Synthetic Digital Polymers for Data Storage Applications

Brittany Prempin Literature Seminar November 17, 2022

In our increasingly digitized world, data storage needs are increasing exponentially and by 2025 global data creation is projected to grow to more than 180 zettabytes.[1] Current data storage technologies such as hard-disk drives, solid-state drives, flash memory devices, and optical storage devices require great amounts of raw materials, space, and energy to store digital information and are all prone to well-known failure mechanisms.[2] Digital polymers (DPs) are a class of sequence-defined, information encoding polymers that present a potential way to meet some of society's data storage needs by storing data densely at a molecular level.

Innovations in the field of DPs have focused primarily focused on the viability of DNA for data storage applications because DNA can be easily synthesized, replicated, and sequenced.[3] However, an alternative is to employ synthetic DPs, i.e., monodisperse synthetic sequenced-defined polymers not based on DNA, in which bits of data are encoded in the monomer sequence and the data are read out or sequenced by mass spectrometry (MS).

Synthetic DPs overcome some of the shortcomings of using DNA as a data storage medium. Synthetic DPs are more durable and are structurally more diverse, which facilitates increased density of data. Whereas one DNA base (or base pair) encodes only two bits of data (i.e., there are four possible values, A, T, C, G), the structural diversity of synthetic DPs means that they can store more data per Dalton of material.[4] In addition, whereas DNA synthesis is error prone when there are long repeats of a single nucleotide, DP synthesis is characterized by much higher fidelity when there are long repeats of the same monomer.[4] The use of synthetic DPs allows researchers to custom design data storage media for their precise needs, rather than retrofit DNA synthesis, replication, and sequencing for a new application.

There are two main methods to increase the storage capacity of synthetic DPs: increasing oligomer lengths and increasing the density of data. Due to the difficulty of synthesizing and sequencing long sequence-defined oligomers, researchers have turned to developing clever polymer architectures to increase the density of data in synthetic DP systems.

Self-immolative polymers are a class of polymers designed to depolymerize upon an external stimulus. Using a synthetic route inspired by Merrifield synthesis of peptides and a sequencing strategy inspired Edman degradation sequencing of proteins, Anslyn and coworkers designed an oligourethane DP to encode a that can be sequenced through self-immolation.[5] The monomers were synthesized by reduction of commercially available N-Fmoc amino acids and were then polymerized by iterative stepwise solid-phase synthesis.[5] When irradiated with microwaves at 70 °C under basic conditions, the oligomers undergo a spontaneous 5-exo trig cyclization beginning with the terminal β-alcohol (Figure 1). The researchers sampled the depolymerizing solution at 20-minute intervals and detected the corresponding cyclic carbamates molecules released sequentially from the oligomers by LC-MS.[5] This LC-MS method is easier than the tandem mass-spectrometry method ordinarily used to sequence synthetic DPs.

Inspired by the work of Meier and coworkers,[6] Anslyn and coworkers used isotopic tagging of a mixture of oligourethane DPs to increase the density of data encoded.[7] In order to sequence the mixture of oligourethane DPs in the correct order, they employed isotopic tagging of the N-terminal monomer of a mixture of eight unique 10-mers. By using a software script called Mol.Encrypter, they were able to encode a 256-bit cypher key into a sample of eight 10-mers. By

using a related script called Mol.Decrypter, they could decode the data read out by LC-MS and recover the cypher key. This DP design exploits monomer diversity and a known depolymerization mechanism to generate a synthetic oligomer well suited for storage of a few hundred bits of data, thus exemplifying how synthetic DPs can be employed for this specific application. Although the solid phase synthesis of these oligomers is not atom economical, this system still leverages simple, well-understood chemistry for facile sequencing.

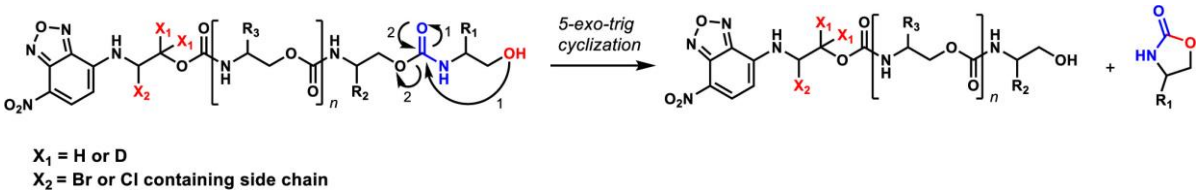

Figure 1 Intramolecular Chain End depolymerization of oligourethanes enables sequencing via sequential loss of 2-oxazolidinone molecules. [7]

A binary tree is a computer science data structure in which the arrangement of nodes in the tree structure encodes the information. Inspired by the binary tree-like structures and monodispersity of dendrimers, Zhang and coworkers used thiol maleimide Michael coupling, a type of click chemistry, to create a dendrimer DP. [8] Each dendrimer is composed of two monomers: a hexyl-substituted monomer that represents the 1-bit and a butyl-substituted monomer that represents the 0-bit. Through a stepwise divergent strategy, the hexyl and butyl monomers were precisely coupled from core to periphery to construct dendrimer DPs having binary tree structures. [8] Each branch point consists of a thioether linkage that is reliably cleaved in tandem MS/MS experiments, generating a recognizable fragmentation pattern.

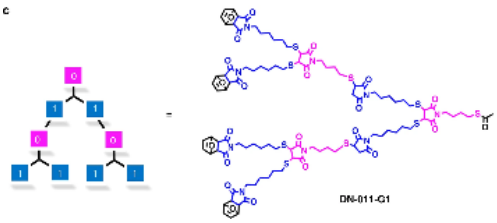

Figure 2 Molecular structure of binary dendrimer DN-011-G1. [8]

For dendrimer DN-011-G1 (Figure 2), there are 6 pathways to cleave each 0-bit or 1-bit monomer off the dendrimer from periphery to core. A 9-character binary path represents each of the 6 fragmentation pathways and each binary path is represented by a row in a 2D data matrix. Through encryption, 720 unique data matrices produced from the binary paths can be encrypted into one unique matrix for each dendrimer. As the dendrimers become larger, their storage capacity increases. The first-generation dendrimer, for example, encodes 6 bytes of data, while the second-generation dendrimer encodes 3672 bytes of data, increasing the data storage capacity by a magnitude of 612. [8] This is DP design exemplifies how structural diversity in synthetic DPs can be leveraged to increase their density of data.

In pursuit of high-density DPs synthesized through atom-economic methods, Tang and coworkers have created a nonlinear miktoarm star polymer that encodes a 2D matrix of data through two different types of click chemistries. [9] By coupling hydroxyl-yne and thiol-ene click chemistries, the researchers polymerized sequence-defined oligo(monothioacetals) (Figure 3). Using primarily commercially available starting materials, they generated a total of 16

combinations of repeat units from four propiolate building blocks with different R groups and four mercaptoalcohol building blocks with different lengths of methylene spacers. [9]

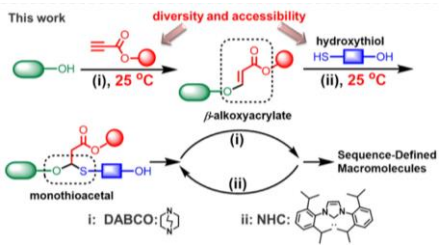

Figure 3 Click chemistry polymerization scheme for oligo(monothioacetals). [9]

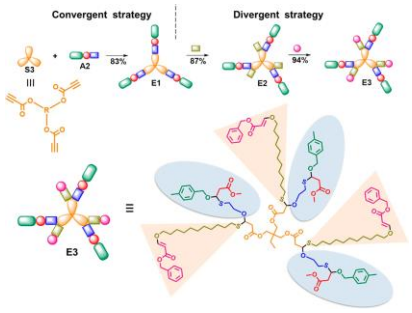

Figure 4 Structure of miktoarm star digital polymer. [9]

Through a combination of convergent and divergent strategies, they synthesized a 6 arm miktoarm star polymer pictured in Figure 4. Like the work of Zhang, [8] this backbone architecture incorporates S–C bonds that are predictably cleaved under ESI MS/MS conditions. This motif enables identification of all the fragment signals of the twenty distinct fragmentation pathways of the miktoarm star DP.

This DP design strategy employs atom-economical synthesis of a 3-dimensional sequence-defined oligomers and generates a 2D matrix of data at a high storage density (0.013 bit/Da). However, this sequencing strategy requires a substantial amount of time and a very skilled mass spectrometrist to analyze the ESI MS/MS data. Even with highly skilled scientists, it took approximately 4 hours to extract the data form the MS spectra. This sequencing strategy would benefit from computational support to automate readout.

Although Anslyn, Zhang, and Tang have shown methods to increase the data storage density of synthetic DPs, much more needs to be done before DPs will replace existing data storage technologies. The data storage capacity of synthetic DPs remains in the hundreds of bits and few strategies have been established to facilitate the sequencing and rewriting of data. As a result, the current applications of DPs are limited to long term data storage, which lends itself well to molecular encryption and anticounterfeiting applications. To move towards a wider scope of synthetic DP applications, more strategies are needed to rewrite and edit data encoded in monomer sequences as well as increasing the speed of the full cycle. For example, it is worth exploring how synthetic DP synthesis could be automated through flow chemistry[10] or how synthetic DPs could be rewritten with dynamic covalent chemistry. [11]

---

1) Statista. Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025 (in zettabytes). *Statista* **2021**.
2) Lunt, B. M. How long is long-term data storage? In *Archiving Conference*, 2011; Society for Imaging Science and Technology: Vol. 2011, pp 29-33.
3) Ceze, L.; Nivala, J.; Strauss, K. Molecular digital data storage using DNA. *Nat. Rev. Genet.* **2019**, *20* (8), 456-466.
4) Erlich, Y.; Zielinski, D. DNA Fountain enables a robust and efficient storage architecture. *Science* **2017**, *355* (6328), 950-954.
5) 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.; et al. Sequencing of Sequence-Defined Oligourethanes via Controlled Self-Immolation. *JACS* **2020**, *142* (6), 2744-2749.

6) Frölich, M.; Hofheinz, D.; Meier, M. A. R. Reading mixtures of uniform sequence-defined macromolecules to increase data storage capacity. *Commun. Chem.* **2020**, *3* (1), 184.
7) Dahlhauser, S. D.; Wight, C. D.; Moor, S. R.; Scanga, R. A.; Ngo, P.; York, J. T.; Vera, M. S.; Blake, K. J.; Riddington, I. M.; Reuther, J. F.; et al. Molecular Encryption and Steganography Using Mixtures of Simultaneously Sequenced, Sequence-Defined Oligourethanes. *ACS Cent. Sci.* **2022**, *8* (8), 1125-1133.
8) Huang, Z.; Shi, Q.; Guo, J.; Meng, F.; Zhang, Y.; Lu, Y.; Qian, Z.; Li, X.; Zhou, N.; Zhang, Z.; et al. Binary tree-inspired digital dendrimer. *Nat. Commun.* **2019**, *10* (1), 1918.
9) 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. *JACS* **2022**, *144* (4), 1672-1680.
10) 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* (10), eabl8614.
11) Skene, W. G.; Lehn, J.-M. P. Dynamers: Polyacylhydrazone reversible covalent polymers, component exchange, and constitutional diversity. *PNAS* **2004**, *101* (22), 8270-8275.