European Polymer Journal 120 (2019) 109260

Contents lists available at ScienceDirect

European Polymer Journal

journal homepage: www.elsevier.com/locate/europolj

Macromolecular Nanotechnology

Molecular access to multi-dimensionally encoded information

Jan Steinkoenig[1], Resat Aksakal[1], Filip Du Prez*

Polymer Chemistry Research Group, Centre of Macromolecular Chemistry (CMaC), Department of Organic and Macromolecular Chemistry, Faculty of Sciences, Ghent University, Krijgslaan 281, S4-bis B-9000 Ghent, Belgium

ARTICLE INFO

Keywords:
Sequence-defined polymers
Barcodes
Data storage
Digital information
Multidimensional
QR codes

ABSTRACT

Polymer scientist have only recently realized that information storage on the molecular level is not only restricted to DNA-based systems. Similar encoding and decoding of data have been demonstrated on synthetic polymers that could overcome some of the drawbacks associated with DNA, such as the ability to make use of a larger monomer alphabet. This feature article describes some of the recent data storage strategies that were investigated, ranging from writing information on linear sequence-defined macromolecules up to layer-by-layer casted surfaces and QR codes. In addition, some strategies to increase storage density are elaborated and some trends regarding future perspectives on molecular data storage from the literature are critically evaluated. This work ends with highlighting the demand for new strategies setting up reliable solutions for future data management technologies.

1. Introduction

Humanity has been trying to preserve information since ancient times, dating back to cave paintings or stone tablets by prehistoric men 40,000 years ago [1]. Today, technological developments allow us to store much more complex data in a range of storage units, such as data centers, databases or even in the form of flash memories. In 2014, the total global digital information was about 4.4 zettabytes (4.4×10^{21} bytes) and it is at least doubling every year [2]. At this growth rate, it is projected that the amount of information will already reach around 400 zettabytes (4×10^{23} bytes) by 2040 [2]. Thus, data storage and information overload is arguably one of the largest technological challenges of the 21st century: To be able to store 400 zettabytes of data on established technologies, flash memory manufacturing alone would require 10^9 kg of silicon wafers [3,4]. The demand greatly exceeds today's global silicon supply that is estimated at 10^7–10^8 kg [3]. Thus, current technology can only handle a fraction of the coming data deluge, which is expected to consume all the world's microchip-grade silicon by 2040.

Before embarking on the discussion about molecular storage media, it is important to introduce the most common types of data storage (see Fig. 1): (*i*) linear binary codes represented by 0 and 1, which are one-dimensional (1D) by nature and (*ii*) two-dimensional (2D) matrices, most commonly represented as Quick Response (QR) codes.

The vast majority of information ranging from computer science to labelling applications such as barcodes, is represented by linear binary codes. Thus, in the first part of the present article, linear 1D molecular storage media, including deoxyribonucleic acid (DNA) and synthetic macromolecules, are discussed. The storage density and long-term stability of these 1D structures are compared and the available read out tools elaborated. In the second part, the particular example of QR codes is described. QR codes are 2D barcodes that can be elaborated from linear 1D oligomers, thus a discussion about this area of research is provided. Finally, in the last part of this article, multi-dimensionality on molecular level as well as on a surface are compared with each other and further elaborated critically.

2. Molecular one-dimensional data storage

In the search for molecular storage media, especially for storing data that needs to be archived (so-called cold data storage) for legal or regulatory reasons – like rarely accessed surveillance videos, medical records, or historical government documents – DNA has attracted much interest in the last decades [5–7] Nature's information storage, which carries the genetic information of all living systems, has the capacity to store much more information per volume compared to a flash memory [8]. While storing messages in DNA was first demonstrated in 1988, the first large-scale example of storing data was reported in 2012, when Church et al. managed to "save" entire books on DNA microchips that were "read" via DNA sequencing [9–12]. In another example, Goldman

* Corresponding author.
E-mail address: Filip.DuPrez@UGent.be (F. Du Prez).
[1] Equally contributing authors.

https://doi.org/10.1016/j.eurpolymj.2019.109260
Received 12 July 2019; Received in revised form 16 September 2019; Accepted 19 September 2019
Available online 19 September 2019
0014-3057/

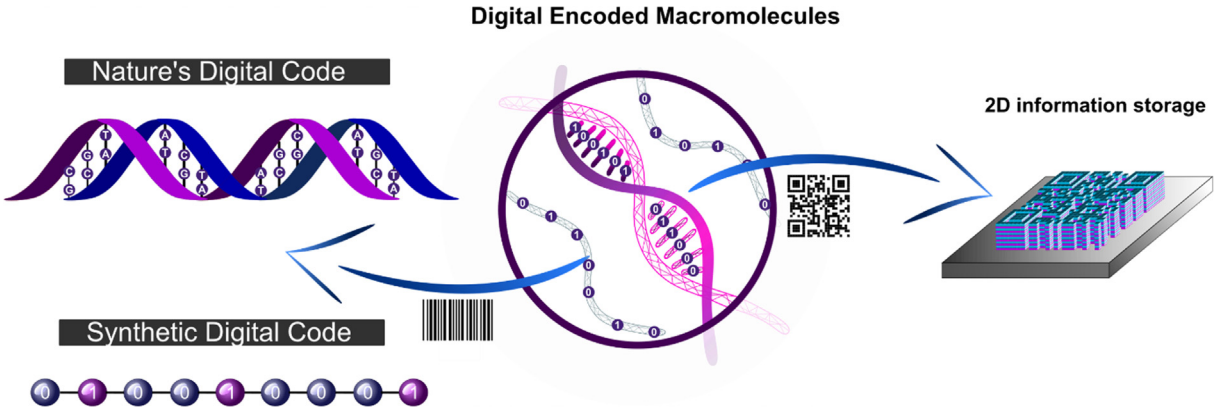

Fig. 1. Representation of various types of molecular data storage. The one-dimensional information storage (molecular barcode) is mainly showcased by DNA and linear synthetic macromolecules. The two-dimensional information storage is more complex but could be illustrated as a QR code on a surface.

et al. demonstrated the DNA capacity by storing all Shakespeare sonnets, a scientific paper in a PDF format, a JPEG image and an extract of Martin Luther King's "I have a dream" speech in an MP3 format; thus storing a total of 757,051 bytes [13]. A new record was set by researchers at Microsoft and the University of Washington, when they stored 200 megabytes of data on a miniscule amount of DNA containing 1.5 billion base pairs [14].

Major progress by dedicated biotechnological improvements in DNA science gives direct access to strands' multiplication (using polymerase chain reaction, for instance), as well as to fast and more and more precise readout tools such as the nanopore technology [15–17]. However, some issues remain in terms of (re-)writing information, the range of digits (DNA entails only 4 nucleobases) and the information capacity per repeating unit (*i.e.* one nucleobase introduces one digit). Other challenges include biochemical constraints such as errors in the sequences, deviations during the sequencing or long-term stability of the DNA-based data storage [18–20]. Although, it was recently demonstrated that in certain buffered conditions, plasmid DNA can be used for long-term error-free data archival (*e.g.* 3 years at −20 °C or 20 days at 65 °C) [21], maintaining ideal preservation conditions are expensive and consume energy in the long term.

Polymer chemists realized recently that the potential of data storage is not restricted to DNA and that some of the aforementioned drawbacks could be in principle overcome by encoding data at the molecular level using synthetic macromolecules [22,23]. In this manner, the storage density can be increased dramatically since monomers constituting synthetic polymers are not limited to the four nucleobase motifs that make up DNA [24]. It should be noted here that contemporary research on xeno nucleic acids (XNA), which are synthetic nucleic acid analogues are not discussed in this perspective article [25]. Moreover, a long-term stability at ambient temperature, could be in principle achieved by carefully adjusting the polymer's structure. Therefore, the recent progress in the synthesis and characterization of sequence-defined polymers accesses their implementation for data storage purposes, potentially representing the basis for the future data management technology (see Fig. 2) [26–30]. These advances, detailed below, include the possibility of 1 bit coding *via* short sequences and also the preparation of longer sequences that allow higher storage density or the development of new (automated) strategies. Each procedure entails various advantages and drawbacks that will be further discussed (see Fig. 2).

Fig. 2 Strategies for data storage on one-dimensional sequence defined synthetic macromolecules. Advantages and shortcomings in terms of final data storage is showcased.

For example, Lutz and co-workers reported the synthesis of eight different oligo(triazole amide) trimers that could encode an information of 3 bits (2^3 combinations) [31]. Using conventional molecular characterization techniques such as proton nuclear magnetic resonance (^{1}H NMR) spectroscopy and matrix-assisted laser desorption ionization coupled to time of flight mass spectrometry (MALDI-ToF MS), the authors were able to identify the sequence of a given trimer, and thus to retrieve the stored information. An important aspect when sequencing is the ability to read the information in the right direction (*i.e.* left to right or right to left). This problem can be exemplified in a simple way using two dimers entailing non-identical building blocks (*e.g.* A-B). Depending on the interpretation, the encoded information could be reconstructed as 1–0 or 0–1, which would corrupt the decoded information. Hence, it was easier in a first attempt, to prepare short oligomer sequences [32]. The same group later reported the synthesis of longer sequences containing an information up to 10 bits, based on the combination of different dyads in a convergent strategy [33]. On the other hand, Barner-Kowollik and co-workers demonstrated that an encoded sequence with varying composition allowed for decoding *via* tandem MS, without prior knowledge of its synthetic history [34]. It should also be noted that even if revealing an unknown chain structure is possible, this reading process remains an extremely difficult and time consuming challenge.

A simple workaround, however, is to design sequences that would fragment, whilst leaving the α- and ω-end groups intact, in addition to pre-indicating the reading direction. For instance, with this strategy Lutz and coworkers significantly improved storage capacity to 8 bytes using a sequence-defined macromolecule entailing eight blocks, in which every block consisted of 8 bits (64 bits in total) [35]. During the sequencing process, the macromolecule was fragmented *via* weak NO−C bonds, leaving a series of intact bytes. These byte-fragments contained additional labelling to indicate their respective position, which later served to identify the order of every byte along the chain. The fragmented bytes were next individually sequenced under MS^3 conditions, revealing the information in the right sequence.

However, major complications arise when larger information storage is targeted, both in writing and reading. While longer chains allow the storage of more information, the synthesis of very long sequence-defined polymers faces practical limitations (*e.g.* sterical constraints when using a solid-support, purity decrease with longer polymer chains etc.) [8,36]. In fact, to allow a storage capacity of 10 MBs *via* comparable approaches (small in terms of day to day usage), a macromolecule consisting of 80 million repeating units would be necessary. However, today's synthetic techniques simply do not allow such chain lengths and therefore render methodologies impractical on a "1 bit per repeating unit" strategy [37].

Thus, strategies that allow improving storage density per repeating unit or per chain are crucial to obtain high storage capacities on similar structures. For instance, Meier and co-workers have recently demonstrated that the capacity of a repeating unit could significantly be improved utilizing multicomponent reactions [38]. Making use of the Ugi

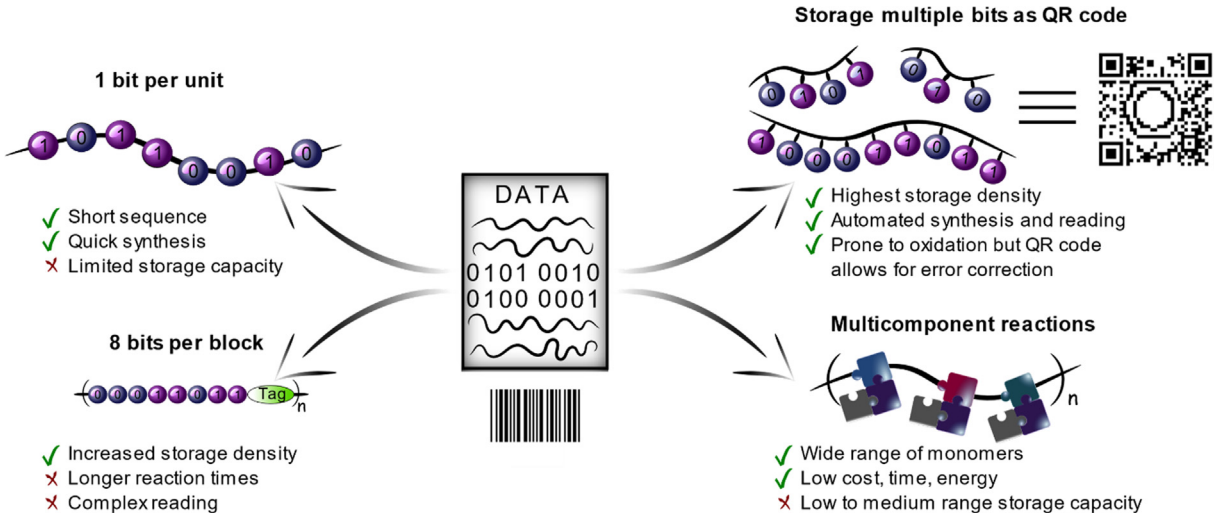

Fig. 2. Strategies for data storage on one-dimensional sequence defined synthetic macromolecules. Advantages and shortcomings in terms of final data storage is showcased.

multicomponent reaction, the encryption and successive decryption of a molecular key was reported, in which the library of available components allowed a total of 500,000 different molecular keys (18 bits), whereas a combination of six of these keys was reported to reach a size of 113 bits. To demonstrate the use of this concept, the keys were adsorbed onto paper and re-isolated for subsequent decryption of an encoded message. In a further study, sequence defined macromolecules were synthesized *via* the Biginelli and the Passerini multicomponent reactions [39]. By employing repeating units of six components at a time from a list of more than 100 available components, a density of up to 24 bits per repeating unit was achieved, demonstrated on a subset of 14 different monomers that were synthesized and evaluated. This study proved that the storage capacity could be significantly increased by the ability of preparing a large library of compounds. Moreover, the high number of different functionalities that can be incorporated along a backbone or side chain is also a crucial factor for the storage capacity improvement. For example, while only 32 different pentamers can be made using 2 different monomers, the amount of different combinations that can theoretically be reached with 20 different building blocks is 3.2×10^6 (*i.e.* base-20 system). Furthermore, using a combination of discrete oligomers, would introduce an additional dimension to work towards increasing data capacity. This point is discussed in the following sections.

3. Particular example of QR codes

QR codes are two-dimensional barcodes, which were designed by the Japanese company Denso Wave in 1994 [40]. In comparison to standard barcodes, QR codes became popular as the stored information density is high, while maintaining high precision, even after code damage [41]. In contrast to one-dimensional barcodes, which are scanned mechanically (*e.g.* by infrared light beam) and translated into a digital output, the design of QR codes is more complex. Generally, QR codes are scanned with a camera (giving a picture) and an internal program converts the image into the digital information. During the digitalization process, the program recognizes three large squares close to three of the four corners of the code, and (sometimes) a smaller square close to the forth corner (see Fig. 3) [41]. Subsequently, the small dots (bits, 1 for blank; 0 for black) are converted into information and validated with an error correcting algorithm. As it is showcased in Fig. 3, a QR code design entails error correction, demask pattern instructions and format corrections. The manual transformation of a QR code is thus tedious and time consuming and less straightforward as for linear barcodes. One factor contributing to the complexity of decoding a QR

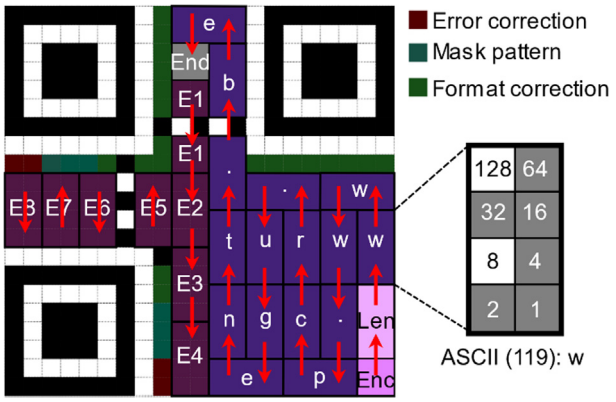

Fig. 3. Design of a QR code highlighting the most important features (*e.g.* error correction, mask pattern and format correction) as well as the data storage. The arrows indicate the reading process starting from the encoding (Enc) description to the length (Len) of the word. Afterwards, the full information can be found, which is stopped by the end (End) code (usually 0000). The rest of the code is attributed to error corrections (E1-E8).

code manually is the demasking process. The mask pattern is encoded in the QR code near the upper left square (for faster readouts and to avoid unreadability after damage, the same information is stored close to the bottom left square) [42]. Particularly, after demasking, the readout is more straightforward as illustrated in Fig. 3: (i) The information read out starts in the bottom right corner with instructions indicating which type of description has been used. The most common description is alphanumeric (*e.g.* links to websites). (ii) The next step is the length determination of the description. In the example in Fig. 3, a website was used containing 16 characters. Thus, the next 16 squares (each 2×4 pixels) encode one character for the readout. And finally, (iii), the information always ends with a 0000 sequence. The rest of the QR code is dedicated to a complicated error correction, which is not further discussed here. However, the error correction enables the read out of codes even if the information encoded on the QR code has been partly destroyed [41].

Hence, although (molecular) 1D barcodes are easier to access than 2D-systems and their readout is straightforward, a slight damage results in irreversible data loss unlike damaged 2D barcodes that still give highly precise readouts.

Du Prez and co-workers have recently exemplified a strategy for QR code writing and reading using 1D-sequence-defined oligomers [43,44].

In their approach, multiple oligomers were synthesized on an adapted peptide synthesizer, which resulted in a machine-readable 33 × 33 QR code. The latter could hold a data capacity of 1089 bits. To achieve this, the QR code was initially converted into a binary string that was encoded into multiple sequences of different lengths. While increasing chain length or number of incorporated functionalities contributes to the complexity of sequencing, being able to code and decode information in short time frames is also crucial. Thus, it is important to be able to computerize the coding and readout process, so that rapid access to data is possible. Accordingly, in this study, the coding and decoding process was fully automated by developing the "Chemcoder" algorithm that converted the obtained sequence fragments into a given QR code without any errors [43]. Another algorithm "chemreader" was developed in parallel in order to reconstruct the oligomer sequences from mass spectrometry data. This example highlights the need of fully automated strategies for faster and accurate molecular data management, in terms of synthesis, coding and readout.

4. Multi-dimensional molecular data storage

As discussed in the previous sections, data storage on 1D structures such as sequence-defined oligomers is a contemporary approach to address the challenges of the upcoming decades. A few drawbacks of encoded information on 1D structures include the reading errors caused by unwanted chemical transformations (such as oxidation and degradation), and a high synthetic effort to re-write information on an existing oligomer. More precisely, any molecular damage to the macromolecular sequence might result in a completely inaccessible readout unless self-healing structures are being implemented. As exemplified with QR codes, such drawbacks can be circumvented by progressing to writing (optical) information in a 2D fashion. Such information encoded surfaces (IES) are accessed by precision spatially resolved surface chemistry, including nanometer precise modification using photochemistry (even below the diffractive limit of light) [45,46]. Accordingly, information can be encoded on surfaces (*e.g.* represented by a 2D barcode such as a QR code) featuring sufficient high spatial resolution on nanometer scale for future readout [47–49]. Recently, the field of 2D data storage has advanced [50] and evolved to such a complexity that careful considerations have to be met. Thus, it is crucial to differentiate between the dimensionality of the data-encoded substrate (*i.e.* linear chains are regarded as 1D, branched structures and surfaces are regarded as 2D) and the dimensionality of the final readout (*i.e.* a barcode is 1D whereas matrices such as QR codes are 2D). Thus, this section is dedicated to the question of dimensionality and is structured as follows: (*i*) First, the data storage on a dendrimer is discussed. Here, the macromolecule's architecture is 2D although the information following one dendron can be regarded as linear (1D). The final readout is a 2D matrix, thus justifying the presence of this example in this section (see Fig. 4A). (*ii*) Then, an example of linear data-encoded sequences casted on a surface using layer-by-layer (LbL) assembly is presented [51]. The complexity in such a case is that the 2D substrate is designed to even form a 3D structure. The final readout, however, will be one-dimensional (following the z-axis) (see Fig. 4B). (iii) In the last part, it is shown that optical information can also be stored by utilizing non-sequence encoded polymers. Here, the data is only stored on the surface accessible via a 2D readout but never on the molecule itself (see Fig. 4C). Not included here – but also relevant in terms of 3D data storage – is a very recent work from Lutz and co-workers, where the electron diffraction of crystallized sequence-encoded oligomers was measured. Here, the information is encoded in a 3D manner (expressed by various lattice distances) [52].

Before embarking on the field of IES, a very recent contribution has been made on sequence-encoded dendrons [53]. Here, a non-linear macromolecular architecture design was utilized to generate a barcode (for a later optical readout). The authors attribute a high storage density to their system due to the Huffman-like structure of the sequence-encoded Dendron [54]. Critically, Huffman codes are never symmetrical, thus the approach of Huang *et al.* might be Huffman's code inspired [53]. Here, a first-generation dendrimer (see Fig. 4A) was synthesized with a monomer based on a binary code. Subsequently, the information of a single branch was decoded using tandem MS. Generally, decoding a full-symmetrical dendrimer (as depicted in Fig. 4A) is impossible as the reading directionality is missing. Here, linear chains have the clear advantage of alpha and omega groups representing the start and the end of the sequence. Thus, the authors opted for a dendron, which was synthesized and subsequently analyzed, but where reading directionality is still not granted. Critically, six divergent MS/MS induced fragmentation pathways leading to six possible codes were identified. The data was then transferred into a matrix barcode. The authors achieved higher information storage capacities in comparison to their linear analogues. However, the dendrimer system does not allow for universal writing of information in a single molecule.

A potential technology for future multi-dimensional encryption is the generation of IES, which are generated by LbL assembly [55]. Here, the surface is dipped into or spin coated with a solution containing a polyelectrolyte (*e.g.* polycation) to generate monolayer (theoretically) of macromolecules on the surface. Induced by the corresponding counter ion of the polyelectrolyte, multi-layer defects might, however, occur. Subsequently, the monolayer surface is dipped into or spin coated with a solution containing the oppositely charged polyelectrolyte (*e.g.* a polyanion following the previously mentioned case). Theoretically, the bilayer should adhere due to Coulomb interactions generating a perfectly negatively charged surface. In practice, however, superlinear growth is observed [56]. Here, both multilayer addition as well as interlayer diffusion contribute to a defected surface. Recently, Lutz and co-workers pioneered a study where they LbL assembled a library of sequence-defined macromolecules on a surface [55]. They reported a successful linear deposition of 16 digitally encoded polyanions resulting in a 70 nm thick film. Importantly, they prevented the superlinear growth, which would ultimately transform the heterogenic 16 layers encoded structure to a more defected layer structure by interlayer diffusion. Furthermore, Lutz and co-worker described the relevance of the sequence decomposition. As 16 layers were assembled, a total number of 16^{16} alternatives could have been created. The film directionality (*i.e.* reading either from bottom-to-top or top-to-bottom) is another parameter that they considered.

Having developed a surface-based encoding, the authors expect a high relevance for future applications. However, unlike QR codes that are x/y-axis encoded information, Lutz's approach is a z-axis encoded approach. Critically, the LbL assembly along the z-axis is thus confronted with minor drawbacks: First, any mechanical damage (*e.g.* scratch) applied to the surface will cause irreversible information loss. As most damages applied to a surface are regional, the information might still be accessible at another, but intact region. Second, there is no elaborated readout technology for LbL surfaces available yet. While digital multilayers have never been characterized to our knowledge, the work of Lutz and co-worker showcases that a successful deposition is possible and future investigations regarding the full readout must be carried out. Generally, most characterizations of layer-deposited surfaces are accessed by atomic force microscopy (AFM, for homogeneity and film thickness) [57], X-ray photoelectron spectroscopy (XPS, for elemental compositions) [58] and ellipsometry (layer thickness) [55]. Non-destructive mass spectrometric methods with depth-profiling and high resolution, as required for a successful information readout, are still in their infancy. A very recent and promising contribution by Whitesides and co-workers on sequence-encode oligopeptides based on self-assembled matrix desorption ionization (SAMDI) mass spectrometry proves however, that surface-based mass spectrometry is feasible using the cleavage of the gold/sulfur linkage by irradiation [59]. The technique requires the attachment of a permanent charge to the sequence in order to ensure successful ionization. Another promising surface technique is the time-of-flight secondary ion mass spectrometry

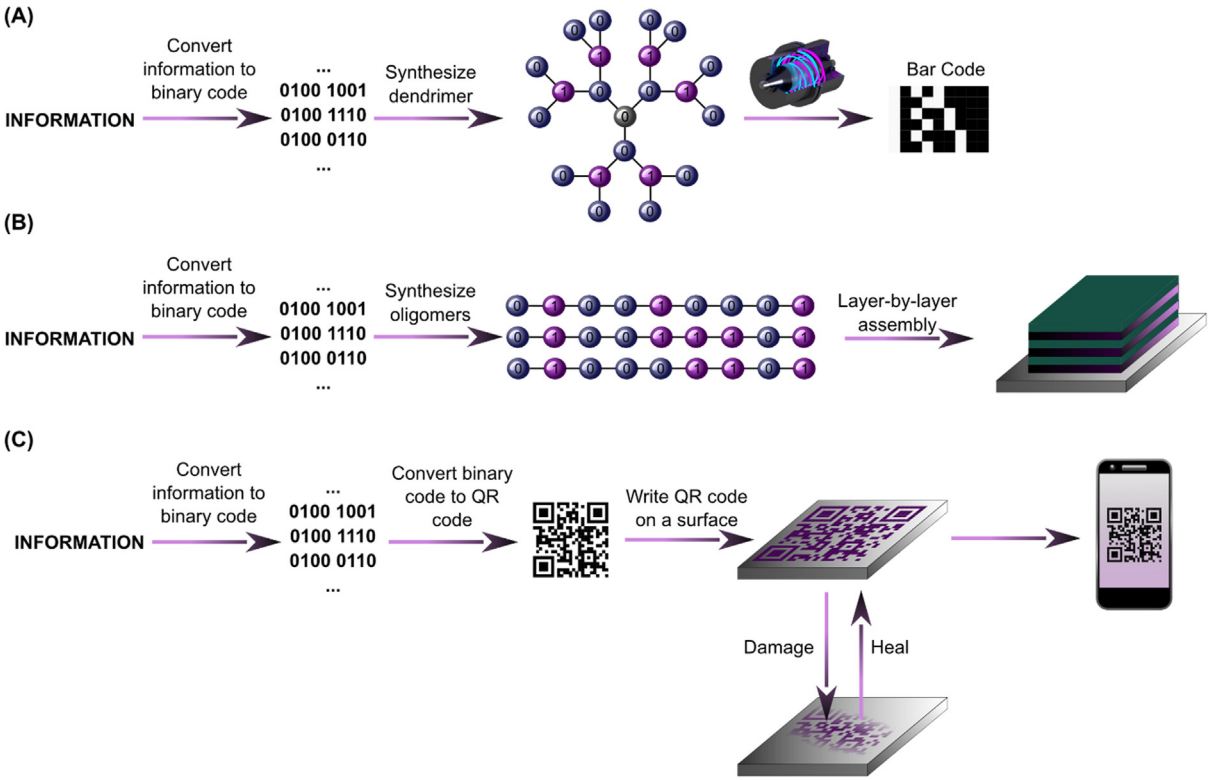

Fig. 4. Various approaches towards a 2D information storage: (A) data-encoded dendrimer design with MS/MS readout; (B) layer-by-layer assembly of sequence-encoded oligomers (depicted in purple) and disperse polyelectrolytes (depicted in cyan). So far, no readout has been reported for this approach. (C) QR code design with optical readout and self-healing characteristics.

(ToF-SIMS), which has a very high x/y axis resolution for molecular fragments [60,61]. However, the formed molecular ion fragments easily as the energy for desorption exceeds the chemical bond strength leading to complex mass spectra. Other milder techniques such as surface-assisted desorption/ionization mass spectrometry coupled to a ToF analyzer (SALDI-ToF) is highly dependent on the surface nanostructure, which interacts with the laser energy and operates as the matrix for ionization. Reported surface nanostructures include HgTe surfaces that might be hampering the LbL assembly [62]. Also, ToF-SIMS instruments are equipped with a controlled depth-profiling but SALDI-ToF technique does not allow for profiling applications. Thus, reading out sequence-encoded oligomers casted on a surface in a LbL approach might be theoretically feasible but will require detailed optimization of the measuring technique.

Spatially resolved surface patterning for information storage in the x/y plane (*e.g.* via a QR code) is challenging. Other than chemical storage where the readout relies on weak chemical bonds during multi-dimensional mass spectrometric fragmentation, optical readout is a physical process and depends on regional differences of surface roughness. Again, the LbL assembly technique provides the best results in surface texture. As Ji and co-workers showcased recently [47], a high quality QR code was generated by LbL deposition of a cationic poly (ethylene imine) and azide-containing poly(acrylic acid) derivative. Subsequently, the azide was irradiated under UV light providing robust cross-links between the individual layers. Furthermore, the authors demonstrated the self-healing characteristics of the thus encoded information after the surface has been subjected to mechanical damage. Here, the surface roughness, which is crucial for the QR code activity, was restored by exposing the surface to a saturated humidity. Ji and co-workers attribute the self-healing to the presence of humidity that favor polyelectrolytes mobility in the un-crosslinked regions and therefore enable the retrieval of the information patterned on the film (represented by the differences in light scattering). It is important to mention here that the polymeric material itself was not information encoded. The information depth resulted from the surface patterning, creating a QR code for subsequent readout. Thus, a combination of x/y plane storage (QR code) and sequence encoding along the z-axis by a LbL assembly could yield unprecedented storage densities with easy initial optical readout (QR code) and non-optical (but complex) readout in z-direction. As discussed earlier, Lutz and co-workers have not implemented an easy readout (*e.g.* by ToF-SIMS) for the LbL assembled system as the entire system was based on sequence-defined oligomers where the monomer units were covalently linked together. In their study, a supramolecular approach might have been a valuable alternative *via* the combination of sequence-defined macromolecules featuring supramolecular assembly.

In a non-sequence defined fashion, Besenius and co-workers reported the controlled growth by sequential deposition of oppositely charged oligopeptides on gold surfaces [61]. As the resulting stacked structures are kinetically trapped [62], it might be possible to sequentially reverse the supramolecular forces and to access an easy readout in z-axis (Fig. 5). This methodology could be translated to sequence-defined polymers in order to design new unprecedented pathways for multi-dimensional data storage.

5. Conclusion

How data are stored in the future is a task of utmost importance that has to be tackled within this decade. Promising technologies and protocols give high precision macromolecules that could be used for encoding information. So far, one-dimensional data storage with linear macromolecules has received major attention. The most advanced candidate, DNA, serves as life's library and was given millions of years to evolve. Scientists see DNA's high potential for cold data storage due to the large research that has been done on DNA synthesis and sequencing. Nowadays, the technology is available to generate data-

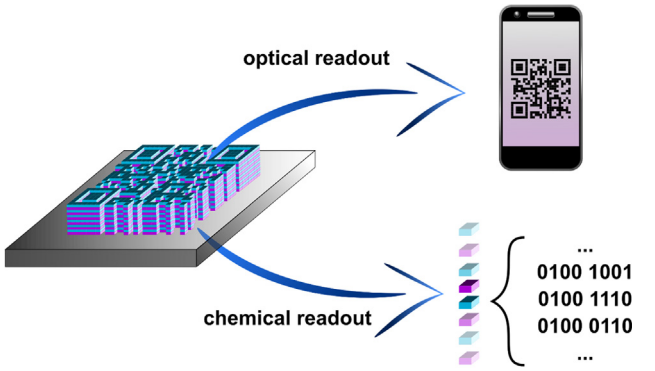

Fig. 5. Perspective of data storage on surface featuring x/y plane matrix (*e.g.* QR code) for optical readout and sequence-defined supramolecular assembly along z-axis for chemical readout.

encoded DNA strands and to multiply them. Some issues (*e.g.* base incompatibilities, reading errors and the limit of the binary code to the four nucleobases) stimulate the demand for fully synthetic and more robust alternatives. Yet, one-dimensional information encoding is susceptible to irreversible data loss upon undesired chemical transformation including oxidation, photoprocesses and hydrolysis. Two-dimensional data storage, on the other hand, is a promising alternative that has raised only little attention so far. Transforming a linear binary code into two-dimensional codes such as QR codes will circumvent some of the one-dimensional analogues' drawbacks. At the same time, the readout of 2D codes is a physical process (thus rapid and precise) whereas the readout of linear sequences still relies on sophisticated and time-consuming physicochemical methods such as mass spectrometry.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

The authors would like to acknowledge financial support from the FWO under EOS-project 30650939. Furthermore, the authors thank Nezha Badi (UGent) for critically revising the manuscript and for fruitful discussions.

References

[1] M. Aubert, A. Brumm, M. Ramli, T. Sutikna, E.W. Saptomo, B. Hakim, M.J. Morwood, G.D. van den Bergh, L. Kinsley, A. Dosseto, Pleistocene cave art from Sulawesi Indonesia, Nature 514 (2014) 223.
[2] M. Hilbert, P. López, The world's technological capacity to store, communicate, and compute information, Science 332 (2011) 60–65.
[3] Z.-W. Xu, Cloud-Sea computing systems: towards thousand-fold improvement in performance per watt for the coming zettabyte era, J. Comput. Sci. Technol. 29 (2014) 177–181.
[4] V. Zhirnov, R.M. Zadegan, G.S. Sandhu, G.M. Church, W.L. Hughes, Nucleic acid memory, Nat. Mater. 15 (2016) 366–370.
[5] L. Ceze, J. Nivala, K. Strauss, Molecular digital data storage using DNA, Nat. Rev. Genet. 20 (2019) 456–466.
[6] C. Bancroft, T. Bowler, B. Bloom, C.T. Clelland, Long-Term storage of information in DNA, Science 293 (2001) 1763–1765.
[7] Y. Erlich, D. Zielinski, DNA fountain enables a robust and efficient storage architecture, Science 355 (2017) 950–954.
[8] M.G.T.A. Rutten, F.W. Vaandrager, J.A.A.W. Elemans, R.J.M. Nolte, Encoding information into polymers, Nat. Rev. Chem. 2 (2018) 365–381.
[9] R. Chepesiuk, Where the chips fall: environmental health in the semiconductor industry, Environ. Health Perspect. 107 (1999).
[10] E.D. Williams, R.U. Ayres, M. Heller, The 1.7 kilogram microchip: energy and material use in the production of semiconductor devices, Environ. Sci. Technol. 36 (2002) 5504–5510.
[11] J. Davis, Microvenus, Art J. 55 (1996) 70–74.
[12] G.M. Church, Y. Gao, S. Kosuri, Next-generation digital information storage in DNA, Science 337 (2012) 1628.
[13] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E.M. LeProust, B. Sipos, E. Birney, Towards practical, high-capacity, low-maintenance information storage in synthesized DNA, Nature 494 (2013) 77.
[14] M. Peplow, Rebooting the molecular computer, ACS Cent. Sci. 2 (2016) 874–877.
[15] A.D. Tyler, L. Mataseje, C.J. Urfano, L. Schmidt, K.S. Antonation, M.R. Mulvey, C.R. Corbett, Evaluation of oxford nanopore's MinION sequencing device for microbial whole genome sequencing applications, Sci. Rep. 8 (2018) 10931.
[16] M. Boukhet, N.F. König, A.A. Ouahabi, G. Baaken, J.-F. Lutz, J.C. Behrends, Translocation of precision polymers through biological nanopores, Macromol. Rapid Commun. 38 (2017) 1700680.
[17] J.J. Kasianowicz, S.M. Bezrukov, On three decades of nanopore sequencing, Nat Biotechnol. 34 (2016) 481–482.
[18] J.J. Schwartz, C. Lee, J. Shendure, Accurate gene synthesis with tag-directed retrieval of sequence-verified DNA molecules, Nat. Methods. 9 (2012) 913–915.
[19] M.G. Ross, C. Russ, M. Costello, A. Hollinger, N.J. Lennon, R. Hegarty, C. Nusbaum, D.B. Jaffe, Characterizing and measuring bias in sequence data, Genome Biol. 14 (2013) R51.
[20] B. Röder, K. Frühwirth, C. Vogl, M. Wagner, P. Rossmanith, Impact of long-term storage on stability of standard DNA for nucleic acid-based methods, J. Clin. Microbiol. 48 (2010) 4260–4262.
[21] H.H. Nguyen, J. Park, S.J. Park, C.-S. Lee, S. Hwang, Y.-B. Shin, T.H. Ha, M. Kim, Long-term stability and integrity of plasmid-based DNA data storage, Polymers 10 (2018) 28.
[22] J.-F. Lutz, M. Ouchi, D.R. Liu, M. Sawamoto, Sequence-controlled polymers, Science 341 (2013) 1238149.
[23] H. Colquhoun, J.-F. Lutz, Information-containing macromolecules, Nat. Chem. 6 (2014) 455–456.
[24] J.-F. Lutz, Coding macromolecules: inputting information in polymers using monomer-based alphabets, Macromolecules 48 (14) (2015) 4759–4767.
[25] M. Schmidt, Xenobiology: a new form of life as the ultimate biosafety tool, BioEssays 32 (2010) 322–331.
[26] N. Badi, J.-F. Lutz, Sequence control in polymer synthesis, Chem. Soc. Rev. 38 (2009) 3383–3390.
[27] J. Mao, W. Zhang, S.Z. Cheng, C. Wesdemiotis, Analysis of monodisperse, sequence-defined, and POSS-functionalized polyester copolymers by MALDI tandem mass spectrometry, Eur. J. Mass Spectrom. 25 (2019) 164–174.
[28] K.S. Wetzel, M.A.R. Meier, Monodisperse, sequence-defined macromolecules as a tool to evaluate the limits of ring-closing metathesis, Polym. Chem 10 (2019) 2716–2722.
[29] S.C. Solleder, R.V. Schneider, K.S. Wetzel, A.C. Boukis, M.A.R. Meier, Recent progress in the design of monodisperse, sequence-defined macromolecules, Macromol. Rapid Commun. 38 (2017) 1600711.
[30] S. Martens, J.O. Holloway, F.E. Du Prez, Click and click-inspired chemistry for the design of sequence-controlled polymers, Macromol. Rapid Commun. 38 (24) (2017) 1700469.
[31] T.T. Trinh, L. Oswald, D. Chan-Seng, J.-F. Lutz, Synthesis of molecularly encoded oligomers using a chemoselective "AB + CD" iterative approach, Macromol. Rapid Commun. 35 (2014) 141–145.
[32] R.K. Roy, A. Meszynska, C. Laure, L. Charles, C. Verchin, J.-F. Lutz, Design and synthesis of digitally encoded polymers that can be decoded and erased, Nat Commun. 6 (2015) 7237.
[33] R.K. Roy, C. Laure, D. Fischer-Krauser, L. Charles, J.-F. Lutz, Convergent synthesis of digitally-encoded poly(alkoxyamine amide)s, Chem. Commun. 51 (2015) 15677–15680.
[34] N. Zydziak, W. Konrad, F. Feist, S. Afonin, S. Weidner, C. Barner-Kowollik, Coding and decoding libraries of sequence-defined functional copolymers synthesized via photoligation, Nat. Commun. 7 (2016) 13672.
[35] A. Al Ouahabi, J.-A. Amalian, L. Charles, J.-F. Lutz, Mass spectrometry sequencing of long digital polymers facilitated by programmed inter-byte fragmentation, Nat. Commun. 8 (2017) 967.
[36] J.O. Holloway, C. Mertens, F. Du Prez, N. Badi, Automated synthesis protocol of sequence-defined oligo-urethane-amides using thiolactone chemistry, Macromol. Rapid Commun. 40 (2019) 1800685.
[37] J.O. Holloway, K.S. Wetzel, S. Martens, F.E. Du Prez, M.A.R. Meier, Direct comparison of solution and solid phase synthesis of sequence-defined macromolecules, Polym. Chem. 10 (2019) 3859–3867.
[38] A.C. Boukis, K. Reiter, M. Frölich, D. Hofheinz, M.A.R. Meier, Multicomponent reactions provide key molecules for secret communication, Nat. Commun. 9 (2018) 1439.
[39] A.C. Boukis, M.A.R. Meier, Data storage in sequence-defined macromolecules via multicomponent reactions, Eur. Polym. J. 104 (2018) 32–38.
[40] K.H. Pandya, H.J. Galiyawala, A survey on QR Codes: in context of research and application, Int. J. Emerg. Technol. Adv. Eng. 4 (2014) 258–262.
[41] A. Grillo, A. Lentini, M. Querini, G.F. Italiano, High capacity colored two dimensional codes, Proc. IMCSIT (2010) 709–716.
[42] G. Singh, S. Rani, A. Saroa, S. Girdhar, J. Singh, A. Arora, D. Aulakh, M. Wriedt, Organosilatranes with thioester-anchored heterocyclic ring assembly: Cu2+ ion binding and fabrication of hybrid silica nanoparticles, RSC Adv. 5 (2015) 65963–65974.
[43] S. Martens, A. Landuyt, P. Espeel, B. Devreese, P. Dawyndt, F. Du Prez, Multifunctional sequence-defined macromolecules for chemical data storage, Nat. Commun. 9 (2018) 4451.
[44] S. Martens, J. Van den Begin, A. Madder, F.E. Du Prez, P. Espeel, Automated synthesis of monodisperse oligomers, featuring sequence control and tailored functionalization, J. Am. Chem. Soc. 138 (2016) 14182–14185.
[45] P. Mueller, M.M. Zieger, B. Richter, A.S. Quick, J. Fischer, J.B. Mueller, L. Zhou,

G.U. Nienhaus, M. Bastmeyer, C. Barner-Kowollik, M. Wegener, Molecular switch for sub-diffraction laser lithography by photoenol intermediate-state cis-trans isomerization, ACS Nano 11 (2017) 6396–6403.
[46] R. Ogaki, M. Alexander, P. Kingshott, Chemical patterning in biointerface science, Mater. Today 13 (2010) 22–35.
[47] X.-C. Chen, W.-P. Huang, K.-F. Ren, J. Ji, Self-healing label materials based on photo-cross-linkable polymeric films with dynamic surface structures, ACS Nano 12 (2018) 8686–8696.
[48] X. Duan, N. Liu, Scanning plasmonic color display, ACS Nano 12 (2018) 8817–8823.
[49] O. Gökçe, C. Mercandetti, E. Delamarche, High-content optical codes for protecting rapid diagnostic tests from counterfeiting, Anal. Chem. 90 (2018) 7383–7390.
[50] D. Karamessini, S. Poyer, L. Charles, J.-F. Lutz, 2D sequence-coded oligourethane barcodes for plastic materials labeling, Macromol. Rapid Commun. 38 (2017) 1700426.
[51] J.J. Richardson, M. Björnmalm, F. Caruso, Technology-driven layer-by-layer assembly of nanofilms, Science 348 (2015) aaa2491.
[52] B.É. Petit, B. Lotz, J.-F. Lutz, About the crystallization of abiotic coded matter, ACS Macro Lett. 8 (7) (2019) 779–782.
[53] Z. Huang, Q. Shi, J. Guo, F. Meng, Y. Zhang, Y. Lu, Z. Qian, X. Li, N. Zhou, Z. Zhang, X. Zhu, Binary tree-inspired digital dendrimer, Nat. Commun. 10 (2019) 1918.
[54] D.A. Huffman, A method for the construction of minimum-redundancy codes, Proc. IRE 40 (1952) 1098–1101.
[55] R. Szweda, M. Tschopp, O. Felix, G. Decher, J.-F. Lutz, Sequences of sequences: spatial organization of coded matter through layer-by-layer assembly of digital polymers, Angew. Chem. Int. Ed. 57 (2018) 15817–15821.
[56] N.S. Zacharia, D.M. DeLongchamp, M. Modestino, P.T. Hammond, Controlling diffusion and exchange in layer-by-layer assemblies, Macromolecules 40 (2007) 1598–1603.
[57] F.J. Giessibl, Advances in atomic force microscopy, Rev. Mod. Phys. 75 (2003) 949–983.
[58] A. Benkaddour, K. Jradi, S. Robert, C. Daneault, Grafting of polycaprolactone on oxidized nanocelluloses by click chemistry, Nanomaterials 3 (2013) 141–157.
[59] B.J. Cafferty, A.S. Ten, M.J. Fink, S. Morey, D.J. Preston, M. Mrksich, G.M. Whitesides, Storage of information using small organic molecules, ACS Cent. Sci. 5 (5) (2019) 911–916.
[60] D. Abt, B.V.K.J. Schmidt, O. Pop-Georgievski, A.S. Quick, D. Danilov, N.Y. Kostina, M. Bruns, W. Wenzel, M. Wegener, C. Rodriguez-Emmenegger, C. Barner-Kowollik, Designing molecular printboards: a photolithographic platform for recordable surfaces, Chem. Eur. J. 21 (2015) 13186–13190.
[61] H. Frisch, E.-C. Fritz, F. Stricker, L. Schmüser, D. Spitzer, T. Weidner, B.J. Ravoo, P. Besenius, Kinetically controlled sequential growth of surface-grafted chiral supramolecular copolymers, Angew. Chem. Int. Ed. 55 (2016) 7242–7246.
[62] E. Krieg, M.M.C. Bastings, P. Besenius, B. Rybtchinski, Supramolecular polymers in aqueous media, Chem. Rev. 116 (2016) 2414–2477.

Jan Steinkoenig completed his Bachelor (2012) and Master (2015) studies at Karlsruhe Institute of Technology (KIT). The master thesis was supervised by Christopher Barner-Kowollik and was dedicated to poly(ionic liquid)s entailing reducing agents. Afterwards, he joined the research team of Barner-Kowollik both at the KIT node and the Queensland University of Technology (QUT) node to complete his PhD studies on the characterization of macromolecules via mass spectrometry. After graduation in 2017, he joined the PCR group of Filip Du Prez in 2018 working on sequence-defined macromolecules. Since 2019 he is a FWO fellow.

Resat Aksakal has received his BSc degree in Chemistry from the University of Düsseldorf, Germany (2012) and has obtained an MSc in Polymer Chemistry at the University of Warwick (2013). He has pursued a PhD degree (2017) in Materials Science under the guidance of Dr. Remzi Becer at Queen Mary University of London on controlled radical polymerisation techniques. He is currently working as a post-doctoral researcher (FWO Fellow) in the group of Prof. Filip Du Prez and his research project focuses on the synthesis of sequence-defined polymers.

Filip Du Prez is since 1999 heading the Polymer Chemistry Research Group (www.PCR.UGent.be) at Ghent University in which about 25 researchers are dealing with the development of new polymer structures, such as sequence-defined polymers, exploration of powerful functionalization methods and the design of dynamic/renewable polymer materials. In 2008, he had a Visiting Professor position at the CAMD Research Center (UNSW, Sydney). He is author of about 300 peer-reviewed publications (h-index 50), 12 patents, 10 book chapters and (co-)chairman of more than 10 (inter)national conferences on polymer chemistry related topics. Since 2018, he is associate editor of Polymer Chemistry.