



INSTITUT DE FRANCE
Académie des sciences

Comptes Rendus

Chimie

Eline Laurent, Jean-Arthur Amalian, Thibault Schutz, Kevin Launay, Jean-Louis Clément, Didier Gigmes, Alexandre Burel, Christine Carapito, Laurence Charles, Marc-André Delsuc and Jean-François Lutz

Storing the portrait of *Antoine de Lavoisier* in a single macromolecule

Volume 24, issue 1 (2021), p. 69-76

Published online: 4 March 2021

<https://doi.org/10.5802/crchim.72>



This article is licensed under the
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.
<http://creativecommons.org/licenses/by/4.0/>



Les Comptes Rendus. Chimie sont membres du
Centre Mersenne pour l'édition scientifique ouverte
www.centre-mersenne.org
e-ISSN : 1878-1543



Preliminary communication / *Communication préliminaire*

Storing the portrait of *Antoine de Lavoisier* in a single macromolecule

Stockage du portrait d'Antoine Lavoisier dans une macromolécule

Eline Laurent^a, Jean-Arthur Amalian^{® b}, Thibault Schutz^a, Kevin Launay^b, Jean-Louis Clément^{® b}, Didier Gimes^{® b}, Alexandre Burel^c, Christine Carapito^{® c}, Laurence Charles^{® *, b}, Marc-André Delsuc^{® *, d} and Jean-François Lutz^{® *, a}

^a Université de Strasbourg, CNRS, Institut Charles Sadron UPR22, 23 rue du Loess, 67034 Strasbourg Cedex 2, France

^b Aix Marseille Université, CNRS, Institut de Chimie Radicalaire, UMR 7273, 23 Av Escadrille Normandie-Niemen, 13397 Marseille Cedex 20, France

^c Université de Strasbourg, CNRS, Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC, 23 rue du Loess, 67034 Strasbourg Cedex 2, France

^d Institut de Génétique et de Biologie Moléculaire et Cellulaire, INSERM, U596, CNRS, UMR7104, Université de Strasbourg, 1 rue Laurent Fries, 67404, Illkirch-Graffenstaden, France

E-mails: eline.laurent@etu.unistra.fr (E. Laurent), jean-arthur.amalian@univ-amu.fr (J.-A. Amalian), thibault.schutz@etu.unistra.fr (T. Schutz), kevin.launay@univ-amu.fr (K. Launay), jean-louis.clement@univ-amu.fr (J.-L. Clément), Didier.Gimes@univ-amu.fr (D. Gimes), a.burel@unistra.fr (A. Burel), ccarapito@unistra.fr (C. Carapito), laurence.charles@univ-amu.fr (L. Charles), madelsuc@unistra.fr (M.-A. Delsuc), jflutz@unistra.fr (J.-F. Lutz)

Abstract. A pixelated portrait of Antoine de Lavoisier was stored in a digitally-encoded poly(phosphodiester). The storage capacity attained in this work (440 bits/chain) is the highest ever reported for a synthetic informational polymer. To reach this very high capacity, a combination of data compression and polymer design was used. For instance, the digital polymer was synthesized by automated phosphoramidite chemistry using an optimized set of nineteen building blocks (8 coded monomers, 10 mass tags and one alkoxyamine linker enabling decryption by tandem mass spectrometry). Consequently, the polymer could be comprehensively decoded by electrospray mass spectrometry and the portrait of the French chemist was recovered.

Résumé. Un portrait pixelisé d'Antoine Lavoisier a été stocké à l'échelle moléculaire dans un poly(phosphodiester) numérique. La capacité de stockage de 440 bits/chaîne obtenue dans ce travail est la plus haute jamais atteinte avec un polymère synthétique. Pour ce faire, un algorithme de

* Corresponding authors.

compression et un design macromoléculaire avancé ont été combinés. En l'occurrence, les polymères ont été synthétisés par chimie de phosphoramidite à l'aide d'une chimiothèque de 19 composés (8 monomères codés, 10 marqueurs de masse et un espaceur alcoxyamine facilitant la lecture par spectrométrie de masse en tandem). Ainsi, le polymère a pu être décodé par spectrométrie de masse et l'image du chimiste a pu être extraite du polymère.

Keywords. Sequence-controlled polymers, Digital polymers, Information-containing macromolecules, Solid-phase synthesis, Mass spectrometry sequencing.

Manuscript received 7th December 2020, revised and accepted 11th February 2021.

1. Introduction

Macromolecular storage is an interesting emerging trend for saving data at the molecular scale [1–3]. In this approach, molecular building blocks (i.e. monomers) are set as basic information units and their arrangement in a polymer chain is exploited as a processable (i.e. writable, readable and editable) information string. This can be achieved with biological informational polymers, such as DNA [4,5], but also using a wide variety of synthetic copolymers, as demonstrated by our group [6–9] and others [10–13]. Since academic research about automated chemical synthesis (i.e. writing) [14] and automated sequencing (i.e. reading) [15] of DNA is much more mature than the one about synthetic informational polymers, DNA data storage currently gives access to much higher storage capacities than synthetic analogues [5]. Yet, it has been recently proposed that the storage properties (e.g. density, capacity and stability) of synthetic polymers might overrule those of DNA on the long term [16]. Indeed, the molecular structure of synthetic macromolecules can be varied almost infinitely to attain optimal properties, whereas the one of DNA is fixed by biological constraints.

Among the different types of synthetic polymers that have been reported for data storage, abiotic poly(phosphodiester)s are a very promising option [16]. These sequence-defined macromolecules are synthesized by automated phosphoramidite chemistry; an approach that is also used for the chemical synthesis of DNA [17]. However, non-natural monomers are used in these syntheses instead of nucleoside phosphoramidites [18]. Consequently, in terms of molecular structure, these poly(phosphodiester)s have nothing in common with nucleic acids at the exception of phosphates formed by phosphorylation and subsequent oxidation. Hence, these polymers allow information

storage but their molecular structure is not restricted by biological constraints. Over past years, we have therefore gradually improved their design to render them more and more suitable for data storage applications. The first generation of digital poly(phosphodiester)s was prepared in 2015 using a binary alphabet (i.e. two different monomers allowing a maximum storage density of 1 bit/monomer) [19]. Initially, the polymers were synthesized manually but later that year, we reported the automated synthesis of longer chains [20]. However, this first generation of digitally-encoded poly(phosphodiester)s was very difficult to decrypt using sequencing tools such as tandem mass spectrometry (MS/MS) [21] or nanopore sequencing [22,23]. This issue was solved in 2017, when we reported the design of long poly(phosphodiester)s that can be decoded by MS/MS using a routine mass spectrometer [24]. To do so, tetramethylpiperidinyloxy (TEMPO)-based alkoxyamine motifs were periodically included in the polymer chains using an appropriate phosphoramidite building block. When subjected to MS/MS conditions, these sites break preferentially, thus leading to a library of coded fragments of defined size. Each fragment is pre-labelled with a mass-tag that permits its identification and therefore, after performing the sequencing of all fragments in pseudo-MS³ conditions, the complete information sequence can be recovered. Yet, a poly(phosphodiester) with a maximum storage capacity of 77 bits/chain was achieved in this work and its decryption required a time-consuming manual interpretation of the MS/MS and MS³ spectra. Although a software named MS-DECODER was developed the same year for the decoding of synthetic digital polymers [25], it could not be applied to these optimized poly(phosphodiester)s because TEMPO-based alkoxyamines lead to intense side peaks that hinder automated decryption. This problem was solved in 2020 using an optimized alkoxyamine motif

named RISC2 (RISC stands for Ring InSide Chain) that minimizes side-products formation [26]. In parallel, we also reported in 2020 expanded monomer alphabets containing 4- or 8-symbols, which allow storage densities of 2 or 3 bits/monomer, respectively [27]. All these recent developments enable the design of high-capacity digital polymers. However, at present, the biggest digital poly(phosphodiester) ever decrypted had a storage capacity of 144 bits/chain, which is still far from the maximum possible. For instance, no compression algorithm was used to encode this polymer and its decoding was not performed automatically.

In this context, we report herein the synthesis and automated decryption of a poly(phosphodiester) of very high storage capacity. This was achieved by combining the four following solutions: (i) using an expanded monomer alphabet of 8-symbols, (ii) using the optimized alkoxyamine RISC2, (iii) developing an appropriate and expanded set of fragment markers, and (iv) using a compression algorithm. In order to illustrate the efficacy of this design, we describe here the preparation of a digital macromolecule that stores the portrait of the most renowned French Chemist, Antoine de Lavoisier (also known as Antoine Lavoisier after the French revolution). A pixelated version of a known engraving of Lavoisier was created, compressed and stored in a digital poly(phosphodiester). The single chain capacity of 440 bits/chain attained in this work is the highest ever reported for a synthetic informational polymer.

2. Results and discussion

The polymer studied in this work was synthesized by solid-phase phosphoramidite chemistry on an automated DNA synthesizer [14]. Synthesis was performed on a crosslinked polystyrene resin. In order to prepare a high-capacity digital polyphosphodiester, three types of phosphoramidite building blocks are necessary: (i) coded comonomers that allow information storage [19], (ii) an alkoxyamine-containing linker that guides fragmentation in MS/MS sequencing [24], and (iii) mass tags that permit to identify mass spectrometry fragments [24]. Figure S1 shows the molecular structure of all the phosphoramidite building blocks used in this work and Figure 1 displays the general molecular structure of the resulting poly(phosphodiester). Digital-encoding

was achieved using an alphabet composed of eight different monomers (**M**₁–**M**₈ in Figure S1 and Figure 1) [27]. For instance, **M**₁, **M**₂, **M**₃, **M**₄, **M**₅, **M**₆, **M**₇, and **M**₈ code for the triads 000, 001, 010, 011, 100, 101, 110 and 111, respectively. As mentioned in the introduction, the recently-reported RISC2 compound was used in this work as alkoxyamine-containing linker [26]. It was included periodically in the chain (i.e. every eight coded monomers) as shown in Figure 1. Furthermore, ten different nucleosides were used as mass tags (T, C, A, G, B, I, F, R, P, D in Figure 1). Seven of them have already been used in previous works [24], while the other three, namely R, P and D, have been investigated for the first time in the present study. In order to enable the unambiguous identification of a chain-fragment, each mass tag shall have an exact molar mass that satisfies strict criteria, as previously described [24]. The new markers R, P and D were therefore carefully selected according to these rules. All the mass tags were incorporated in the polymer chains using the corresponding phosphoramidite monomers (Figure S1) with the exception of T that comes from the preloaded polystyrene support.

This molecular alphabet was used herein to synthesize a macromolecule that stores the pixelated portrait of Antoine de Lavoisier (Figure 2). The black & white Lavoisier picture was coded on a 20 × 22 grid (440 bits) with 1 bit coding for black pixels and 0 for white pixels. The picture, considered as a bit stream, was coded using an arithmetic coding compressing scheme developed by “project Nayuki” (code available in supporting information document) [28]. The series of 440 bits was first linearized and a checksum bit was added to it, and zero-padded to a length multiple of 8. Then a frequency table for 8 bits sequences was computed and used for arithmetic coding to compress the bit stream to 264 bits. This compressed sequence was then expressed as a sequence of 88 coded monomers, 10 alkoxyamine spacers and 10 mass tags (108 building blocks in total), as shown in Figure 2. Following previously established-conventions [20,24,27], the reading direction of the polymer was set opposite to the synthesis direction. The mass tags sequence is therefore read in the following order (from left to right in Figure 2): no tag, D, P, R, F, I, B, G, A, C, T. Following the acronym conventions set in Figure 1, the primary structure of the synthesized polymer is as follow:

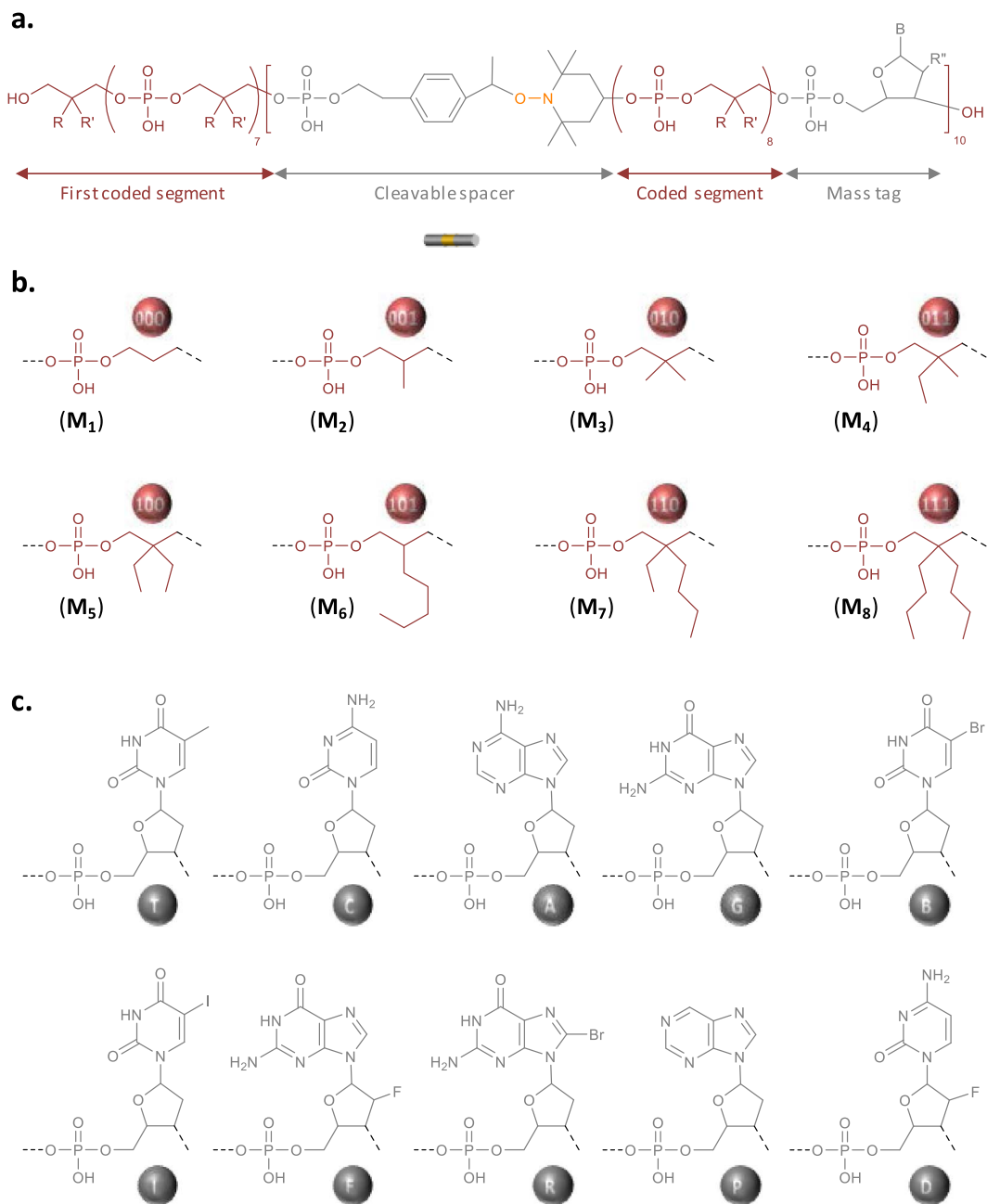


Figure 1. Design of the digital polymer. (a) General molecular structure of the digitally-encoded poly(phosphodiester) synthesized in this work. (b) Molecular structure of the eight different synthons that permit to code binary information in the chains. (c) Molecular structure of the ten different mass tags, which facilitate the decryption of the digital sequence by mass spectrometry.

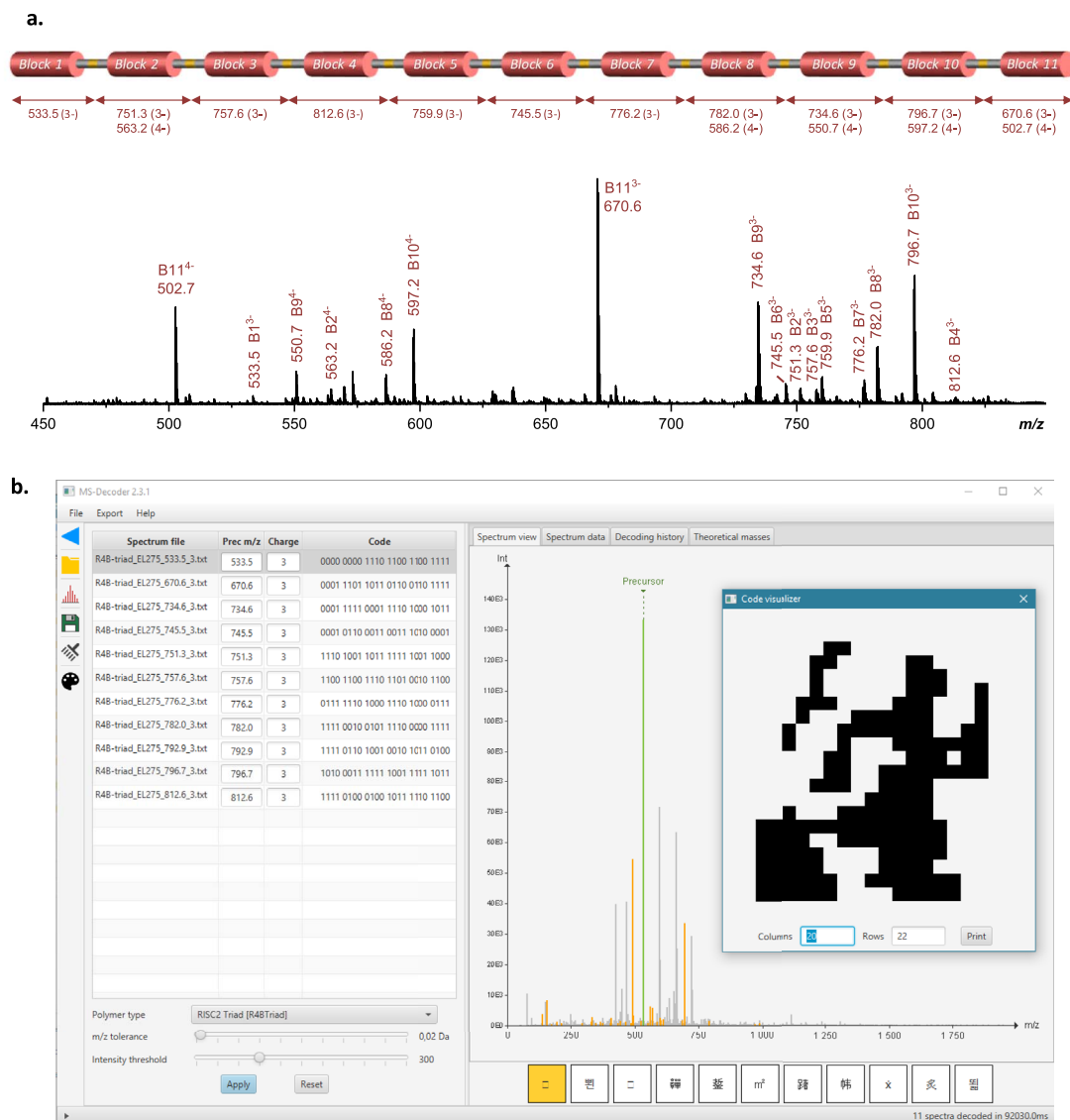


Figure 3. Polymer decryption. (a) ESI-MS spectrum recorded for the Lavoisier-containing polymer in negative mode, using a cone voltage of 60 V to induce cleavage of all alkoxyamine bonds. (b) Screenshot of the MS-DECODER interface for sequence annotation and reading.

indicates that the main population of the multimodal distribution has a molecular weight which is close to the expected theoretical value. Of note, the dn/dc of the recorded polymer was only roughly estimated and therefore the measured molecular weight values are not meant to be absolute. Overall, ESI-MS, MALDI-TOF-MS and SEC results tend to indicate that the targeted polymer was synthesized, even though it is not pure. All signature fragments

of the targeted sequence were confirmed by ESI-MS experiments performed with the cone voltage set to 60 V to induce cleavage of all alkoxyamine bonds: as shown in Figure 3a (and Table S1), the eleven coded blocks of the polymer were all individually observed in ESI-MS. This proves undoubtedly that the targeted primary structure was prepared. Consequently, all the sequence fragments were then subjected to further CID fragmentation and individually sequenced

(Figures S4–S7)¹. It shall be remarked that the polydispersity of the initial polymer does not affect sequencing because the ion sampled for fragmentation is selectively chosen based on its m/z value. The obtained spectra were analyzed both manually and by the MS-DECODER software [25]. For this, the algorithm of the software was upgraded for decoding the 8-symbols alphabet and the 11 mass-tags that were used in this work. Overall, the complete information sequence was deciphered in about one minute, which is drastically faster than manual sequencing. During analysis, the bit stream to be decoded was determined from the observed monomer sequence, the frequency table was used to decompress the message, and the checksum bit was verified. The 440 bits thus obtained can then be displayed as a 20×22 picture (Figure 3b).

3. Conclusion

In summary, the synthesis and mass spectrometry decoding of an informational poly(phosphodiester) allowing an unprecedented storage capacity of 440 bits/chain was described in this work. This was attained by optimizing the molecular design of the polymer and by employing an appropriate compression algorithm. In terms of polymer design, a set of nineteen building blocks (eight coded monomers, ten mass tags and one cleavable spacer) was used to synthesize this polymer. The polymer was then decoded by multistage mass spectrometry. The intact macromolecule could be observed in MALDI-MS as a singly charged macro-anion but the highly charged species generated in ESI dissociated during its transfer in the vacuum side of the instrument. Nevertheless, the library of spontaneously-formed in-source fragments confirmed that the polymer was synthesized and allowed complete deciphering of the information sequence. Furthermore, thanks to the use of the optimized alkoxyamine spacer RISC2, the reading of the polymer could be automated using the MS-DECODER software. As a proof-of-principle, the portrait of Antoine de Lavoisier was stored at

the molecular scale in the present work. Overall, this study is an homage to one of the founding fathers of modern chemistry and underlines that the limits of chemically-synthesized informational polymers have not yet been reached [29,30].

Acknowledgments

These results were obtained with the support of the French National Research Agency (ANR project 00111001, grant numbers ANR-16-CE29-0004-01, ANR-16-CE29-0004-02 and CE29-0004-03). LC acknowledges support from Spectropole, the Analytical Facility of Aix-Marseille University, by allowing a special access to the instruments purchased with European Funding (FEDER OBJ2142-3341). The SEC results shown in the publication were obtained with the help of the polymer characterization service of the Institut Charles Sadron. The authors thank Mélanie Legros and Catherine Foussat (Institut Charles Sadron) for helpful discussions.

Supplementary data

Supporting information for this article is available on the journal's website under <https://doi.org/10.5802/crchim.72> or from the author.

References

- [1] J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science*, 2013, **341**, article no. 1238149.
- [2] H. Colquhoun, J.-F. Lutz, *Nat. Chem.*, 2014, **6**, 455-456.
- [3] M. G. T. A. Rutten, F. W. Vaandrager, J. A. A. W. Elemans, R. J. M. Nolte, *Nat. Rev. Chem.*, 2018, **2**, 365-381.
- [4] V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church, W. L. Hughes, *Nat. Mater.*, 2016, **15**, 366-370.
- [5] L. Ceze, J. Nivala, K. Strauss, *Nat. Rev. Genet.*, 2019, **20**, 456-466.
- [6] T. T. Trinh, L. Oswald, D. Chan-Seng, J.-F. Lutz, *Macromol. Rapid Commun.*, 2014, **35**, 141-145.
- [7] R. K. Roy, A. Meszynska, C. Laure, L. Charles, C. Verchin, J.-F. Lutz, *Nat. Commun.*, 2015, **6**, article no. 7237.
- [8] U. S. Gunay, B. E. Petit, D. Karamessini, A. Al Ouahabi, J.-A. Amalian, C. Chendo, M. Bouquey, D. Gigmès, L. Charles, J.-F. Lutz, *Chem*, 2016, **1**, 114-126.
- [9] G. Cavallo, A. Al Ouahabi, L. Oswald, L. Charles, J.-F. Lutz, *J. Am. Chem. Soc.*, 2016, **138**, 9417-9420.
- [10] A. C. Boukis, M. A. R. Meier, *Eur. Polym. J.*, 2018, **104**, 32-38.
- [11] S. Martens, A. Landuyt, P. Espeel, B. Devreese, P. Dawyndt, F. Du Prez, *Nat. Commun.*, 2018, **9**, article no. 4451.

¹Since the macromolecule undergoes in-source fragmentation during ESI-MS analysis, the sequencing of each fragment is performed at a MS² stage and not in a pseudo-MS³ as described in [24,26] and [27].

- [12] Z. Huang, Q. Shi, J. Guo, F. Meng, Y. Zhang, Y. Lu, Z. Qian, X. Li, N. Zhou, Z. Zhang, X. Zhu, *Nat. Commun.*, 2019, **10**, article no. 1918.
- [13] J. M. Lee, M. B. Koo, S. W. Lee, H. Lee, J. Kwon, Y. H. Shim, S. Y. Kim, K. T. Kim, *Nat. Commun.*, 2020, **11**, article no. 56.
- [14] M. H. Caruthers, *Science*, 1985, **230**, 281-285.
- [15] J. Shendure, S. Balasubramanian, G. M. Church, W. Gilbert, J. Rogers, J. A. Schloss, R. H. Waterston, *Nature*, 2017, **550**, 345-353.
- [16] J.-F. Lutz, *Macromolecules*, 2015, **48**, 4759-4767.
- [17] S. L. Beaucage, R. P. Iyer, *Tetrahedron*, 1992, **48**, 2223-2311.
- [18] N. Appukutti, C. J. Serpell, *Polym. Chem.*, 2018, **9**, 2210-2226.
- [19] A. Al Ouahabi, L. Charles, J.-F. Lutz, *J. Am. Chem. Soc.*, 2015, **137**, 5629-5635.
- [20] A. Al Ouahabi, M. Kotera, L. Charles, J.-F. Lutz, *ACS Macro Lett.*, 2015, **4**, 1077-1080.
- [21] J. A. Amalian, A. Al Ouahabi, G. Cavallo, N. F. König, S. Poyer, J. F. Lutz, L. Charles, *J. Mass Spectrom.*, 2017, **52**, 788-798.
- [22] M. Boukhet, N. F. König, A. A. Ouahabi, G. Baaken, J.-F. Lutz, J. C. Behrends, *Macromol. Rapid Commun.*, 2017, **38**, article no. 1700680.
- [23] C. Cao, L. F. Krapp, A. Al Ouahabi, N. F. König, N. Cirauqui, A. Radenovic, J.-F. Lutz, M. D. Peraro, *Sci. Adv.*, 2020, **6**, article no. eabc2661.
- [24] A. Al Ouahabi, J.-A. Amalian, L. Charles, J.-F. Lutz, *Nat. Commun.*, 2017, **8**, article no. 967.
- [25] A. Burel, C. Carapito, J.-F. Lutz, L. Charles, *Macromolecules*, 2017, **50**, 8290-8296.
- [26] K. Launay, J.-A. Amalian, E. Laurent, L. Oswald, A. Al Ouahabi, A. Burel, F. Dufour, C. Carapito, J.-L. Clément, J.-F. Lutz, L. Charles, D. Gigmes, *Angew. Chem., Int. Ed.*, 2021, **60**, 917-926.
- [27] E. Laurent, J.-A. Amalian, M. Parmentier, L. Oswald, A. Al Ouahabi, F. Dufour, K. Launay, J.-L. Clément, D. Gigmes, M.-A. Delsuc, L. Charles, J.-F. Lutz, *Macromolecules*, 2020, **53**, 4022-4029.
- [28] Nayuki, "Projet Nayuki", 2018, <https://www.nayuki.io/page/reference-arithmetic-coding>.
- [29] J.-F. Lutz, *Isr. J. Chem.*, 2020, **60**, 151-159.
- [30] J.-F. Lutz, *ACS Macro Lett.*, 2020, **9**, 185-189.