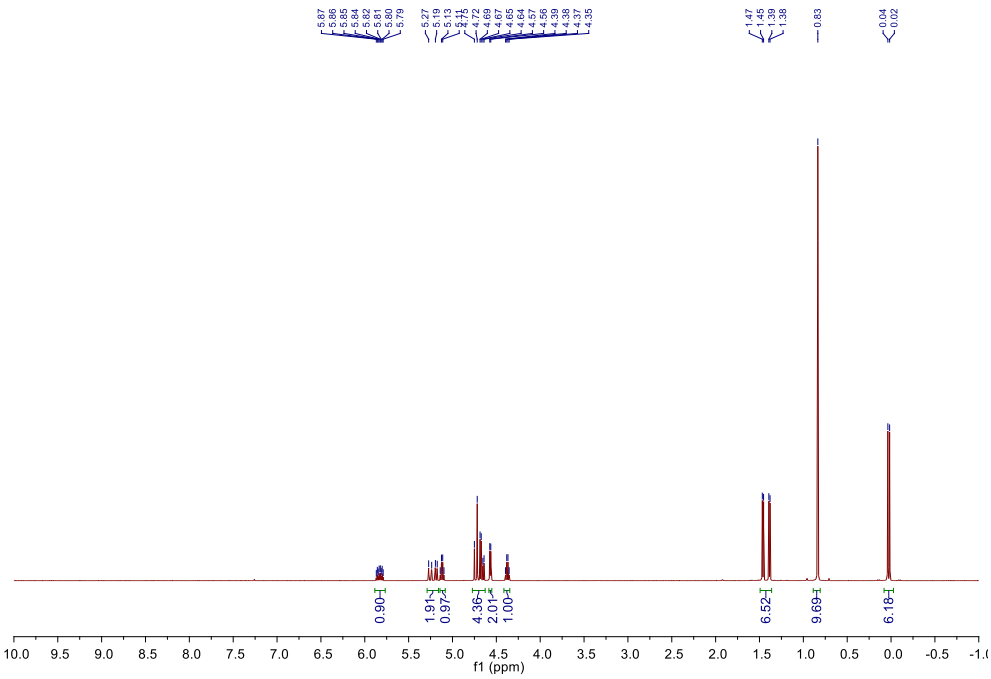


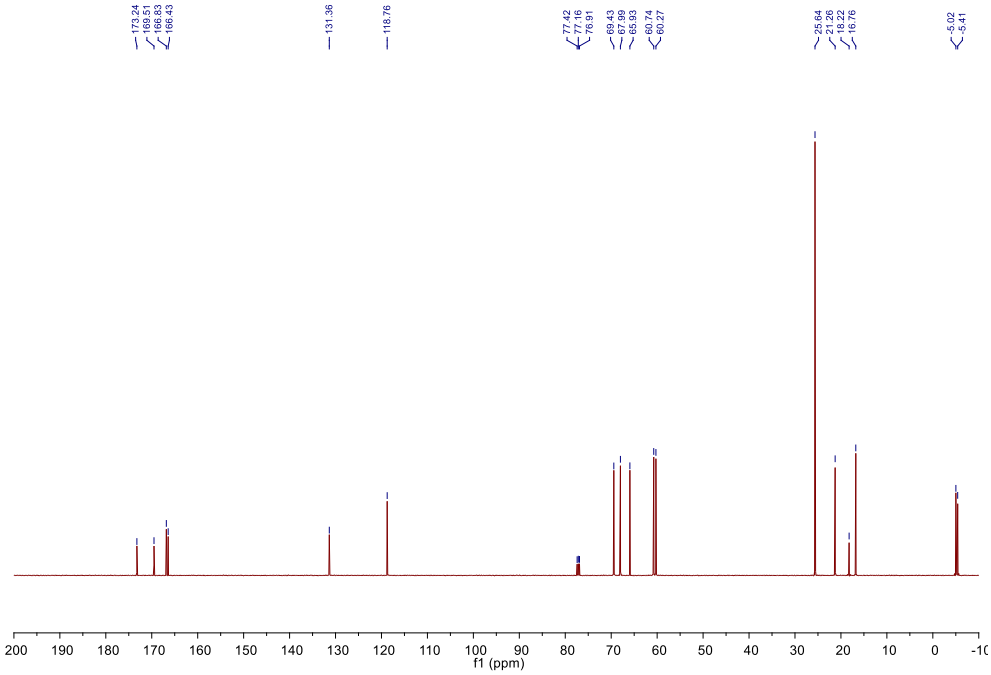
^{1}H and ^{13}C NMR spectra of **LGLG**.



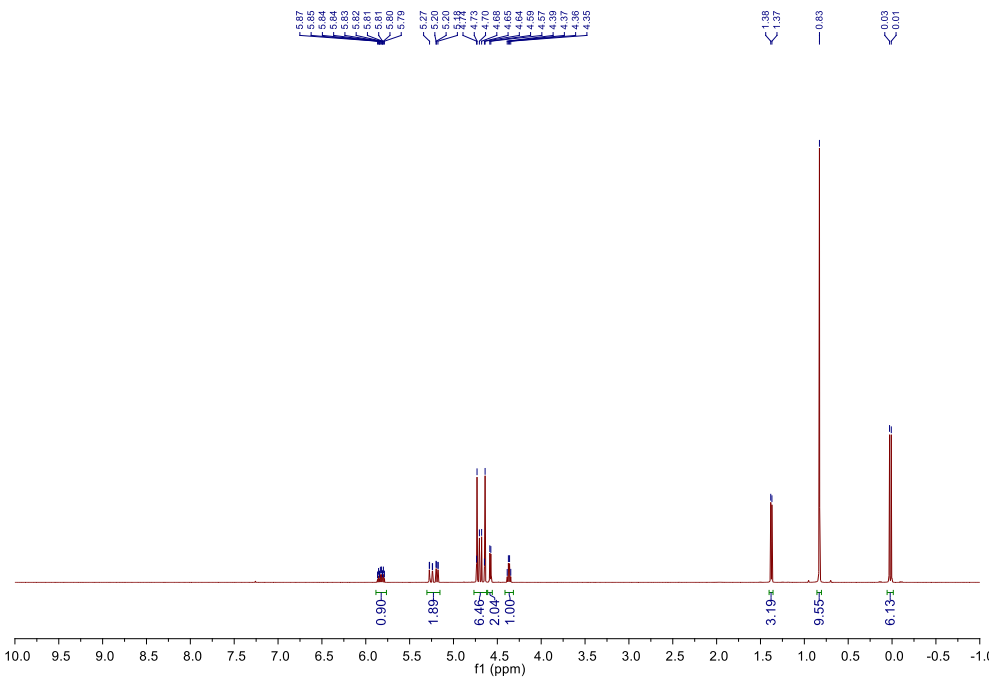


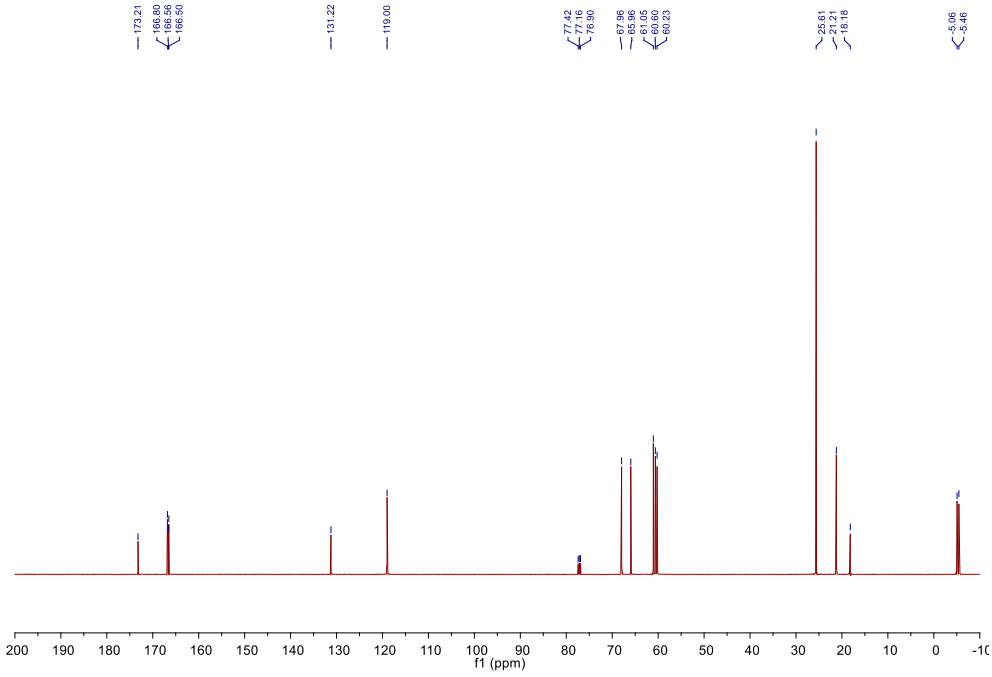
^{1}H and ^{13}C NMR spectra of **LGGL**.



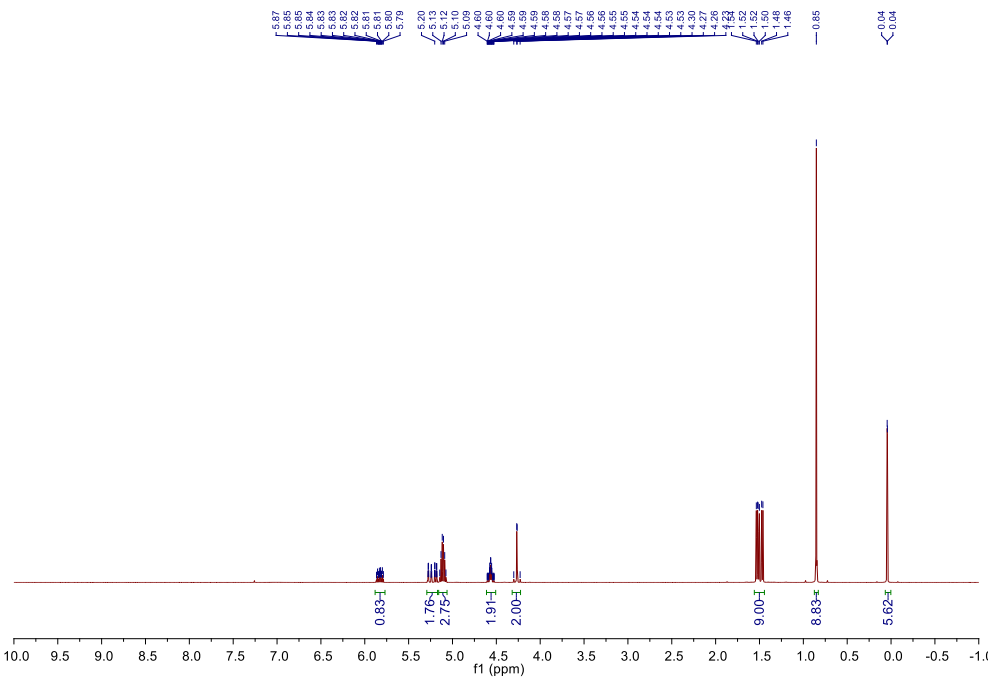


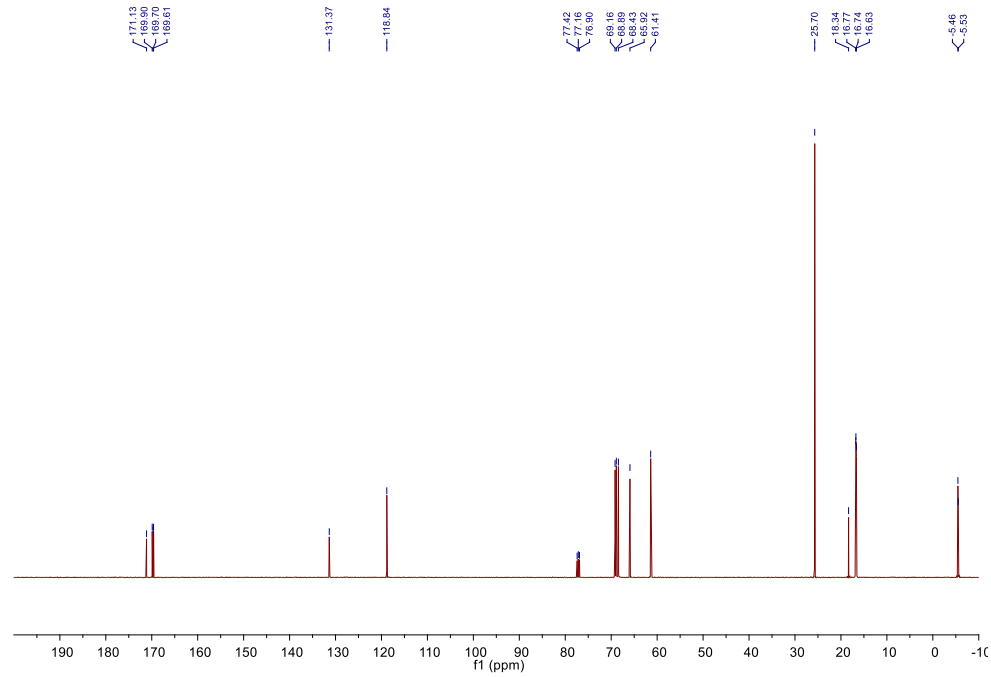
^{1}H and ^{13}C NMR spectra of **LGGG**.



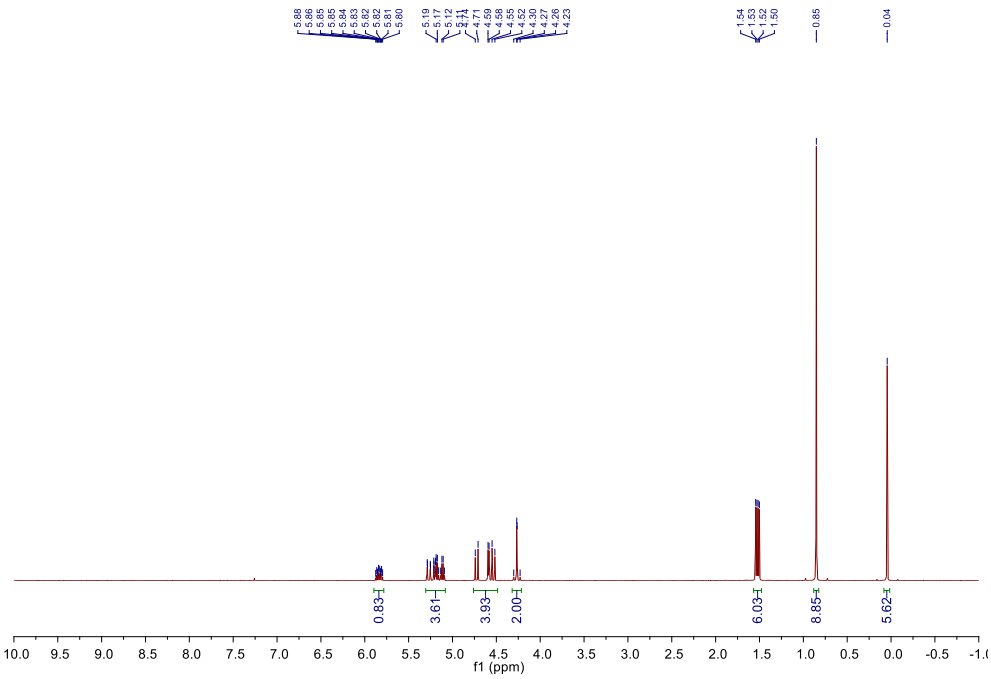


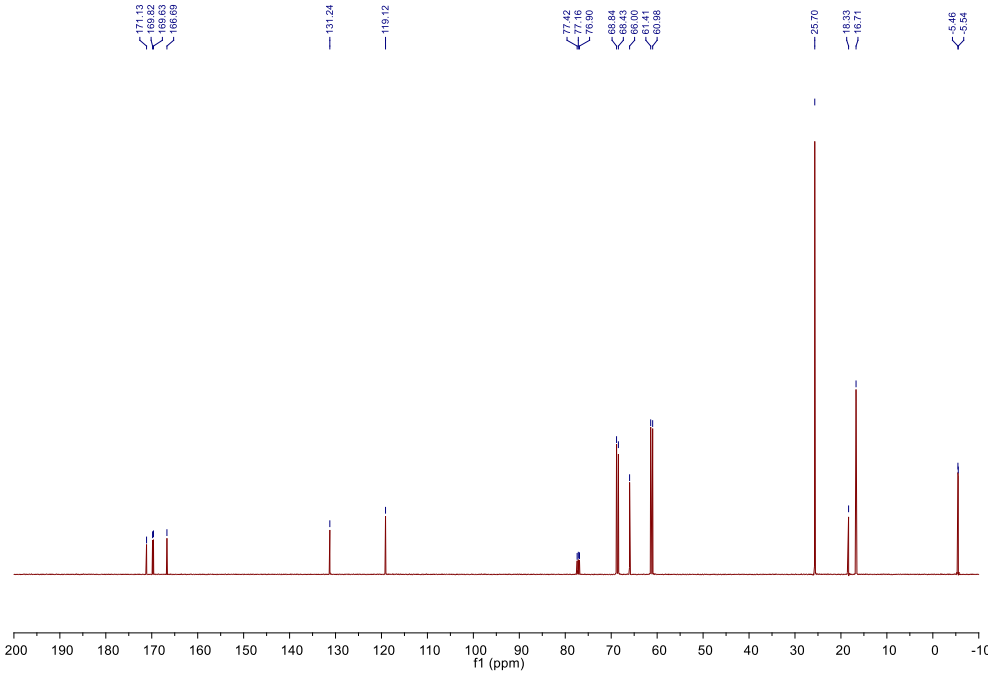
^{1}H and ^{13}C NMR spectra of **GLLL**.



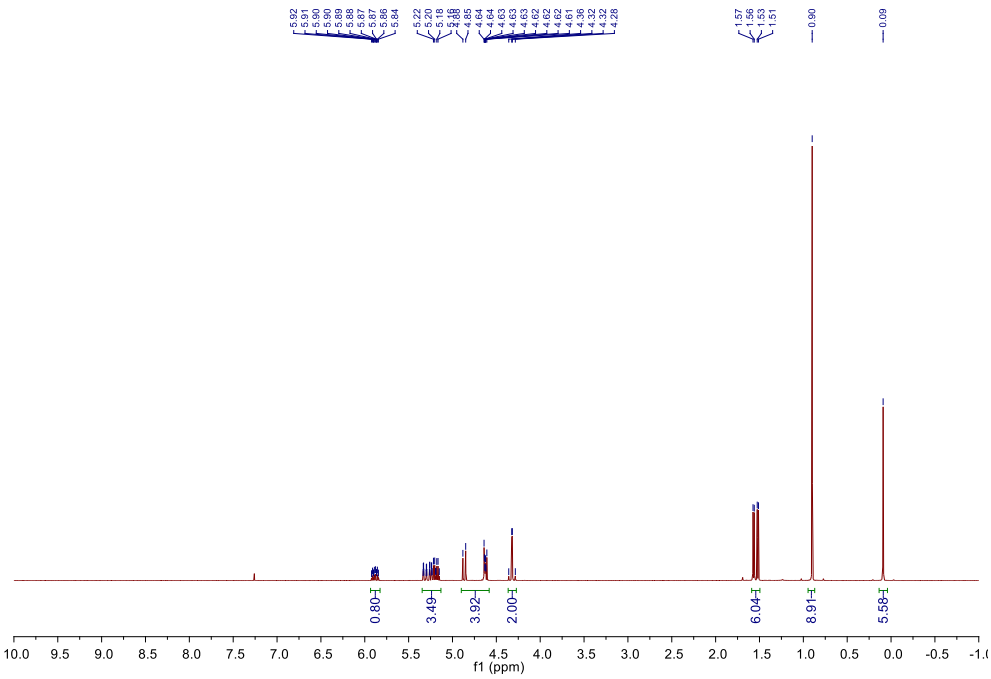


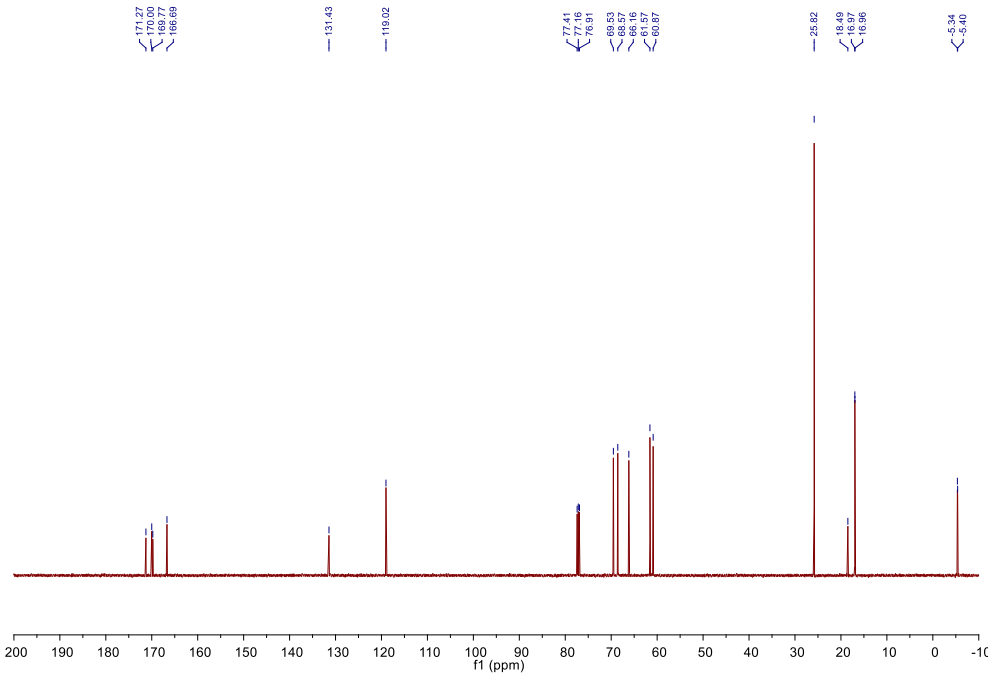
^{1}H and ^{13}C NMR spectra of **GLLG**.



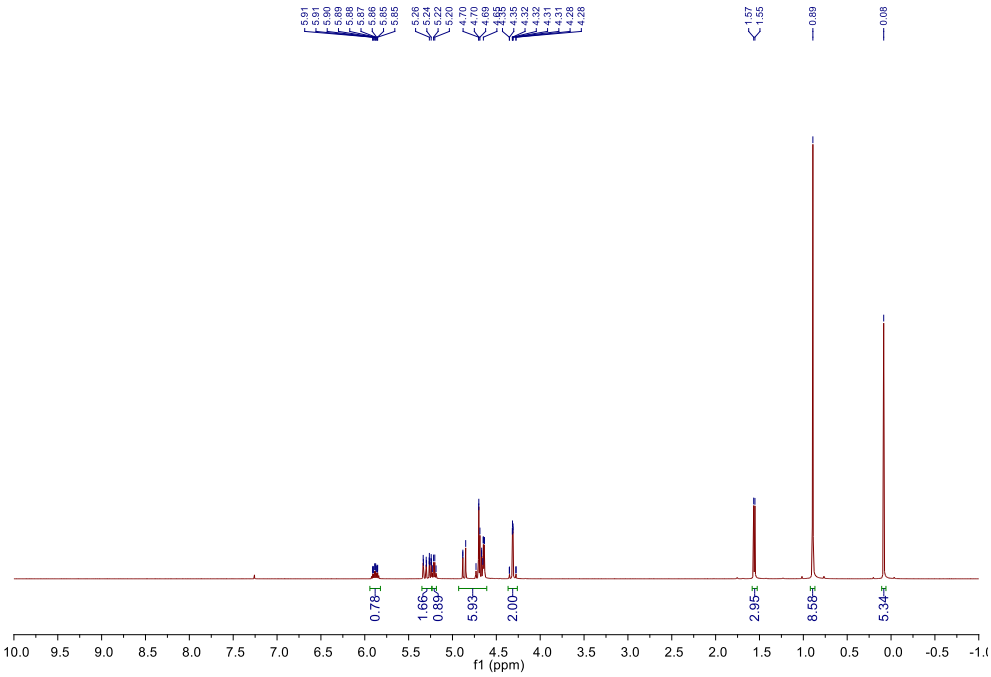


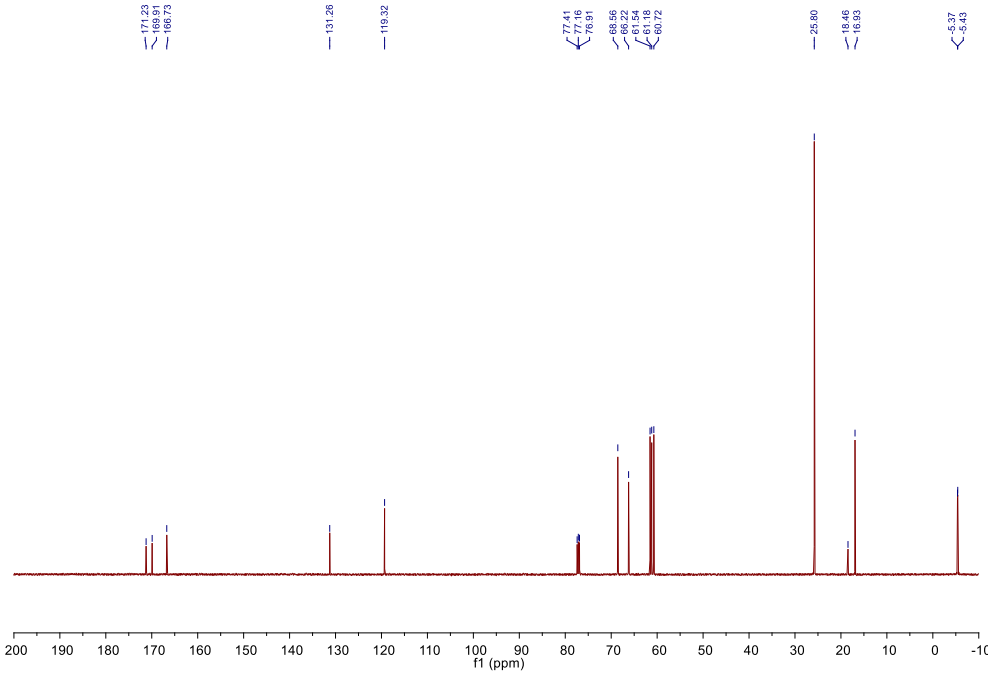
^{1}H and ^{13}C NMR spectra of **GLGL**.



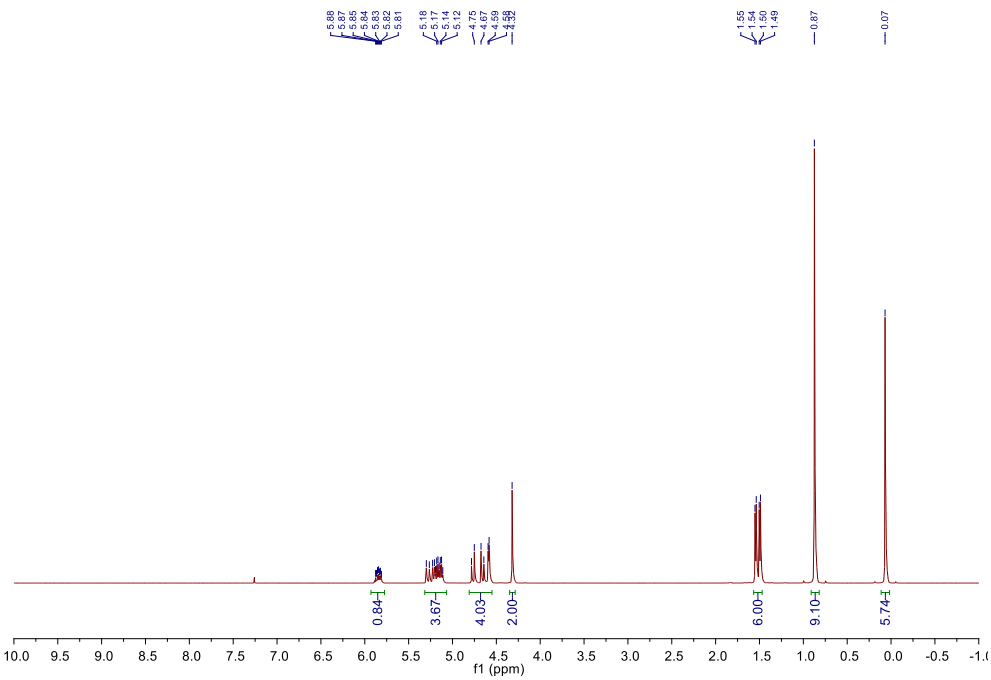


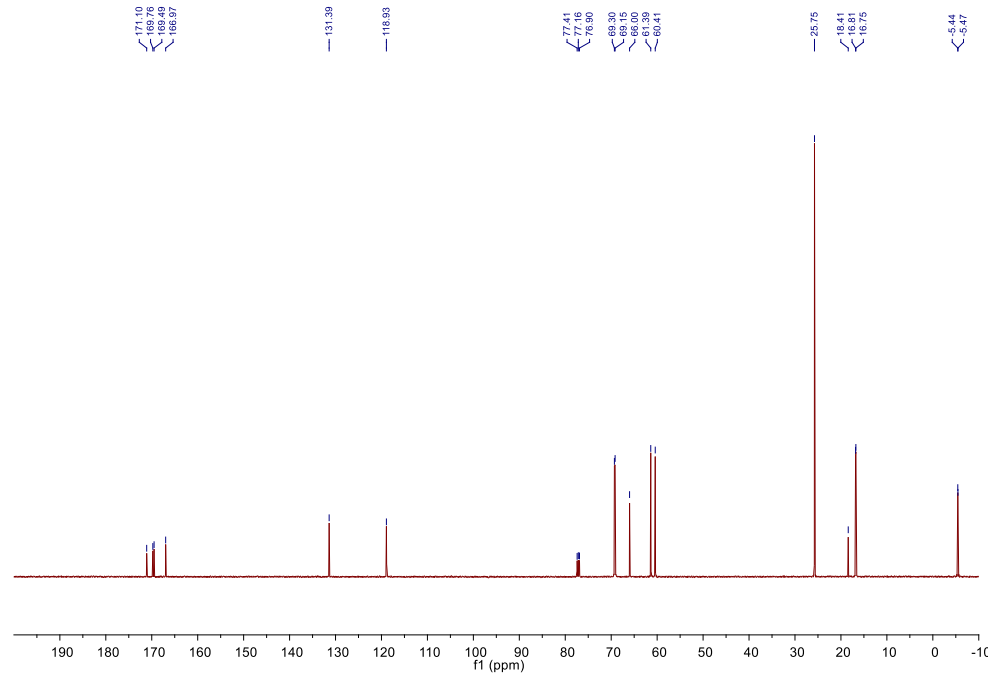
^{1}H and ^{13}C NMR spectra of **GLGG**.



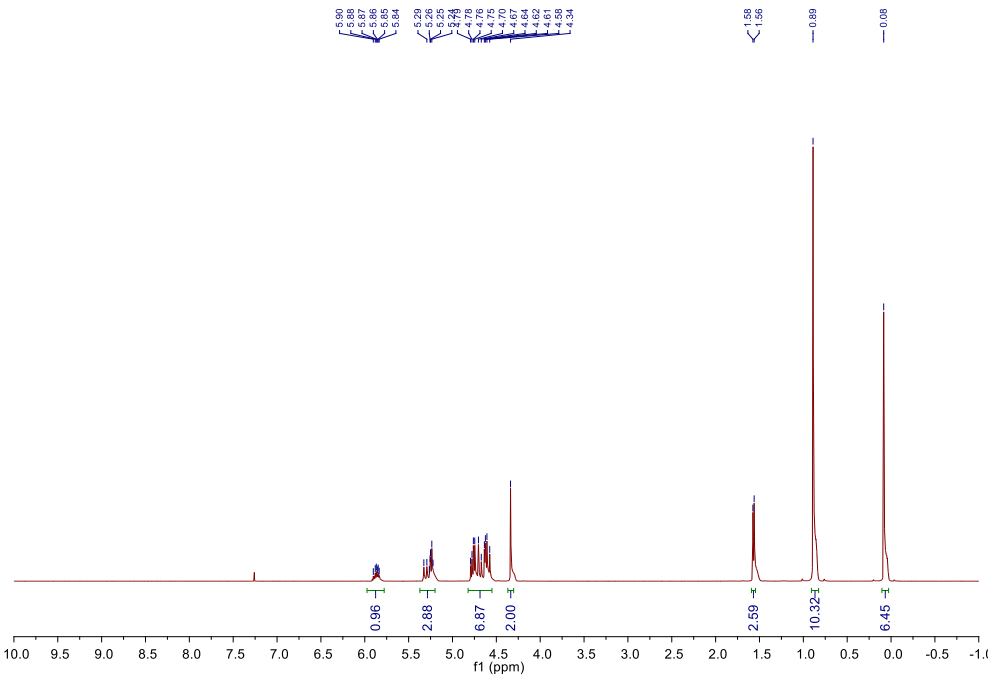


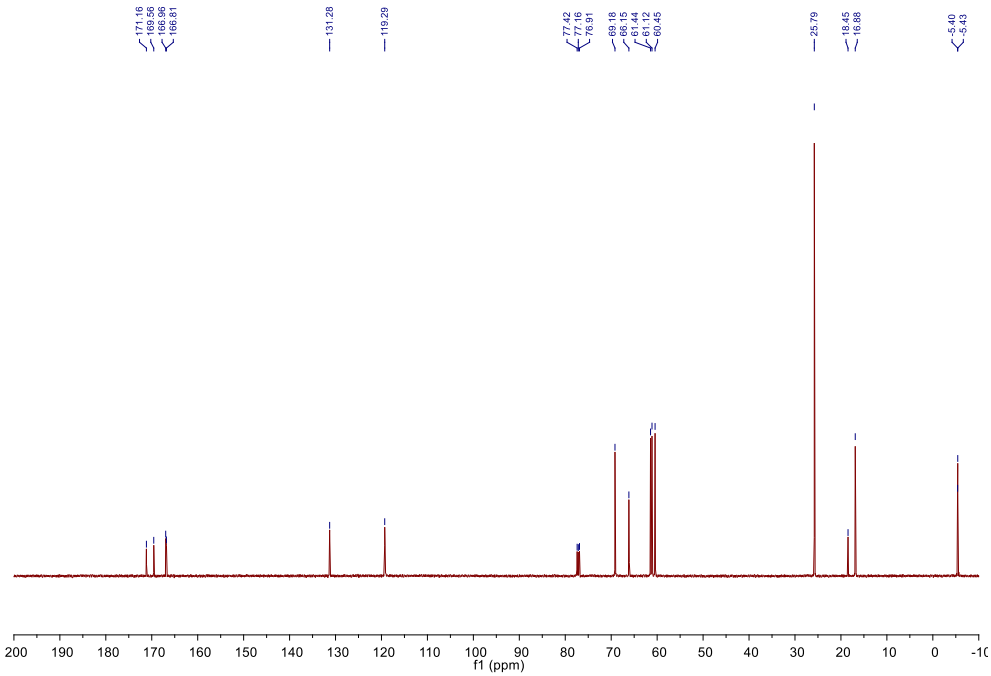
^{1}H and ^{13}C NMR spectra of **GGLL**.



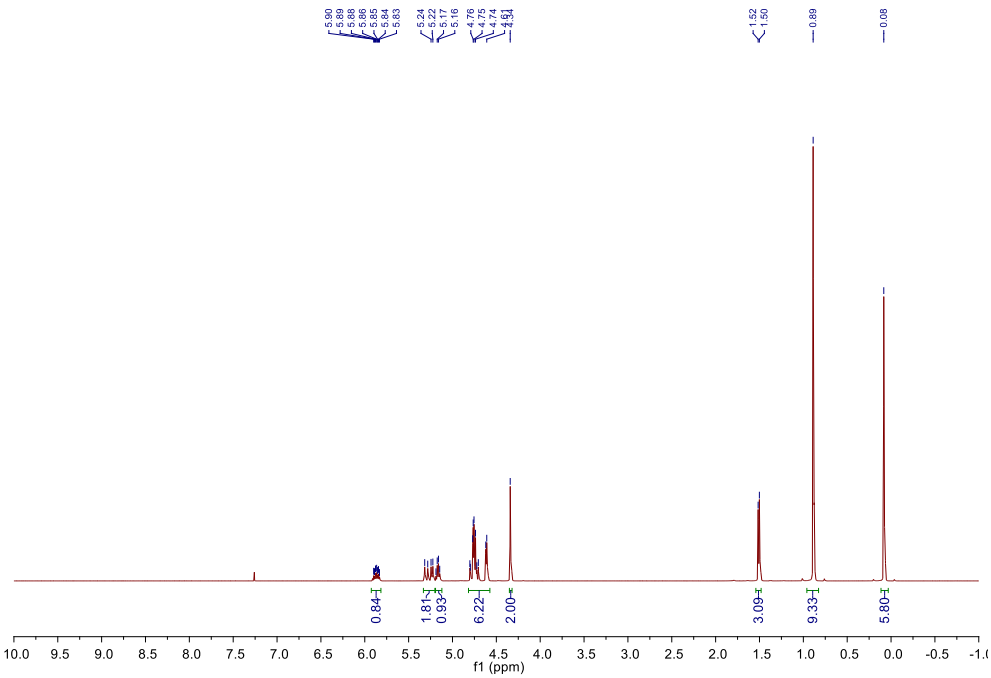


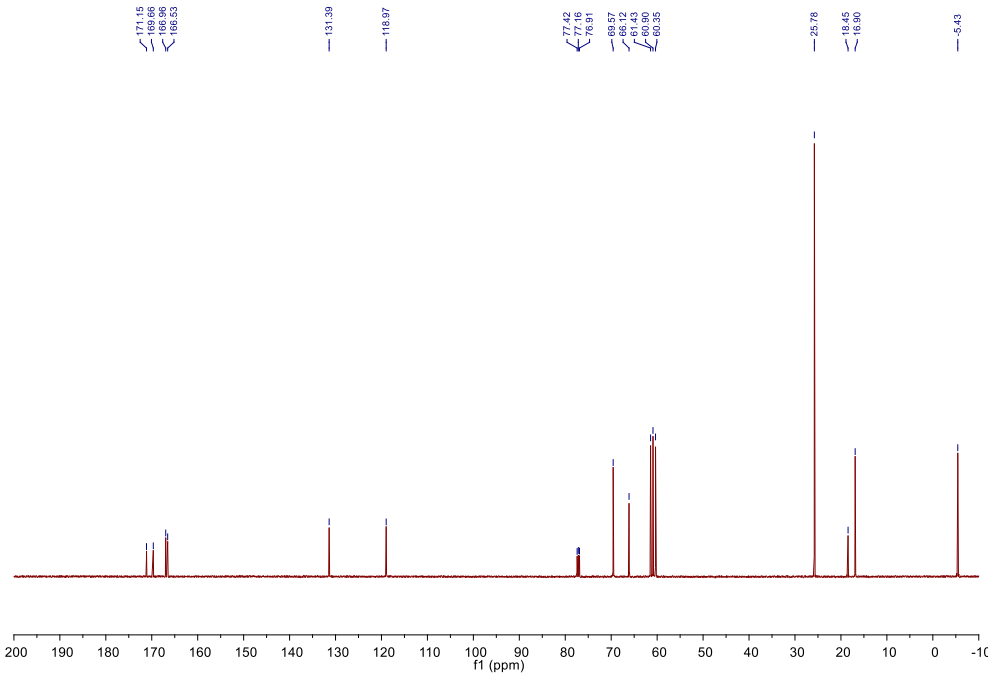
^{1}H and ^{13}C NMR spectra of **GGLG**.



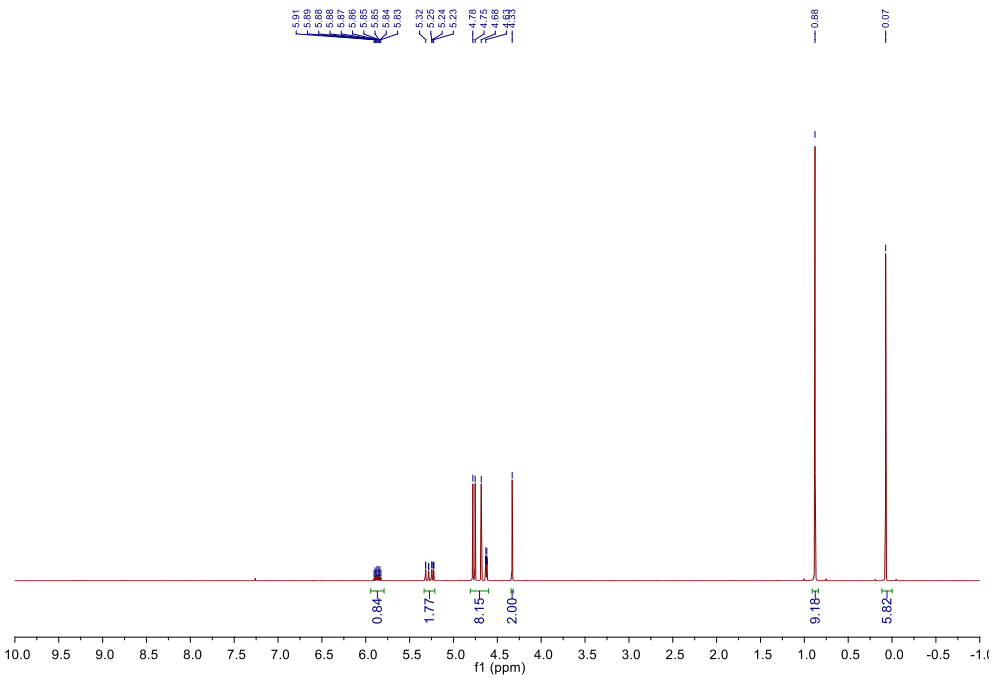


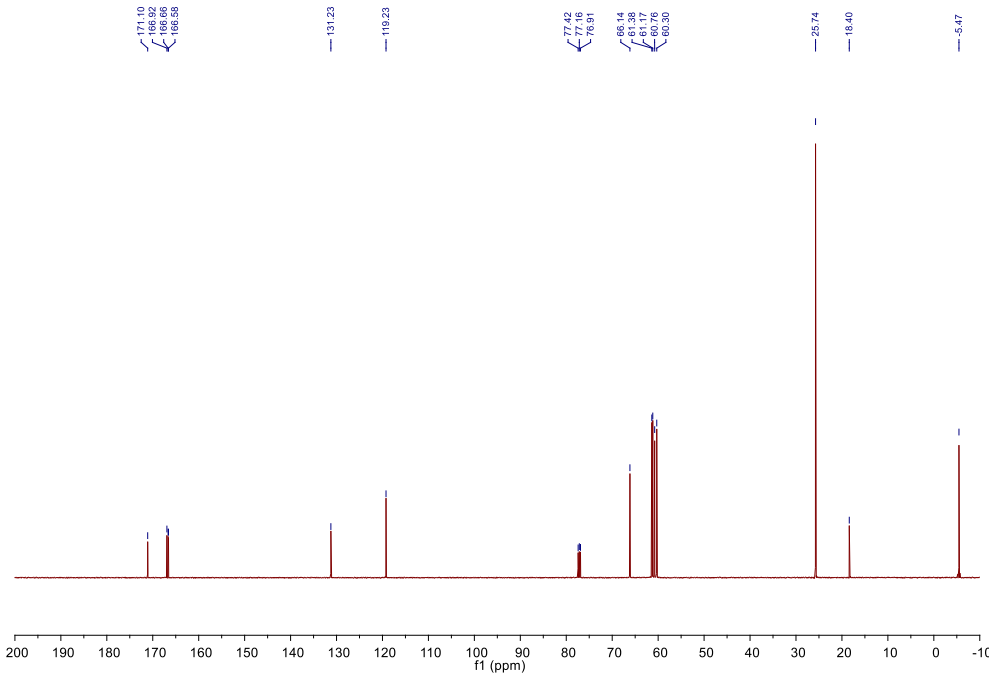
^{1}H and ^{13}C NMR spectra of **GGGL**.





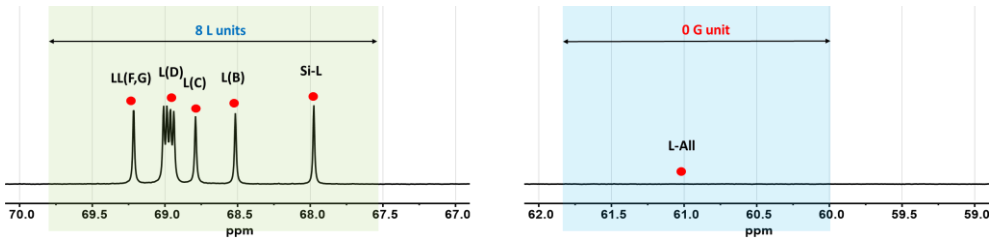
^{1}H and ^{13}C NMR spectra of **GGGG**.



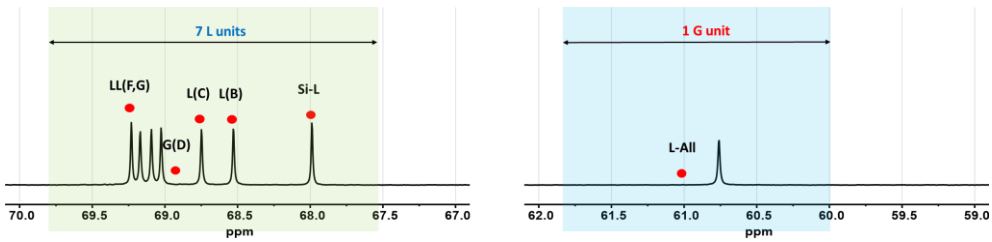


A.4.2 ^{13}C NMR sequencing results of octameric sequence-defined oLGs

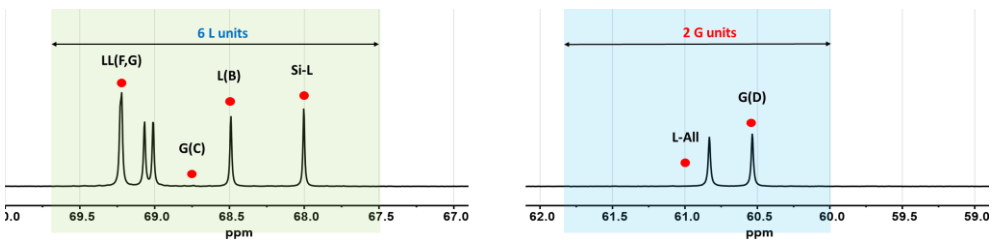
Decoding from ^{13}C NMR spectrum of Si-LLLLLLLL-All.



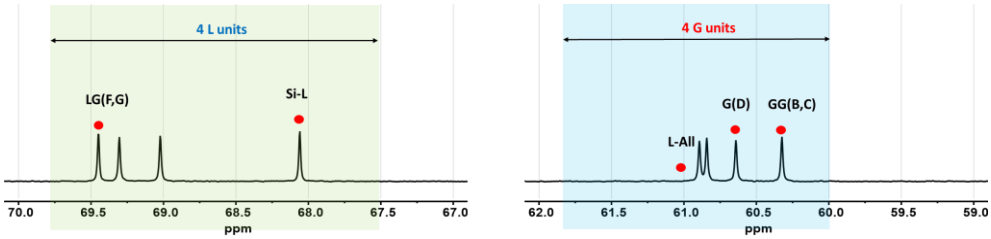
Decoding from ^{13}C NMR spectrum of Si-LLLGLLLL-All.



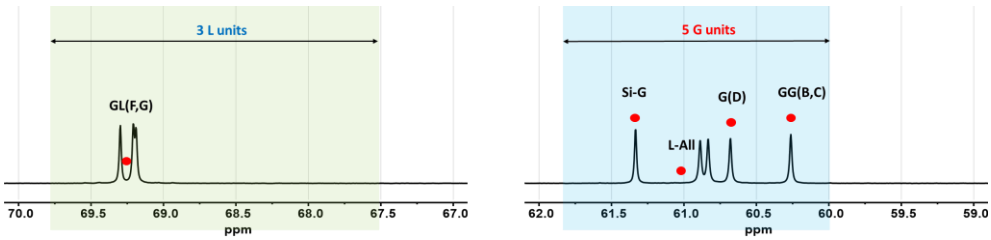
Decoding from ^{13}C NMR spectrum of Si-LLGGLLLL-All.



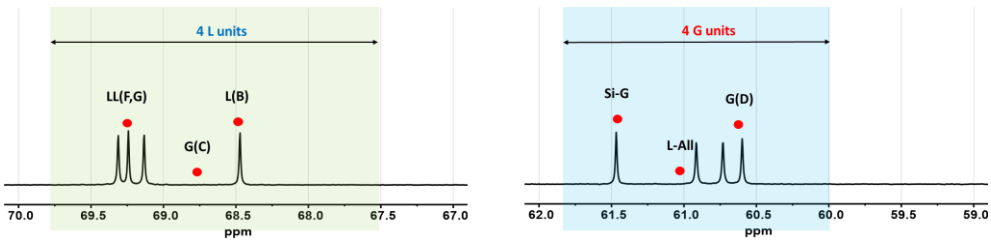
Decoding from ^{13}C NMR spectrum of Si-LGGGLLGL-All.



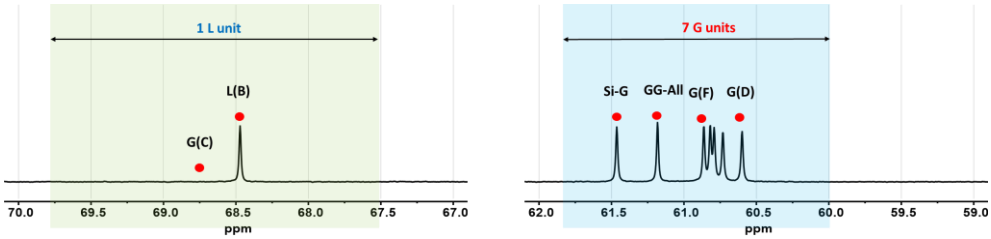
Decoding from ^{13}C NMR spectrum of Si-GGGGLGLL-All.



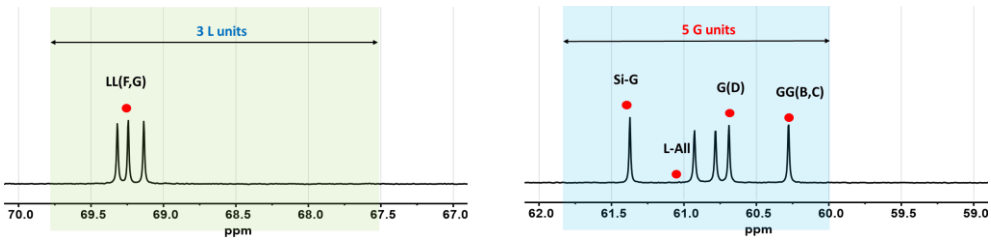
Decoding from ^{13}C NMR spectrum of Si-GLGGGLLL-All.



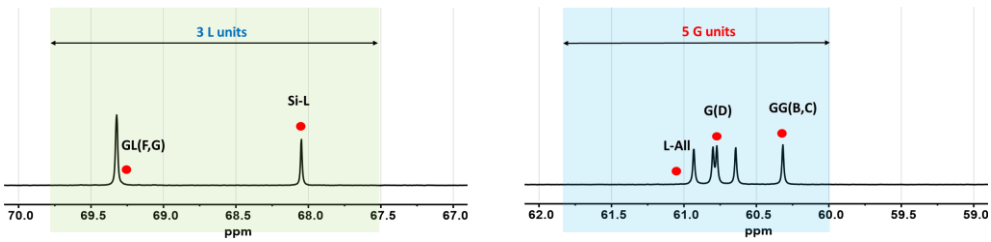
Decoding from ^{13}C NMR spectrum of Si-GLGGGGGG-All.



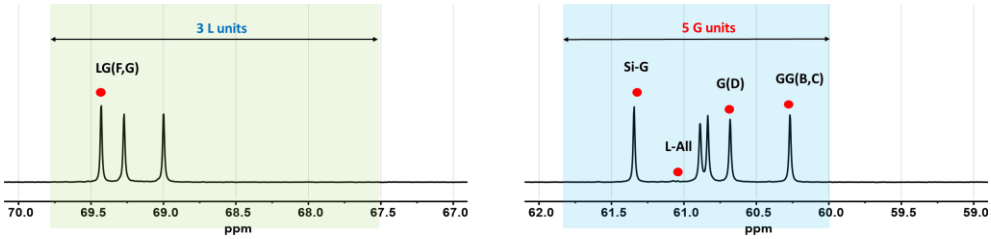
Decoding from ^{13}C NMR spectrum of Si-GGGGGLLL-All.



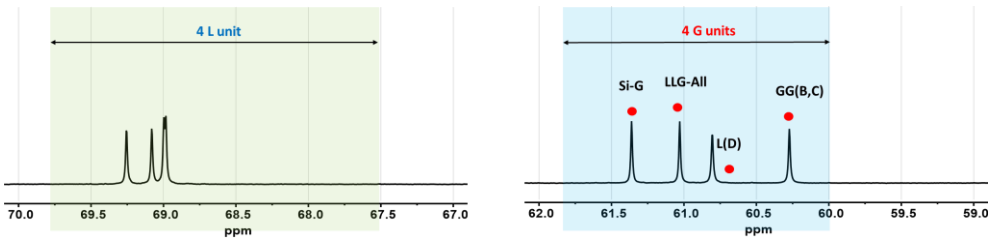
Decoding from ^{13}C NMR spectrum of Si-LGGGGGLL-All.



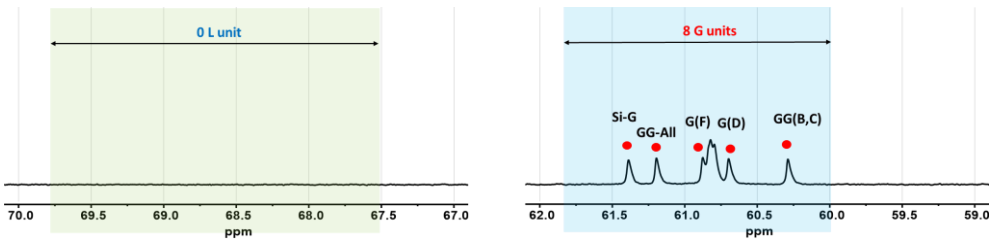
Decoding from ^{13}C NMR spectrum of Si-GGGGLLGL-All.



Decoding from ^{13}C NMR spectrum of Si-GGGLLLLG-All.



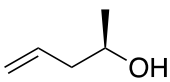
Decoding from ^{13}C NMR spectrum of Si-GGGGGGGG-All.



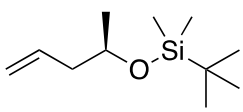
Chapter 5: Synthesis of Enantiomeric ω-Substituted Hydroxyalkanoates from Terminal Epoxides: Building Blocks for Sequence-defined polyesters and Macromolecular Engineering

A.5.1 Characterization of HAs

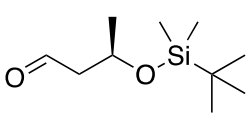
Synthesis of **1**



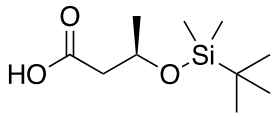
(*R*)-pent-4-en-2-ol. Colorless oil (6.70 g, 63.8%) ^{1}H NMR (500MHz, $CDCl_3$): δ 5.87 – 5.76 (m, 1H), 5.13 (dd, *J* = 8.3, 7.5 Hz, 2H), 3.90 – 3.79 (m, 1H), 2.30 – 2.12 (m, 2H), 1.73 (s, 1H), 1.20 (d, *J* = 6.2 Hz, 3H).



(*R*)-*tert*-butyldimethyl(pent-4-en-2-yloxy)silane. Colorless oil (13.6 g, 87.0%) ^{1}H NMR (500MHz, $CDCl_3$): δ 5.81 (ddt, J = 17.4, 10.3, 7.2 Hz, 1H), 5.09 – 4.98 (m, 2H), 3.84 (h, *J* = 6.1 Hz, 1H), 2.18 (pd, *J* = 13.6, 6.6 Hz, 2H), 1.13 (d, J = 6.1 Hz, 3H), 0.89 (d, *J* = 2.8 Hz, 9H), 0.05 (s, 6H).

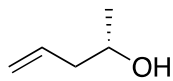


(*R*)-3-((*tert*-butyldimethylsilyl)oxy)butanal. Colorless oil (10.7 g, 78.0%) ^{1}H NMR (500MHz, $CDCl_3$): δ 9.79 (s, 1H), 4.42 – 4.29 (m, 1H), 2.62 – 2.40 (m, 2H), 1.23 (d, *J* = 6.2 Hz, 3H), 0.88 (d, *J* = 17.0 Hz, 9H), 0.07 (d, *J* = 7.7 Hz, 6H).

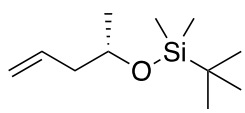


(*R*)-3-((*tert*-butyldimethylsilyl)oxy)butanoic acid (1). Colorless oil (10.1 g, 86.8%) ^{1}H NMR (500MHz, $CDCl_3$): δ 10.71 (br, 1H), 4.33 – 4.23 (m, 1H), 2.49 (dd, *J* = 5.8, 2.1 Hz, 2H), 1.24 (dd, *J* = 6.1, 2.1 Hz, 3H), 0.90 – 0.86 (m, 9H), 0.08 (dd, *J* = 6.9, 2.2 Hz, 6H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 177.83, 65.79, 44.62, 25.96, 25.91, 25.84, 23.87, 18.08, -4.41, -4.97. HRMS (ESI): calcd for $C_{10}H_{22}O_3Si$-H^- [**M**-H]$^-$: 217.1260 Da; found: 217.1267 Da.

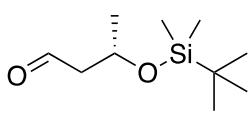
Synthesis of **2**



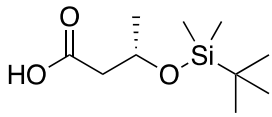
(*S*)-pent-4-en-2-ol.. Colorless oil (6.42 g, 61.1%) ^{1}H NMR (500MHz, $CDCl_3$): δ 5.87 – 5.76 (m, 1H), 5.13 (dd, J = 8.3, 7.5 Hz, 2H), 3.90 – 3.79 (m, 1H), 2.30 – 2.12 (m, 2H), 1.73 (s, 1H), 1.20 (d, J = 6.2 Hz, 3H).



(*S*)-*tert*-butyldimethyl(pent-4-en-2-yloxy)silane. Colorless oil (12.8 g, 85.9 %) ^{1}H NMR (500MHz, $CDCl_3$): δ 5.81 (ddt, *J* = 17.4, 10.3, 7.2 Hz, 1H), 5.09 – 4.98 (m, 2H), 3.84 (h, *J* = 6.1 Hz, 1H), 2.18 (pd, *J* = 13.6, 6.6 Hz, 2H), 1.13 (d, *J* = 6.1 Hz, 3H), 0.89 (d, *J* = 2.8 Hz, 9H), 0.05 (s, 6H).

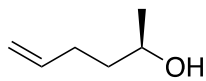


(*S*)-3-((*tert*-butyldimethylsilyl)oxy)butanal. Colorless oil (10.3 g, 80.3%) ^{1}H NMR (500MHz, $CDCl_3$): δ 9.79 (s, 1H), 4.42 – 4.29 (m, 1H), 2.62 – 2.40 (m, 2H), 1.23 (d, *J* = 6.2 Hz, 3H), 0.88 (d, *J* = 17.0 Hz, 9H), 0.07 (d, *J* = 7.7 Hz, 6H).

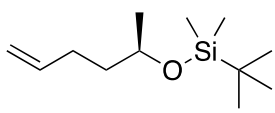


(*S*)-3-((*tert*-butyldimethylsilyl)oxy)butanoic acid (2). Colorless oil (11.1 g, 92.2%) ^{1}H NMR (500MHz, $CDCl_3$): δ 10.93 (br, 1H), 4.33 – 4.23 (m, 1H), 2.49 (dd, *J* = 5.8, 2.1 Hz, 2H), 1.24 (dd, *J* = 6.1, 2.1 Hz, 3H), 0.90 – 0.86 (m, 9H), 0.08 (dd, J = 6.9, 2.2 Hz, 6H). ^{13}C NMR (126 MHz, $cdcl_3$) δ 177.95, 65.79, 44.67, 26.02, 25.96, 25.92, 25.84, 25.72, 23.89, 18.08, -4.42, -4.96. HRMS (ESI): calcd for $C_{10}H_{22}O_3Si$-H^- [**M**-H]$^-$: 217.1260 Da; found: 217.1264 Da.

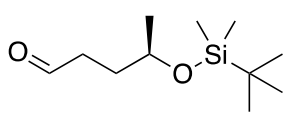
Synthesis of **3**



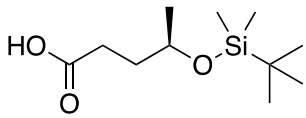
(*R*)-hex-5-en-2-ol. Colorless oil (8.20 g, 55.8%) ^{1}H NMR (500MHz, $CDCl_3$): δ 5.82 (ddt, *J* = 16.9, 10.2, 6.7 Hz, 1H), 5.08 – 4.92 (m, 2H), 3.87 – 3.75 (m, 1H), 2.23 – 2.04 (m, 2H), 1.76 – 1.68 (m, 1H), 1.60 – 1.44 (m, 2H), 1.19 (t, *J* = 6.6 Hz, 3H).



(*R*)-*tert*-butyl(hex-5-en-2-yloxy)dimethylsilane. Colorless oil (15.1 g, 86.0%) ^{1}H NMR (500MHz, $CDCl_3$): δ 5.82 (ddt, *J* = 16.9, 10.2, 6.6 Hz, 1H), 5.07 – 4.89 (m, 2H), 3.86 – 3.74 (m, 1H), 2.22 – 1.97 (m, 2H), 1.62 – 1.40 (m, 2H), 1.13 (d, *J* = 6.1 Hz, 3H), 0.96 – 0.84 (m, 9H), 0.03 (d, *J* = 13.5 Hz, 6H).

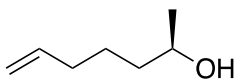


(*R*)-4-((*tert*-butyldimethylsilyl)oxy)pentanal. Colorless oil (12.3 g, 80.7%) ^{1}H NMR (500MHz, $CDCl_3$): δ 9.76 (d, J = 1.0 Hz, 1H), 3.85 (dq, *J* = 12.0, 6.1 Hz, 1H), 2.57 – 2.41 (m, 2H), 1.86 – 1.58 (m, 2H), 1.13 (d, *J* = 6.1 Hz, 3H), 0.86 (s, 9H), 0.03 (d, *J* = 6.3 Hz, 6H).

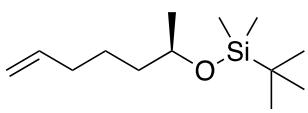


(*R*)-4-((*tert*-butyldimethylsilyl)oxy)pentanoic acid (3). Colorless oil (11.5 g, 87.2%) ^{1}H NMR (500MHz, $CDCl_3$): δ 11.23 (s, 1H), 3.96 – 3.79 (m, 2H), 2.60 – 2.32 (m, 2H), 1.88 – 1.61 (m, 2H), 1.15 (d, *J* = 6.1 Hz, 3H), 0.88 (s, 9H), 0.05 (d, *J* = 2.3 Hz, 6H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 179.51, 67.61, 34.10, 30.29, 25.97, 25.83, 23.73, 18.19, -4.26, -4.71. HRMS (ESI): calcd for $C_{11}H_{24}O_3Si$-H^- [**M**-H]$^-$: 231.1417 Da; found: 231.1422 Da.

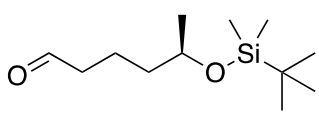
Synthesis of **5**



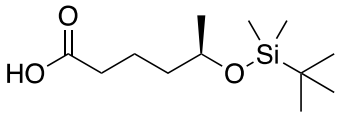
(*R*)-hept-6-en-2-ol. Colorless oil (9.60 g, 69.0%) ^{1}H NMR (500MHz, $CDCl_3$): δ 5.81 (tt, *J* = 9.8, 6.7 Hz, 1H), 4.98 (dd, J = 28.5, 13.7 Hz, 2H), 3.81 (dd, *J* = 11.4, 5.6 Hz, 1H), 2.06 (t, *J* = 10.3 Hz, 2H), 1.56 – 1.33 (m, 5H), 1.19 (d, *J* = 6.1 Hz, 3H).



(*R*)-*tert*-butyl(hept-6-en-2-yloxy)dimethylsilane. Colorless oil (17.6 g, 91.7%) ^{1}H NMR (500MHz, $CDCl_3$): δ 5.87 – 5.74 (m, 1H), 5.04 – 4.90 (m, 2H), 3.83 – 3.74 (m, 1H), 2.09 – 2.00 (m, 2H), 1.54 – 1.29 (m, 4H), 1.16 – 1.08 (m, 3H), 0.93 – 0.85 (m, 9H), 0.08 – 0.01 (m, 6H).

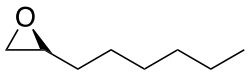


(*R*)-5-((*tert*-butyldimethylsilyl)oxy)hexanal. Colorless oil (17.6 g, 88.7%) ^{1}H NMR (500MHz, $CDCl_3$): δ 9.74 (t, *J* = 1.8 Hz, 1H), 3.79 (dq, *J* = 12.1, 6.1 Hz, 1H), 2.45 – 2.38 (m, 2H), 1.78 – 1.54 (m, 2H), 1.49 – 1.35 (m, 2H), 1.15 – 1.08 (m, 3H), 0.89 – 0.83 (m, 9H), 0.06 – 0.00 (m, 6H).

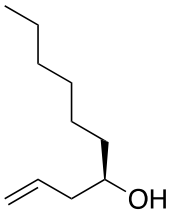


(*R*)-5-((*tert*-butyldimethylsilyl)oxy)hexanoic acid (5). Colorless oil (17.6 g, 93.7%) ^{1}H NMR (500MHz, $CDCl_3$): δ 3.86 – 3.75 (m, 1H), 2.41 – 2.30 (m, 2H), 1.78 – 1.56 (m, 2H), 1.53 – 1.37 (m, 2H), 1.18 – 1.09 (m, 3H), 0.92 – 0.84 (m, 9H), 0.08 – 0.01 (m, 6H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 180.46, 68.31, 38.96, 34.26, 26.02, 25.97, 23.86, 21.07, 18.25, -4.27, -4.63. HRMS (ESI): calcd for $C_{12}H_{26}O_3Si$-H^- [**M**-H]$^-$: 245.1573 Da; found: 245.1579 Da.

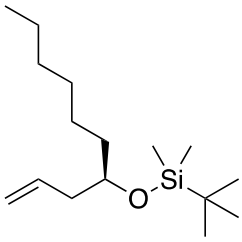
Synthesis of **6**



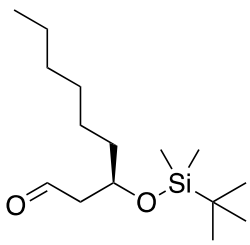
(*R*)-2-Hexyloxirane. Colorless oil (18.5 g, 44.0%) ^{1}H NMR (500MHz, $CDCl_3$): δ 2.90 (tt, *J* = 5.8, 3.1 Hz, 1H), 2.74 (t, *J* = 4.5 Hz, 1H), 2.46 (dd, *J* = 5.1, 2.7 Hz, 1H), 1.70 – 1.19 (m, 10H), 0.89 (t, *J* = 6.7 Hz, 3H).



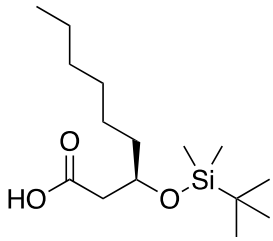
(*R*)-Dec-1-en-4-ol. Colorless oil (12.0 g, 70.2%) ^{1}H NMR (500MHz, $CDCl_3$): δ 5.90 – 5.76 (m, 1H), 5.19 – 5.08 (m, 2H), 3.69 – 3.59 (m, 1H), 2.36 – 2.07 (m, 2H), 1.57 (s, 1H), 1.51 – 1.39 (m, 3H), 1.33 (d, *J* = 32.6 Hz, 7H), 0.88 (t, *J* = 6.6 Hz, 3H).



(*R*)-*tert*-Butyl(dec-1-en-4-yloxy)dimethylsilane. Colorless oil (19.12 g, 86.2%) ^{1}H NMR (500MHz, $CDCl_3$): δ 5.82 (ddt, *J* = 17.5, 10.4, 7.2 Hz, 1H), 5.03 (dd, *J* = 9.0, 8.0 Hz, 2H), 3.68 (p, *J* = 5.7 Hz, 1H), 2.27 – 2.12 (m, 2H), 1.46 – 1.20 (m, 10H), 0.88 (d, *J* = 9.0 Hz, 12H), 0.04 (s, 6H).



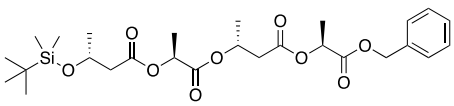
(*R*)-3-((*tert*-butyldimethylsilyl)oxy)nonanal. Colorless oil (13.1 g, 67.8%) ^{1}H NMR (500MHz, $CDCl_3$): δ 9.81 (t, *J* = 2.5 Hz, 1H), 4.17 (p, *J* = 5.9 Hz, 1H), 2.51 (dd, *J* = 5.6, 2.5 Hz, 2H), 1.56 – 1.45 (m, 2H), 1.36 – 1.21 (m, 8H), 0.92 – 0.84 (m, 12H), 0.08 – 0.03 (m, 6H).



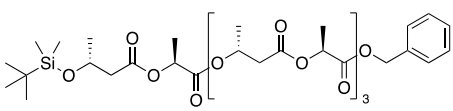
(*R*)-3-((*tert*-butyldimethylsilyl)oxy)nonanoic acid (6). Colorless oil (13.6 g, 98.0%) ^{1}H NMR (500MHz, $CDCl_3$): δ 10.57(br, 1H), 4.08 (dt, *J* = 13.8, 6.8 Hz, 2H), 2.51 (qd, *J* = 15.2, 5.4 Hz, 4H), 1.61 – 1.46 (m, 4H), 1.28 (s, 16H), 0.88 (d, *J* = 10.4 Hz, 23H), 0.09 (d, *J* = 5.9 Hz, 11H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 177.56, 69.54, 42.30, 37.51, 31.91, 29.43, 25.89, 25.18, 22.72, 18.11, 14.21, -4.40, -4.74. HRMS (ESI): calcd for $C_{15}H_{32}O_3Si$-H^- [**M**-H]$^-$: 287.2043 Da; found: 287.2049 Da.

Synthesis of (**1·L**)$_n$

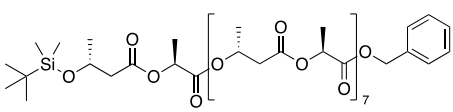
1·L. Colorless oil (8.8 g, 92.1%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.39 – 7.31 (m, 5H), 5.23 – 5.05 (m, 3H), 4.33 – 4.22 (m, 1H), 2.50 (ddd, *J* = 20.5, 14.8, 6.5 Hz, 2H), 1.49 (d, *J* = 7.1 Hz, 3H), 1.19 (d, *J* = 6.1 Hz, 3H), 0.85 (s, 9H), 0.04 (d, *J* = 12.6 Hz, 6H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 171.23, 170.81, 135.47, 128.73, 128.54, 128.29, 77.41, 77.16, 76.91, 68.65, 67.11, 65.91, 44.54, 25.88, 23.99, 18.10, 17.08, -4.37, -4.88.



(1·L)$_2$. Colorless oil (5.6 g, 91.4%); ^{1}H NMR (500MHz, $CDCl_3$): δ 7.39 – 7.30 (m, 4H), 5.30 (h, *J* = 6.4 Hz, 1H), 5.21 – 5.11 (m, 3H), 5.01 (q, *J* = 7.1 Hz, 1H), 4.30 – 4.23 (m, 1H), 2.79 – 2.40 (m, 4H), 1.50 (d, *J* = 7.1 Hz, 3H), 1.45 (d, *J* = 7.1 Hz, 3H), 1.29 (d, *J* = 6.3 Hz, 3H), 1.21 (d, *J* = 6.1 Hz, 3H), 0.87 – 0.83 (m, 9H), 0.07 – 0.02 (m, 6H).

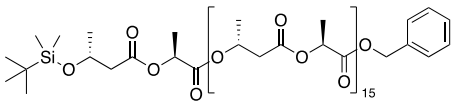


(1·L)$_4$. Colorless oil (4.1 g, 92.4%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41 – 7.29 (m, 5H), 5.35 – 5.24 (m, 3H), 5.22 – 5.09 (m, 3H), 5.08 – 4.98 (m, 3H), 4.31 – 4.21 (m, 1H), 2.80 – 2.39 (m, 8H), 1.51 – 1.43 (m, 12H), 1.35 – 1.19 (m, 12H), 0.85 (s, 9H), 0.04 (d, *J* = 13.6 Hz, 6H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 170.07, 169.77, 169.47, 169.39, 169.35, 135.42, 128.75, 128.57, 128.31, 77.41, 77.16, 76.91, 69.03, 68.70, 68.55, 68.53, 68.46, 67.19, 65.91, 44.52, 40.27, 25.88, 24.01, 19.75, 19.71, 18.10, 16.99, 16.95, 16.89, 16.87, -4.37, -4.87.

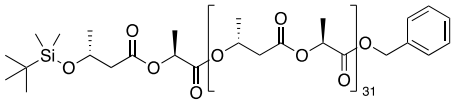


(1·L)$_8$. Colorless oil (3.1 g, 88.5%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 – 7.27 (m, 5H), 5.29 (h, *J* = 6.3 Hz, 7H), 5.22 – 5.10 (m, 3H), 5.07 – 4.96 (m, 7H), 4.31 – 4.20 (m, 1H), 2.58 (dddd, *J* = 63.6, 20.5, 15.3, 6.2 Hz, 16H), 1.47 (dt, *J* = 12.0, 6.1 Hz, 24H), 1.35 – 1.20 (m, 24H), 0.85 (s, 9H), 0.04 (d, *J* = 13.6 Hz, 6H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 171.12, 170.47, 170.06, 169.75, 169.45, 169.35, 135.41, 128.73, 128.55, 128.30, 77.41, 77.16, 76.91, 69.06, 68.98, 68.56, 68.51, 67.17, 65.89, 44.50, 40.25, 40.22, 25.87, 24.00, 19.73, 18.09, 16.86, -4.38, -4.88; M_n and *Đ* (GPC): 1660 Da and 1.01; MS (MALDI-TOF): calcd for $C_{69}H_{102}O_{33}Si+Na^+$

[**M**+Na]$^+$: 1509.60 Da; found: 1509.79 Da.



(1·L)$_{16}$. Colorless oil (2.4 g, 86.1%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41 – 7.29 (m, 5H), 5.29 (h, J = 6.4 Hz, 15H), 5.21 – 5.11 (m, 3H), 5.07 – 4.97 (m, 15H), 4.32 – 4.22 (m, 1H), 2.80 – 2.38 (m, 32H), 1.51 – 1.43 (m, 48H), 1.35 – 1.19 (m, 48H), 0.85 (s, 9H), 0.04 (d, J = 13.7 Hz, 6H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 169.76, 169.37, 128.74, 128.56, 128.31, 77.41, 77.16, 76.91, 69.03, 68.69, 68.54, 68.45, 67.18, 65.90, 44.50, 40.25, 40.22, 25.88, 24.01, 19.75, 19.71, 18.10, 16.99, 16.95, 16.88, -4.37, -4.88; M_n and $Đ$ (GPC): 3210 Da and 1.03; MS (MALDI-TOF): calcd for $C_{125}H_{182}O_{65}Si$+Na^+ [**M**+Na]$^+$: 2774.06 Da; found: 2774.58 Da.



(1·L)$_{32}$. Colorless oil (1.8 g, 80.7%); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 – 7.30 (m, 5H), 5.29 (h, J = 6.4 Hz, 31H), 5.22 – 5.11 (m, 3H), 5.02 (dq, J = 14.2, 7.1 Hz, 31H), 4.31 – 4.22 (m, 1H), 2.79 – 2.39 (m, 64H), 1.52 – 1.43 (m, 96H), 1.35 – 1.19 (m, 96H), 0.85 (s, 9H), 0.04 (d, J = 13.8 Hz, 6H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 169.76, 169.37, 128.74, 128.56, 128.31, 77.41, 77.16, 76.91, 69.03, 68.69, 68.54, 68.45, 67.18, 65.90, 44.50, 40.22, 25.88, 24.01, 19.75, 16.99, 16.95, 16.88, -4.37, -4.88; M_n and $Đ$ (GPC): 7100 Da and 1.03; MS (MALDI-TOF): calcd for $C_{237}H_{342}O_{129}Si$+Na^+ [**M**+Na]$^+$: 5302.99 Da; found: 5302.81 Da.

국문초록

분자량이 단일하고 서열이 정의된 폴리에스터의 합성과 정보 저장 매체로의 활용

이정민
고분자화학전공
화학부
서울대학교

분자량, 단량체 서열 및 입체배열이 완벽하게 통제된 서열 특이적 고분자 합성은 고분자 화학 분야의 중요한 과제이다. 서열이 정의된 고분자는 폴대머, 촉매 및 항균 물질 등 광범위한 영역에 활용될 잠재성을 가지고 있다. 특히, 이는 단량체 서열을 디지털 정보로 변환함으로써 정보 저장 매체로 사용될 수 있다. 단계적 반복 합성법과 반복 지수 성장법 등 균일한 거대분자를 만들기 위한 여러가지 전략들이 개발되었지만 규모 확장성, 고분자 길이 또는 반복되는 서열과 같은 한계가 있었다. 필자는 교차 수렴법을 통하여 서열이 정의된 폴리에스터를 합성하고 생성된 고분자의 비주기적 서열을 이용하여 정보를 저장하는 연구들을 하였다.

본 논문에서는 크기 배제 크로마토그래피 (prep-SEC) 를 이용한 정제 방법과 함께 교차 수렴법을 통하여 큰 분자량을 갖는 서열 특이적 폴리페닐락타이드-락타이드 공중합체 (PcLs) 를 합성하였다. 해당 방법은 최소한의 화학 반응으로 이진법으로 인코딩된 PcL의 확장 가능한 합성을 용이하게 하였다. 64비트 PcL에 저장된 정보는 MALDI-TOF 텐덤 질량

분석기의 단일 측정으로 디코딩할 수 있었다. 또한, 분해성 시퀀싱 방법은 큰 분자량의 128-비트 PcL의 단량체 서열 분석을 가능하게 하였다.

교차 수렴 방법을 통한 서열 특이적 고분자의 합성은 단계적 경제적 합성이지만 대용량 정보 저장을 위해서는 대규모 화학 반응을 수반하는 다수의 서열 특이적 고분자의 합성이 요구된다. 해당 문제는 유동 화학을 이용한 서열 특이적 폴리락타이드-글라이콜라이드 공중합체 (PLGAs) 의 반자동적 합성법 개발을 통하여 해결하였다. 해당 가속화된 합성법은 배치 반응과 비교하여 훨씬 더 적은 시간 내에 896-비트 크기의 비트맵 이미지를 14개의 PLGA 고분자 사슬에 인코딩할 수 있게 하였다. 또한, 고분자 사슬의 식별자로 작동하는 8-비트의 주소 코드의 도입은 혼합물 상태인 여러 PLGA 고분자 사슬의 텐덤 질량 시퀀싱을 가능하게 하였다.

서열이 정의된 고분자의 서열을 분석하는 방법의 대부분은 텐덤 질량 분석과 자기희생 시퀀싱 등의 파괴적 시퀀싱 방법인데 이는 필연적으로 분석 때마다 고분자를 소비해야한다. 따라서, 탄소 동위원소 핵자기 공명 분광법 (^{13}C NMR spectroscopy) 를 통하여 순수 거울상 이성질체인 올리고락타이트-글라이콜라이드 공중합체 (oLGs) 와 올리고만델라이드-페닐락타이드 공중합체 (oMPs) 의 서열을 분석하는 비파괴적인 시퀀싱 방법을 개발하였다. oMP와 oLG 혼합물의 서열은 서열을 나타내는 피크의 겹치지 않는 화학적 이동 영역으로 인해 단일 ^{13}C NMR 측정으로 해독할 수 있었다. 유동 화학 합성을 통하여 192-비트의 비트맵을 순수 거울상 이성질체인 oMP와 oLG에 인코딩하였고 12개의 oMP와 oLG 등몰 혼합물을 온전하게 디코딩하였다.

생분해성 폴리하이드록시알카노에이트 (PHAs)는 탄화수소 기반의 플라스틱의 대체제로 많은 관심을 받고 있지만 생물학적·화학적 방법을

통해 생성되는 PHAs의 제한적인 화학 구조로 인해 응용 범위가 제한적이다. 따라서, 필자는 원하는 원자 구성, 입체화학적 형태, 관능기를 지니는 하이드록시알카노에이트 (HAs) 단량체를 구하기 쉬운 말단 에폭사이드 분자 및 말단 알킨 분자로부터 생산하는 합성 과정을 확립하였다. 해당 방법으로 생성된 HAs 단량체는 서열 특이적 PHAs를 생성하는 빌딩 블록으로 사용되어 해당 고분자의 분자량, 단량체 서열, 입체 배열 및 기능적 부분을 완전히 통제할 수 있었다. 추가적으로, 서열 제한적 PHAs의 거대 분자 공학을 통해 고분자의 결정성과 열적 특성을 조절할 수 있었다.

이러한 연구들을 통해 필자는 대량으로 큰 분자량의 서열이 정의된 폴리에스터를 합성하는 방법을 확립하고 유동 화학을 도입하여 효율성을 높일 수 있었다. 서열 특이적 고분자는 정보 저장 매체로서 활용하였고 인코딩과 디코딩 전략을 발전시켰다. 또한, 서열 제한적 고분자를 통한 거대 분자 공학의 가능성을 제시하였다. 필자는 이러한 결과들이 고분자의 구조-기능 상관관계 분석과 같은 고분자 화학의 근본적인 문제들을 해결하고 광범위한 연구 분야에 사용되는 혁신적인 재료 개발에 기여할 것이라 기대한다.

주요어: 서열 특이적 고분자, 교차 수렴법, 연속 흐름 합성, 정보 저장, 시퀀싱, 거대 분자 공학

학번: 2017-23953

감사의 글

박사 과정을 마무리하면서 제가 본 학위 논문을 쓸 수 있게 도움을 주신 모든 분들께 감사 인사 드립니다. 우선 그 누구보다도 저의 지도교수님이신 김경택 교수님께 감사합니다. 약 6년 동안 옆에서 교수님을 보면서 연구를 대하는 태도와 열정을 배울 수 있었고 연구 외적으로도 다양한 것들을 배울 수 있었습니다. 교수님의 지도 덕분에 전공 지식도 그리 뛰어나지 못했던 제가 고분자 화학 분야에 지속적으로 관심을 갖고 그 결과 독립적인 연구원이 될 수 있었다고 생각합니다. 그리고 저의 첫 사수로서 연구하는 방법을 알려주신 라윤주 선배님, 유기 합성을 할 수 있도록 기초부터 지도를 해주신 김지원 선배님께도 감사합니다. 또한, 교차 수렴법을 통한 서열 특이적 고분자 합성 연구를 함께 수행한 이슬우, 이희림, 구모범 연구원; 유동 화학을 이용한 서열 특이적 고분자의 반자동적 합성 연구를 함께 수행한 권준호, 이수정, 장희정 연구원; 핵자기 공명분광법을 통한 비파괴적인 시퀀싱 방법 개발 연구를 함께 수행한 장희정 연구원; 화학 구조가 자유로운 하이드록시알카노에이트 단량체 합성법 개발 연구를 함께 수행한 김도균, 송정은 연구원에게 진심으로 감사합니다. 혼자만의 연구가 아니었기에 책임감을 갖고 열심히 할 수 있었고 수많은 공동 연구원들 덕분에 짧은 학위 기간 동안 다양한 연구를 수행할 수 있었습니다. 저의 박사 수료 심사 및 박사 학위 논문 심사에 참석해주셔서 많은 조언을 해주신 손병혁 교수님, 이연 교수님, 홍종인 교수님, 최태림 교수님, 서명은 교수님께도 감사드립니다.

감사하게도 저는 연구 외적으로도 저희 고분자 합성 실험실 구성원들의 많은 도움을 받았습니다. 우선 학부 후배이자 대학원 동기인 슬우에게 가장

감사합니다. 연구와 관련된 일뿐만 아니라 모든 사적인 부분까지도 마음 편하게 공유할 수 있었고 덕분에 힘이 들 때는 버틸 수 있었으며 즐거운 대학원 생활을 보낼 수 있었습니다. 그리고 저의 부사수인 희정이에게 감사합니다. 아무리 힘든 일이라도 맡은 일은 책임을 지는 자세에 제가 믿고 연구를 포함한 연구실 생활을 할 수 있었습니다. 또한, 저의 일임에도 선한 마음으로 최선을 다해서 저를 도와주신 수정 누나에게도 깊이 감사드립니다. 저와 많은 디스커션을 하고 함께 샌디에고 학회에 다녀온 도균이, 바쁜 시기에도 저를 도와주시고 조언해주신 정은 누나, 영어 교정이 필요할 때마다 기꺼이 도와준 밸린, 졸업하는 과정에서 도움을 주신 지원이 형한테도 감사합니다. 덕분에 학위 과정 동안 좋은 추억들을 쌓을 수 있었습니다. 이외에도 저와 함께 실험실 생활을 하면서 도움을 주신 모든 분들께 감사드립니다(라윤주, 정문곤, 조아라, 정은선, 박한슬, 윤미선, 김준영, 마현지, 권용범, 하성민 선배, 권준호, 손우림, 이희림, 구모범, 정지수, 강준우, 정우열, 박수빈, 이재학).

개인적으로는 제가 어떤 선택을 하든 응원해주시고 금전적으로 힘들 때마다 저의 생활을 지원해주신 부모님께 진심으로 감사드리고 집에 갈 때마다 저를 반겨주는 저의 가족(엄마, 아빠, 동생, 테리)에게 감사합니다. 그리고 5년 반 동안 언제나 곁에 있으면서 아무리 안좋은 일이 있어도 행복한 대학원 생활을 할 수 있게 해준 지원이에게 감사합니다. 또한, 제가 필요할 때마다 이야기 상대가 되어준 종영이에게도 감사합니다. 마지막으로 가끔씩 만나더라도 대학원 생활을 할 힘이 나게 해준 학부 동기 및 후배들, 고등학교 친구들에게 감사합니다. 다시 한번 저의 소중한 인연이 되어주신 모든 분들에게 감사합니다.

2022 년 8 월 5 일
이정민 드림