

Nondestructive Sequencing of Enantiopure Oligoesters by Nuclear Magnetic Resonance Spectroscopy

Jeong Min Lee,[†] Heejeong Jang,[†] Seul Woo Lee, and Kyoung Taek Kim*

Cite This: *JACS Au* 2022, 2, 2108–2118

Read Online

ACCESS |

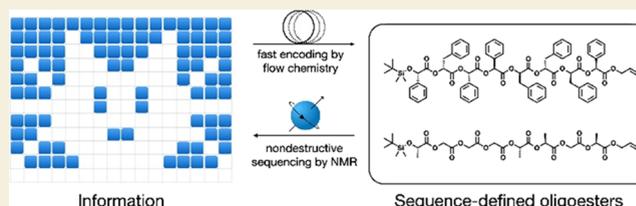
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Sequence-defined synthetic oligomers and polymers are promising molecular media for permanently storing digital information. However, the information decoding process relies on 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-mandelic-*co*-D-phenyl lactic acid)s (oMPs), and oligo(L-lactic-*co*-glycolic acid)s (oLGs) 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 oMP and oLG, synthesized through semiautomated 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.

KEYWORDS: sequence-defined oligoester, nondestructive sequencing, ^{13}C NMR spectroscopy, continuous flow synthesis, information storage



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 biomacromolecules such as DNA has evolved from curiosity-driven research to a promising alternative to existing information storage technologies.^{5–8} DNAs can store digital information in their chemical structures using only a few atoms per bit, the unit of digital information.^{9–11} 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.¹² Compared to DNA, sequence-defined polymers are beneficial thanks to the simpler chemical structures of monomers, ease of synthesis, and enhanced stability in various conditions.^{13,14} Therefore, sequence-defined oligomers and polymers have attracted recent interest as information storage media.^{15–18}

However, realization of these 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, encoding information by chemical synthesis imposes significant cost and time constraints for large-scale information storage. To overcome this challenge, parallel synthesis of sequence-defined macromolecules by automated and continuous processes is necessary to accelerate the rate of chemical encoding.^{22–26}

Decoding the information stored in sequence-defined macromolecules also presents a challenge. Most of the current methods for decoding the information stored in sequence-defined polymers 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 polymers must be developed for macromolecular media to become practically suitable for information storage.^{32–37} Recently, nanopore sequencing of

Received: July 8, 2022

Revised: August 4, 2022

Accepted: August 4, 2022

Published: August 15, 2022



Scheme 1. Synthesis of Sequence-Defined Oligoesters

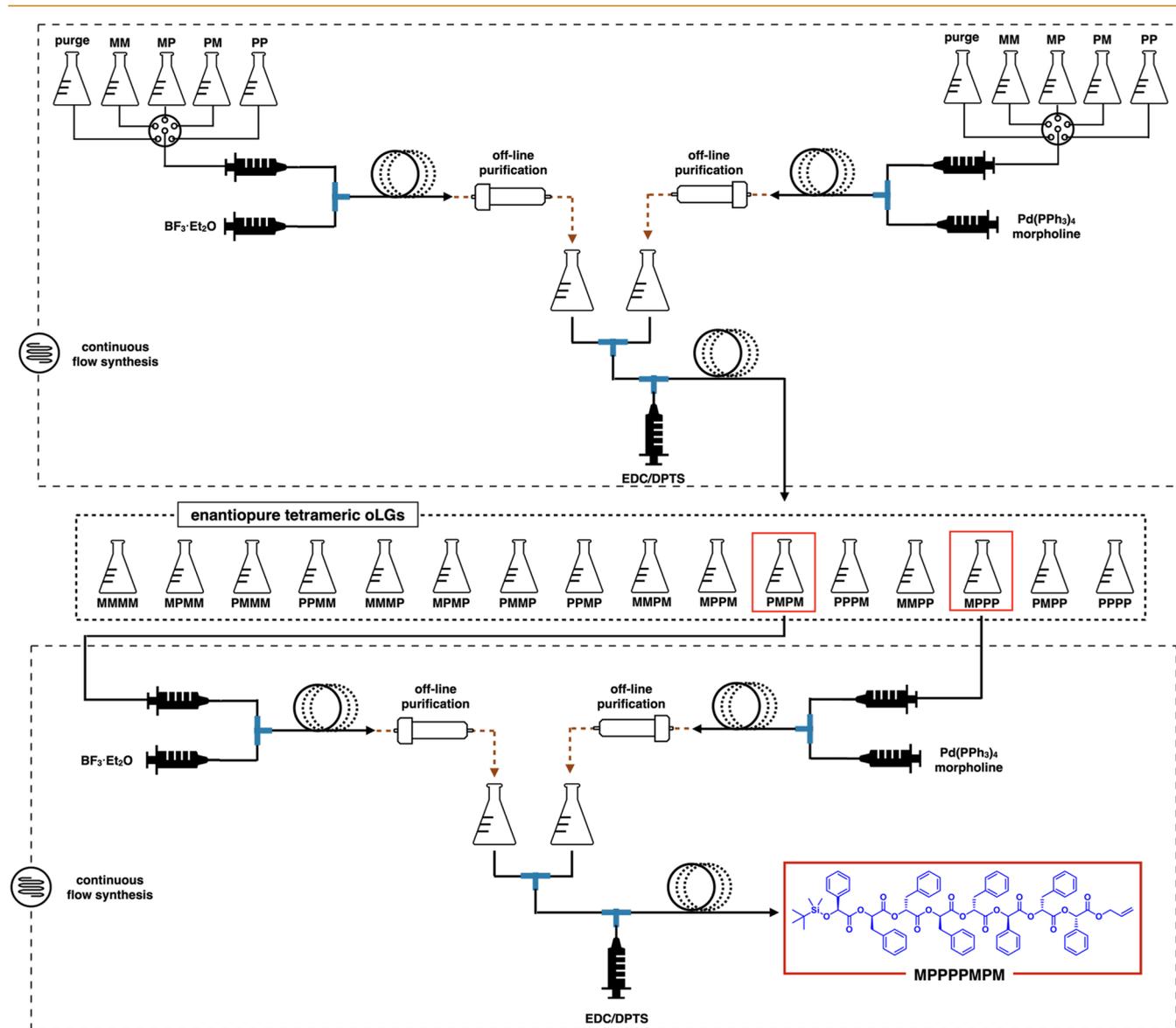
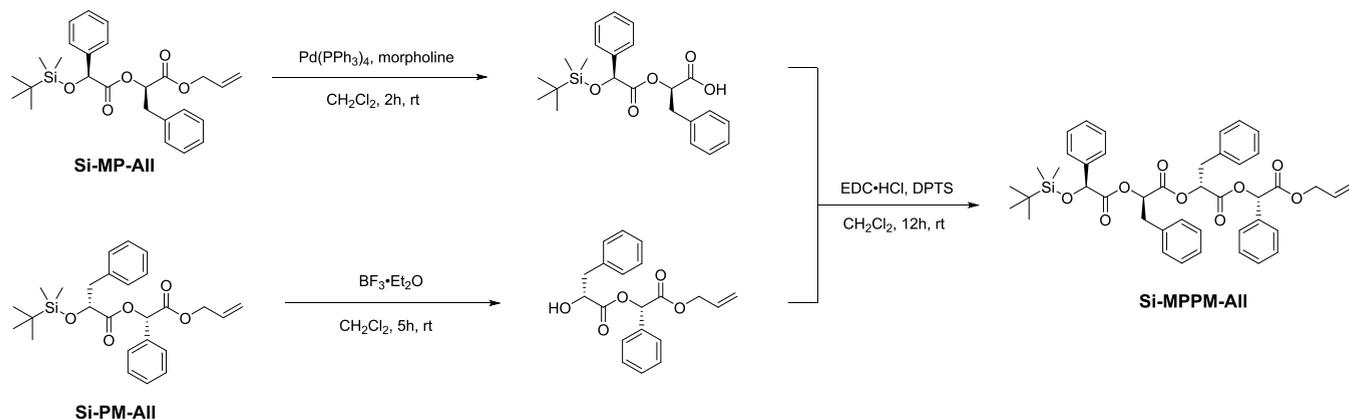


Figure 1. Flow chemistry for encoding 8-bit information in sequence-defined oMPs. A continuous flow synthesis of four dyads yielded 16-tetrad library. Subsequently, a sequence-defined octameric oMP, MPPPPMPM, could be obtained by cross convergence of MPPP and PMPM using a semiautomated flow system.

oligonucleotides has been adopted for decoding sequence-defined oligomers.^{38–40}

On the other hand, NMR spectroscopy, which can detect the structural differences around the atom of interest, is widely used

for the analysis of microstructures (arrangements of enantiomeric repeating units along the polymer backbone) of polymers. Especially, ^{13}C NMR has been used to measure comonomer distributions and tacticity of stereoregular polymers such as vinyl polymers and poly(L/D-lactide)s.^{41–44} 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;^{45–47} this is especially true for high-molecular-weight polymers with molecular weight distribution. Previously, Meyer and co-workers reported the sequencing of the repeating fragments composed of lactic, glycolic, or caproic acid by ^1H and ^{13}C NMR.^{48,49} Despite these previous works, full sequencing of oligoesters up to octamers has not been achieved.

Here, we report 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-mandelic-*co*-D-phenyl lactic acid)s (oMPs), and oligo(L-lactic-*co*-glycolic acid)s (oLGs) was constructed by semiautomated 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 oMP and oLG 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 oMP and oLG. 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.

RESULTS AND DISCUSSION

Accelerated Synthesis of Sequence-Defined Octamers of α -Hydroxy Acids by Flow Chemistry

We employed a cross-convergent approach to synthesize the sequence-defined oligomers in a step-economical manner.^{50–54} 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 of the possible sequences of a minimal number of repeating units are prerequisites for the cross-convergent approach (Scheme 1). The permutation in which L-mandelic acid (M) represents 0 and D-phenyl lactic acid (P) represents 1 yielded four dyads (MM, MP, PM, and PP), which cover all possible sequences of the dimers of M and P having protective groups for the hydroxyl and carboxyl end groups. Cross convergence of these dyads would produce 16 tetramers of M and P (tetrads) covering all of the possible sequences during the divergent stage of the synthesis, such that the number of possible products is maximized.

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 the number of sequence-defined products. To encode information in oligoesters at a rate higher than that of conventional batch processes, we used the semiautomated

method to synthesize all the possible tetrads in a continuous flow process²⁶ (Figure 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 ($\text{BF}_3\cdot\text{Et}_2\text{O}$) and the other for the allyl transfer reaction to morpholine catalyzed by tetrakis-(triphenylphosphine)palladium(0) ($\text{Pd}(\text{PPh}_3)_4$), 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 1 M 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 due 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 offline 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 semiautomated process allowed the synthesis of a 16-tetrad library to be completed on a gram scale within 24 h.

The resulting tetrads of M and P 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 (200 mg) were injected into the flow reactor. This was followed by offline purification via column chromatography. This flow process produced an octameric oMP encoding the target 8-bit information within 1 h; this period included the offline purification step. The purified oMPs, acquired with similar yields of tetrads, were fully characterized by ^1H and ^{13}C NMR spectroscopy techniques and matrix-assisted laser desorption and ionization-time-of-flight (MALDI-TOF) mass spectrometry, which confirmed their purity.

Based on the nonoverlapping of the peaks of L-lactic acid (L) and glycolic acid (G) units with those of the M and P units in ^{13}C NMR spectroscopy, we also synthesized a sequence-defined oligoester of L and G as an 8-bit-storing molecular medium using the flow chemistry setup described above in the same condition. The injection of four dyads of L and G having the same protective groups as those of the dyads of M and P into the flow chemistry setup yielded 16 tetrads on the 0.5 g scale in 24 h. The purified tetrads, obtained in yields of 88–92%, were subsequently converged to the targeted octameric oLGs using the flow reactor employed for the synthesis of the oMPs. The overall yields of oLGs from dyads were above 75%.

Nondestructive Sequencing of oMPs and oLGs by NMR Spectroscopy

The use of enantiomeric α -hydroxy acids as monomers for the cross convergence to oMPs and oLGs renders the resulting oligoesters to exhibit an absolutely defined stereochemical configuration. This simplified the analysis of the NMR signals, as

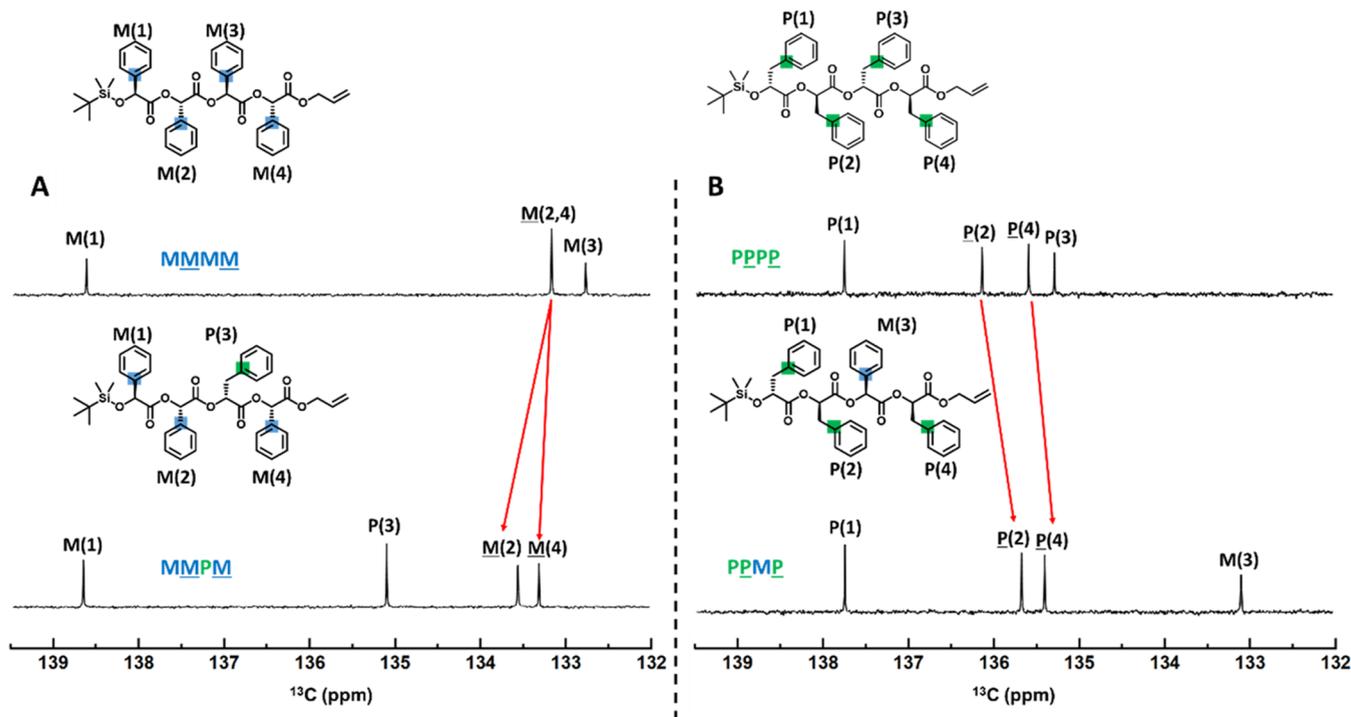


Figure 2. ¹³C NMR spectra of homotetrads and heterotetrads of M and P units. (A) ¹³C NMR spectra of MMMM and MMMP. The overlapped *ipso*-carbons of M units at positions 2 and 4 (133.16 ppm) were deshielded (133.55 and 133.30 ppm, respectively) when neighboring with the P unit. (B) ¹³C NMR spectra of PPPP and PPMP. The *ipso*-carbons of P units at positions 2 and 4 (136.13 and 135.58 ppm, respectively) were shielded (135.67 and 135.39 ppm, respectively) when neighboring with the M unit.

it removed the complexities arising from the splitting of the peaks caused by the uncertainty of the chemistry and stereochemical configuration of the neighboring monomers.

We first examined the ¹³C NMR spectra of the tetrads of M and P 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 *ipso*-carbons of the aromatic substituents of M and P 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 1 to 4 in the direction from the Si-terminus to the All-terminus. The peaks corresponding to the *ipso*-carbons of the four M units of the homotetrad MMMM from the Si-terminus to the All-terminus appeared in the order M(1), M(2), M(4), and M(3) at the chemical shifts of 138.64, 133.16, 133.16, and 132.76 parts per million (ppm), respectively. Similarly, the ¹³C NMR spectrum of the tetrad PPPP showed a set of *ipso*-carbon peaks of the P repeating units in the order P(1), P(2), P(4), and P(3) at 137.76, 136.13, 135.58, and 135.28 ppm, respectively.

In contrast to the NMR spectra of the homotetrads, MMMM and PPPP, those of the heterotetrads indicated that the *ipso*-carbon of the phenyl group of the tetrads experiences an electronic shielding/deshielding effect arising from the presence of aromatic substituents in the vicinity. The *ipso*-carbons of the M units were deshielded when they were neighbors with P units but shielded when they were neighbors with M units. Similar trends were observed in the case of the ¹³C NMR peaks of the *ipso*-carbons of the P units, which were shifted downfield when the repeating units were neighbors with P units and upfield when

the repeating units were neighbors with M units. For example, the ¹³C NMR spectrum of MMMP shows the downfield shift of the *ipso*-carbon peaks of M(2) and M(4), in contrast to the spectrum of MMMM owing to the presence of P(3) (Figure 2A). In the case of tetrad PPMP, the *ipso*-carbon peaks corresponding to P(2) and P(4) were upfield-shifted, in contrast to the spectrum of PPPP, because of the presence of M(3) (Figure 2B). 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 S1).

The sequencing of octameric oMPs by ¹³C 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. To identify a subtle change in the sequences of oMPs, we investigated a series of ¹³C NMR spectra of oMPs having one M unit moving from the 2-position to 7-position. Despite the identical composition of monomers, these oMPs clearly exhibited the *ipso*-carbon peaks corresponding to all repeating units at distinguished chemical shifts (Figure S13). Encouraged by these results, we established the rules for decoding the sequence of oMPs based on the electronic shielding/deshielding effects caused by the neighboring monomers. These decoding rules are summarized in Figure 3A.

Based on these rules, we attempted NMR sequencing of an oMP, Si-MPMPMPM-All, as shown in Figure 3B. The ¹³C 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

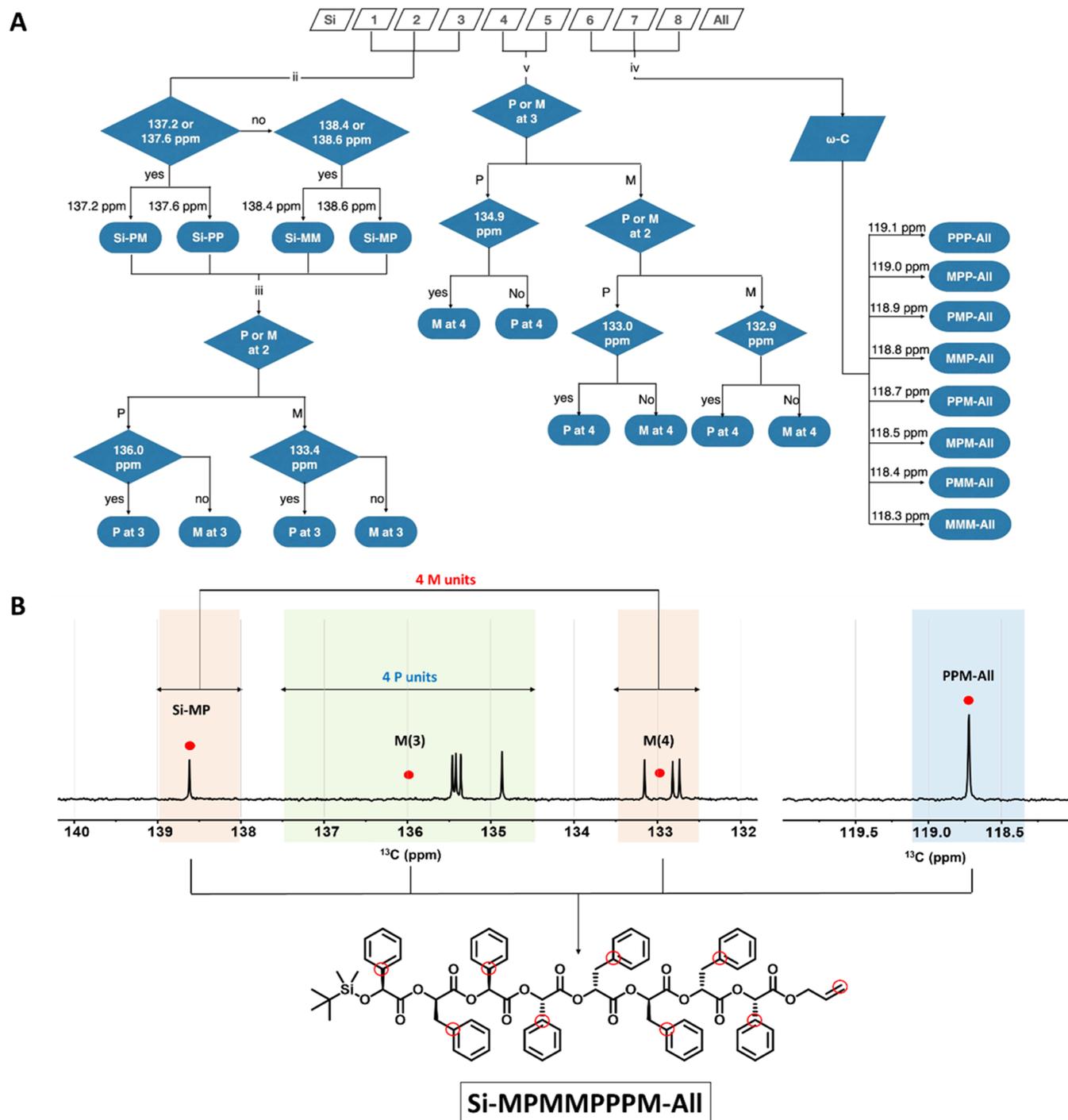


Figure 3. NMR sequencing of octameric oMPs. (A) Decoding diagrams of sequence-defined octameric oMPs. (B) Decoding of octameric oMPs based on the ^{13}C NMR spectrum and deciphered chemical structure of oMP. Red circles indicate the chemical shift that should be checked for sequencing.

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 a MALDI-TOF/TOF mass spectrometer (Figure S15). 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 oLG could be decoded by analyzing the ^{13}C NMR spectrum to determine the chemical shifts of the peaks corresponding to the α -carbons of the L and G units composing the sequence-defined octamer. The positions of the monomers of the oLG were assigned from A to H, starting from the Si-terminus to the All-terminus. After confirming the decodability of a single G unit at different positions of oLG, we composed the rules for decoding the sequence of oLGs based on

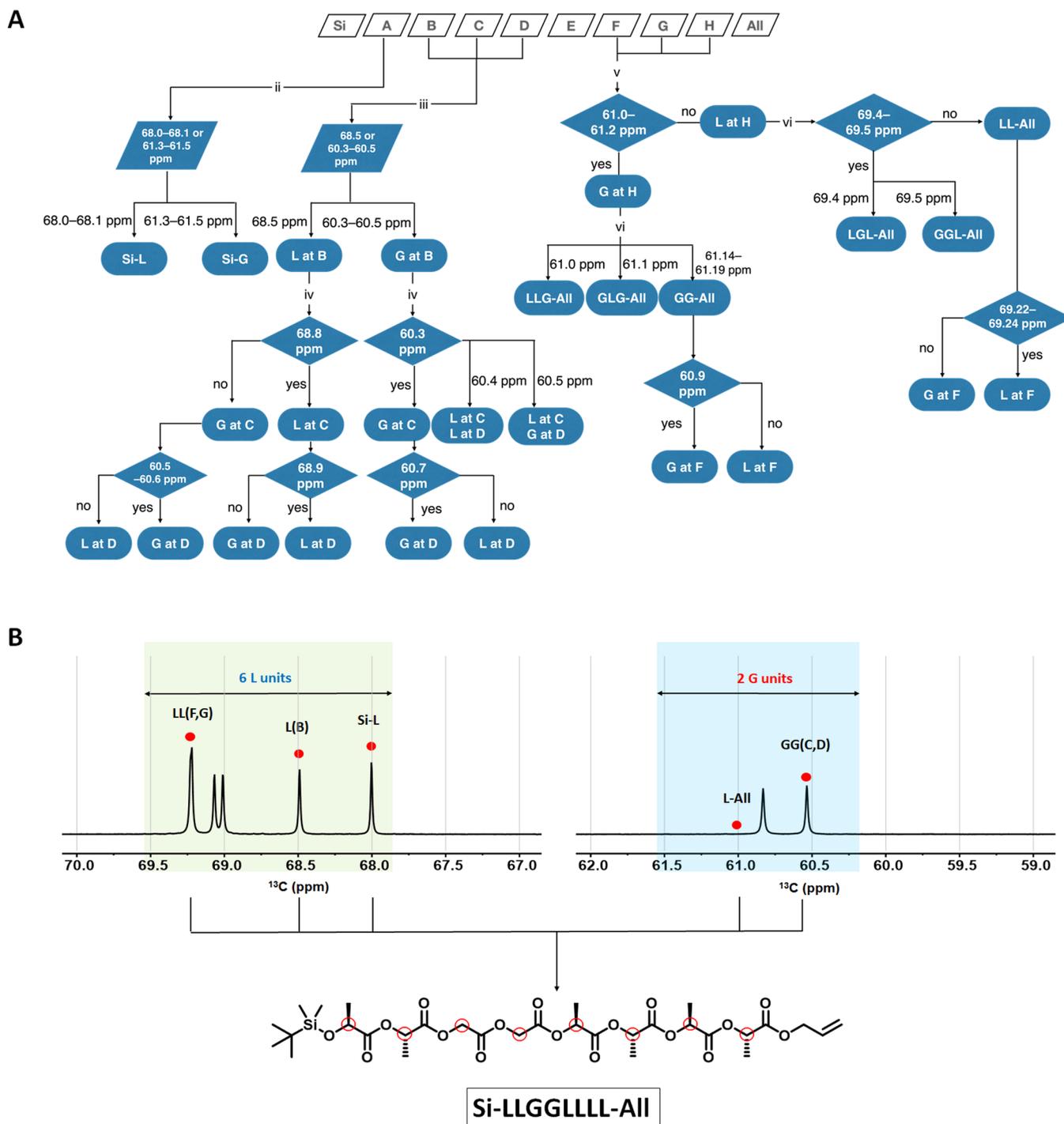


Figure 4. NMR sequencing of octameric oLGs. (A) Decoding diagrams of sequence-defined octameric oLGs. (B) Decoding of octameric oLGs based on the ^{13}C NMR spectrum and the deciphered chemical structure of oLG. Red circles indicate the checkpoint for sequencing.

the peak of the α -carbon. These decoding rules are summarized in Figure 4A.

As a demonstration, Si-LLGLLLL-All was sequenced based on the decoding rules, as shown in Figure 4B. The number of peaks present in the 68.0–70.0 ppm range suggests that the oLG is composed of six L and two 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 suggests 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 in the 61.0–61.2 ppm region. The absence of a peak in the 69.4–69.5 ppm range suggests that the seventh repeating unit is L. Finally, the appearance of a peak at 69.2 ppm is indicative of an L unit at position F. Therefore, the final sequence of the oLG was determined to be Si-LLGLLLL-All, which is in keeping with the proposed structure of oLG. Moreover, tandem mass sequencing performed using an electrospray ionization (ESI) mass spectrometer yielded a sequence identical to that obtained from NMR sequencing of oLG (Figure S16).

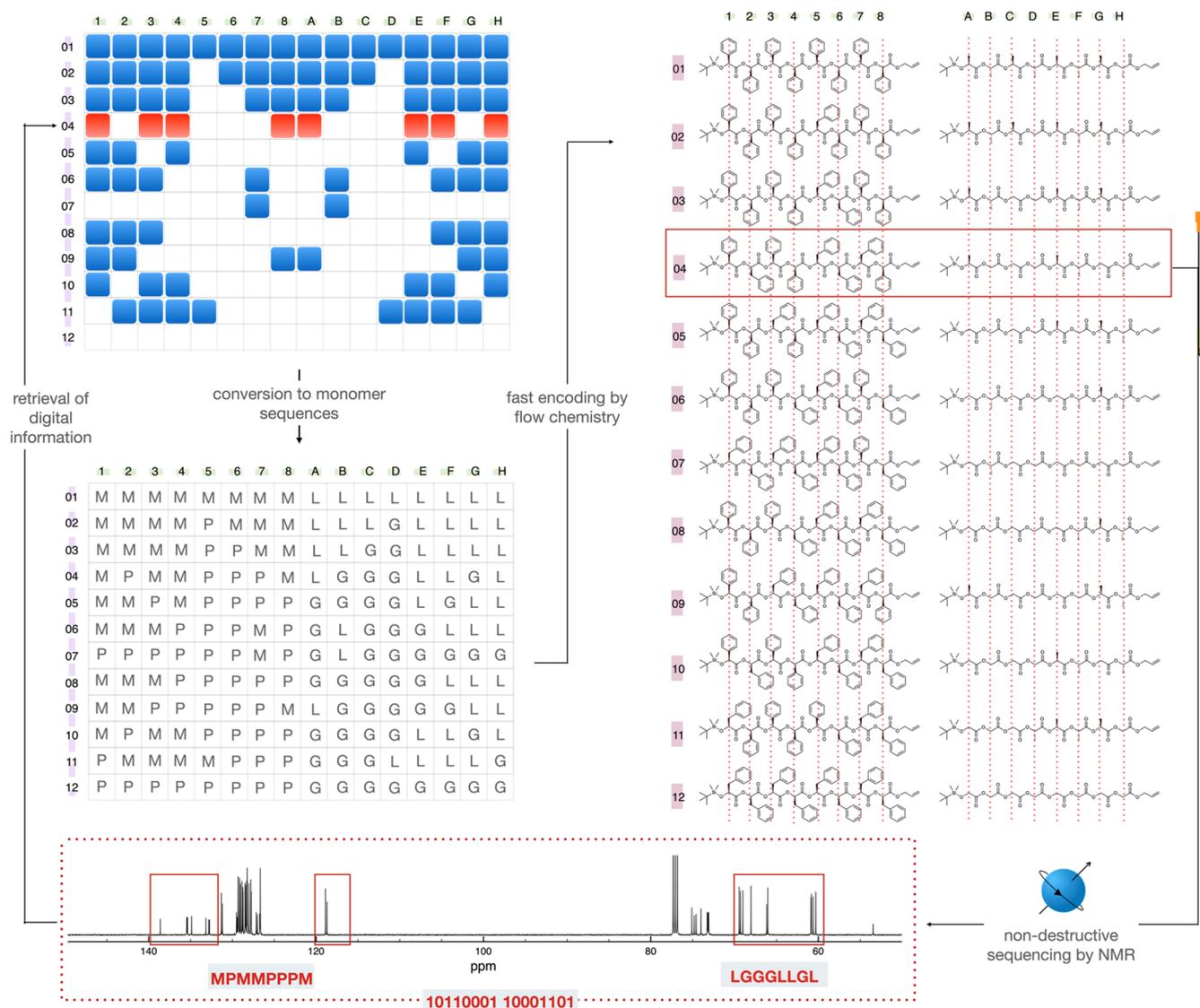


Figure 5. Encoding and decoding process of enantiopure oligoesters. The converted bitmap image was encoded into 12 sets of sequence-defined octameric oMP and oLG by a semiautomated flow process. Nondestructive decoding of the 04 mixture revealed the absolute sequence of oMP and oLG, which could be retrieved to digital information (highlighted to a red rectangles in a bitmap image).

Our rules for decoding the sequences of oligoesters based on their ¹³C NMR spectra indicate that there is no duplication between the 256 possible ¹³C NMR spectra of oMPs or the 256 possible spectra of oLGs. 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 these presynthesized oligoesters covering all of 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.

Encoding and Decoding of Digital Information in Oligoesters

We chose two pairs of enantiomeric α -hydroxy acids, *L*-mandelic acid/*D*-phenyl lactic 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 S17). 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 ¹³C 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 5). 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 semiautomated 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

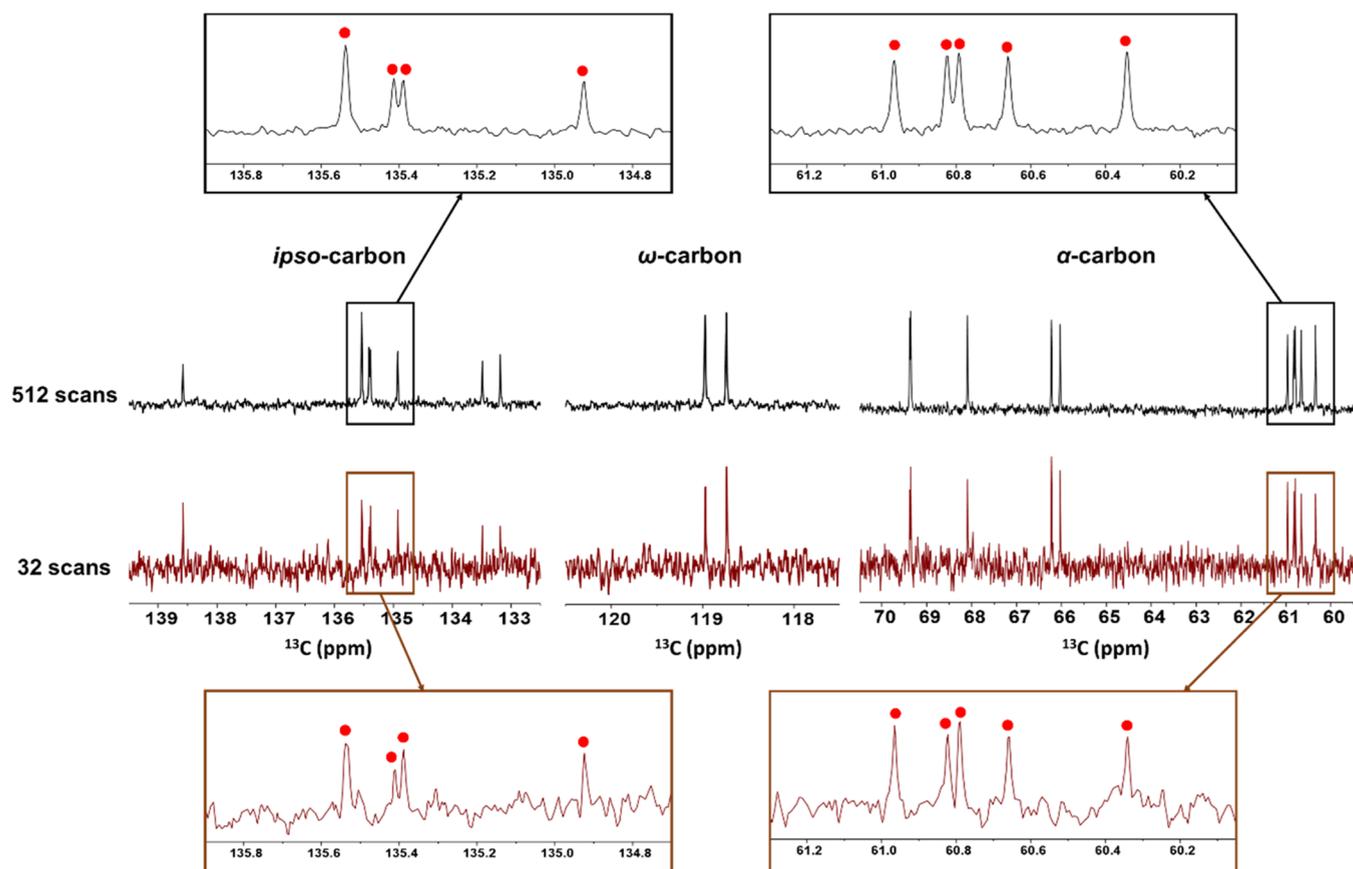


Figure 6. ^{13}C NMR spectra in a range of *ipso*-carbons (left), ω -carbon (center), and α -carbons (right) of the octameric mixture **09** with different numbers of scans. The peaks in the spectrum with 32 scans (maroon) could be distinguishable and matched with the reference spectrum (black).

(~1%) than the time required to complete the synthesis of the same set of oligoesters by repeating the conventional batch reaction and purification processes. Following the decoding rules, the 12 sets of oMPs and oLGs could be completely decoded (Figure S18).

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 of 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 an NMR spectrometer (Varian, 125 MHz for ^{13}C) to acquire the spectrum. The sequencing of oMP and oLG was completed using a single spectrum containing the *ipso*-carbon peaks of oMPs and the α -carbon peaks of oLG, which did not overlap. The 12 spectra of the mixtures of oMP and oLG could be decoded completely, and the stored information could be fully retrieved within 1 h by comparing the recorded spectra acquired by the minimum number of scans by NMR (32 scans for 110 s) with the existing reference spectra of oMPs and oLGs (Figure 6).

Oligoesters could be degraded by hydrolysis or the epimerization of α -proton in solution, in particular at high temperatures or in the presence of basic catalysts.^{55,56} The end groups of oMPs and oLGs used in this study are protected by TBDMS and an allyl ester group, which prevents degradation by hydrolysis. To investigate stability, the oligoesters were contained in NMR tubes with a cap and parafilm seal and without any additives. The decoding and retrieval of the stored image could be achieved without any reading errors even after 10

months using the same NMR samples (Figure S19). We also note that the storage capacity in the mixture of oMP and oLG could be expanded by the correlation of two sequences. For example, 4-bit information can be encoded by the correlation of position 1 of oMP and position A of oLG, which makes the mixture of oMP and oLG to store 32 bits.

CONCLUSIONS

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 semiautomated 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.

METHODS

Materials

L-Lactic acid ($\geq 98\%$), L-mandelic acid ($\geq 99\%$), allyl bromide, *tert*-butyldimethylsilyl chloride, trifluoroborane etherate, tetrakis(triphenylphosphine)palladium (0), and morpholine were purchased from Sigma-Aldrich and used without further purification. Glycolic acid ($\geq 98\%$) and 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride were purchased from Tokyo Chemical Industry and used without purification. D-Phenyl lactic acid ($\geq 95\%$) was purchased from AK Scientific, Inc. and used without purification. Dichloromethane was distilled over CaH₂ under N₂.

General Instruments

A Legato 101 syringe pump was purchased from Kd Scientific. A 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. An Omnitfit EZ column was purchased from Diba Industries Inc. Gastorr AG-42-01 was purchased from GL Sciences. A 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 Sfar silica column cartridge. Hexane and ethyl acetate were used as eluents. ¹H NMR and ¹³C NMR spectra were recorded on a Varian INOVA 500 MHz NMR spectrometer in CDCl₃. 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.

Continuous Flow Synthesis for Sequence-Defined Tetrads

All permutations of enantiopure oMP and oLG tetrads were generated by a continuous flow system consisting of a synthetic step and cleaning step, which is operated via six-way syringe pumps connected with a programmable controller. About 2 mmol of a dyad (1 M in DCM) and BF₃·Et₂O (7 M in DCM) were injected at 0.1 mL/min and mixed through a 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 tetrahydrofuran (THF)) and Pd(PPh₃)₄ and morpholine mixture (0.03 M and 1.05 M in THF) were injected at 0.1 mL/min and mixed through the T-mixer. The mixture was allowed to react in the reaction loop (volume of 2 mL). Deprotected dyads, hydroxyl and carboxylic acids, 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 the reaction loop (volume of 6 mL). After the synthesis step, the DCM solvent (6 mL) was purged to the flow reactor for cleaning the reaction loop. Synthesis and cleaning cycles 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 the Semiautomated Flow System

The octameric oMPs and oLGs were obtained by flow synthesis following the same conditions described for the synthesis of tetrads. About 200 mg of a tetrad (0.25–0.5 M in DCM) and BF₃·Et₂O (7 M in DCM) were injected at 0.1 mL/min and mixed through a 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(PPh₃)₄ and morpholine mixture (0.03 and 1.05 M in THF) were injected at 0.1 mL/min and mixed through the 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 the reaction loop (volume of 9 mL). The collected octameric oligoester was purified with automated flash column chromatography (85–91% yield).

¹³C NMR Sequencing of Octameric oMPs

(Rule i) The number of M units in 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 134.5–137.5 ppm range is identical to the number of P units (Figure S2). (Rule ii) The monomer units at positions 1 and 2 can be determined based on the peak of the *ipso*-carbon that is deshielded most by a silicon atom present in its proximity (Figure S3). In the 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 S4). The peak corresponding to P(3) appears at 133.4 ppm in the case of 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 S5). (Rule v) The monomer units on positions 4 and 5 can be predicted by an indirect method (Figure S6). 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.

¹³C NMR Sequencing of Octameric oLGs

(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 S7 and S8). (Rule ii) The monomer unit at position A can be determined based on the presence of the peak of the α -carbon at a specific chemical shift (Figures S9 and S10). 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 in specific regions (Figures S9 and S10). 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 positions C and D are predicted by an indirect method (Figure S11). 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 in the case of neighboring G(C), L(C) and L(D), or L(C) and G(D), respectively. When the G units are at positions 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 α -carbon at the specific chemical shift (Figure S12). The peak corresponding to the α -carbon of G(H) appears at 61.0–2 ppm. (Rule vi) The monomer units at positions F and G are predicted by an indirect method (Figure S12). The α -carbon of G(H) appears at 61.0, 61.1, or 61.2 ppm in 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 in 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.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacsau.2c00388>.

Characterization data for sequence-defined oligoesters, ¹³C NMR sequencing results of oMP and oLG mixtures, and copies of NMR spectra (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Kyoung Taek Kim – Department of Chemistry, Seoul National University, Seoul 08826, Korea; orcid.org/0000-0002-4915-7602; Email: ktkim72@snu.ac.kr

Authors

Jeong Min Lee – Department of Chemistry, Seoul National University, Seoul 08826, Korea; orcid.org/0000-0002-7531-6558

Heejeong Jang – Department of Chemistry, Seoul National University, Seoul 08826, Korea

Seul Woo Lee – Department of Chemistry, Seoul National University, Seoul 08826, Korea; orcid.org/0000-0001-7014-0365

Complete contact information is available at <https://pubs.acs.org/doi/10.1021/jacsau.2c00388>

Author Contributions

[†]J.M.L. and H.J. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Research Foundation (NRF) grant funded by the Korean Government (MSIT) (NRF-2020R1A5A1019631 and NRF-2022R1A2C3013240). K.T.K. thanks Seoul National University (SNU) for the support by the Creative-Pioneering Researchers Program (305-20200046). The authors are grateful to Jonghyuk Im at SNU for the NMR consultation.

■ 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) Church, G. M.; Gao, Y.; Kosuri, S. Next-generation digital information storage in DNA. *Science* **2012**, *337*, 1628.
- (6) 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.
- (7) Ceze, L.; Nivala, J.; Strauss, K. Molecular digital data storage using DNA. *Nat. Rev. Genet.* **2019**, *20*, 456–466.
- (8) Rutten, M. G. T. A.; Vaandrager, F. W.; Elemans, J. A.; Nolte, R. J. Encoding information into polymers. *Nat. Rev. Chem.* **2018**, *2*, 365–381.
- (9) Zhirnov, V.; Zadegan, R. M.; Sandhu, G. S.; Church, G. M.; Hughes, W. L. Nucleic acid memory. *Nat. Mater.* **2016**, *15*, 366–370.
- (10) Kosuri, S.; Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **2014**, *11*, 499–507.
- (11) 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*, No. 2383.
- (12) Grass, R. N.; Heckel, R.; Puddu, M.; Paunesco, 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.
- (13) Charles, L.; Lutz, J.-F. Design of Abiological Digital Poly (phosphodiester)s. *Acc. Chem. Res.* **2021**, *54*, 1791–1800.
- (14) 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.
- (15) 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*, No. 4451.
- (16) 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.
- (17) 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*, No. 56.
- (18) 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.
- (19) Laurent, E.; Amalian, J.-A.; Parmentier, M.; Oswald, L.; Al Ouahabi, A.; Dufour, F.; Launay, K.; Clément, J.-L.; Gignes, D.; Delsuc, M.-A.; Charles, L.; Lutz, J.-F. High-capacity digital polymers: storing images in single molecules. *Macromolecules* **2020**, *53*, 4022–4029.
- (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. R. Dual sequence definition increases the data storage capacity of sequence-defined macromolecules. *Commun. Chem.* **2020**, *3*, No. 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. U.S.A.* **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*, No. 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*, No. 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.; Charles, L.; Dignes, D. 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.; Zhu, X. Binary tree-inspired digital dendrimer. *Nat. Commun.* **2019**, *10*, No. 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) 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.

(34) 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.

(35) 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.

(36) 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.

(37) Kretschmer, J.; David, T.; Dračinský, M.; Socha, O.; Jirak, D.; Vít, M.; Jurok, R.; Kuchař, M.; Císařová, I.; Polasek, M. Paramagnetic encoding of molecules. *Nat. Commun.* **2022**, *13*, No. 3179.

(38) Jain, M.; Koren, S.; Miga, K. H.; Quick, J.; Rand, A. C.; Sasaki, T. A.; Tyson, J. R.; Beggs, A. D.; Dilthey, A. T.; Fiddes, I. T.; Malla, S.; Marriott, H.; Nieto, T.; O'Grady, J.; Olsen, H. E.; Pedersen, B. S.; Rhie, A.; Richardson, H.; Quinlan, A. R.; Snutch, T. P.; Tee, L.; Paten, B.; Phillippy, A. M.; Simpson, J. T.; Loman, N. J.; Loose, M. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **2018**, *36*, 338–345.

(39) 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. U.S.A.* **2010**, *107*, 12080–12085.

(40) Cao, C.; Krapp, L. F.; Al Ouahabi, A.; König, N. F.; Cirauqui, N.; Radenovic, A.; Lutz, J.-F.; Dal Peraro, M. Aerolysin nanopores decode digital information stored in tailored macromolecular analytes. *Sci. Adv.* **2020**, *6*, No. eabc2661.

(41) Randall, J. C. *Polymer Sequence Determination: Carbon-13 NMR Method*; Academic Press: New York, 1977.

(42) 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.

(43) Ovitt, T. M.; Coates, G. W. Stereochemistry of lactide polymerization with chiral catalysts: new opportunities for stereo-control using polymer exchange mechanisms. *J. Am. Chem. Soc.* **2002**, *124*, 1316–1326.

(44) 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.

(45) 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.

(46) Cais, R. E.; 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.

(47) Kasperczyk, J. Microstructural analysis of poly[(L,L-lactide)-co-(glycolide)] by ¹H and ¹³C n.m.r. spectroscopy. *Polymer* **1996**, *37*, 201–203.

(48) Stayshich, R. M.; Meyer, T. Y. New insights into poly (lactic-glycolic acid) microstructure: using repeating sequence copolymers to decipher complex NMR and thermal behavior. *J. Am. Chem. Soc.* **2010**, *132*, 10920–10934.

(49) Weiss, R. M.; Jones, E. M.; Shafer, D. E.; Stayshich, R. M.; Meyer, T. Y. Synthesis of repeating sequence copolymers of lactic, glycolic, and caprolactic acids. *J. Polym. Sci., Part A: Polym. Chem.* **2011**, *49*, 1847–1855.

(50) 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.

(51) 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.

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

(53) 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.

(54) 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 dononacotamer and its shorter homologues. *J. Am. Chem. Soc.* **2016**, *138*, 4210–4218.

(55) Gerlt, J. A.; Kozarich, J. W.; Kenyon, G. L.; Gassman, P. G. Electrophilic catalysis can explain the unexpected acidity of carbon acids in enzyme-catalyzed reactions. *J. Am. Chem. Soc.* **1991**, *113*, 9667–9669.

(56) Botvin, V.; Karaseva, S.; Khasanov, V. Depolymerization of lactic acid oligomers into lactide: Epimerization, stereocomplex formation, and nature of interactions of oligomers. *Polym. Degrad. Stab.* **2020**, *182*, No. 109382.