



High-Capacity Digital Polymers: Storing Images in Single Molecules

Eline Laurent, Jean-Arthur Amalian, Marie Parmentier, Laurence Oswald, Abdelaziz Al Ouahabi, Florent Dufour, Kevin Launay, Jean-Louis Clément, Didier Gigmes, Marc-André Delsuc, et al.

► To cite this version:

Eline Laurent, Jean-Arthur Amalian, Marie Parmentier, Laurence Oswald, Abdelaziz Al Ouahabi, et al.. High-Capacity Digital Polymers: Storing Images in Single Molecules. *Macromolecules*, 2020, 53 (10), pp.4022-4029. 10.1021/acs.macromol.0c00666 . hal-02862054

HAL Id: hal-02862054

<https://hal.science/hal-02862054>

Submitted on 30 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

High-capacity Digital Polymers: Storing Images in Single Molecules

*Eline Laurent,¹ Jean-Arthur Amalian,² Marie Parmentier,¹ Laurence Oswald,¹ Abdelaziz Al Ouahabi,¹
Florent Dufour,¹ Kevin Launay,² Jean-Louis Clément,² Didier Gignes,² Marc-André Delsuc,³ Laurence
Charles*² and Jean-François Lutz*¹*

¹ Université de Strasbourg, CNRS, Institut Charles Sadron UPR22, 23 rue du Loess, 67034 Strasbourg
Cedex 2, France. E-mail : jflutz@unistra.fr

² Aix Marseille Université, CNRS, UMR 7273, Institute of Radical Chemistry, 13397, Marseille Cedex
20, France. E-mail : laurence.charles@univ-amu.fr

³ Institut de Génétique et de Biologie Moléculaire et Cellulaire, INSERM, U596, CNRS, UMR7104,
Université de Strasbourg, 1 rue Laurent Fries, 67404, Illkirch-Graffenstaden, France.

Published in *Macromolecules*, 53, 4022-4029 (2020)

<https://doi.org/10.1021/acs.macromol.0c00666>

ABSTRACT: High-capacity digital poly(phosphodiester)s were synthesized by stepwise automated phosphoramidite chemistry. Coding libraries containing either 4 or 8 phosphoramidite monomers of different mass were used to encode the polymers, thus enabling storage densities of 2 or 3 bits/monomer, respectively. In addition, a monomer containing a cleavable alkoxyamine and nucleotide mass tags enabling fragment identification were included in the chains to allow their decryption by electrospray pseudo-MS³ sequencing. As a proof of concept, black and white images with sizes ranging from 80 to 144 pixels were encoded in single polymer chains and decoded by mass spectrometry. Six different polymers were prepared in this work; three with the 4-monomers alphabet and three with the 8-monomers alphabet. In all cases, uniform macromolecules were obtained and deciphered. Yet, the bulkiest monomers of the 8-symbols alphabet required optimized protocols for uniform polymer synthesis. A macromolecular storage capacity of 144 bits per chain was obtained in this work, which is the highest capacity ever attained for a synthetic informational polymer.

INTRODUCTION

Digital polymers are macromolecules that store binary information in form of a controlled monomer sequence.¹⁻⁴ Originally, macromolecular storage has been mainly demonstrated with DNA.⁵⁻⁶ Shortly after, it was evidenced that digital information can be stored in various synthetic polymers, which are listed in order of discovery: poly(triazole amide)s,⁷ poly(alkoxyamide amide)s,⁸ poly(phosphodiester)s,⁹ polyurethanes,¹⁰ poly(alkoxyamine phosphodiester)s,¹¹ polymers obtained through multicomponent reactions,¹²⁻¹³ poly(urethane amide)s,¹⁴ linear¹⁵⁻¹⁶ and dendritic¹⁷ poly(succinimide thioether)s, poly(*N*-substituted urethane)s,¹⁸ polyesters¹⁹ and peptoid duplexes.²⁰ However, all these digital polymers, including DNA, are synthesized by stepwise chemical synthesis,²¹ which significantly limits chain-length. Consequently, only a small piece of binary information (*e.g.* a word) can usually be stored in a single molecule. The storage of more significant amount of data (*e.g.* sentence, text, image or video) requires the spatial organization of oligomer or polymer libraries, *e.g.* in micro-chips, mixtures, crystals or self-assembled materials.^{6, 22-25}

In order to increase the storage capacity of a single molecule, three different parameters can be adjusted: (i) chain-length, (ii) storage density (i.e. volume of a coded monomer, number of encoded bits per monomer) and (iii) data compression. The former two aspects can be optimized by molecular design, whereas the latter one is a classical information theory problem. As mentioned above, chain-length is a limiting factor in iterative polymer synthesis. Most of the reported works deal with oligomers in the range of 3- to 20-mer.⁴ Our group has reported the synthesis of 106-mer digital poly(phosphodiester) in 2015 and the sequencing of a 77-mer in 2017.²⁶⁻²⁷ The synthesis of 128-mer digitally-encoded polyester was reported this year.¹⁹ For DNA, sizes in the range 115- to 200-mer have been prepared on micro-chips but without performing surface-cleavage and characterization of the formed macromolecules.⁶ Overall, it seems very difficult to go beyond these limits without the help of ligation tools. Thus, recent research has also focused on the optimization of storage density. For DNA, most reported binary encryptions have

been done using the natural quaternary genetic alphabet. DNA has in theory a maximum storage density of 2 bits/monomer, but in practice a lower one due to biochemical constraints.⁶ Even though expanded genetic alphabets have been reported,²⁸ their possible use in data storage applications has not been demonstrated yet. Comparatively, synthetic polymer chemistry offers more opportunities for adjusting storage density.²⁹ First of all, the volume occupied by a coded monomer unit can be significantly smaller in some synthetic macromolecules than in DNA.³⁰ In addition, storage density can be increased by the use of expanded alphabets, since the number of monomers is easily adjustable in synthetic copolymers.²⁹ A few examples of this type have been reported, with theoretical storage densities ranging from 2 to 24 bits/monomer.^{12, 14, 18, 31} However, only oligomers have been prepared in these prior works, thus giving access to moderate single-molecule storage capacities. In the present work, we combined chain-length and storage density control to achieve high-capacity digital polymers. These information-containing macromolecules were prepared by automated phosphoramidite chemistry^{9, 26} using 4- and 8-symbols monomer alphabets. The storage and decryption of small pixelated images is demonstrated herein.

RESULTS AND DISCUSSION

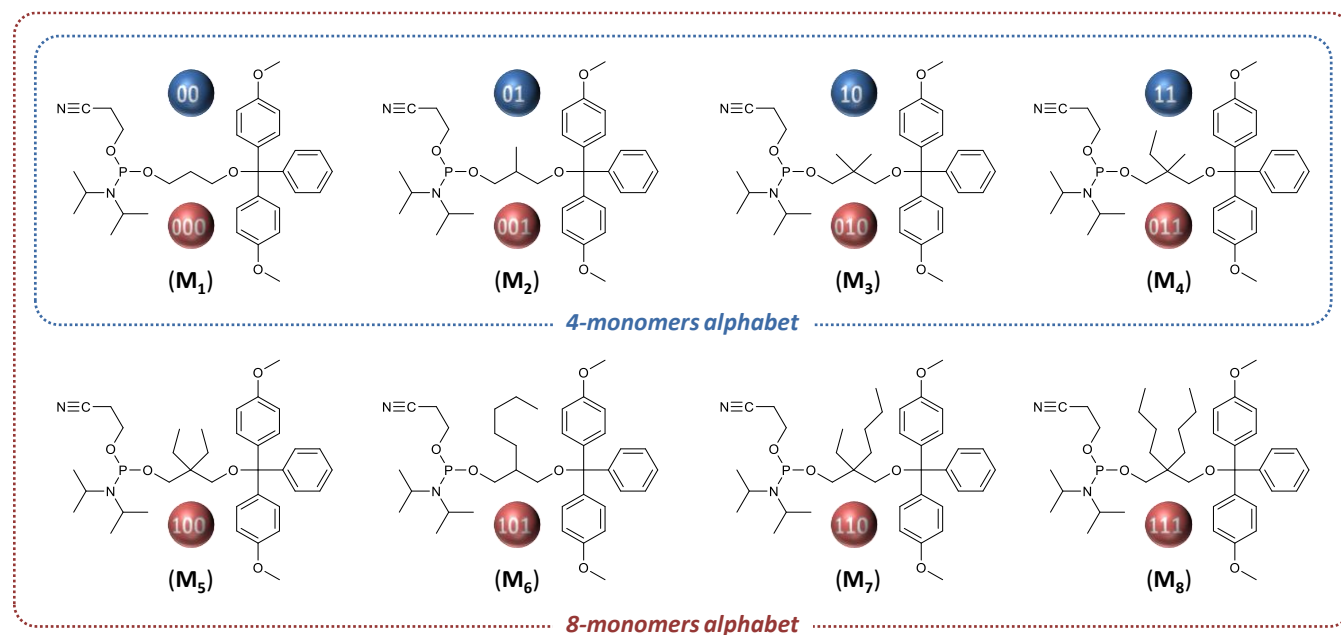
General Polymer Design. The digital polymers studied in this work are sequence-defined poly(phosphodiester)s.^{9, 32-35} They were prepared by stepwise phosphoramidite chemistry on a DNA synthesizer.^{26, 36} This chemistry requires monomers containing a reactive phosphoramidite group and a dimethoxytrityl (DMT)-protected primary alcohol. The stepwise polymer synthesis is performed on a porous glass solid support and involves four successive reactions to attach one monomer: (i) detritylation of the support (*deprotection*) (ii) reaction of the phosphoramidite group of the monomer with the OH group of the support (*coupling*), (iii) quenching of the unreacted OH groups (*capping*), and (iv) transformation of the resulting phosphite into a phosphate (*oxidation*). Repeating this four-step cycle n times allows synthesis of uniform polymers with an exact chain-length of n units.

In order to synthesize digital poly(phosphodiester)s, coded phosphoramidite monomers allowing binary encoding shall be selected. Typically, monomers of different molar mass are used, since the corresponding polymers can be easily decoded by mass spectrometry sequencing.²⁹ In most previous works, two phosphoramidite monomers including a propyl and 2,2-dimethyl propyl moieties have been used as binary alphabet,^{9, 26} even though alternative alphabets have also been reported.³⁷⁻³⁸ In all cases, two monomers were used, thus enabling a maximum storage density of 1 bit/monomer. In the present work, expanded alphabets containing either 4 or 8 monomers were developed (Chart 1), thus allowing storage densities of 2 and 3 bits/monomer, respectively. To achieve these alphabets, a series of eight phosphoramidite monomers **M1-M8**, with molar mass gradually increasing with an increment of either 14 or 28 Da, was studied. Following an arbitrary convention established in our laboratory, the lightest monomer codes for 0-bit(s) whereas the heaviest one codes for 1-bit(s). Hence, in the 4-monomers alphabet, monomers **M1**, **M2**, **M3** and **M4** represent the binary dyads 00, 01, 10 and 11, respectively. In the 8-monomers alphabet, monomers **M1**, **M2**, **M3**, **M4**, **M5**, **M6**, **M7**, and **M8** represent the binary triads 000, 001, 010, 011, 100, 101, 110 and 111, respectively.

Synthetic digital polymers are principally decrypted by tandem mass spectrometry (MS/MS) sequencing.³⁹⁻⁴⁰ Yet, the MS/MS spectra of standard poly(phosphodiester)s are complicated to interpret if the information sequence is longer than a couple of bytes.⁴¹ However, it was shown in a previous work that mass spectrometry sequencing can be facilitated by the use of inter-byte linkers containing fragile alkoxyamine bonds.²⁷ This previous design was therefore retained in the present work. The phosphoramidite monomer **L** (Scheme S1) including a tetramethylpiperidinyloxy (TEMPO)-based alkoxyamine was incorporated between segments of the digital polymers. This design also requires the use of specific mass tags allowing segment identification. Nucleotides of variable mass and composition were used for that purpose, as previously reported.²⁷ The arbitrary convention established in our laboratory is that the segment sequence shall be decrypted following the order: no tag, F, G, A, C, T. The

mass tag T comes from the preloaded porous glass solid-support, whereas the tags F, A, G and C are included in the chain using the corresponding phosphoramidite nucleosides (Scheme S1). In summary, the digital polymers studied in this work were synthesized using either 8 (4 coded symbols, 1 inter-segment linker and 3 mass tags) or 13 (8 coded symbols, 1 inter-segment linker and 4 mass tags) phosphoramidite monomers.

Chart 1. Molecular structure of the coded phosphoramidite monomers used in this work.



Preparation and evaluation of the monomers. The synthesis of monomers M_1 and M_3 was described in a previous work.⁹ Monomers M_2 and M_4 - M_8 were prepared in two steps from the corresponding commercial diols, as shown in Scheme S2. In the first step, the DMT-monoprotected diols were typically obtained in fair yields, whereas the second step leading to phosphoramidite formation was near quantitative. All monomers were isolated and characterized by ¹H NMR, ¹³C NMR and high-resolution electrospray mass spectrometry (ESI-HRMS). To our knowledge, monomers M_2 and M_4 - M_8 were never tested in automated phosphoramidite chemistry. Thus, a series of model oligomers was first synthesized

in order to evaluate the coupling efficiency of these monomers (Table 1). These oligomers were prepared using standard phosphoramidite protocols, including a coupling time of 96 seconds. For monomers **M**₂ and **M**₄-**M**₇, uniform pentamers were obtained, as evidenced by ESI-HRMS measurements (Figure S1-S5). However, for monomers including bulkier side-chain substituents such as **M**₈, the targeted structures could only be detected in small amounts in polydisperse samples (Figures S6). It is known that steric hindrance may influence the coupling efficiency of phosphoramidite monomers.⁴² For instance, RNA building blocks usually require longer coupling times than their DNA counterparts because of the bulkiness of the ribose protecting group.⁴³ Thus, the coupling efficiency of **M**₇ and **M**₈ was investigated by UV monitoring of DMT release during the synthesis of thymine-terminated model homo-octamers.²⁶

⁴⁴ It was observed that the use a freshly-prepared acetonitrile solution of **M**₇ leads to near quantitative yields. For **M**₈, a freshly-prepared acetonitrile solution leads however to a global yield of about 98%. Furthermore, final yields decrease if the monomers solutions are kept inside the automated synthesizer at room temperature for some days before performing synthesis. Thus, the use of fresh solutions, prepared directly after cold monomer storage, is recommended. In addition, for the bulkiest monomer **M**₈, different protocols were investigated to increase coupling efficiency. For instance, extending coupling times to 5 minutes, repeating twice the coupling/oxidation sequence or increasing initial **M**₈ concentration are all valid strategies for reaching near quantitative coupling.

Table 1. Characterization of model oligomers.

Sequence ^a	<i>M</i> ^b [Da]	<i>m/z</i> _{th}	<i>m/z</i> ^c
M ₂ ·M ₂ ·M ₂ ·M ₂ ·M ₂ -T	1002.2095	500.0975	500.0971
M ₄ ·M ₄ ·M ₄ ·M ₄ ·M ₄ -T	1142.3660	570.1757	570.1768
M ₅ ·M ₅ ·M ₅ ·M ₅ ·M ₅ -T	1212.4433	605.2149	605.2152
M ₆ ·M ₆ ·M ₆ ·M ₆ ·M ₆ -T	1282.5225	640.2540	640.2559
M ₇ ·M ₇ ·M ₇ ·M ₇ ·M ₇ -T	1352.6007	675.2931	675.2925
M ₈ ·M ₈ ·M ₈ ·M ₈ ·M ₈ -T	1492.7573	745.3714	745.3725

^a The monomer sequences are listed in reverse order as compared to synthesis, according to a previously-set sequencing convention.²⁷ ^b Monoisotopic mass. ^c Measured in negative mode ESI-HRMS as [M-2H]²⁻ species.

Coding with the 4-monomers alphabet. Digital polymers were first synthesized using the alphabet involving the four monomers **M₁-M₄**. (*i.e.* storage density of 2 bits/monomer). As described in the introduction, an important objective of this work is to demonstrate that a single molecule can be of sufficient capacity to store a complete data file. Thus, as a proof of concept, small pixelated images were stored herein in single polymer chains. Black and white images were created in a grid of 8 columns and 10 rows (Table 2). The images were converted into a binary sequence using the simple convention white = 0, black = 1. The binary string was then converted into a coded monomer sequence according to the dyad-language described above. It shall be noted that lossless data compression was not applied in this work, even though it could increase single-chain capacity. Indeed, the objective of this study is to assess the possibilities of molecular design (and not of algorithms) to improve the storage properties of single molecules.

Table 2 shows the information sequence of the polymers that were synthesized with the four monomers alphabet. Three different images were encrypted in single polymer chains: a flask (in polymer **P1**) as well as the atomic symbols of phosphorus (**P2**) and oxygen (**P3**). As explained above, the polymers were synthesized using 8 different phosphoramidite building blocks: 4 coding monomers **M₁-M₄**, one linker **L**, and 3 mass tags A, G and C. Following previous design,²⁷ the information sequence was included in the chain opposite to the sense of synthesis. While the synthesis starts at the T-extremity (*i.e.* moiety coming from the solid support), the decodable sequence is read from the other extremity. The linker **L** was included in the chain in between blocks of eight coded monomers (*i.e.* 16 bits). Although **L** could be theoretically added in between shorter or longer information segments, this particular design is favorable for obtaining homogeneously-charged fragments in mass spectrometry sequencing.²⁷ During synthesis, the incorporation of **L** was immediately followed by the addition of a mass tag, following the order C, A, G. No mass tag was included in the final segment.

Table 2. Description of the digital polymers prepared with the 4-monomers alphabet.

	Image	Binary sequence ^a	Monomer sequence ^b	C ^c	DP ^d	M ^e [Da]
P1		01111110·00100100· 00100100·00100100· 00100100·00100100· 01000010·10000001· 10000001·01111110	M ₂ ·M ₄ ·M ₄ ·M ₃ ·M ₁ ·M ₃ ·M ₂ ·M ₁ ·L- M ₁ ·M ₃ ·M ₂ ·M ₁ ·M ₁ ·M ₃ ·M ₂ ·M ₁ ·G-L- M ₁ ·M ₃ ·M ₂ ·M ₁ ·M ₁ ·M ₃ ·M ₂ ·M ₁ ·A-L- M ₂ ·M ₁ ·M ₁ ·M ₃ ·M ₃ ·M ₁ ·M ₁ ·M ₂ ·C-L- M ₃ ·M ₁ ·M ₁ ·M ₂ ·M ₂ ·M ₄ ·M ₄ ·M ₃ -T	80	39	8794.9999
P2		11111111·10000001· 10111001·10100101· 10100101·10111001· 10100001·10100001· 10000001·11111111	M ₄ ·M ₄ ·M ₄ ·M ₄ ·M ₃ ·M ₁ ·M ₁ ·M ₂ -L- M ₃ ·M ₄ ·M ₃ ·M ₂ ·M ₃ ·M ₃ ·M ₂ ·M ₂ ·G-L- M ₃ ·M ₃ ·M ₂ ·M ₂ ·M ₃ ·M ₄ ·M ₃ ·M ₂ ·A-L- M ₃ ·M ₃ ·M ₁ ·M ₂ ·M ₃ ·M ₃ ·M ₁ ·M ₂ ·C-L- M ₃ ·M ₁ ·M ₁ ·M ₂ ·M ₄ ·M ₄ ·M ₄ ·M ₄ -T	80	39	9159.4068
P3		11111111·10000001· 10011001·10100101· 10100101·10100101· 10100101·10011001· 10000001·11111111	M ₄ ·M ₄ ·M ₄ ·M ₄ ·M ₃ ·M ₁ ·M ₁ ·M ₂ -L- M ₃ ·M ₂ ·M ₃ ·M ₂ ·M ₃ ·M ₃ ·M ₂ ·M ₂ ·G-L- M ₃ ·M ₃ ·M ₂ ·M ₂ ·M ₃ ·M ₃ ·M ₂ ·M ₂ ·A-L- M ₃ ·M ₃ ·M ₂ ·M ₂ ·M ₃ ·M ₂ ·M ₃ ·M ₂ ·C-L- M ₃ ·M ₁ ·M ₁ ·M ₂ ·M ₄ ·M ₄ ·M ₄ ·M ₄ -T	80	39	9131.3755

^a For clarity, separation dots are displayed every 8 bits to indicate the lines of the grids. ^b Dots and hyphens do not have a molecular meaning. Dots are included between monomers for clarity only. Hyphens are included for signaling the cleavable linkers as well as the terminal thymine nucleotide coming from the preloaded solid-support. ^c Storage capacity in bits/chain. ^d Degree of polymerization including coded monomers, linkers and mass tags. ^e Monoisotopic mass of the polymer.

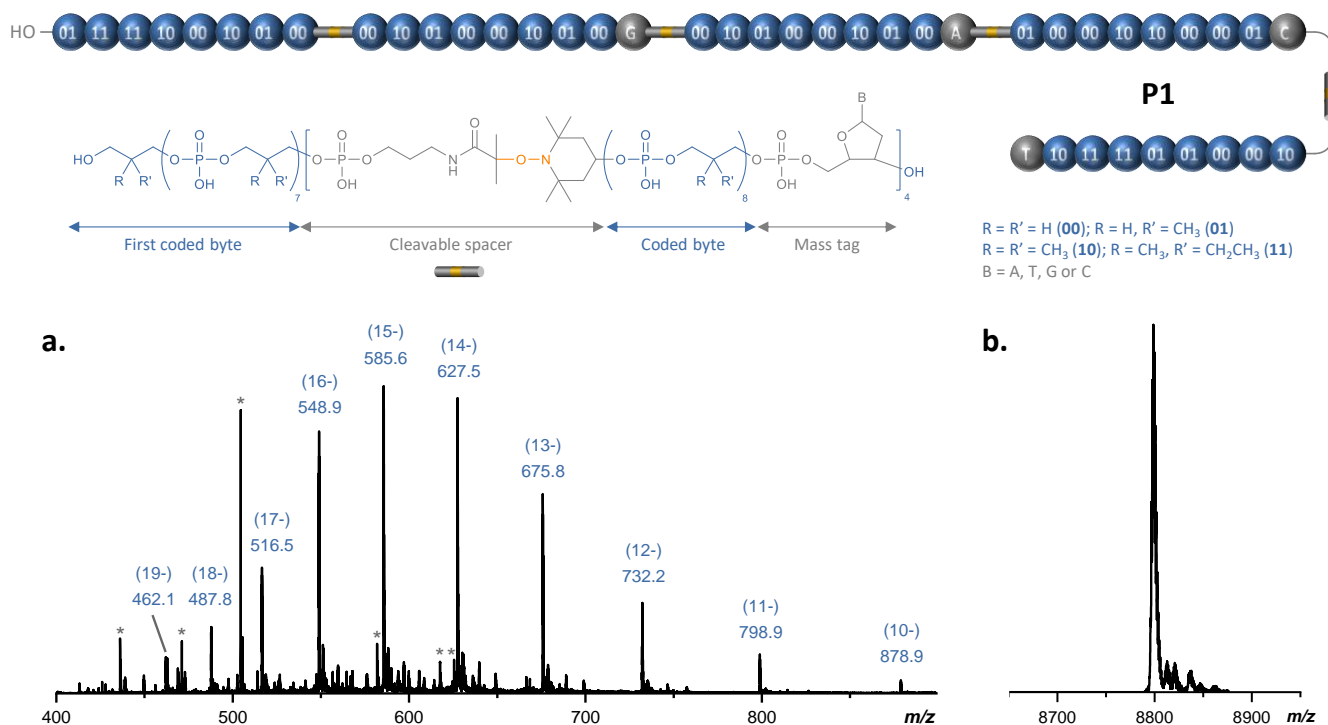


Figure 1. Negative mode ESI-MS analysis of digital polymer P1: (a) original spectrum including the different charge states of the polymer. The grey stars denote in-source fragments. (b) Charge deconvolution of the mass spectrum. The general molecular structure and the digital sequence of the polymer are shown above the spectrum.

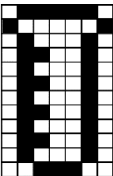
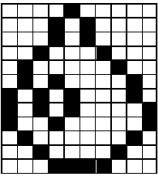
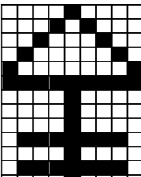
All the polymers of Table 2 were prepared on an automated synthesizer using standard phosphoramidite protocols.²⁶ After synthesis, the digital polymers were cleaved, purified (Table S1) and characterized by mass spectrometry (Figure 1 and Figures S7-S8). In all cases, the targeted polymers were detected as dominant species by negative mode ESI mass spectrometry. For example, Figure 1 shows the ESI mass spectra of **P1** before and after charge deconvolution. These signals could be easily exploited for sequencing the polymers using a previously-described pseudo-MS³ procedure.²⁷ In this approach, collision-induced dissociation (CID) is first used to stimulate programmed chain-fragmentation on the weak alkoxyamine sites (*i.e.* **L** units). These MS/MS conditions result in the formation of a library of sub-chain segments, which can be identified by their mass tags (Table S2 and Figures S9-S11). In the present case, these fragments all contain 16 bits of information, as schematized in Figure 1. Afterwards, each chain-fragment is again subjected to CID and broken down into monomer sub-pieces (*i.e.* MS³ conditions), thus enabling its sequencing (Figures S9-S11). The complete sequence of the initial polymer chain can then be reconstructed with the help of the mass tags.²⁷ For instance, the pixelated images displayed in Table 2 could all be extracted from polymers **P1-P3** using this pseudo-MS³ strategy.⁴⁵

Coding with the 8-monomers alphabet. Digital polymers were also synthesized using the 8-monomers alphabet displayed in chart 1. Table 3 shows the digital encryption and the primary structure of the different polymers prepared with this expanded alphabet. Images of different format were stored in these macromolecules, including a test tube (in polymer **P4**), a drop (**P5**) and the alchemy symbol of phosphorus (**P6**). Thirteen phosphoramidite monomers (8 coded symbols, 1 inter-segment linker and 4 mass tags) were used to prepare these polymers. The rules of encryption and decryption were the same as described in the previous section. The only difference is the number of coded monomers.

As discussed above, sterically-hindered monomers may lead to incomplete coupling steps when using standard automated phosphoramidite protocols. This aspect can be particularly problematic for the

synthesis of longer sequences, which typically requires high coupling efficiency. Indeed, preliminary experiments indicated that standard automated protocols lead to ill-defined polymers. Thus, three modified protocols were screened for the synthesis of polymer **P4**: longer coupling times for monomers **M5-M8** (*protocol a*), doubled coupling/oxidation steps for monomers **M5-M8** (*protocol b*) and increased monomer concentration for monomers **M4-M8** (*protocol c*). In order to assess the efficiency of these different protocols, the synthesis of **P4** was interrupted after the preparation of the first (Figure S12), second (Figure S13) and third (Figure S14) segment and the formed sequences were characterized by ESI-MS. *Protocol a* was found to be the less efficient of all. Indeed, pronounced defects were already noticed in the first segment when these conditions were applied. Although *protocol b* allowed sequence-defined synthesis for the first two segments, some defects were observed starting at the third segment. Only *protocol c* allowed synthesis of monodisperse polymer **P4**, as shown in Figure 2a. Thus, these conditions were selected for the synthesis **P5-P6**. As shown in Figures S15-S16, uniform polymers were obtained in all cases (Table S1). Furthermore, polymers **P4-P6** were all decrypted by pseudo-MS³ sequencing. Following the strategy described in the previous section, they were first subjected to CID and decomposed into a programmed segment library (Figure 2b and Figures S18-S19). Then, all segments were individually sequenced as shown in Figures S17-S19 and the complete information sequences were reconstructed following the order dictated by the mass tags. These results show that single molecules containing 84 (**P4**), 120 (**P5**) and 144 (**P6**) bits of information can be comprehensively sequenced in a single pseudo-MS³ mass spectrometry measurement. To our knowledge, 144 bits per chain is the highest storage capacity ever reported for a single synthetic polymer chain.

Table 3. Description of the digital polymers prepared with the 8-monomers alphabet.

	Image	Binary sequence ^a	Monomer sequence ^b	C ^c	DP ^d	M ^e [Da]
P4		0111110-1000001- 0100010-0110010- 0100010-0110010- 0100010-0110010- 0100010-0110010- 0100010-0011100	M ₄ ·M ₈ ·M ₃ ·M ₁ ·M ₃ ·M ₅ ·M ₃ -L- M ₄ ·M ₂ ·M ₂ ·M ₁ ·M ₅ ·M ₇ ·M ₃ ·A-L- M ₃ ·M ₂ ·M ₂ ·M ₅ ·M ₅ ·M ₃ ·C-L- M ₄ ·M ₂ ·M ₂ ·M ₁ ·M ₅ ·M ₄ ·M ₅ -T	84	33	6852.1265
P5		0000100000-0001010000- 0001010000-0010001000- 0100000100-0101000010- 1010100010-1010100001- 1001000001-0100000010- 0010000100-0001111000	M ₁ ·M ₃ ·M ₁ ·M ₁ ·M ₃ ·M ₅ ·M ₁ ·M ₂ -L- M ₃ ·M ₁ ·M ₂ ·M ₁ ·M ₅ ·M ₂ ·M ₁ ·M ₂ ·G-L- M ₁ ·M ₆ ·M ₁ ·M ₃ ·M ₆ ·M ₃ ·M ₂ ·M ₃ ·A-L- M ₆ ·M ₁ ·M ₄ ·M ₂ ·M ₁ ·M ₂ ·M ₃ ·M ₁ ·C-L- M ₂ ·M ₁ ·M ₅ ·M ₂ ·M ₁ ·M ₂ ·M ₈ ·M ₁ -T	120	47	9075.3129
P6		000010000-000101000- 001000100-010000010- 100000001-111111111- 000010000-000010000- 000010000-011111110- 000010000-011111110- 000010000-000010000- 000010000-000010000	M ₁ ·M ₃ ·M ₁ ·M ₁ ·M ₆ ·M ₁ ·M ₂ ·M ₁ -L- M ₅ ·M ₃ ·M ₁ ·M ₃ ·M ₅ ·M ₁ ·M ₂ ·M ₈ ·F-L- M ₈ ·M ₈ ·M ₁ ·M ₃ ·M ₁ ·M ₁ ·M ₃ ·M ₁ ·G-L- M ₁ ·M ₃ ·M ₁ ·M ₄ ·M ₈ ·M ₇ ·M ₁ ·M ₃ ·A-L- M ₁ ·M ₄ ·M ₈ ·M ₇ ·M ₁ ·M ₃ ·M ₁ ·M ₁ ·C-L- M ₃ ·M ₁ ·M ₁ ·M ₃ ·M ₁ ·M ₁ ·M ₃ ·M ₁ -T	144	57	11367.1299

^a Separation dots indicate the lines of the graphical matrix. ^b Dots are included between monomers for clarity only. Hyphens are included for signaling the cleavable linkers as well as the terminal thymine nucleotide. ^c Storage capacity in bits/chain. ^d Degree of polymerization including coded monomers, linkers and mass tags. ^e Monoisotopic mass of the polymer.

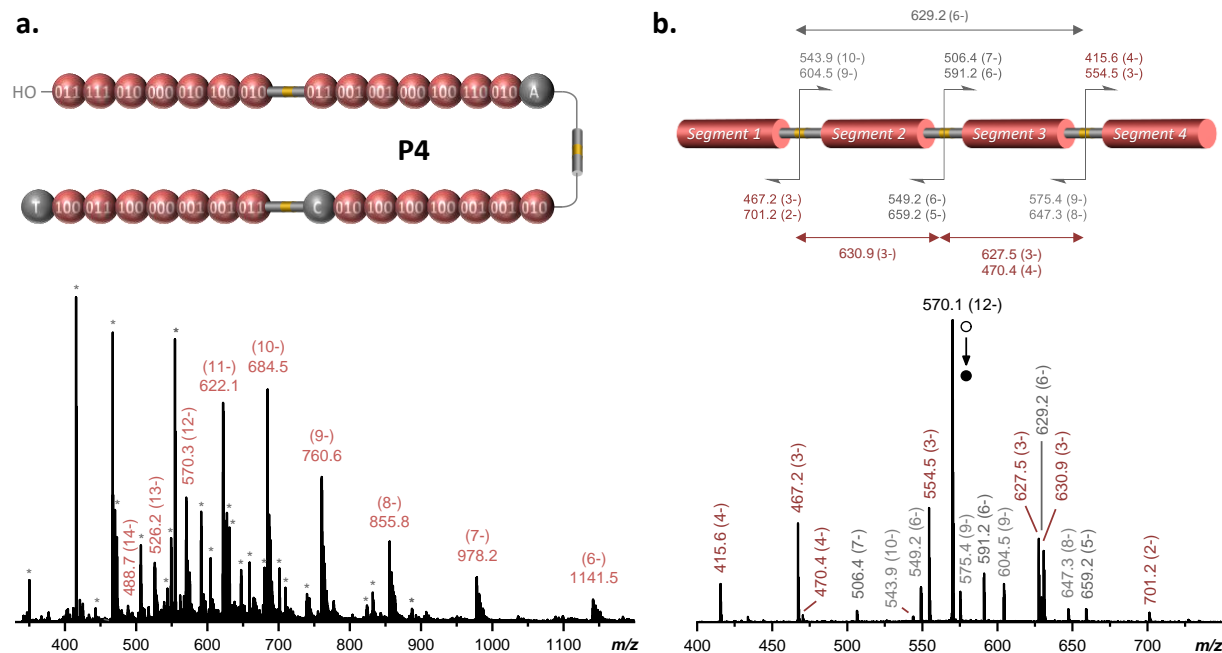


Figure 2. (a) Negative mode ESI-MS analysis of digital polymer **P4**. The grey stars denote internal fragments. (b) MS/MS spectrum (8 eV, laboratory frame) of the same polymer obtained by collision-induced dissociation of the $[\mathbf{P4-12H}]^{12-}$ precursor ion, which carries on average three charges per segment.

CONCLUSION

High-capacity digital poly(phosphodiester)s were synthesized in this work. These information-containing macromolecules were prepared using expanded alphabets containing either 4 or 8 coded phosphoramidite monomers. The main advantage of this design is that it permits to combine high chain-length and high storage density. Indeed, automated phosphoramidite synthesis allows synthesis of relatively long polymer chains (33- to 57-mers in the present work), while the use of expanded alphabets enable to squeeze more information in a given monomer (2 bits or 3 bits/monomer in the present work). Both alphabets enabled synthesis of uniform digital macromolecules, even though the bulkiest monomers of the 8-symbols alphabet required optimized coupling protocols to be incorporated in high yields in the coded chains. Furthermore, cleavable inter-segment linkers and segment mass-tags were included in the polymers in order to induce programmed chain fragmentation in ESI pseudo-MS³ sequencing. As a consequence, all coded samples could be easily decoded by mass spectrometry. As a proof of concept, small black and white images were stored in single macromolecules and decrypted using this simple molecular design. For example, a 144 pixels image was included in a 57-mer poly(phosphodiester). Such a storage capacity is to our knowledge the highest one ever reported for a synthetic copolymer. Nevertheless, the design reported in this article is not yet at its limit. Indeed, it is most certainly possible to increase further single-chain storage capacity by increasing chain length (at least until a 100-mer), by expanding further the monomer alphabet (*e.g.* to 16 symbols) and, of course, by using well-known lossless compression algorithms (*e.g.* run-length encoding, Huffman coding). Altogether, this work demonstrates that a complete data file of significant size can be stored in a single molecule. This is particularly relevant for tracking technologies in which molecular traces are used; for example, in anti-counterfeiting or traceability applications.⁴⁶

Funding Sources

These results were obtained with the support of the French National Research Agency (ANR project 00111001, grant numbers ANR-16-CE29-0004-01 and ANR-16-CE29-0004-02).

ACKNOWLEDGMENT

L.C. 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).

REFERENCES

1. Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-Controlled Polymers. *Science* **2013**, *341*, 1238149.
2. Lutz, J.-F. Defining the Field of Sequence-Controlled Polymers. *Macromol. Rapid Commun.* **2017**, *38*, 1700582.
3. Colquhoun, H.; Lutz, J.-F. Information-containing macromolecules. *Nat. Chem.* **2014**, *6*, 455-456.
4. Rutten, M. G. T. A.; Vaandrager, F. W.; Elemans, J. A. A. W.; Nolte, R. J. M. Encoding information into polymers. *Nat. Rev. Chem.* **2018**, *2*, 365–381.
5. Zhirnov, V.; Zadegan, R. M.; Sandhu, G. S.; Church, G. M.; Hughes, W. L. Nucleic acid memory. *Nat. Mater.* **2016**, *15*, 366-370.
6. Ceze, L.; Nivala, J.; Strauss, K. Molecular digital data storage using DNA. *Nat. Rev. Genet.* **2019**, *20*, 456-466.
7. Trinh, T. T.; Oswald, L.; Chan-Seng, D.; Lutz, J.-F. Synthesis of Molecularly Encoded Oligomers Using a Chemoselective “AB + CD” Iterative Approach. *Macromol. Rapid Commun.* **2014**, *35*, 141-145.

8. Roy, R. K.; Meszynska, A.; Laure, C.; Charles, L.; Verchin, C.; Lutz, J.-F. Design and Synthesis of Digitally-Encoded Polymers that can be Decoded and Erased. *Nat. Commun.* **2015**, *6*, 7237.
9. 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.
10. Gunay, U. S.; Petit, B. E.; Karamessini, D.; Al Ouahabi, A.; Amalian, J.-A.; Chendo, C.; Bouquey, M.; Gigmes, D.; Charles, L.; Lutz, J.-F. Chemoselective Synthesis of Uniform Sequence-Coded Polyurethanes and Their Use as Molecular Tags. *Chem* **2016**, *1*, 114-126.
11. Cavallo, G.; Al Ouahabi, A.; Oswald, L.; Charles, L.; Lutz, J.-F. Orthogonal Synthesis of “Easy-to-Read” Information-Containing Polymers Using Phosphoramidite and Radical Coupling Steps. *J. Am. Chem. Soc.* **2016**, *138*, 9417-9420.
12. Boukis, A. C.; Meier, M. A. R. Data storage in sequence-defined macromolecules via multicomponent reactions. *Eur. Polym. J.* **2018**, *104*, 32-38.
13. Holloway, J. O.; Van Lijsebetten, F.; Badi, N.; Houck, H. A.; Du Prez, F. E. From Sequence-Defined Macromolecules to Macromolecular Pin Codes. *Adv. Sci.* **2020**, *7*, 1903698.
14. Martens, S.; Landuyt, A.; Espeel, P.; Devreese, B.; Dawyndt, P.; Du Prez, F. Multifunctional sequence-defined macromolecules for chemical data storage. *Nat. Commun.* **2018**, *9*, 4451.
15. Ding, K.; Zhang, Y.; Huang, Z.; Liu, B.; Shi, Q.; Hu, L.; Zhou, N.; Zhang, Z.; Zhu, X. Easily encodable/decodable digital polymers linked by dithiosuccinimide motif. *Eur. Polym. J.* **2019**, *119*, 421-425.
16. Liu, B.; Shi, Q.; Hu, L.; Huang, Z.; Zhu, X.; Zhang, Z. Engineering digital polymer based on thiol–maleimide Michael coupling toward effective writing and reading. *Polym. Chem.* **2020**, *11*, 1702-1707.
17. Huang, Z.; Shi, Q.; Guo, J.; Meng, F.; Zhang, Y.; Lu, Y.; Qian, Z.; Li, X.; Zhou, N.; Zhang, Z.; Zhu, X. Binary tree-inspired digital dendrimer. *Nat. Commun.* **2019**, *10*, 1918.
18. Mondal, T.; Greff, V.; Petit, B. É.; Charles, L.; Lutz, J.-F. Efficient Protocol for the Synthesis of “N-Coded” Oligo- and Poly(N-Substituted Urethanes). *ACS Macro Lett.* **2019**, *8*, 1002-1005.

19. Lee, J. M.; Koo, M. B.; Lee, S. W.; Lee, H.; Kwon, J.; Shim, Y. H.; Kim, S. Y.; Kim, K. T. High-density information storage in an absolutely defined aperiodic sequence of monodisperse copolyester. *Nat. Commun.* **2020**, *11*, 56.
20. Leguizamon, S. C.; Scott, T. F. Sequence-selective dynamic covalent assembly of information-bearing oligomers. *Nat. Commun.* **2020**, *11*, 784.
21. Lutz, J.-F.; Lehn, J.-M.; Meijer, E. W.; Matyjaszewski, K. From Precision Polymers to Complex Materials and Systems. *Nat. Rev. Mater.* **2016**, *1*, 16024.
22. Steinkoenig, J.; Aksakal, R.; Du Prez, F. Molecular access to multi-dimensionally encoded information. *Eur. Polym. J.* **2019**, *120*, 109260.
23. 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.
24. Arcadia, C. E.; Kennedy, E.; Geiser, J.; Dombroski, A.; Oakley, K.; Chen, S.-L.; Sprague, L.; Ozmen, M.; Sello, J.; Weber, P. M.; Reda, S.; Rose, C.; Kim, E.; Rubenstein, B. M.; Rosenstein, J. K. Multicomponent molecular memory. *Nat. Commun.* **2020**, *11*, 691.
25. Nagy, L.; Kuki, Á.; Nagy, T.; Vadkerti, B.; Erdélyi, Z.; Kárpáti, L.; Zsuga, M.; Kéki, S. Encoding Information into Polyethylene Glycol Using an Alcohol-Isocyanate “Click” Reaction. *Int. J. Mol. Sci.* **2020**, *21*.
26. 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.
27. Al Ouahabi, A.; Amalian, J.-A.; Charles, L.; Lutz, J.-F. Mass spectrometry sequencing of long digital polymers facilitated by programmed inter-byte fragmentation. *Nat. Commun.* **2017**, *8*, 967.
28. Hoshika, S.; Leal, N. A.; Kim, M.-J.; Kim, M.-S.; Karalkar, N. B.; Kim, H.-J.; Bates, A. M.; Watkins, N. E.; SantaLucia, H. A.; Meyer, A. J.; DasGupta, S.; Piccirilli, J. A.; Ellington, A. D.; SantaLucia, J.; Georgiadis, M. M.; Benner, S. A. Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* **2019**, *363*, 884.
29. Lutz, J.-F. Coding Macromolecules: Inputting Information in Polymers Using Monomer-Based Alphabets. *Macromolecules* **2015**, *48*, 4759-4767.

30. Petit, B. É.; Lotz, B.; Lutz, J.-F. About the Crystallization of Abiotic Coded Matter. *ACS Macro Lett.* **2019**, *8*, 779-782.
31. Cavallo, G.; Poyer, S.; Amalian, J.-A.; Dufour, F.; Burel, A.; Carapito, C.; Charles, L.; Lutz, J.-F. Cleavable Binary Dyads: Simplifying Data Extraction and Increasing Storage Density in Digital Polymers. *Angew. Chem., Int. Ed.* **2018**, *57*, 6266-6269.
32. Appukutti, N.; Serpell, C. J. High definition polyphosphoesters: between nucleic acids and plastics. *Polym. Chem.* **2018**, *9*, 2210-2226.
33. Vybornyi, M.; Vyborna, Y.; Häner, R. DNA-inspired oligomers: from oligophosphates to functional materials. *Chem. Soc. Rev.* **2019**, *48*, 4347-4360.
34. Edwardson, T. G. W.; Carneiro, K. M. M.; Serpell, C. J.; Sleiman, H. F. An Efficient and Modular Route to Sequence-Defined Polymers Appended to DNA. *Angew. Chem., Int. Ed.* **2014**, *53*, 4567-4571.
35. de Rochambeau, D.; Sun, Y.; Barlog, M.; Bazzi, H. S.; Sleiman, H. F. Modular Strategy To Expand the Chemical Diversity of DNA and Sequence-Controlled Polymers. *J. Org. Chem.* **2018**, *83*, 9774-9786.
36. Caruthers, M. H. Gene synthesis machines: DNA chemistry and its uses. *Science* **1985**, *230*, 281-285.
37. König, N. F.; Al Ouahabi, A.; Poyer, S.; Charles, L.; Lutz, J.-F. A Simple Post-Polymerization Modification Method for Controlling Side-Chain Information in Digital Polymers. *Angew. Chem., Int. Ed.* **2017**, *56*, 7297-7301.
38. König, N. F.; Al Ouahabi, A.; Oswald, L.; Szweda, R.; Charles, L.; Lutz, J.-F. Photo-editable macromolecular information. *Nat. Commun.* **2019**, *10*, 3774.
39. Mutlu, H.; Lutz, J.-F. Reading Polymers: Sequencing of Natural and Synthetic Macromolecules. *Angew. Chem., Int. Ed.* **2014**, *53*, 13010-13019.
40. Wesdemiotis, C. Multidimensional Mass Spectrometry of Synthetic Polymers and Advanced Materials. *Angew. Chem., Int. Ed.* **2017**, *56*, 1452-1464.

41. Amalian, J. A.; Al Ouahabi, A.; Cavallo, G.; König, N. F.; Poyer, S.; Lutz, J. F.; Charles, L. Controlling the structure of sequence-defined poly(phosphodiester)s for optimal MS/MS reading of digital information. *J. Mass Spectrom.* **2017**, *52*, 788-798.
42. Beaucage, S. L.; Iyer, R. P. Advances in the Synthesis of Oligonucleotides by the Phosphoramidite Approach. *Tetrahedron* **1992**, *48*, 2223-2311.
43. Caruthers, Marvin H. A brief review of DNA and RNA chemical synthesis. *Biochem. Soc. Trans.* **2011**, *39*, 575-580.
44. Voirin, E.; Behr, J.-P.; Kotera, M. Versatile synthesis of oligodeoxyribonucleotide–oligospermine conjugates. *Nat. Protoc.* **2007**, *2*, 1360-1367.
45. For all coded polymers including DNA, basic syntax rules shall be known for performing sequencing. In the present work, for example, the molar mass of the coding synthons, the encoding direction, and the number of columns and rows in the coding matrix shall be known for decryption.
46. Arppe, R.; Sørensen, T. J. Physical unclonable functions generated through chemical methods for anti-counterfeiting. *Nat. Rev. Chem.* **2017**, *1*, 0031.

High-capacity Digital Polymers: Storing Images in Single Molecules

*Eline Laurent,¹ Jean-Arthur Amalian,² Marie Parmentier,¹ Laurence Oswald,¹ Abdelaziz Al Ouahabi,¹ Florent Dufour,¹ Kevin Launay,² Jean-Louis Clément,² Didier Gigmes,² Marc-André Delsuc,³ Laurence Charles^{*2} and Jean-François Lutz^{*1}*

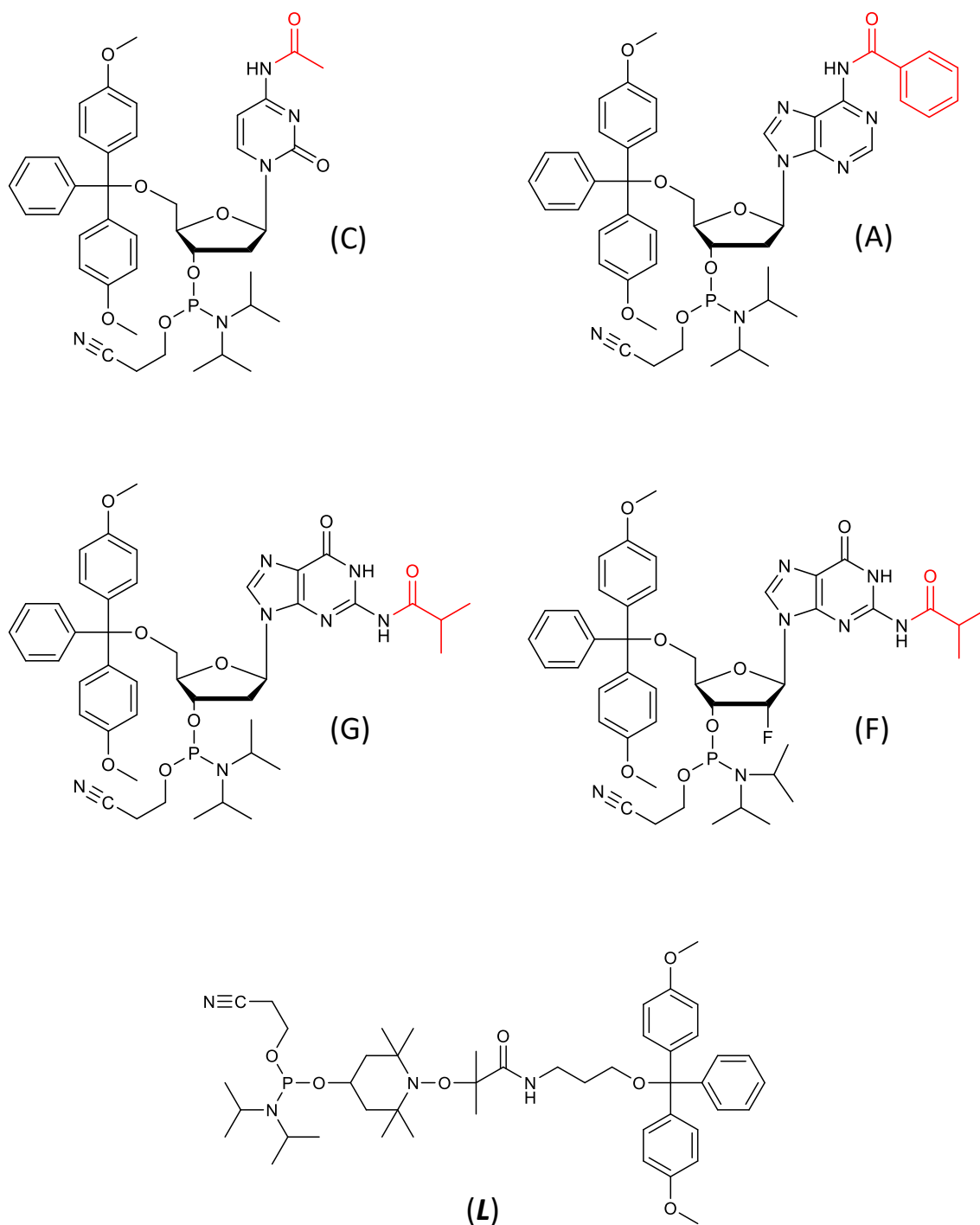
¹ Université de Strasbourg, CNRS, Institut Charles Sadron UPR22, 23 rue du Loess, 67034 Strasbourg Cedex 2, France. ² Aix Marseille Université, CNRS, UMR 7273, Institute of Radical Chemistry, 13397, Marseille Cedex 20, France. ³ Institut de Génétique et de Biologie Moléculaire et Cellulaire, INSERM, U596, CNRS, UMR7104, Université de Strasbourg, 1 rue Laurent Fries, 67404, Illkirch-Graffenstaden, France.

A. Materials and reagents:	2
B. Experimental Procedures	4
B.1. Synthesis of the phosphoramidite monomers	4
B.1.1. Synthesis of I2	4
B.1.2. Synthesis of M2	5
B.1.3. Synthesis of I4	5
B.1.4. Synthesis of M4	6
B.1.5. Synthesis of I5	6
B.1.6. Synthesis of M5	7
B.1.7. Synthesis of I6	7
B.1.8. Synthesis of M6	8
B.1.9. Synthesis of I7	8
B.1.10. Synthesis of M7	9
B.1.11. Synthesis of I8	10
B.1.12. Synthesis of M8	10
B.2. Automated solid-phase syntheses of polymers	11
B.3. Coupling efficiency tests	12
C. Measurements and Analysis	13
D. Supplementary Figures	14
E. References	33

A. Materials and reagents:

2-Butyl-2-ethyl-1,3-propanediol (>98.0%, TCI), 2-tert-butylpropane-1,3-diol (98%, Alfa Aesar), 2,2-di-n-butyl-1,3-propanediol (96%, Alfa Aesar), 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (97 %, ABCR), cyclohexane (Carlo Erba), anhydrous dichloromethane (99.8 %, anh. DCM, cont. amylene as stabilizer, Sigma-Aldrich), *N,N*-diisopropylethylamine (>99.0 %, DIPEA, TCI), 4,4'-dimethoxytrityl chloride (DMTr-Cl, ChemGenes), ethyl acetate (EtOAc, Carlo Erba), 2-ethyl-2-methyl-1,3-propanediol (>97%, TCI), 2,2-diethyl-1,3-propanediol (>98.0%, TCI), hexane (Carlo Erba), 2-methyl-1,3-propanediol (99%, TCI), 2-n-pentylpropane-1,3-diol (97%, Alfa Aesar), anhydrous pyridine (99.8 %, anh. pyr, in Sure/Seal™, Sigma-Aldrich), silica gel (high-purity grade, pore size 60 Å, 230 - 400 mesh, for flash chromatography, SiO₂, Fluka), sodium chloride (for brine solution, ESCO), anhydrous sodium sulfate (99.6 %, VWR), and triethylamine (97 %, Et₃N, Acros Organics), were used as received without further purification. Anhydrous THF was obtained using a dry solvent station GT S100. All air-sensitive reactions have been carried out under argon atmosphere. The phosphoramidites monomers **M**₁ and **M**₃ and the linker **L** were prepared following previously-reported procedures and stored at -18°C.¹⁻²

Reagents for automated phosphoramidite synthesis: anhydrous acetonitrile (phosphoramidite diluent & dry washings, ChemGenes), acetonitrile (≥99.9 %, washings, Roth), activation reagent (0.25 M 5-ethylthio tetrazole in MeCN, ChemGenes), Cap A (acetic anhydride/pyridine/THF, ChemGenes), Cap B (10 % *N*-methylimidazole in THF, ChemGenes), DMT removal reagent (3 w% TCA in DCM, Roth), drying traps (small, 10 - 15 mL, ChemGenes), dT-CPG 1000 (1 μmol in cartridge, 1000 Å pore size, Glen Research), glen-pak DNA purification cartridge (10 nmole - 1.0 μmole, Glen Research) and oxidizing solution (0.02 M iodine/pyridine/H₂O/THF, ChemGenes) were used as received. The nucleoside phosphoramidites (Scheme S1) dA-CE phosphoramidite, Ac-dC-CE phosphoramidite, dG-CE phosphoramidite, and 2'-F-G-CE phosphoramidite (Glen Research) were also used as received. All the phosphoramidites monomers were stored in the freezer at -18 °C.

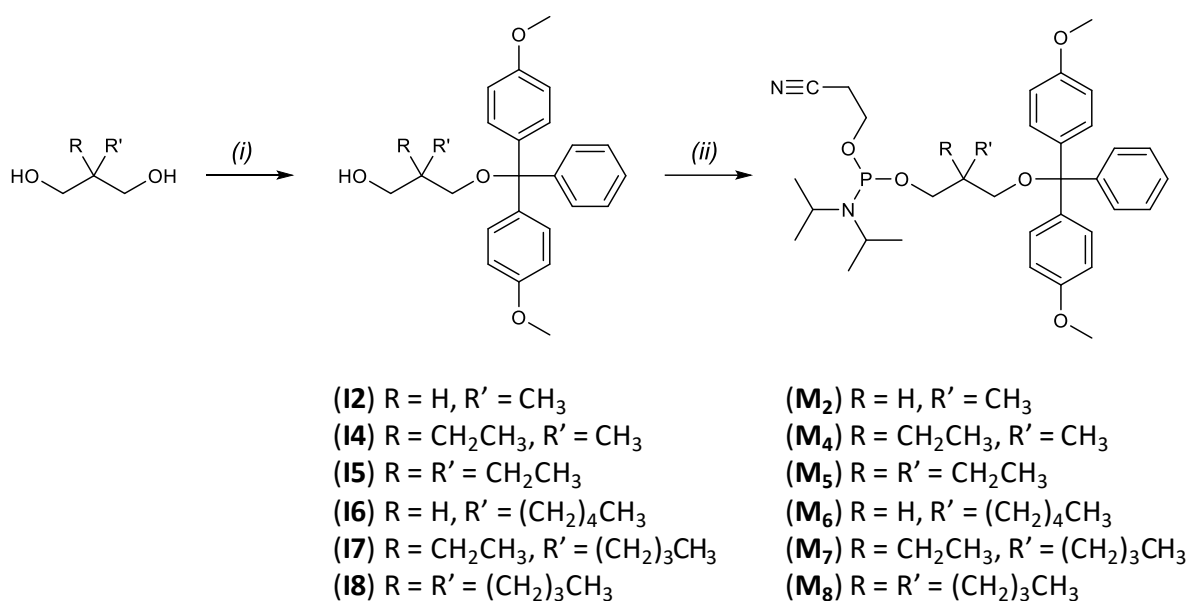


Scheme S1. Molecular structure of the phosphoramidite mass tags and linker used in this work. The red protecting groups on the purine and pyrimidine rings are removed during cleavage, leading to primary amines.

B. Experimental Procedures

B.1. Synthesis of the phosphoramidite monomers

The phosphoramidite monomers **M**₂, and **M**₄-**M**₈ were prepared following the synthetic route shown in Scheme S2. In the first step, a commercial diol is monoprotected by a DMT moiety, thus leading to intermediates **I**₂ and **I**₄-**I**₈. The remaining alcohol group is then reacted with 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite in a second step.



Scheme S2. General route for the synthesis of monomers **M**₂ and **M**₄-**M**₈. Experimental conditions: (i) 4,4'-dimethoxytrityl chloride, pyridine/THF, RT; (ii) 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite, DIPEA, DCM.

B.1.1. *Synthesis of I*₂.

To a stirred solution of 2-methyl-1,3-propanediol (4.73 g, 52.5 mmol, 1.0 eq.) in 66 mL pyridine and 96 mL THF anhydrous 4,4'-dimethoxytrityl chloride (17.81 g, 52.5 mmol, 1.0 eq.) was added in four intervals of each 45 min at RT under argon atmosphere. After the fourth addition, the mixture was stirred at RT for 1 h. The reaction was stopped with the addition of 26 mL of methanol and the mixture was evaporated to dryness. The residue was mixed with 135 mL NaHCO₃ (sat., aq.) and was extracted with EtOAc (2 x 135 mL). The combined organic layer was washed with H₂O (1 x 200 mL), brine (1 x 200 mL), dried over Na₂SO₄, filtered and the solvent was removed. The crude product was purified by column chromatography on silica

gel (hexane / EtOAc: 9/1 with 1% Et₃N) to obtain 10.82 g of **I2** (52%) as a yellowish foam. HRMS *m/z*: [M+Na]⁺ calculated for C₂₄H₂₈O₄Na⁺ 415.1880; found 415.1881. ¹H NMR (400.13 MHz, CDCl₃, δ, ppm): 0.86 (d, 3H, HCCH₃), 2.04 (m, 1H, HCCH₃), 2.36 (m, 1H, CH₂ODMT), 3.02 (m, 1H, HOCH₂), 3.22 (m, 1H, CH₂ODMT), 3.59 (m, 2H, HOCH₂), 3.79 (s, 6H, Ar_{DMT}OCH₃), 6.83 (d, 4H, Ar_{DMT}H), 7.18-7.24 (m, 1H, Ar_{DMT}H), 7.26-7.34 (m, 6H, Ar_{DMT}H), 7.39-7.45 (m, 2H, Ar_{DMT}H). ¹³C NMR (100.62 MHz, CDCl₃, δ, ppm): 13.9, 36.2, 55.3, 67.7, 67.8, 86.5, 113.3, 126.9, 128.0, 128.2, 130.1, 136.2, 136.3, 145.1, 158.6.

B.1.2. Synthesis of **M2**.

10.82 g of **I2** (27.5 mmol, 1.0 eq.) were dissolved in 53 mL anhydrous DCM. Then, DIPEA (19.21 mL, 110.2 mmol, 4.0 eq.) was added. The solution was cooled to 0 °C under an argon atmosphere. 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (6.65 g, 28.1 mmol, 1.02 eq.) dissolved in 9 mL anhydrous DCM was added at 0 °C. The reaction mixture was stirred at 0 °C for 30 min, then allowed to reach RT and stirred for 1 h. The reaction mixture was extracted with 53 mL NaHCO₃ (sat., aq.). The aqueous phase was washed with 40 mL DCM. The combined organic layer was dried over Na₂SO₄, filtered and the solvent was removed. The crude product was purified by column chromatography on silica gel (hexane / EtOAc: 9/1 with 1% Et₃N) to obtain 14.96 g of **M2** (92%) as a colorless foam. HRMS *m/z*: [M+H]⁺ calculated for C₃₄H₄₆N₂O₅P⁺ 593.3139; found 593.3141. ¹H NMR (400.13 MHz, CDCl₃, δ, ppm): 0.97 (dd, 3H, HCCH₃), 1.07-1.23 (m, 12H, N(CH(CH₃)₂)₂), 2.00-2.13 (m, 1H, HCCH₃), 2.50-2.58 (m, 2H, CH₂CH₂C≡N), 2.92-3.10 (m, 2H, CH₂OP, CH₂ODMT), 3.40-3.65 (m, 4H, CH₂OP, CH₂ODMT, CH₂CH₂C≡N), 3.67-3.77 (m, 2H, N(CH(CH₃)₂)₂), 3.79 (s, 6H, Ar_{DMT}OCH₃), 6.81 (d, 4H, Ar_{DMT}H), 7.16-7.22 (m, 1H, Ar_{DMT}H), 7.26-7.34 (m, 6H, Ar_{DMT}H), 7.40-7.45 (m, 2H, Ar_{DMT}H). ¹³C NMR (100.62 MHz, CDCl₃, δ, ppm): 14.6, 20.4, 24.6, 24.7, 35.9, 43.2, 55.3, 58.4, 65.3, 66.2, 85.8, 113.1, 117.7, 126.7, 127.8, 128.4, 130.3, 136.7, 145.5, 158.5. ³¹P NMR (161.96 MHz, CDCl₃, δ, ppm): 147.4.

B.1.3. Synthesis of **I4**.

To a stirred solution of 2-ethyl-2-methyl-1,3-propandiol (2 g, 16.9 mmol, 1.0 eq.) in 30 mL pyridine and 60 mL THF anhydrous 4,4'-dimethoxytrityl chloride (6.3 g, 18.6 mmol, 1.1 eq.) was added in three intervals of each 60 min at RT under argon atmosphere. After the third addition, the mixture was stirred at RT overnight. The reaction was stopped with the addition of 15 mL of methanol and the mixture was evaporated to dryness. The residue was mixed with 100 mL NaHCO₃ (sat., aq.) and was extracted with EtOAc (2 x 50 mL). The combined organic

layer was washed with H₂O (1 x 100 mL), brine (1 x 100 mL), dried over Na₂SO₄, filtered and the solvent was removed. The crude product was purified by column chromatography on silica gel (Cyclohexane / EtOAc: 8/2 with 1% Et₃N) to obtain 4.27 g of **I4** (60%) as a colorless foam. HRMS *m/z*: [M+Na]⁺ calculated for C₂₇H₃₂O₄Na⁺ 443.2193; found 443.2192. ¹H NMR (400.13 MHz, CDCl₃, δ, ppm): 0.67-0.8 (m, 6H, CCH₃, CCH₂CH₃), 1.30 (m, 2H, CCH₂CH₃), 2.21 (m, 1H, HOCH₂), 2.91 (m, 2H, CH₂ODMT), 3.41 (m, 2H, HOCH₂), 3.7 (s, 6H, Ar_{DMT}OCH₃), 6.75 (d, 4H, Ar_{DMT}H), 7.06-7.16 (m, 1H, Ar_{DMT}H), 7.17-7.27 (m, 6H, Ar_{DMT}H), 7.28-7.37 (m, 2H, Ar_{DMT}H). ¹³C NMR (100.62 MHz, CDCl₃, δ, ppm): 7.6, 18.76, 26.79, 39.12, 55.26, 69.7, 70.38, 86.19, 113.22, 126.8, 127.9, 129.3, 130, 139.5, 144.8, 158.52.

B.1.4. Synthesis of **M4**.

2.5 g of **I4** (5.9 mmol, 1.0 eq.) were dissolved in 40 mL anhydrous DCM. Then, DIPEA (4.2 mL, 23.8 mmol, 4.0 eq.) was added. The solution was cooled to 0 °C under an argon atmosphere. 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (1.5 g, 6.5 mmol, 1.1 eq.) dissolved in 5 mL anhydrous DCM was added dropwise at 0 °C with continuous stirring. The reaction flask was then allow reaching back room temperature and stirred for 1 hour. The mixture was evaporating to dryness. The crude product was dissolved in ethyl acetate and then directly purified by column chromatography on silica gel (cyclohexane / EtOAc: 6/4 with 1% Et₃N) to obtain 3.1 g of **M4** (84%) as a colorless oil. HRMS *m/z*: [M+H]⁺ calculated for C₃₆H₅₀N₂O₅P⁺ 621.3452; found 621.3454. ¹H NMR (400.13 MHz, CDCl₃, δ, ppm) : 0.71(td, 3H, H₂CCH₃), 0.86 (d, 3H, HCCH₃), 1.13-1.20 (m, 12H, N(CH(CH₃)₂)₂), 1.34-1.42 (m, 2H, CCH₂CH₃), 2.69 (m, 2H, CH₂CH₂C≡N), 2.92-3.0(m, 2H, CH₂ODMT), 3.48-3.53 (m, 1H, CH₂OP), 3.63 (m, 3H, CH₂OP, CH₂CH₂C≡N), 3.74-3.77 (m, 2H, N(CH(CH₃)₂)₂), 3.78(s, 6H, Ar_{DMT}OCH₃), 6.88 (d,4H, Ar_{DMT}H), 7.20 (m, 1H, Ar_{DMT}H), 7.32 (m, 6H, Ar_{DMT}H), 7.48 (d, 2H, Ar_{DMT}H). ¹³C NMR (100.62 MHz, CDCl₃, δ, ppm): 7.11, 18.60, 18.92, 26.64, 26.82, 39.29, 54.58, 54.87, 65.95, 66.56, 66.91, 85.43, 112.2, 112.85, 126.48, 127.57, 128.21, 129.85, 136.44, 145.74, 158.58. ³¹P NMR (161 MHz, CDCl₃, δ, ppm): 147.15.

B.1.5. Synthesis of **I5**.

To a stirred solution of 2,2-diethyl-1,3-propanediol (2 g, 15.1mmol, 1.0 eq.) in 30 mL pyridine and 60 mL THF anhydrous 4,4'-dimethoxytrityl chloride (5.1 g, 16.6 mmol, 1.1 eq.) was added in three intervals of each 60 min at RT under argon atmosphere. After the third addition, the mixture was stirred at RT overnight. The reaction was stopped with the addition

of 15 mL of methanol and the mixture was evaporated to dryness. The residue was mixed with 100 mL NaHCO₃ (sat., aq.) and was extracted with EtOAc (2 x 50 mL). The combined organic layer was washed with H₂O (1 x 100 mL), brine (1 x 100 mL), dried over Na₂SO₄, filtered and the solvent was removed. The crude product was purified by column chromatography on silica gel (cyclohexane / EtOAc: 8/2 with 1% Et₃N) to obtain 3.9 g of **I5** (59%) as a colorless oil. HRMS *m/z*: [M+Na]⁺ calculated for C₂₈H₃₄O₄Na⁺ 457.2349; found 457.2347. ¹H NMR (400.13 MHz, CDCl₃, δ, ppm): 0.63 (t, 6H, C(CH₂CH₃)₂), 1.15-1.30 (m, 4H, C(CH₂CH₃)₂), 3.38 (m, 2H, CH₂ODMT), 3.67 (m, 2H, HOCH₂), 3.69 (s, 6H, Ar_{DMT}OCH₃), 6.75 (d, 4H, Ar_{DMT}H), 7.06-7.17 (m, 1H, Ar_{DMT}H), 7.18-7.28 (m, 6H, Ar_{DMT}H), 7.29-7.37 (m, 2H, Ar_{DMT}H). ¹³C NMR (100.62 MHz, CDCl₃, δ, ppm): 7.2, 14.1, 21, 22.9, 41.5, 55, 60.2, 66.4, 86, 113, 126.64, 127.78, 128.16, 130, 136, 149.7, 158.43.

B.1.6. Synthesis of **M5**.

2 g of **I5** (5.9 mmol, 1.0 eq.) were dissolved in 40 mL anhydrous DCM. Then, DIPEA (3.2 mL, 18.4 mmol, 4.0 eq.) was added. The solution was cooled to 0 °C under an argon atmosphere. 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (1.1 g, 4.6 mmol, 1 eq.) dissolved in 5 mL anhydrous DCM was added dropwise at 0 °C with continuous stirring. The reaction flask was then allowed reaching back room temperature and stirred for 1 hour. The mixture was evaporated to dryness. The crude product was purified by column chromatography on silica gel (cyclohexane / EtOAc: 6/4 with 1% Et₃N) to obtain 2.7 g of **M5** (91.5%) as a colorless oil. HRMS *m/z*: [M+H]⁺ calculated for C₃₇H₅₂N₂O₅P⁺ 635.3608; found 635.3611. ¹H NMR (400.13 MHz, CDCl₃, δ, ppm) : 0.68 (t, 6H, C(CH₂CH₃)₂), 1.17-1.21 (m, 12H, N(CH(CH₃)₂)₂), 1.29-1.42 (m, 4H, C(CH₂CH₃)₂), 2.57 (m, 2H, CH₂CH₂C≡N), 2.90-2.96(m, 2H, CH₂ODMT), 3.48-3.52 (m, 1H, CH₂OP), 3.56-3.66 (m, 3H, CH₂OP, CH₂CH₂C≡N), 3.74-3.80 (m, 2H, N(CH(CH₃)₂)₂), 3.81(s, 6H, Ar_{DMT}OCH₃), 6.84 (m, 4H, Ar_{DMT}H), 7.23 (m, 1H, Ar_{DMT}H), 7.32 (m, 6H, Ar_{DMT}H), 7.48 (d, J = 7.44 Hz, 2H, Ar_{DMT}H). ¹³C NMR (100.62 MHz, CDCl₃, δ, ppm): 7.11, 18.60, 18.92, 26.64, 26.82, 39.29, 55.17, 58.33, 60.38, 64.00, 66.12, 85.30, 112.84, 117.67, 126.51, 127.57, 128.42, 130.30, 136.51, 145.39, 158.30. ³¹P NMR (CDCl₃, δ, ppm): 147.07.

B.1.7. Synthesis of **I6**.

To a stirred solution of 2-n-pentylpropane-1,3-diol (2.50 g, 17.1 mmol, 1.0 eq.) in 21 mL pyridine and 31 mL THF anhydrous 4,4'-dimethoxytrityl chloride (5.79 g, 17.1 mmol, 1.0 eq.) was added in four intervals of each 45 min at RT under argon atmosphere. After the fourth

addition, the mixture was stirred at RT for 1 h. The reaction was stopped with the addition of 8.5 mL of methanol and the mixture was evaporated to dryness. The residue was mixed with 44 mL NaHCO₃ (sat., aq.) and was extracted with EtOAc (2 x 44 mL). The combined organic layer was washed with H₂O (1 x 65 mL), brine (1 x 65 mL), dried over Na₂SO₄, filtered and the solvent was removed. The crude product was purified by column chromatography on silica gel (hexane / EtOAc: 95/5 to 80/20 with 1% Et₃N) to obtain 5.14 g of **I6** (67%) as a colorless oil. HRMS *m/z*: [M+Na]⁺ calculated for C₂₉H₃₆O₄Na⁺ 471.2506; found, 471.2513. ¹H NMR (400.1 MHz, Acetone-D₆, δ, ppm): 0.85 (t, 3H, CH(CH₂)₄CH₃), 1.15-1.45 (m, 8H, CH(CH₂)₄CH₃), 1.73 (m, 1H, CH(CH₂)₄CH₃), 3.13 (m, 2H, CH₂ODMT), 3.37 (t, 1H, HOCH₂), 3.61 (m, 2H, HOCH₂), 3.78 (s, 6H, Ar_{DMT}OCH₃), 6.87 (d, 4H, Ar_{DMT}H), 7.19-7.24 (m, 1H, Ar_{DMT}H), 7.27-7.37 (m, 6H, Ar_{DMT}H), 7.44-7.50 (m, 2H, Ar_{DMT}H). ¹³C NMR (100.62 MHz, Acetone-D₆, δ, ppm): 14.4, 23.3, 27.4, 29.1, 32.9, 42.7, 55.5, 63.9, 64.4, 86.5, 113.8, 127.4, 128.5, 129.0, 130.9, 137.4, 146.7, 159.5.

B.1.8.Synthesis of **M6**.

1.56 g of **I6** (3.47 mmol, 1.0 eq.) were dissolved in 7 mL anhydrous DCM. Then, DIPEA (2.42 mL, 13.9 mmol, 4.0 eq.) was added. The solution was cooled to 0 °C under an argon atmosphere. 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (0.84 g, 3.50 mmol, 1.02 eq.) dissolved in 2 mL anhydrous DCM was added at 0 °C. The reaction mixture was stirred at 0 °C for 30 min, then allowed to reach RT and stirred for 1 h. The reaction mixture was extracted with 7 mL NaHCO₃ (sat., aq.). The aqueous phase was washed with 7 mL DCM. The combined organic layer was dried over Na₂SO₄, filtered and the solvent was removed. The crude product was purified by column chromatography on silica gel (hexane / EtOAc: 95/5 to 90/10 with 1% Et₃N) to obtain 1.87 g of **M6** (83%) as a colorless oil. HRMS *m/z*: [M+H]⁺ calculated for C₃₈H₅₄N₂O₅P⁺ 649.3765; found 649.3770. ¹H NMR (400.1 MHz, Acetone-D₆, δ, ppm): 0.88 (t, 3H, CH(CH₂)₄CH₃), 1.10-1.45 (m, 20H, N(CH(CH₃)₂)₂, CH(CH₂)₄CH₃), 1.87 (m, 1H, CH(CH₂)₄CH₃), 2.69 (m, 2H, CH₂CH₂C≡N), 3.12 (m, 2H, CH₂ODMT), 3.56-3.65 (m, 2H, CH₂OP), 3.66-3.85 (m, 4H, N(CH(CH₃)₂)₂, CH₂CH₂C≡N), 3.79 (s, 6H, Ar_{DMT}OCH₃), 6.88 (d, 4H, Ar_{DMT}H), 7.18-7.24 (m, 1H, Ar_{DMT}H), 7.27-7.38 (m, 6H, Ar_{DMT}H), 7.45-7.52 (m, 2H, Ar_{DMT}H). ¹³C NMR (100.62 MHz, Acetone-D₆, δ, ppm): 14.4, 20.8, 23.3, 24.9, 25.0, 27.3, 29.1, 32.9, 41.6, 43.7, 55.5, 59.3, 63.8, 64.8, 86.5, 113.8, 118.9, 127.4, 128.5, 129.0, 131.0, 137.3, 146.6, 159.5. ³¹P NMR (161.96 MHz, Acetone-D₆, δ, ppm): 147.0.

B.1.9.Synthesis of **I7**.

To a stirred solution of 2-butyl-2-ethyl-1,3-propanediol (4.01 g, 25.0 mmol, 1.0 eq.) in 31 mL pyridine and 45 mL THF anhydrous 4,4'-dimethoxytrityl chloride (8.47 g, 25.0 mmol, 1.0 eq.)

was added in four intervals of each 45 min at RT under argon atmosphere. After the fourth addition, the mixture was stirred at RT for 1 h. The reaction was stopped with the addition of 12 mL of methanol and the mixture was evaporated to dryness. The residue was mixed with 67 mL NaHCO₃ (sat., aq.) and was extracted with EtOAc (2 x 67 mL). The combined organic layer was washed with H₂O (1 x 100 mL), brine (1 x 100 mL), dried over Na₂SO₄, filtered and the solvent was removed. The crude product was purified by column chromatography on silica gel (hexane / EtOAc: 95/5 to 80/20 with 1% Et₃N) to obtain 7.24 g of **I7** (63%) as a colorless oil. HRMS *m/z*: [M+Na]⁺ calculated for C₃₀H₃₈O₄Na⁺ 485.2662; found, 485.2662. ¹H NMR (400.13 MHz, CDCl₃, δ, ppm): 0.73 (t, 3H, CCH₂CH₃), 0.87 (t, 3H, C(CH₂)₃CH₃), 0.95-1.46 (m, 8H, C(CH₂)₃CH₃, CCH₂CH₃), 3.00 (s, 2H, CH₂ODMT), 3.46 (m, 2H, HOCH₂), 3.58 (s, 1H, HOCH₂), 3.79 (s, 6H, Ar_{DMT}OCH₃), 6.83 (d, 4H, Ar_{DMT}H), 7.18-7.24 (m, 1H, Ar_{DMT}H), 7.27-7.35 (m, 6H, Ar_{DMT}H), 7.39-7.45 (m, 2H, Ar_{DMT}H). ¹³C NMR (100.62 MHz, CDCl₃, δ, ppm): 7.4, 14.2, 23.6, 23.7, 25.1, 30.6, 41.5, 55.3, 67.5, 68.1, 86.3, 113.3, 126.9, 128.0, 128.3, 130.2, 136.1, 145.0, 158.6.

B.1.10. Synthesis of M7.

3.02 g of **I7** (6.52 mmol, 1.0 eq.) were dissolved in 12.5 mL anhydrous DCM. Then, DIPEA (4.54 mL, 26.0 mmol, 4.0 eq.) was added. The solution was cooled to 0 °C under an argon atmosphere. 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (1.57 g, 6.65 mmol, 1.02 eq.) dissolved in 3 mL anhydrous DCM was added at 0 °C. The reaction mixture was stirred at 0 °C for 30 min, then allowed to reach RT and stirred for 1 h. The reaction mixture was extracted with 12 mL NaHCO₃ (sat., aq.). The aqueous phase was washed with 10 mL DCM. The combined organic layer was dried over Na₂SO₄, filtered and the solvent was removed. The crude product was purified by column chromatography on silica gel (hexane / EtOAc: 95/5 with 1% Et₃N) to obtain 3.33 g of **M7** (77%) as a yellowish oil. HRMS *m/z*: [M+H]⁺ calculated for C₃₉H₅₆N₂O₅P⁺ 663.3921; found 663.3920. ¹H NMR (400.13 MHz, CDCl₃, δ, ppm): 0.65 (t, 3H, CCH₂CH₃), 0.83 (t, 3H, C(CH₂)₃CH₃), 0.88-1.43 (m, 20H, N(CH(CH₃)₂)₂, C(CH₂)₃CH₃, CCH₂CH₃), 2.50-2.58 (m, 2H, CH₂CH₂C≡N), 2.84-2.96 (m, 2H, CH₂ODMT), 3.41-3.66 (m, 4H, CH₂OP, CH₂CH₂C≡N), 3.70-3.78 (m, 2H, N(CH(CH₃)₂)₂), 3.79 (s, 6H, Ar_{DMT}OCH₃), 6.81 (d, 4H, Ar_{DMT}H), 7.16-7.23 (m, 1H, Ar_{DMT}H), 7.27-7.36 (m, 6H, Ar_{DMT}H), 7.39-7.48 (m, 2H, Ar_{DMT}H). ¹³C NMR (100.62 MHz, CDCl₃, δ, ppm): 7.3, 14.2, 20.6, 23.8, 24.0, 24.7, 24.8, 24.9, 30.9, 41.9, 43.3, 55.3, 58.4, 64.2, 66.0, 85.5, 113.0, 117.8, 126.7, 127.7, 128.6, 130.5, 136.7, 145.5, 158.5. ³¹P NMR (161.96 MHz, CDCl₃, δ, ppm): 147.2.

B.1.11. Synthesis of **I8**.

To a stirred solution of 2,2-di-n-butyl-1,3-propanediol (4.70 g, 25.0 mmol, 1.0 eq.) in 31 mL pyridine and 45 mL THF anhydrous 4,4'-dimethoxytrityl chloride (8.47 g, 25.0 mmol, 1.0 eq.) was added in four intervals of each 45 min at RT under argon atmosphere. After the fourth addition, the mixture was stirred at RT for 1 h. The reaction was stopped with the addition of 12 mL of methanol and the mixture was evaporated to dryness. The residue was mixed with 67 mL NaHCO₃ (sat., aq.) and was extracted with EtOAc (2 x 67 mL). The combined organic layer was washed with H₂O (1 x 100 mL), brine (1 x 100 mL), dried over Na₂SO₄, filtered and the solvent was removed. The crude product was purified by column chromatography on silica gel (hexane / EtOAc: 95/5 to 50/50 with 1% Et₃N) to obtain 6.58 g of **I8** (54%) as a colorless oil. HRMS m/z : [M+Na]⁺ calculated for C₃₂H₄₂O₄Na⁺ 513.2975; found, 513.2981. ¹H NMR (400.1 MHz, Acetone-D₆, δ , ppm): 0.84 (t, 6H, C[(CH₂)₃CH₃]₂), 0.91-1.32 (m, 12H, C[(CH₂)₃CH₃]₂), 2.94 (s, 2H, CH₂ODMT), 3.34 (t, 1H, HOCH₂), 3.52 (s, 2H, HOCH₂), 3.78 (s, 6H, Ar_{DMT}OCH₃), 6.87 (d, 4H, Ar_{DMT}H), 7.18-7.24 (m, 1H, Ar_{DMT}H), 7.26-7.40 (m, 6H, Ar_{DMT}H), 7.47-7.52 (m, 2H, Ar_{DMT}H). ¹³C NMR (100.62 MHz, Acetone-D₆, δ , ppm): 14.5, 24.3, 25.5, 31.9, 42.3, 55.4, 65.3, 65.6, 86.2, 113.7, 127.3, 128.4, 129.0, 131.0, 137.3, 146.6, 159.4.

B.1.12. Synthesis of **M8**.

2.80 g of **I8** (5.70 mmol, 1.0 eq.) were dissolved in 11 mL anhydrous DCM. Then, DIPEA (3.97 mL, 22.8 mmol, 4.0 eq.) was added. The solution was cooled to 0 °C under an argon atmosphere. 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (1.38 g, 5.80 mmol, 1.02 eq.) dissolved in 2.5 mL anhydrous DCM was added at 0 °C. The reaction mixture was stirred at 0 °C for 30 min, then allowed to reach RT and stirred for 1 h. The reaction mixture was extracted with 11 mL NaHCO₃ (sat., aq.). The aqueous phase was washed with 10 mL DCM. The combined organic layer was dried over Na₂SO₄, filtered and the solvent was removed. The crude product was purified by column chromatography on silica gel (hexane / EtOAc: 95/5 to 90/10 with 1% Et₃N) to obtain 3.40 g of **M8** (86%) as a colorless oil. HRMS m/z : [M+H]⁺ calculated for C₄₁H₆₀N₂O₅P⁺ 691.4234; found 691.4237. ¹H NMR (400.1 MHz, Acetone-D₆, δ , ppm): 0.84 (t, 6H, C[(CH₂)₃CH₃]₂), 0.93-1.39 (m, 24H, N(CH(CH₃)₂)₂, C[(CH₂)₃CH₃]₂), 2.72 (t, 2H, CH₂CH₂C \equiv N), 2.96 (m, 2H, CH₂ODMT), 3.52-3.90 (m, 6H, CH₂OP, N(CH(CH₃)₂)₂, CH₂CH₂C \equiv N), 3.79 (s, 6H, Ar_{DMT}OCH₃), 6.88 (d, 4H, Ar_{DMT}H), 7.18-7.25 (m, 1H, Ar_{DMT}H), 7.27-7.40 (m, 6H, Ar_{DMT}H), 7.47-7.52 (m, 2H, Ar_{DMT}H). ¹³C NMR (100.62 MHz, Acetone-D₆, δ , ppm): 14.4, 20.8, 24.3, 24.9, 25.0, 25.4, 31.8, 42.3, 43.8, 55.5, 59.3, 66.7, 66.9, 86.2, 113.7, 118.9, 127.4, 128.4, 129.2, 131.1, 137.2, 146.6, 159.5. ³¹P NMR (161.96 MHz, Acetone-D₆, δ , ppm): 146.9.

B.2. Automated solid-phase syntheses of polymers

The polymers were synthesized under argon in rigorously dry conditions by automated solid-phase phosphoramidite method on an Expedite DNA Synthesizer (Perseptive Biosystem 8900) and on an ABI DNA Synthesizer (Applied BioSystems 3900). The typical method involves four steps (i) deprotection, (ii) coupling, (iii) oxidation and (iv) capping which are performed in each cycle, as described previously (see also below for details).³ Three modified protocols (*Protocols a-c*) have also been investigated in this work (see also below for details). In all cases, the phosphoramidite monomers **M**₁-**M**₈, the linker *L* and the nucleoside phosphoramidites A, C, G and F were dissolved in anhydrous acetonitrile under argon (100 mM or 50 mM solution on the Expedite and ABI instrument, respectively), placed in the synthesizer with all the needed reagents (activation reagent, Cap A, Cap B, DMT removal reagent) and primed twice. The solid support dT-CPG (1 μ mol scale) was placed in the synthesizer and the sequences were performed with DMT-On mode. Thus the last dimethoxytrityl (DMT) protecting group was kept. Once the synthesis was done the solid support column was removed from the synthesizer and the DMT-protected sequence was cleaved from the support. A solution of 28% ammonia and methylamine (1/1, v/v) was used to cleave the DMT-protected polymers from the solid-support at room temperature for 30 min and purified on a Glen-Pak DNA purification cartridge following the same procedure as described before.³ This procedure permits to separate the DMT-terminated targeted structures from the truncated sequences deactivated by the capping reaction. Then, the terminal DMT moiety of the desired sequence-coded polymers was removed directly on the Glen-Pak column and washed out by solvent elution.³ The isolated polymers solutions were lyophilized, weight and the recovery yields are summarized in Table S1.

Standard protocol:

In the standard protocol, the coupling step takes 96 seconds and can be described as follows:

- a. Flush the lines with solvents (maximum speed)
- b. Monomer is activated with 5-(ethylthio)-1H-tetrazole solution (8 s)
- c. Reaction of the activated monomer with the sequence linked to the support (88 s)
- d. Flush the lines with solvents (maximum speed)

Modified protocol a:

In this method the coupling step was elongated as described below:

- a. Flush the lines with solvents (maximum speed)

- b. Monomer is activated with 5-(ethylthio)-1H-terazole solution (75 s)
- c. Reaction of the activated monomer with the sequence linked to the support (225 s)
- d. Flush the lines with solvents (maximum speed)

Modified protocol b:

This modified method includes six steps: (i) deprotection, (ii) coupling, (iii) oxidation, (iv) coupling, (v) oxidation and (vi) capping. The coupling and oxidation steps are doubled in order to obtain a better coupling efficiency. The duration of each coupling step is 96 seconds as in the standard method.

Modified protocol c:

In this method, the concentration of some monomers is increased by a factor of 1.2 (e.g. 120 mM instead of 100 mM on the Expedite system).

Table S1. Yields of the digital polymers synthesized in the present work

	m^a [mg]	Yield^b [%]
P1	7.1	81
P2	7.1	78
P3	6.8	74
P4	4.9	71
P5	7.5	83
P6	9.7	85

^a Solid amount isolated after purification. ^b Calculated from the CPG loading (1 μ mol column)

B.3. Coupling efficiency tests

The efficiency tests were performed during the synthesis of model octamers with the general structure $M_x \cdot M_x \cdot M_x \cdot M_x \cdot M_x \cdot M_x \cdot M_x \cdot M_x \cdot T$, where M_x is a monomer from the series **M₁-M₈**. For estimating the coupling efficiency of a given monomer, the DMT moiety deprotected during the iterative process was collected for each step and analyzed by UV spectroscopy. The efficiency was estimated by measuring the third and penultimate fractions, as previously

described.³ The third fraction is studied because it is the first one in which a monomer is coupled to a monomer (the first fraction corresponds to the deprotection of the CPG support and the second one to the coupling of a monomer to the support). The fractions were collected into a 5 mL volumetric flask, completed to 5 mL with dichloromethane. A 250 μ L volume was transferred to another 5 mL volumetric flask and completed to 5 mL with a solution of 3% trichloroacetic acid in dichloromethane (v/v). The absorbance was measured at 504 nm with dichloromethane as the reference solvent. The absorbance is usually below 1 for 1- μ mol-scale synthesis.

C. Measurements and Analysis

C.1. Mass spectrometry

All samples were dissolved in H₂O/ACN (50:50, v/v) solution containing 0.1% formic acid, further diluted in a methanolic solution of ammonium acetate (3 mM), and introduced at a 10 μ L min⁻¹ flow rate in the electrospray ionization (ESI) source operated in the negative mode (capillary voltage: -2.27 kV) under a desolvation gas (N₂) flow of 100 L h⁻¹ heated at 35°C. The cone voltage was -10 V or -20 V for both MS and MS² experiments while it was adjusted between -30 V and -50 V to induce in-source fragmentation in MS³ experiments. Collision-induced dissociation (CID) was performed in the ion trap device using argon as the collision gas after selection of the primary (in MS²) or secondary (in pseudo-MS³) precursor ion in the quadrupole mass analyzed of the instrument. Data analyses were conducted using the MassLynx 4.1 program provided by Waters.

C.2. Nuclear Magnetic Resonance (NMR)

All NMR were recorded on a Bruker Avance 400 spectrometer equipped with an Ultrashield magnet. Chemical shifts (δ) are reported in parts per million (ppm) against solvent residual signal (¹H NMR, CDCl₃: δ = 7.26 ppm; ¹³C NMR, CDCl₃: δ = 77.16 ppm; ¹H NMR, (CD₃)₂CO: δ = 2.05 ppm; ¹³C NMR, (CD₃)₂CO: δ = 29.84 ppm. ¹H NMR spectra were recorded at 400.13 MHz, ¹³C NMR spectra at 100.62 MHz and ³¹P NMR (161.96 MHz). The NMR solvents deuterated chloroform (99,8 %, chloroform-d1) and deuterated acetone (99,8 %, acetone-d6) were purchased from Aldrich.

D. Supplementary Figures

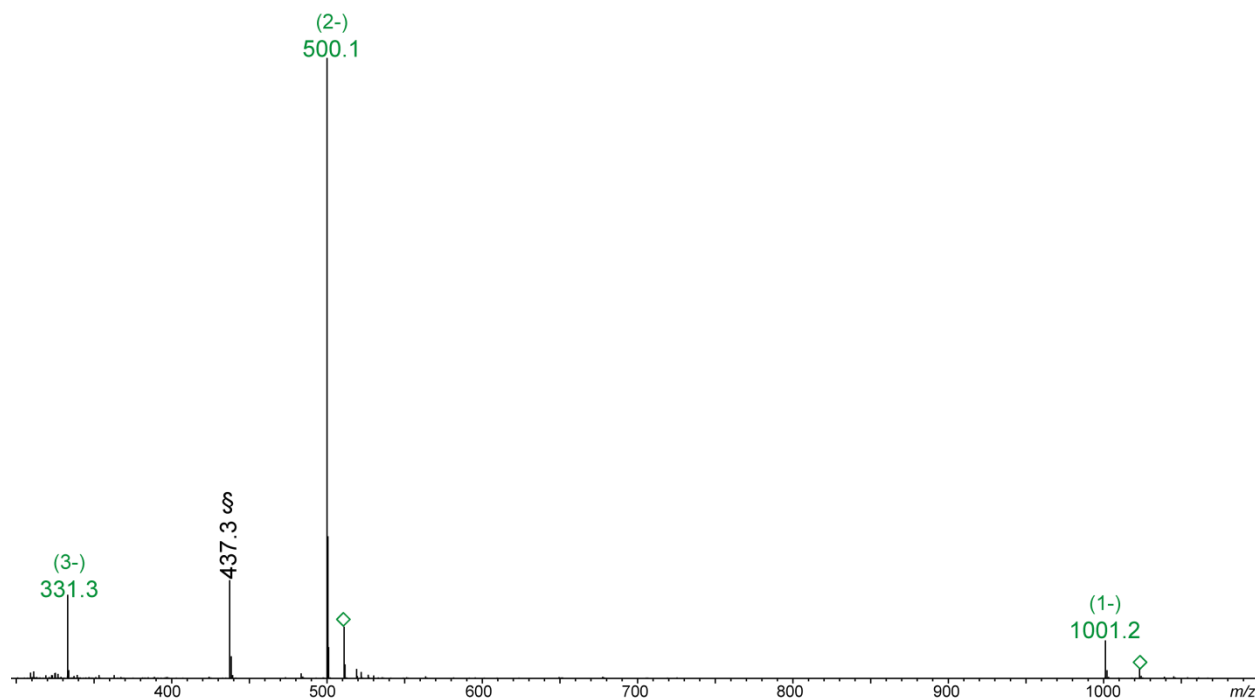


Figure S1. Negative mode ESI-MS spectrum of the model pentamer $M_2 \cdot M_2 \cdot M_2 \cdot M_2 \cdot M_2\text{-T}$, with ion charge state indicated into parentheses. Diamonds designate ionic forms detected after H/Na and/or H/K exchanges. §: peak from the chemical background.

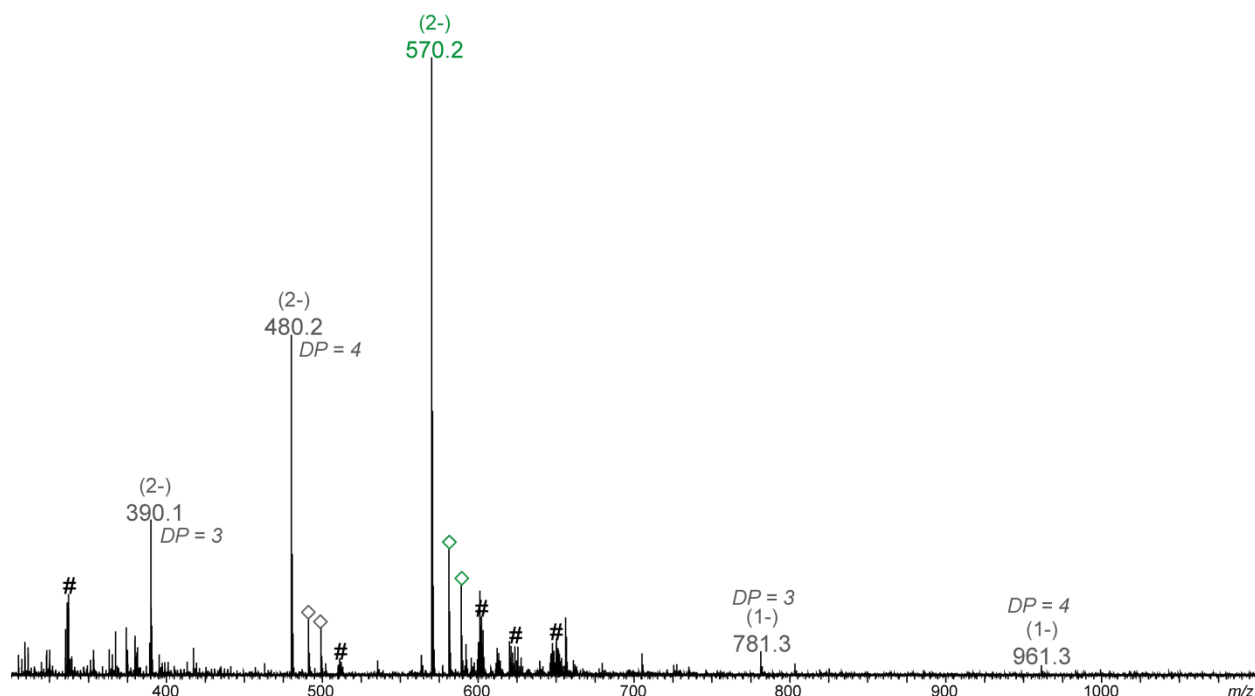


Figure S2. Negative mode ESI-MS spectrum of the model pentamer $M_4 \cdot M_4 \cdot M_4 \cdot M_4 \cdot M_4\text{-T}$, with ion charge state indicated into parentheses. Signals annotated in grey correspond to oligomers of lower degree of polymerization (DP). Diamonds designate ionic forms detected after H/Na and/or H/K exchanges. #: clusters of trichloroacetic acid.

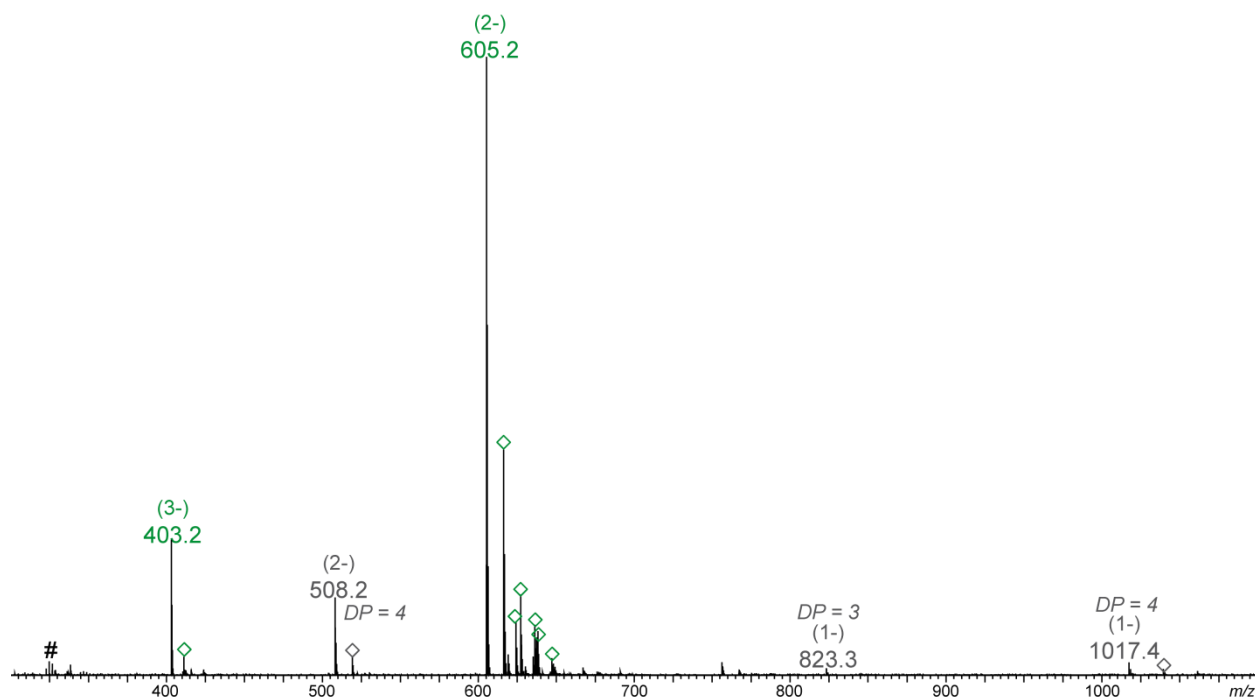


Figure S3. Negative mode ESI-MS spectrum of the model pentamer $M_5 \cdot M_5 \cdot M_5 \cdot M_5 \cdot M_5$ -T, with ion charge state indicated into parentheses. Signals annotated in grey correspond to oligomers of lower degree of polymerization (DP). Diamonds designate ionic forms detected after H/Na and/or H/K exchanges. #: clusters of trichloroacetic acid.

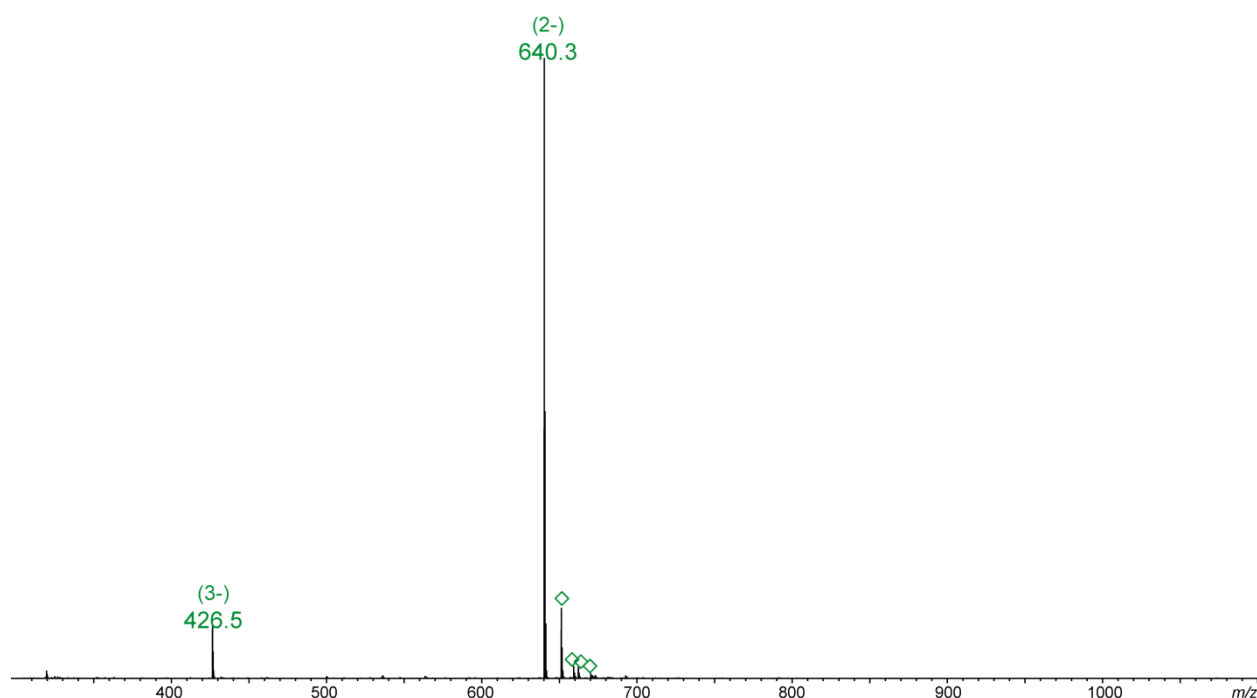


Figure S4. Negative mode ESI-MS spectrum of the model pentamer $M_6 \cdot M_6 \cdot M_6 \cdot M_6 \cdot M_6$ -T, with ion charge state indicated into parentheses. Diamonds designate ionic forms detected after H/Na and/or H/K exchanges.

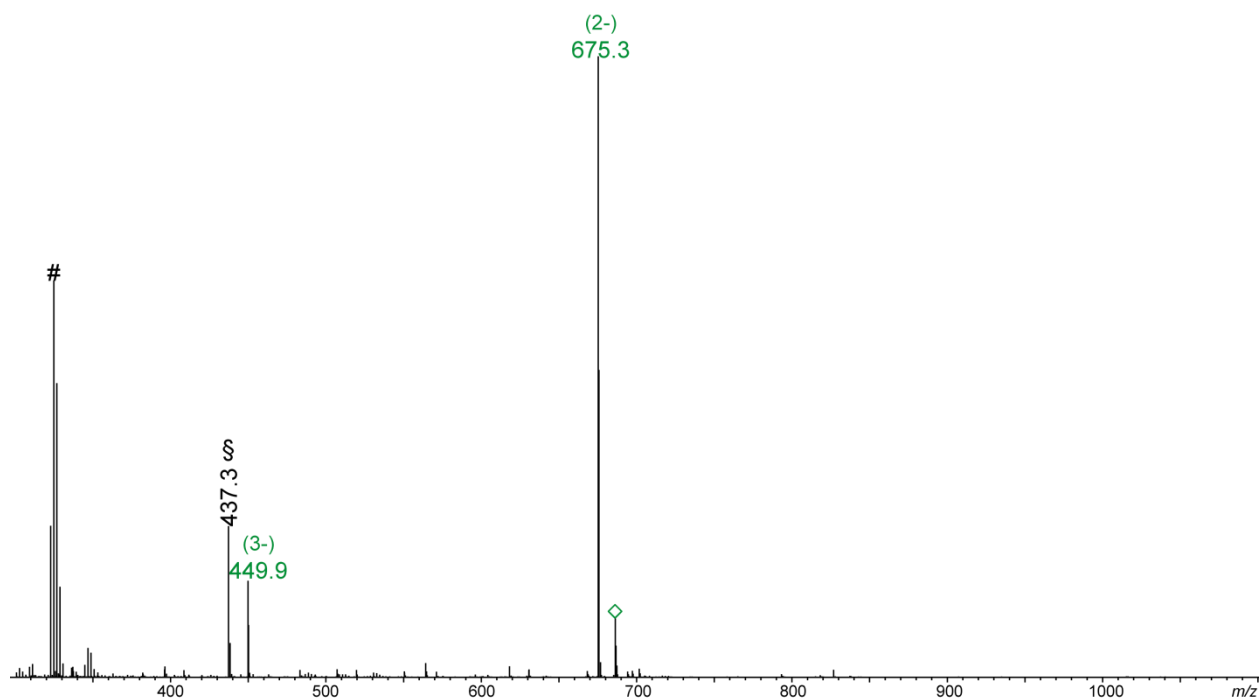


Figure S5. Negative mode ESI-MS spectrum of the model pentamer $M_7 \cdot M_7 \cdot M_7 \cdot M_7 \cdot M_7\text{-T}$, with ion charge state indicated into parentheses. Signals annotated in grey correspond to oligomers of lower degree of polymerization (DP). Diamonds designate ionic forms detected after H/Na and/or H/K exchanges. #: clusters of trichloroacetic acid. §: peak from the chemical background.

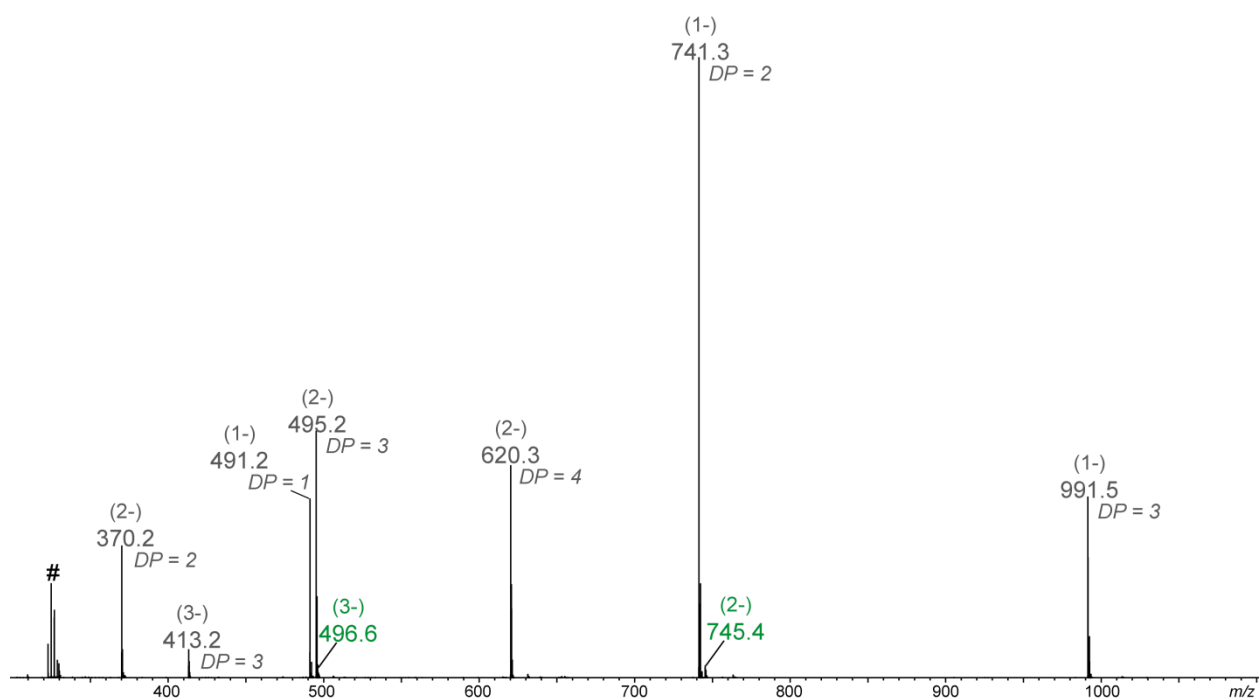


Figure S6. Negative mode ESI-MS spectrum of the model pentamer $M_8 \cdot M_8 \cdot M_8 \cdot M_8 \cdot M_8\text{-T}$, with ion charge state indicated into parentheses. Signals annotated in grey correspond to oligomers of lower degree of polymerization (DP). #: clusters of trichloroacetic acid.

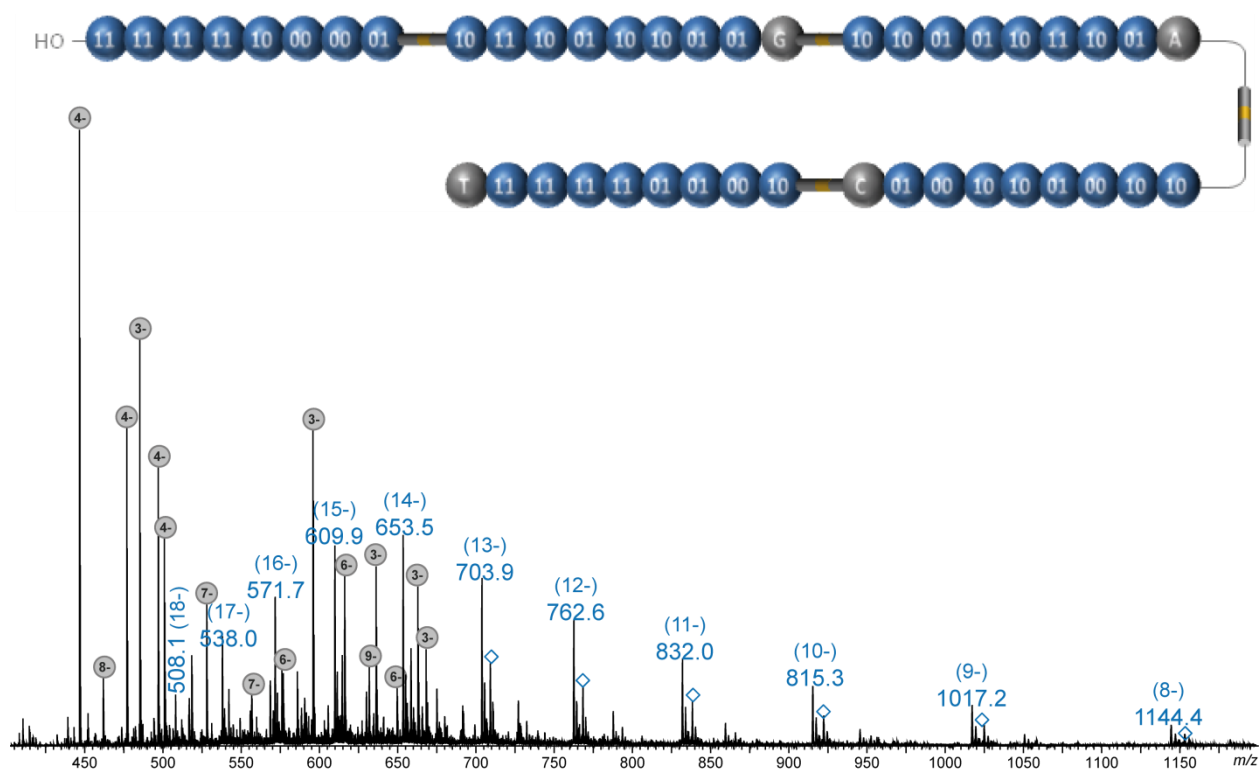


Figure S7. Negative mode ESI-MS spectrum of **P2**. Signals of the intact polymer (in blue) are distributed over different charge states, as indicated into parentheses. Diamonds designate ionic forms detected after H/Na and/or H/K exchanges while signals annotated with charge states in grey circles correspond to in-source fragments.

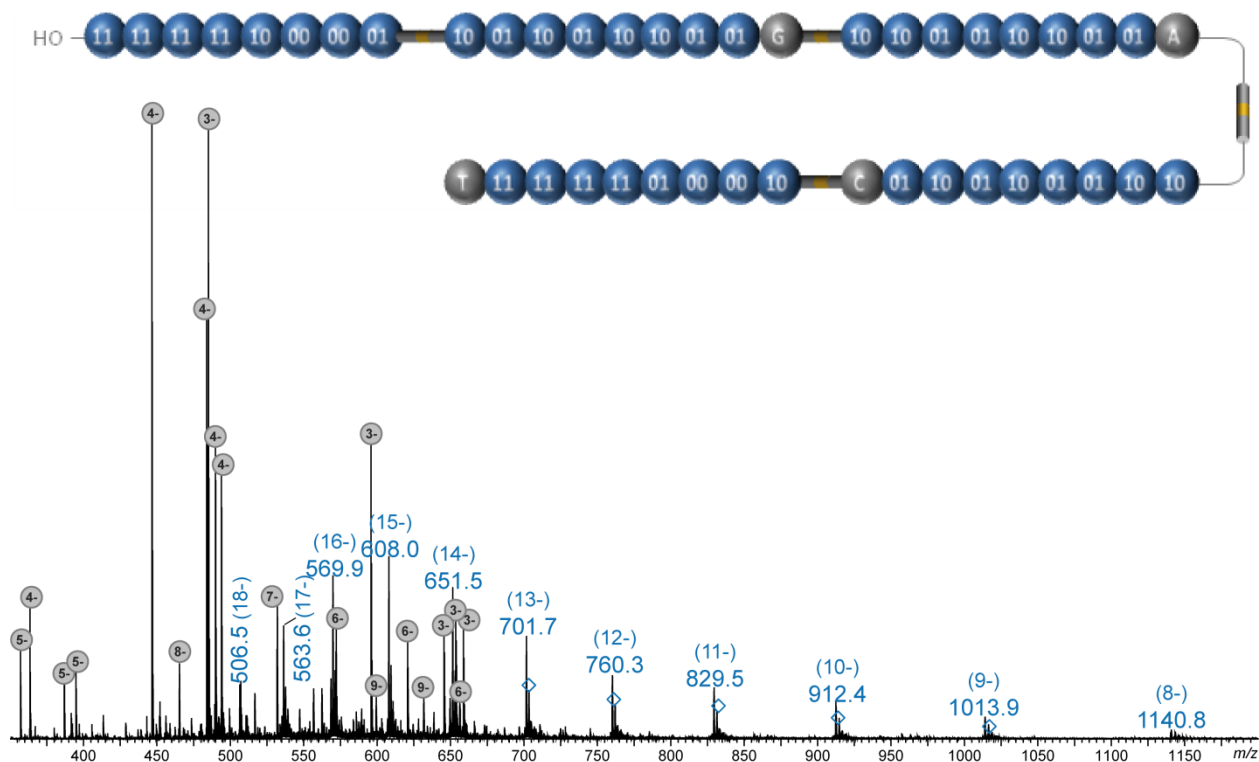


Figure S8. Negative mode ESI-MS spectrum of **P3**. Signals of the intact polymer (in blue) are distributed over different charge states, as indicated into parentheses. Diamonds designate ionic forms detected after H/Na and/or H/K exchanges while signals annotated with charge states in grey circles correspond to in-source fragments.

Sequencing methodology

The pseudo-MS³ methodology developed to recover the whole sequence of these segment-truncated digital polymers relies of three main experimental steps, using negative mode ESI.

Step 1. In the MS mode, the charge state distribution of the polymer indicates the number of its segments because blocks containing 8 coding units were observed to carry 3 charges on average. Accordingly, the preferential z charge state is most often a multiple of 3, from which the number n of segments can be deduced as $n = z/3$.

Step 2. CID is performed for the precursor at the $z=3n$ charge state: because negative charges are evenly distributed, this lowers the complexity of the MS/MS data. This first CID step allows alkoxyamine bonds to be cleaved and yields fragments that contain from 1 to $(n-1)$ segments and a preferential charge state which is a multiple of 3. Fragments containing a single segment are hence at a 3- charge state, and their m/z value depends on both their coding monomer composition and their tag. Since tags were selected to provide a unique mass signature, each 3- fragment can be assigned to a single tag based on its m/z value. As shown in Table S2 built for the 4-monomers alphabet, a given m/z value may be found multiple times but always in a single tag column. For example, m/z 616.1 is found 4 times in the A tag column (but never in any other tag columns), corresponding to different coding unit composition: $6M_1+1M_2+1M_4$, $6M_1+2M_3$, $5M_1+2M_2+1M_3$, and $4M_1+4M_2$. This means that data from Table S2 cannot be systematically used to determine comonomer composition (which will further be determined during segment sequencing) but unambiguously allows identification of the tag hold by the segment, and hence its initial location in the original chain. Of note, m/z data are reported in Table S2 with only one decimal digit for the sake of clarity, but all experiments are performed using a high resolution orthogonal acceleration TOF mass analyzer enabling accurate mass measurements with four significant decimal digits for all m/z values.

Step 3. The cone voltage is raised in order to induce in-source production of fragments, amongst which triply charged fragments are successively subjected to CID. So-obtained pseudo-MS³ spectra are then analyzed based on dissociation rules established for poly(phosphodiester)s to recover each segment sequence.² Briefly, cleavage of all phosphate bonds gives rise to α -containing (a_i^{z-} , b_i^{z-} , c_i^{z-} , d_i^{z-}) or ω -containing (w_i^{z-} , x_i^{z-} , y_i^{z-} , z_i^{z-}) fragments. Because of the reactivity of the carbon-centered radical found in the ω end-group of all ions containing a single segment (but the last one), ω -containing fragments further readily dissociate. Nevertheless, considering only α -containing fragments allows full sequence coverage. Finally, all sequences are ordered based on rules defined with the tag system to reveal the whole polymer sequence.

Step 1 and step 2 are detailed in the case of **P1** (Figure S9a) and step 3 is illustrated for blind sequencing of the first segment of **P1** (Figure S9b).

Table S2. List of m/z values of triply charged fragments containing a single segment, as a function of their coding monomeric composition (4-monomers alphabet) and their tag.

M ₁	M ₂	M ₃	M ₄	No	G	A	C	T	M ₁	M ₂	M ₃	M ₄	No	G	A	C	T
8	0	0	0	415.0	602.8	597.4	589.4	525.7	3	1	2	2	466.4	654.2	648.8	640.8	577.1
7	1	0	0	419.7	607.4	602.1	594.1	530.4	3	1	1	3	471.1	658.8	653.5	645.5	581.8
7	0	1	0	424.4	612.1	606.8	598.8	535.1	3	1	0	4	475.8	663.5	658.2	650.2	586.5
7	0	0	1	429.1	616.8	611.4	603.4	539.8	3	0	5	0	461.8	649.5	644.1	636.1	572.5
6	2	0	0	424.4	612.1	606.8	598.8	535.1	3	0	4	1	466.4	654.2	648.8	640.8	577.1
6	1	1	0	429.1	616.8	611.4	603.4	539.8	3	0	3	2	471.1	658.8	653.5	645.5	581.8
6	1	0	1	433.7	621.5	616.1	608.1	544.4	3	0	2	3	475.8	663.5	658.2	650.2	586.5
6	0	2	0	433.7	621.5	616.1	608.1	544.4	3	0	1	4	480.5	668.2	662.8	654.8	591.1
6	0	1	1	438.4	626.1	620.8	612.8	549.1	3	0	0	5	485.1	672.8	667.5	659.5	595.8
6	0	0	2	443.1	630.8	625.5	617.5	553.8	2	6	0	0	443.1	630.8	625.5	617.5	553.8
5	3	0	0	429.1	616.8	611.4	603.4	539.8	2	5	1	0	447.8	635.5	630.1	622.1	558.4
5	2	1	0	433.7	621.5	616.1	608.1	544.4	2	5	0	1	452.4	640.1	634.8	626.8	563.1
5	2	0	1	438.4	626.1	620.8	612.8	549.1	2	4	2	0	452.4	640.1	634.8	626.8	563.1
5	1	2	0	438.4	626.1	620.8	612.8	549.1	2	4	1	1	457.1	644.8	639.5	631.5	567.8
5	1	1	1	443.1	630.8	625.5	617.5	553.8	2	4	0	2	461.8	649.5	644.1	636.1	572.5
5	1	0	2	447.8	635.5	630.1	622.1	558.4	2	3	3	0	457.1	644.8	639.5	631.5	567.8
5	0	3	0	443.1	630.8	625.5	617.5	553.8	2	3	2	1	461.8	649.5	644.1	636.1	572.5
5	0	2	1	447.8	635.5	630.1	622.1	558.4	2	3	1	2	466.4	654.2	648.8	640.8	577.1
5	0	1	2	452.4	640.1	634.8	626.8	563.1	2	3	0	3	471.1	658.8	653.5	645.5	581.8
5	0	0	3	457.1	644.8	639.5	631.5	567.8	2	2	4	0	461.8	649.5	644.1	636.1	572.5
4	4	0	0	433.7	621.5	616.1	608.1	544.4	2	2	3	1	466.4	654.2	648.8	640.8	577.1
4	3	1	0	438.4	626.1	620.8	612.8	549.1	2	2	2	2	471.1	658.8	653.5	645.5	581.8
4	3	0	1	443.1	630.8	625.5	617.5	553.8	2	2	1	3	475.8	663.5	658.2	650.2	586.5
4	2	2	0	443.1	630.8	625.5	617.5	553.8	2	2	0	4	480.5	668.2	662.8	654.8	591.1
4	2	1	1	447.8	635.5	630.1	622.1	558.4	2	1	5	0	466.4	654.2	648.8	640.8	577.1
4	2	0	2	452.4	640.1	634.8	626.8	563.1	2	1	4	1	471.1	658.8	653.5	645.5	581.8
4	1	3	0	447.8	635.5	630.1	622.1	558.4	2	1	3	2	475.8	663.5	658.2	650.2	586.5
4	1	2	1	452.4	640.1	634.8	626.8	563.1	2	1	2	3	480.5	668.2	662.8	654.8	591.1
4	1	1	2	457.1	644.8	639.5	631.5	567.8	2	1	1	4	485.1	672.8	667.5	659.5	595.8
4	1	0	3	461.8	649.5	644.1	636.1	572.5	2	1	0	5	489.8	677.5	672.2	664.2	600.5
4	0	4	0	452.4	640.1	634.8	626.8	563.1	2	0	6	0	471.1	658.8	653.5	645.5	581.8
4	0	3	1	457.1	644.8	639.5	631.5	567.8	2	0	5	1	475.8	663.5	658.2	650.2	586.5
4	0	2	2	461.8	649.5	644.1	636.1	572.5	2	0	4	2	480.5	668.2	662.8	654.8	591.1
4	0	1	3	466.4	654.2	648.8	640.8	577.1	2	0	3	3	485.1	672.8	667.5	659.5	595.8
4	0	0	4	471.1	658.8	653.5	645.5	581.8	2	0	2	4	489.8	677.5	672.2	664.2	600.5
3	5	0	0	438.4	626.1	620.8	612.8	549.1	2	0	1	5	494.5	682.2	676.9	668.8	605.2
3	4	1	0	443.1	630.8	625.5	617.5	553.8	2	0	0	6	499.1	686.9	681.5	673.5	609.8
3	4	0	1	447.8	635.5	630.1	622.1	558.4	1	7	0	0	447.8	635.5	630.1	622.1	558.4
3	3	2	0	447.8	635.5	630.1	622.1	558.4	1	6	1	0	452.4	640.1	634.8	626.8	563.1
3	3	1	1	452.4	640.1	634.8	626.8	563.1	1	6	0	1	457.1	644.8	639.5	631.5	567.8
3	3	0	2	457.1	644.8	639.5	631.5	567.8	1	5	2	0	457.1	644.8	639.5	631.5	567.8
3	2	3	0	452.4	640.1	634.8	626.8	563.1	1	5	1	1	461.8	649.5	644.1	636.1	572.5
3	2	2	1	457.1	644.8	639.5	631.5	567.8	1	5	0	2	466.4	654.2	648.8	640.8	577.1
3	2	1	2	461.8	649.5	644.1	636.1	572.5	1	4	3	0	461.8	649.5	644.1	636.1	572.5
3	2	0	3	466.4	654.2	648.8	640.8	577.1	1	4	2	1	466.4	654.2	648.8	640.8	577.1
3	1	4	0	457.1	644.8	639.5	631.5	567.8	1	4	1	2	471.1	658.8	653.5	645.5	581.8
3	1	3	1	461.8	649.5	644.1	636.1	572.5	1	4	0	3	475.8	663.5	658.2	650.2	586.5

Table S2. Cont'd.

M ₁	M ₂	M ₃	M ₄	No	G	A	C	T	M ₁	M ₂	M ₃	M ₄	No	G	A	C	T
1	3	4	0	466.4	654.2	648.8	640.8	577.1	0	4	4	0	471.1	658.8	653.5	645.5	581.8
1	3	3	1	471.1	658.8	653.5	645.5	581.8	0	4	3	1	475.8	663.5	658.2	650.2	586.5
1	3	2	2	475.8	663.5	658.2	650.2	586.5	0	4	2	2	480.5	668.2	662.8	654.8	591.1
1	3	1	3	480.5	668.2	662.8	654.8	591.1	0	4	1	3	485.1	672.8	667.5	659.5	595.8
1	3	0	4	485.1	672.8	667.5	659.5	595.8	0	4	0	4	489.8	677.5	672.2	664.2	600.5
1	2	5	0	471.1	658.8	653.5	645.5	581.8	0	3	5	0	475.8	663.5	658.2	650.2	586.5
1	2	4	1	475.8	663.5	658.2	650.2	586.5	0	3	4	1	480.5	668.2	662.8	654.8	591.1
1	2	3	2	480.5	668.2	662.8	654.8	591.1	0	3	3	2	485.1	672.8	667.5	659.5	595.8
1	2	2	3	485.1	672.8	667.5	659.5	595.8	0	3	2	3	489.8	677.5	672.2	664.2	600.5
1	2	1	4	489.8	677.5	672.2	664.2	600.5	0	3	1	4	494.5	682.2	676.9	668.8	605.2
1	2	0	5	494.5	682.2	676.9	668.8	605.2	0	3	0	5	499.1	686.9	681.5	673.5	609.8
1	1	6	0	475.8	663.5	658.2	650.2	586.5	0	2	6	0	480.5	668.2	662.8	654.8	591.1
1	1	5	1	480.5	668.2	662.8	654.8	591.1	0	2	5	1	485.1	672.8	667.5	659.5	595.8
1	1	4	2	485.1	672.8	667.5	659.5	595.8	0	2	4	2	489.8	677.5	672.2	664.2	600.5
1	1	3	3	489.8	677.5	672.2	664.2	600.5	0	2	3	3	494.5	682.2	676.9	668.8	605.2
1	1	2	4	494.5	682.2	676.9	668.8	605.2	0	2	2	4	499.1	686.9	681.5	673.5	609.8
1	1	1	5	499.1	686.9	681.5	673.5	609.8	0	2	1	5	503.8	691.5	686.2	678.2	614.5
1	1	0	6	503.8	691.5	686.2	678.2	614.5	0	2	0	6	508.5	696.2	690.9	682.9	619.2
1	0	7	0	480.5	668.2	662.8	654.8	591.1	0	1	7	0	485.1	672.8	667.5	659.5	595.8
1	0	6	1	485.1	672.8	667.5	659.5	595.8	0	1	6	1	489.8	677.5	672.2	664.2	600.5
1	0	5	2	489.8	677.5	672.2	664.2	600.5	0	1	5	2	494.5	682.2	676.9	668.8	605.2
1	0	4	3	494.5	682.2	676.9	668.8	605.2	0	1	4	3	499.1	686.9	681.5	673.5	609.8
1	0	3	4	499.1	686.9	681.5	673.5	609.8	0	1	3	4	503.8	691.5	686.2	678.2	614.5
1	0	2	5	503.8	691.5	686.2	678.2	614.5	0	1	2	5	508.5	696.2	690.9	682.9	619.2
1	0	1	6	508.5	696.2	690.9	682.9	619.2	0	1	1	6	513.2	700.9	695.5	687.5	623.9
1	0	0	7	513.2	700.9	695.5	687.5	623.9	0	1	0	7	517.8	705.5	700.2	692.2	628.5
0	8	0	0	452.4	640.1	634.8	626.8	563.1	0	0	8	0	489.8	677.5	672.2	664.2	600.5
0	7	1	0	457.1	644.8	639.5	631.5	567.8	0	0	7	1	494.5	682.2	676.9	668.8	605.2
0	7	0	1	461.8	649.5	644.1	636.1	572.5	0	0	6	2	499.1	686.9	681.5	673.5	609.8
0	6	2	0	461.8	649.5	644.1	636.1	572.5	0	0	5	3	503.8	691.5	686.2	678.2	614.5
0	6	1	1	466.4	654.2	648.8	640.8	577.1	0	0	4	4	508.5	696.2	690.9	682.9	619.2
0	6	0	2	471.1	658.8	653.5	645.5	581.8	0	0	3	5	513.2	700.9	695.5	687.5	623.9
0	5	3	0	466.4	654.2	648.8	640.8	577.1	0	0	2	6	517.8	705.5	700.2	692.2	628.5
0	5	2	1	471.1	658.8	653.5	645.5	581.8	0	0	1	7	522.5	710.2	704.9	696.9	633.2
0	5	1	2	475.8	663.5	658.2	650.2	586.5	0	0	0	8	527.2	714.9	709.6	701.6	637.9
0	5	0	3	480.5	668.2	662.8	654.8	591.1									

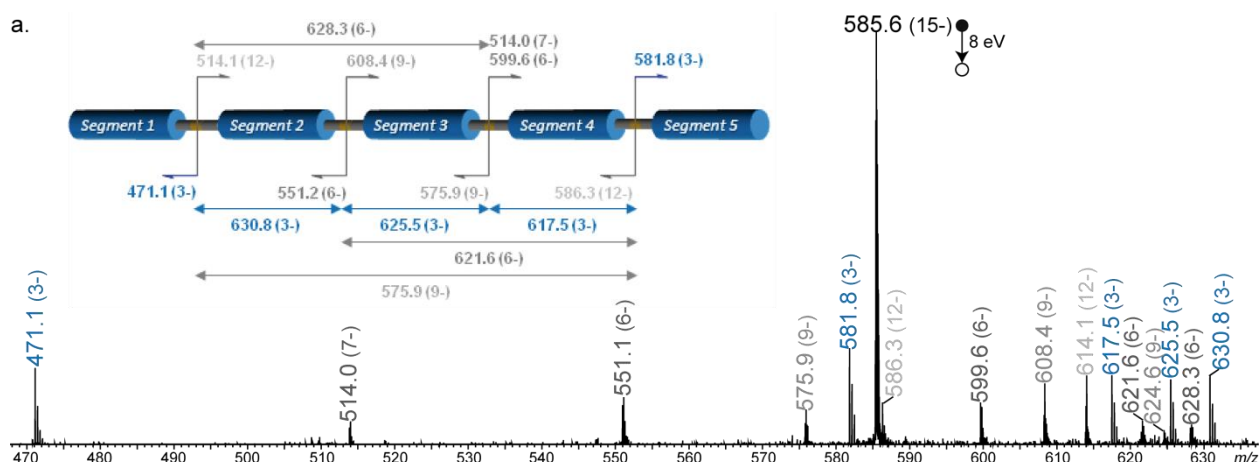


Figure S9. Sequencing of **P1**. Based on data recorded in the MS mode, step 1 consists of determining the number of segments in **P1** from its preferential charge state. Data reported in Figure 1a show maximal abundance for **P1** with $z = 15$, indicating $n = z/3 = 5$ segments. Accordingly, (a) CID (collision energy in the laboratory frame) is performed for $[\mathbf{P1}-15\mathbf{H}]^{15-}$ at m/z 585.6 to induce homolysis of alkoxyamine bonds, yielding fragments that contain from 1 to 4 segments and a preferential charge state which is a multiple of 3. Fragments containing a single segment (annotated in blue) can be easily identified based on their isotopic pattern since they are at a 3- charge state. Their m/z value is then studied with reference to Table S2:

- the m/z 471.1 value can only be found in the column entitled “No” → this fragment contains the first segment, which coding unit composition is $4\mathbf{M}_1+0\mathbf{M}_2+0\mathbf{M}_3+4\mathbf{M}_4$, $3\mathbf{M}_1+1\mathbf{M}_2+1\mathbf{M}_3+3\mathbf{M}_4$, $3\mathbf{M}_1+0\mathbf{M}_2+3\mathbf{M}_3+2\mathbf{M}_4$, $2\mathbf{M}_1+3\mathbf{M}_2+0\mathbf{M}_3+3\mathbf{M}_4$, $2\mathbf{M}_1+2\mathbf{M}_2+2\mathbf{M}_3+2\mathbf{M}_4$, $2\mathbf{M}_1+1\mathbf{M}_2+4\mathbf{M}_3+1\mathbf{M}_4$, $2\mathbf{M}_1+0\mathbf{M}_2+6\mathbf{M}_3+0\mathbf{M}_4$, $1\mathbf{M}_1+4\mathbf{M}_2+1\mathbf{M}_3+2\mathbf{M}_4$, $1\mathbf{M}_1+3\mathbf{M}_2+3\mathbf{M}_3+1\mathbf{M}_4$, $1\mathbf{M}_1+2\mathbf{M}_2+5\mathbf{M}_3+0\mathbf{M}_4$, $0\mathbf{M}_1+6\mathbf{M}_2+0\mathbf{M}_3+2\mathbf{M}_4$, $0\mathbf{M}_1+5\mathbf{M}_2+1\mathbf{M}_3+1\mathbf{M}_4$, or $0\mathbf{M}_1+4\mathbf{M}_2+4\mathbf{M}_3+0\mathbf{M}_4$;
- the m/z 581.8 value can only be found in the column entitled “T” → this fragment contains the last (fifth) segment, which coding unit composition is $4\mathbf{M}_1+0\mathbf{M}_2+0\mathbf{M}_3+4\mathbf{M}_4$, $3\mathbf{M}_1+1\mathbf{M}_2+1\mathbf{M}_3+3\mathbf{M}_4$, $3\mathbf{M}_1+0\mathbf{M}_2+3\mathbf{M}_3+2\mathbf{M}_4$, $2\mathbf{M}_1+3\mathbf{M}_2+0\mathbf{M}_3+3\mathbf{M}_4$, $2\mathbf{M}_1+2\mathbf{M}_2+2\mathbf{M}_3+2\mathbf{M}_4$, $2\mathbf{M}_1+1\mathbf{M}_2+4\mathbf{M}_3+1\mathbf{M}_4$, $2\mathbf{M}_1+0\mathbf{M}_2+6\mathbf{M}_3+0\mathbf{M}_4$, $1\mathbf{M}_1+4\mathbf{M}_2+1\mathbf{M}_3+2\mathbf{M}_4$, $1\mathbf{M}_1+3\mathbf{M}_2+3\mathbf{M}_3+1\mathbf{M}_4$, $1\mathbf{M}_1+2\mathbf{M}_2+5\mathbf{M}_3+0\mathbf{M}_4$, $0\mathbf{M}_1+6\mathbf{M}_2+0\mathbf{M}_3+2\mathbf{M}_4$, $0\mathbf{M}_1+5\mathbf{M}_2+1\mathbf{M}_3+1\mathbf{M}_4$, or $0\mathbf{M}_1+4\mathbf{M}_2+4\mathbf{M}_3+0\mathbf{M}_4$;
- the m/z 617.5 value can only be found in the column entitled “C” → this fragment contains the fourth segment, which coding unit composition is $6\mathbf{M}_1+0\mathbf{M}_2+0\mathbf{M}_3+2\mathbf{M}_4$, $5\mathbf{M}_1+1\mathbf{M}_2+1\mathbf{M}_3+1\mathbf{M}_4$, $5\mathbf{M}_1+0\mathbf{M}_2+3\mathbf{M}_3+0\mathbf{M}_4$, $4\mathbf{M}_1+3\mathbf{M}_2+0\mathbf{M}_3+1\mathbf{M}_4$, $4\mathbf{M}_1+2\mathbf{M}_2+2\mathbf{M}_3+0\mathbf{M}_4$, $3\mathbf{M}_1+4\mathbf{M}_2+1\mathbf{M}_3+0\mathbf{M}_4$, or $2\mathbf{M}_1+6\mathbf{M}_2+0\mathbf{M}_3+0\mathbf{M}_4$;
- the m/z 625.5 value can only be found in the column entitled “A” → this fragment contains the third segment, which coding unit composition is $6\mathbf{M}_1+0\mathbf{M}_2+0\mathbf{M}_3+2\mathbf{M}_4$, $5\mathbf{M}_1+1\mathbf{M}_2+1\mathbf{M}_3+1\mathbf{M}_4$, $5\mathbf{M}_1+0\mathbf{M}_2+3\mathbf{M}_3+0\mathbf{M}_4$, $4\mathbf{M}_1+3\mathbf{M}_2+0\mathbf{M}_3+1\mathbf{M}_4$, $4\mathbf{M}_1+2\mathbf{M}_2+2\mathbf{M}_3+0\mathbf{M}_4$, $3\mathbf{M}_1+4\mathbf{M}_2+1\mathbf{M}_3+0\mathbf{M}_4$, or $2\mathbf{M}_1+6\mathbf{M}_2+0\mathbf{M}_3+0\mathbf{M}_4$;
- the m/z 630.8 value can only be found in the column entitled “G” → this fragment contains the second segment, which coding unit composition is $6\mathbf{M}_1+0\mathbf{M}_2+0\mathbf{M}_3+2\mathbf{M}_4$, $5\mathbf{M}_1+1\mathbf{M}_2+1\mathbf{M}_3+1\mathbf{M}_4$, $5\mathbf{M}_1+0\mathbf{M}_2+3\mathbf{M}_3+0\mathbf{M}_4$, $4\mathbf{M}_1+3\mathbf{M}_2+0\mathbf{M}_3+1\mathbf{M}_4$, $4\mathbf{M}_1+2\mathbf{M}_2+2\mathbf{M}_3+0\mathbf{M}_4$, $3\mathbf{M}_1+4\mathbf{M}_2+1\mathbf{M}_3+0\mathbf{M}_4$, or $2\mathbf{M}_1+6\mathbf{M}_2+0\mathbf{M}_3+0\mathbf{M}_4$.

These results permit to construct the dissociation scheme shown in inset and to identify other fragments (in grey) that contain more than one segment. Each of the five primary fragments at the 3- charge state are then subjected to the second activation stage (so-called pseudo-MS³) to determine the coding sequence of the single segment they contain, as detailed for the first fragment at m/z 471.1 in part (b).

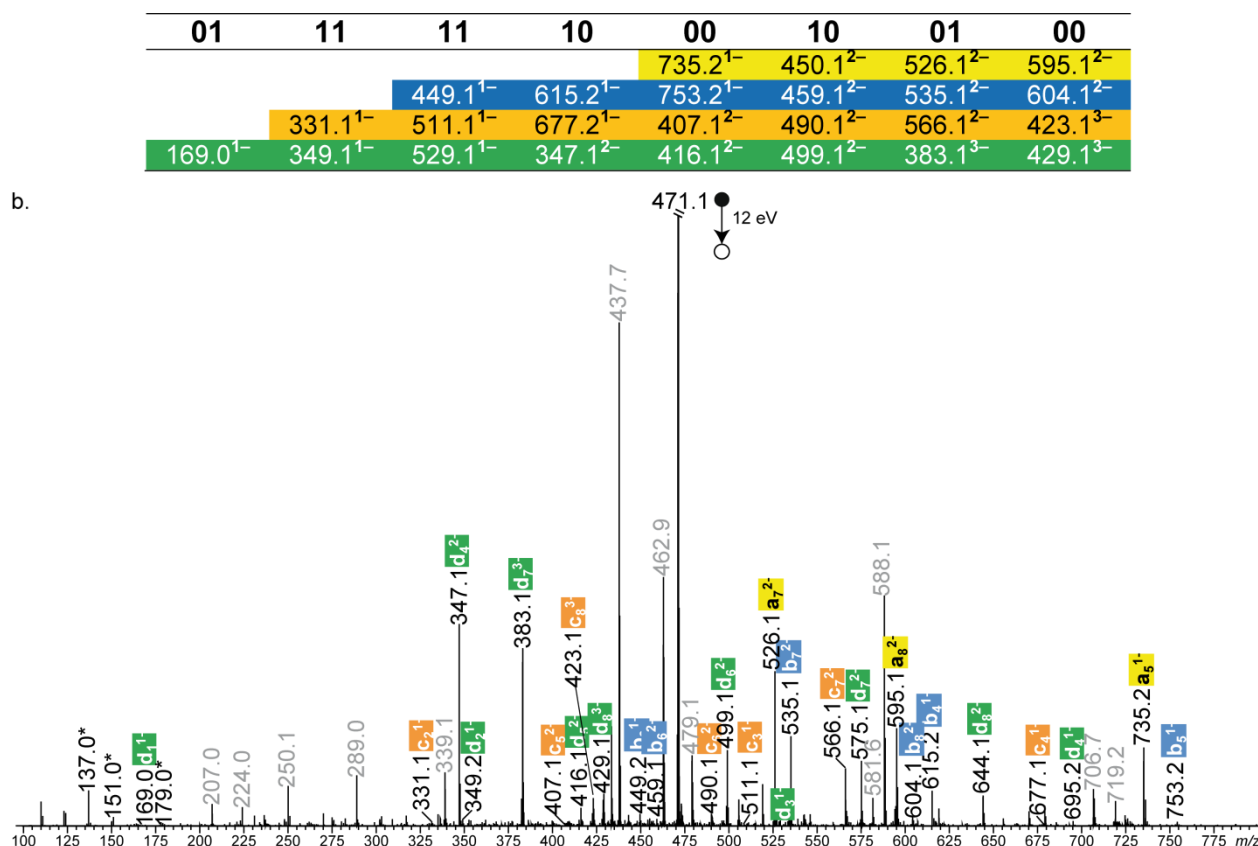


Figure S9(Cont'd). Sequencing of **P1**. **(b)** Pseudo-MS³ spectrum recorded for the m/z 471.1 primary fragment in order to recover the sequence of the first segment. Useful ions to be considered for sequencing are d_i^{z-} (in green), c_i^{z-} (in orange), b_i^{z-} (in blue) and a_i^{z-} (in yellow). Deprotonated monomers are designated by an asterisk and peaks annotated in grey are secondary products. Collision energy is given in the laboratory frame. For the sake of clarity, the y-axis has been expanded by a factor of 2.5). Sequencing is performed with d_i^{z-} ions (most often all detected), using other α -containing fragments (a_i^{z-} , b_i^{z-} , c_i^{z-}) for validation. The first member of the series, d_1^{1-} , is expected at m/z 155.0, m/z 169.0, m/z 183.0 or m/z 197.1 depending of the first coding unit being 00 (M_1), 01 (M_2), 10 (M_3), or 11 (M_4), respectively. Amongst these possible values, a peak is only observed at m/z 169.0 \rightarrow sequence starts as α -01. Each member of the d_i^{z-} series is then searched by adding the mass of one or the other unit (138.0 Da for 00, 152.0 Da for 01, 166.0 Da for 10, 180.1 Da for 11) to the m/z value of the previous congener (*e.g.*, d_2^{1-} found at m/z 349.1 = 169.0 + 180.1 \rightarrow α -01-11; d_3^{1-} found at m/z 529.1 = 349.1 + 180.1 \rightarrow α -01-11-11; ...) while other α -containing fragments are expected at m/z values such as $c_i^{z-} = d_i^{z-} - 18$ Da, $b_i^{z-} = c_i^{z-} - 62$ Da, and $a_i^{z-} = b_i^{z-} - 18$ Da. As illustrated by the top sequence coverage table, this blind sequencing approach reveals the first segment sequence as 01-11-11-10-00-10-01-00.

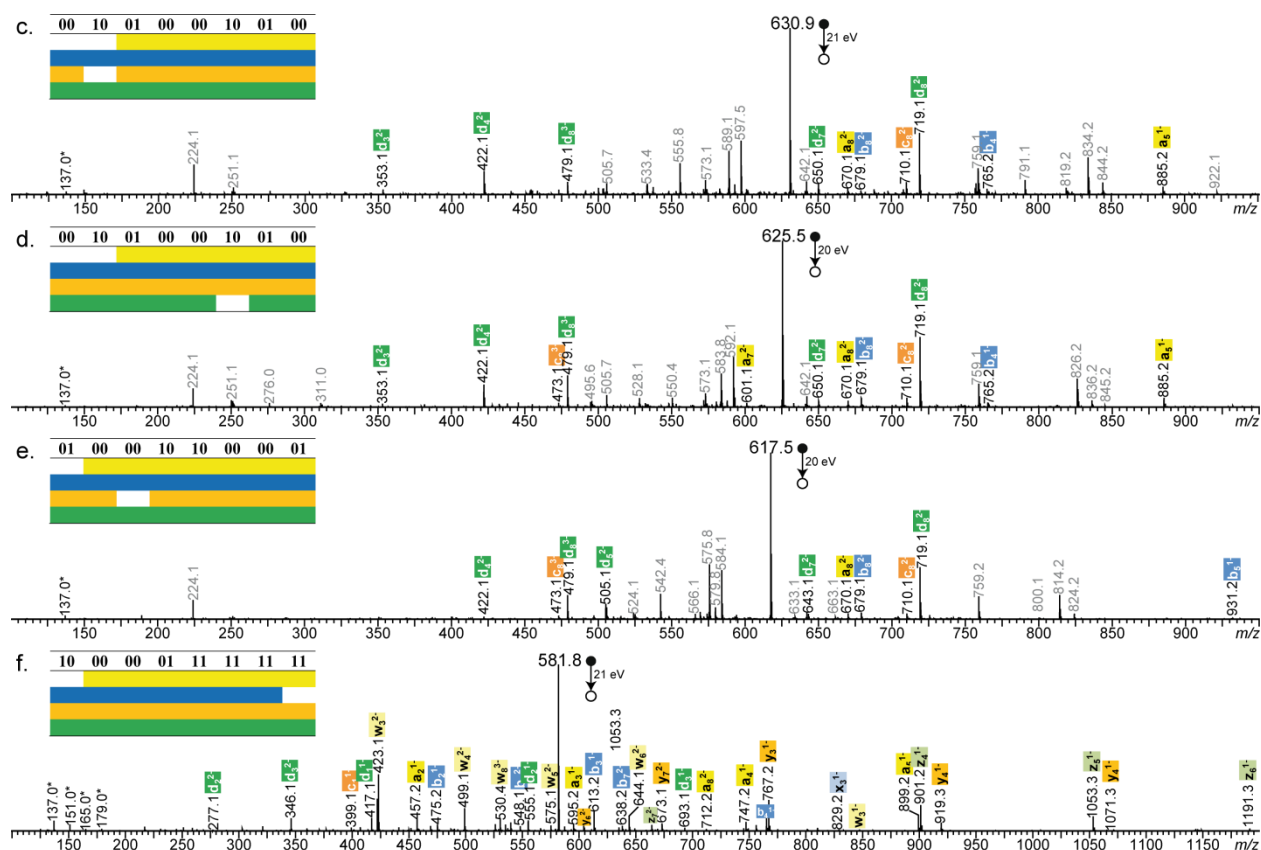


Figure S9(Cont'd). Sequencing of P1. Pseudo-MS³ spectra for (c) the second segment at m/z 630.8, (d) the third segment at m/z 625.5, (e) the fourth segment at m/z 617.5, and (f) the fifth segment at m/z 581.8. Fragments d₁⁻ (green), c₁⁻ (orange), b₁⁻ (blue) and a₁⁻ (yellow) used for sequencing are not all annotated for the sake of clarity but their detection is summarized in the associated sequence coverage table. Deprotonated monomers are designated by an asterisk and peaks annotated in grey are secondary products. All collision energies are given in the laboratory frame.

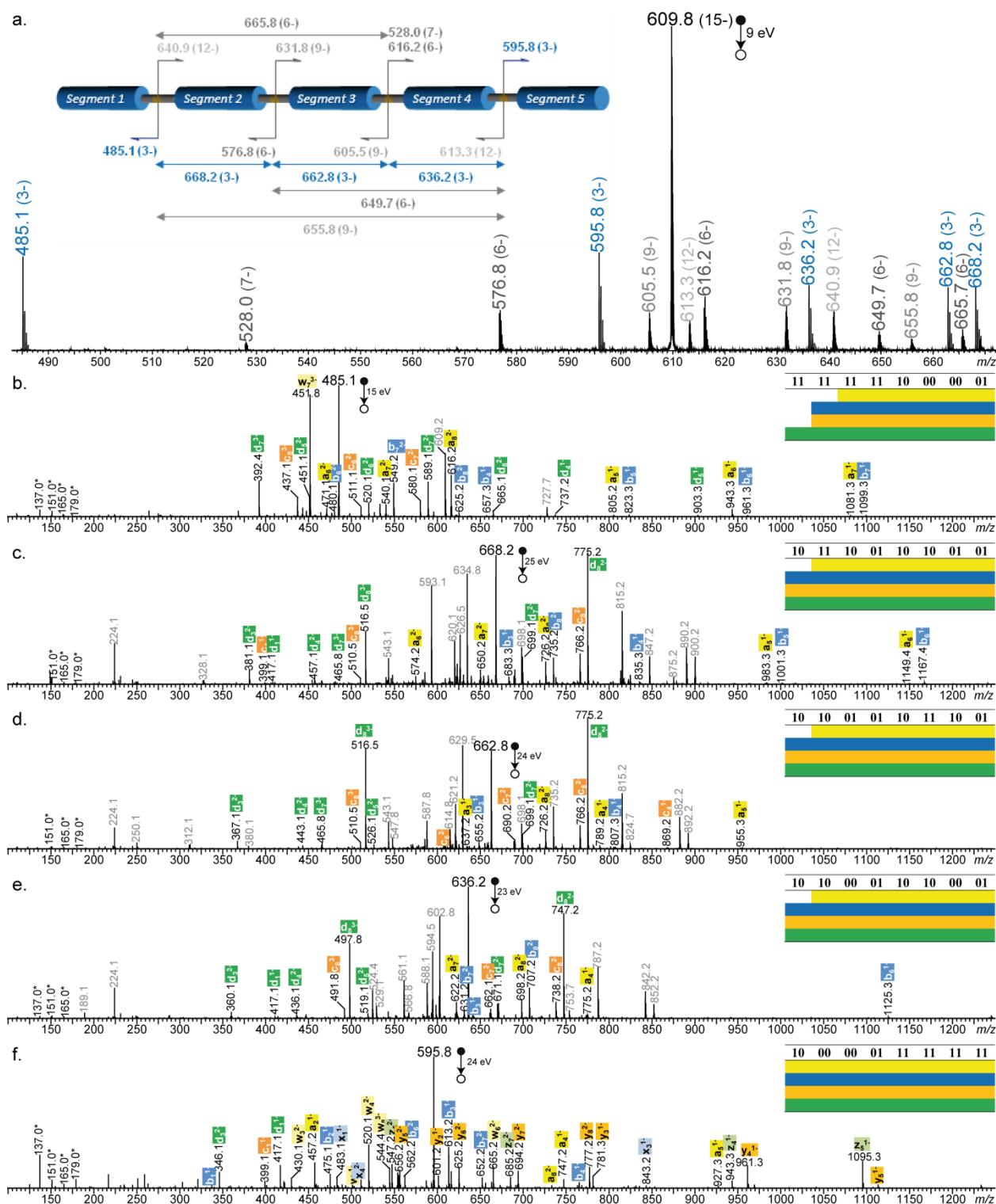


Figure S10. Sequencing of **P2**. (a) ESI-MS² spectrum of the [M-15H]¹⁵⁻ precursor ion at m/z 609.8, where m/z values measured for triply charged byte-fragments (in blue) reveal their initial location in the polymeric chain (see Table S2). Other fragment assignment is indicated in the dissociation scheme shown in inset. Pseudo-MS³ spectra for (b) the first segment at m/z 485.1, (c) the second segment at m/z 668.2, (d) the third segment at m/z 662.8, (e) the fourth segment at m/z 636.2, and (f) the fifth segment at m/z 595.8. Fragments d_i^{2-} (green), c_i^{1-} (orange), b_i^{1-} (blue) and a_i^{1-} (yellow) used for sequencing are not all annotated for the sake of clarity but their detection is summarized in the associated sequence coverage table. Deprotonated monomers are designated by an asterisk and peaks annotated in grey are secondary products. All collision energies are given in the laboratory frame.

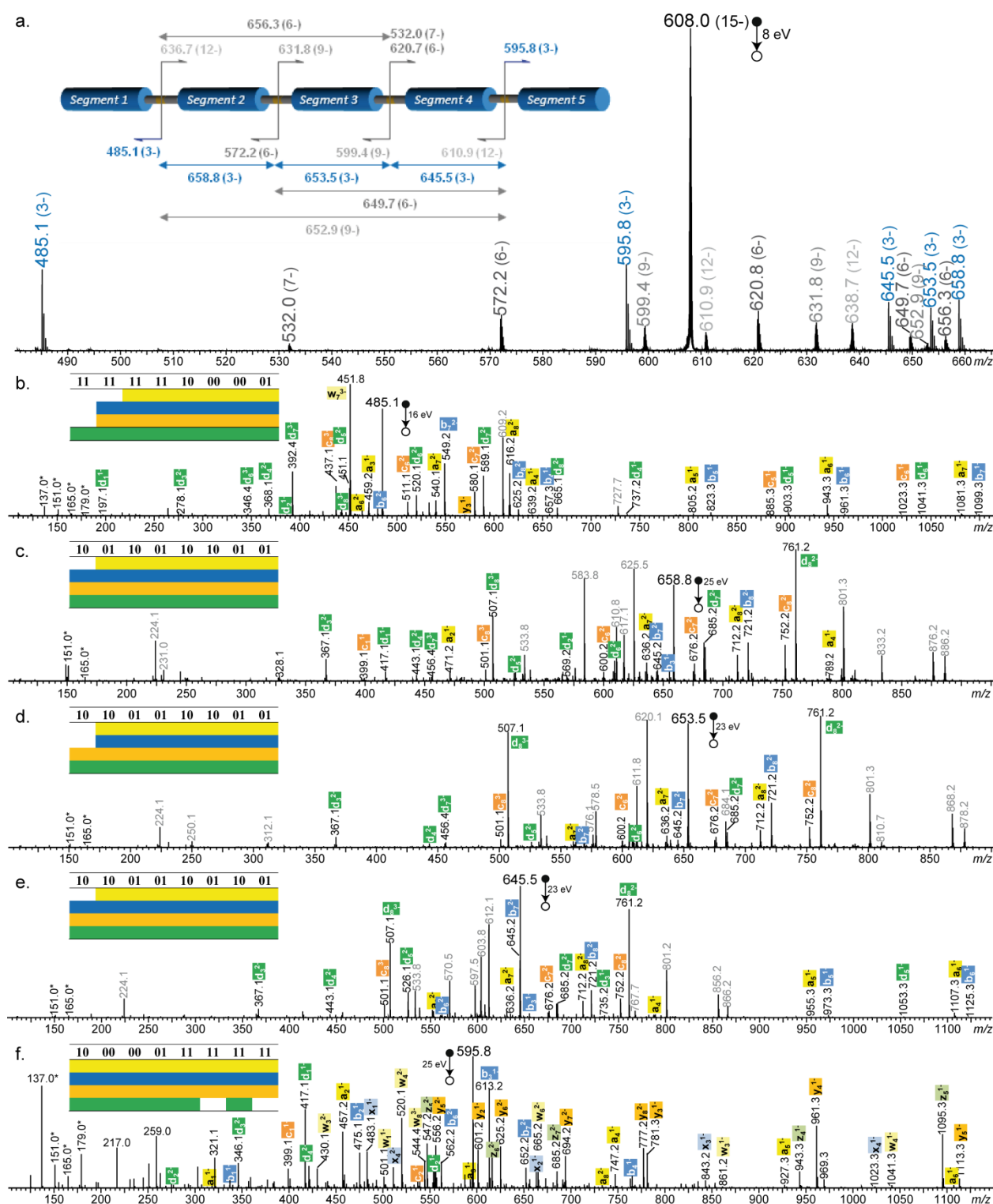


Figure S11. Sequencing of **P3**. (a) ESI-MS² spectrum of the [M-15H]¹⁵⁻ precursor ion at m/z 608.0, where m/z values measured for triply charged byte-fragments (in blue) reveal their initial location in the polymeric chain (see Table S2). Other fragment assignment is indicated in the dissociation scheme shown in inset. Pseudo-MS³ spectra for (b) the first segment at m/z 485.1, (c) the second segment at m/z 658.8, (d) the third segment at m/z 653.5, (e) the fourth segment at m/z 645.5, and (f) the fifth segment at m/z 595.8. Fragments d_i²⁻ (green), c_i¹⁻ (orange), b_i¹⁻ (blue) and a_i¹⁻ (yellow) used for sequencing are not all annotated for the sake of clarity but their detection is summarized in the associated sequence coverage table. Deprotonated monomers are designated by an asterisk and peaks annotated in grey are secondary products. All collision energies are given in the laboratory frame.

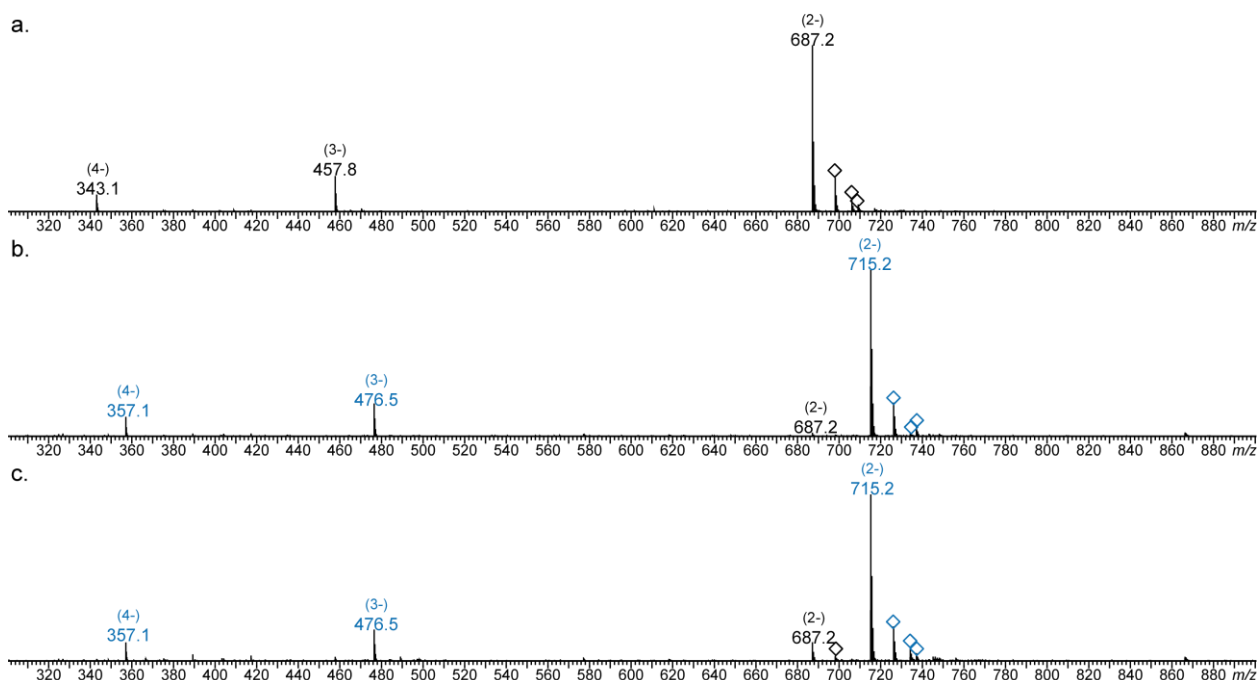


Figure S12. Negative mode MS spectra obtained for the first synthesized segment of **P4**, $M_4 \cdot M_2 \cdot M_2 \cdot M_1 \cdot M_5 \cdot M_4 \cdot M_5 \cdot T$ (1432.4 Da), prepared with (a) protocol a, (b) protocol b or (c) protocol c. The intact oligomer (in blue) and defective species (in black) are distributed over different charge states, as indicated into parentheses. Diamonds designate ionic forms detected after H/Na and/or H/K exchanges while signals annotated with charge states in grey circles correspond to in-source fragments.

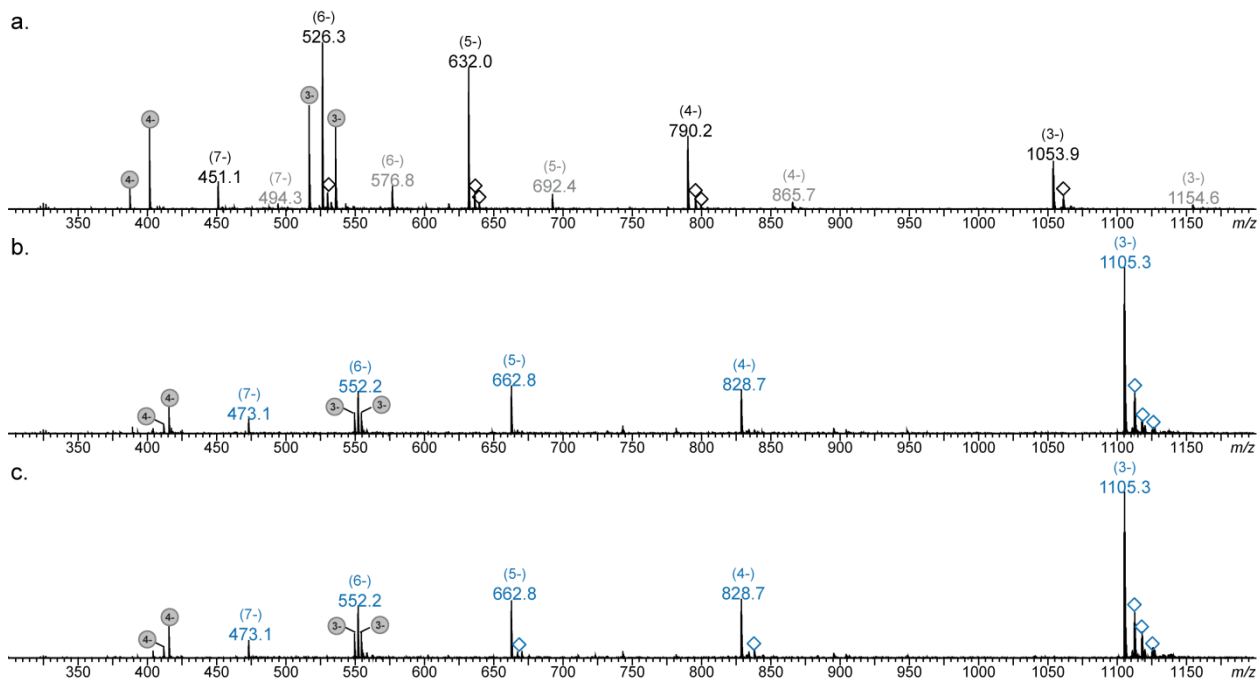


Figure S13. Negative mode MS spectra obtained for the first two synthesized segments of **P4**, $M_3 \cdot M_2 \cdot M_2 \cdot M_5 \cdot M_5 \cdot M_3 \cdot C \cdot L \cdot M_4 \cdot M_2 \cdot M_2 \cdot M_1 \cdot M_5 \cdot M_4 \cdot M_5 \cdot T$ (3318.0 Da), prepared with (a) protocol a, (b) protocol b or (c) protocol c. The intact oligomer (in blue) and defective species (in black and grey) are distributed over different charge states, as indicated into parentheses. Diamonds designate ionic forms detected after H/Na and/or H/K exchanges while signals annotated with charge states in grey circles correspond to in-source fragments.

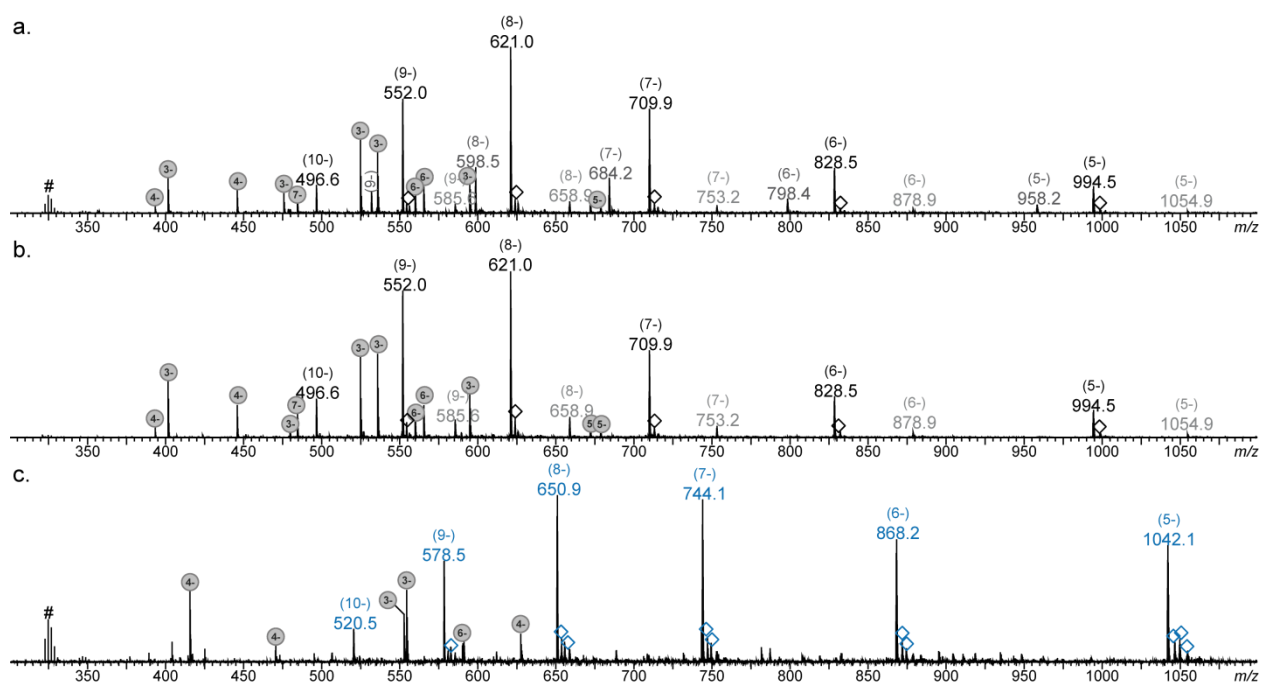


Figure S14. Negative mode MS spectra obtained for the first three synthesized segments of **P4**, $M_4 \cdot M_2 \cdot M_2 \cdot M_1 \cdot M_5 \cdot M_7 \cdot M_3 - A - L - M_3 \cdot M_2 \cdot M_2 \cdot M_5 \cdot M_5 \cdot M_5 \cdot M_3 - C - L - M_4 \cdot M_2 \cdot M_2 \cdot M_1 \cdot M_5 \cdot M_4 \cdot M_5 - T$ (5213.5485 Da), prepared with (a) protocol a, (b) protocol b or (c) protocol c. The intact oligomer (in blue) and defective species (in black and grey) are distributed over different charge states, as indicated into parentheses. Diamonds designate ionic forms detected after H/Na and/or H/K exchanges while signals annotated with charge states in grey circles correspond to in-source fragments. #: clusters of trichloroacetic acid.

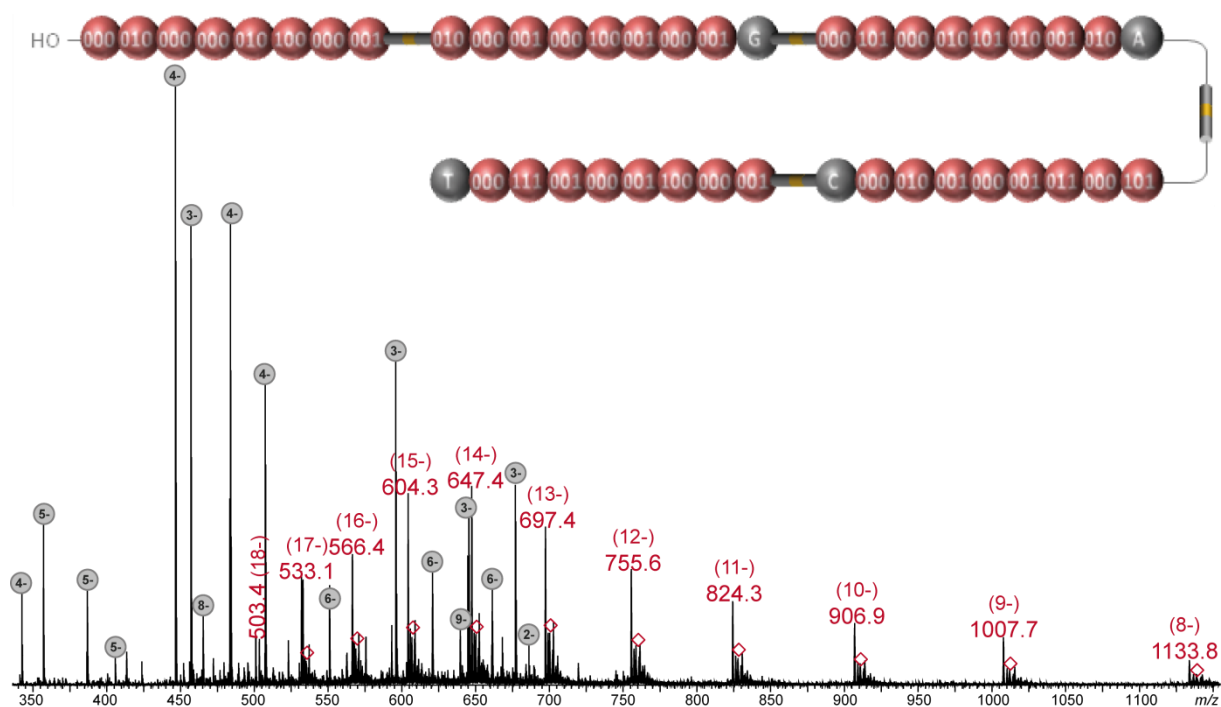


Figure S15. Negative mode ESI-MS spectrum of **P5**. Signals of the intact polymer (in red) are distributed over different charge states, as indicated into parentheses. Diamonds designate ionic forms detected after H/Na and/or H/K exchanges while signals annotated with charge states in grey circles correspond to in-source fragments.

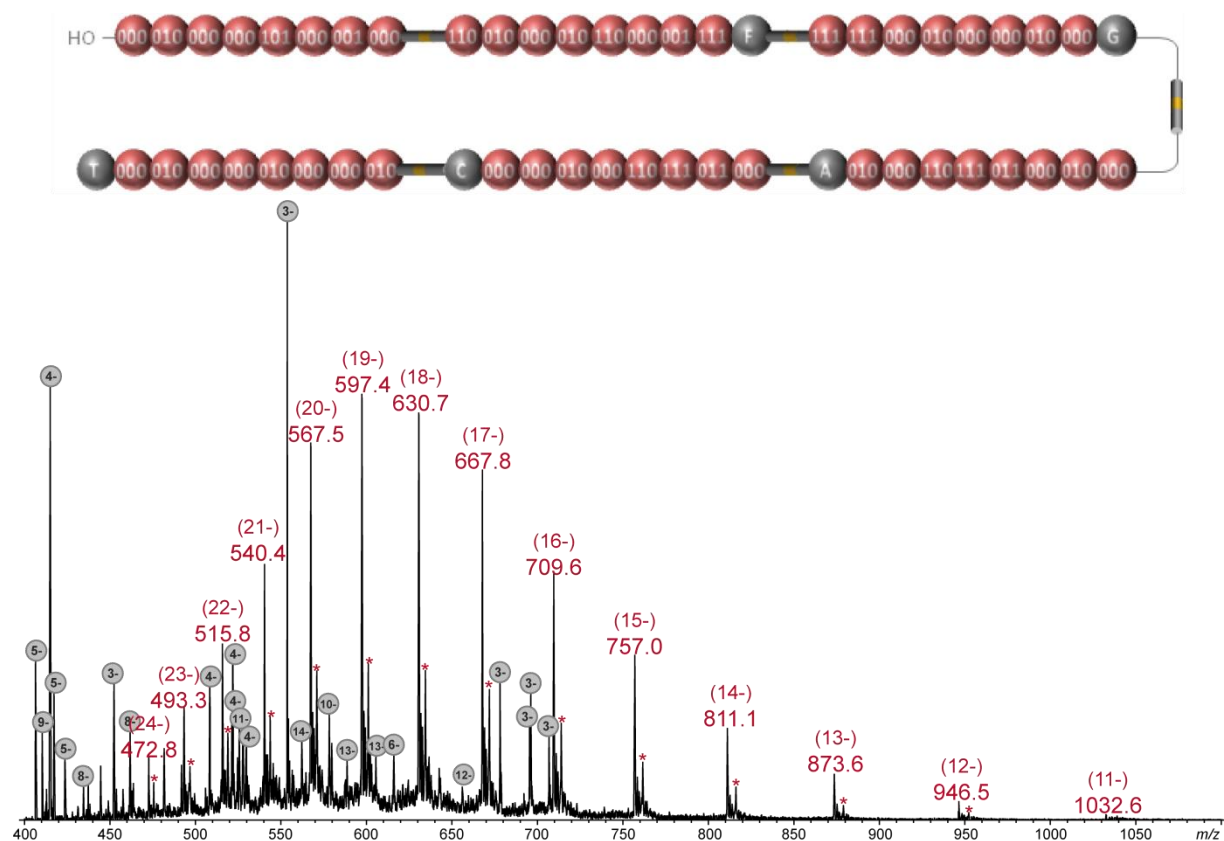


Figure S16. Negative mode ESI-MS spectrum of **P6**. Signals of the intact polymer (in red) are distributed over different charge states, as indicated into parentheses. Peaks annotated with an asterisk correspond to the partially deprotected polymer. Signals annotated with charge states in grey circles correspond to in-source fragments.

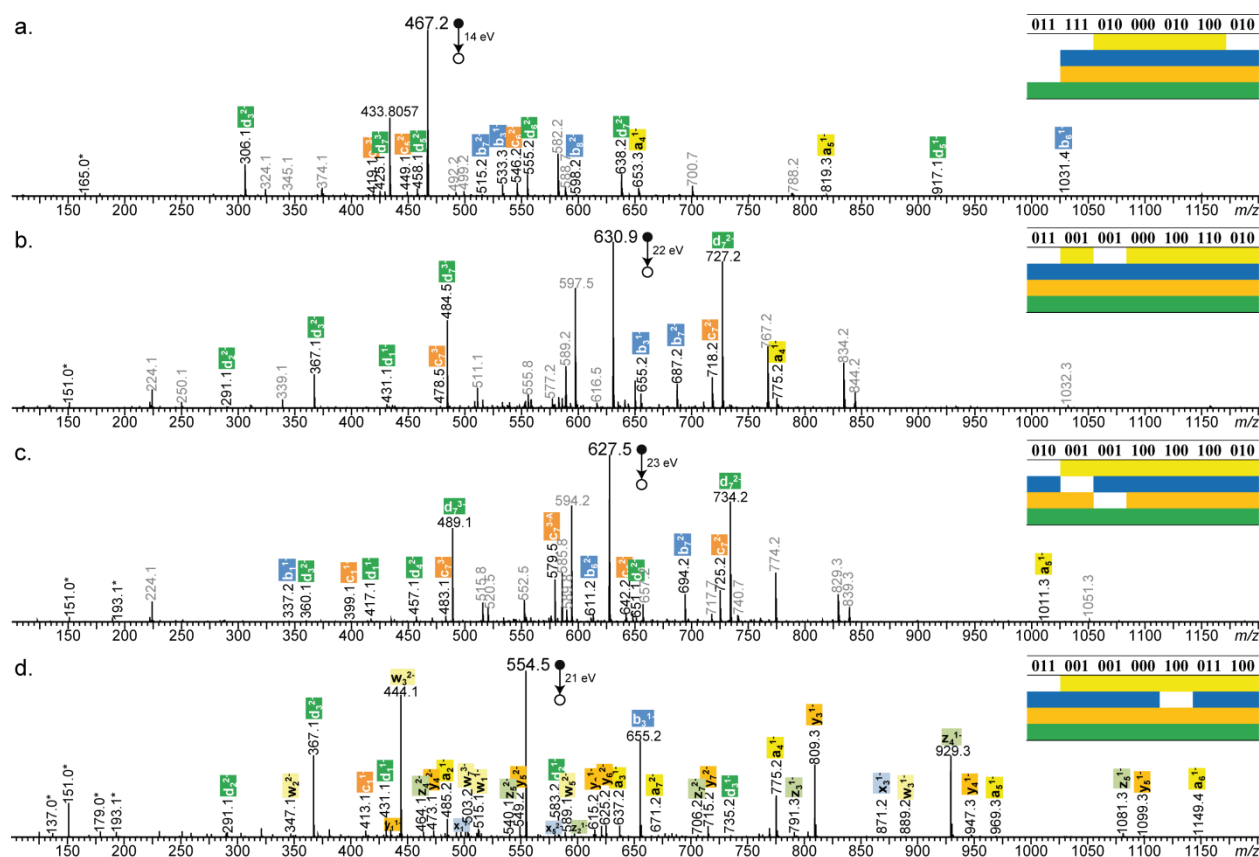


Figure S17. Sequencing of **P4**. Pseudo- MS^3 spectra for (a) the first segment at m/z 467.2, (b) the second segment at m/z 630.9, (c) the third segment at m/z 627.5, and (d) the fourth segment at m/z 554.5. Fragments d_1^{z-} (green), c_1^{z-} (orange), b_1^{z-} (blue) and a_1^{z-} (yellow) used for sequencing are not all annotated for the sake of clarity but their detection is summarized in the associated sequence coverage table. Deprotonated monomers are designated by an asterisk and peaks annotated in grey are secondary products. All collision energies are given in the laboratory frame.

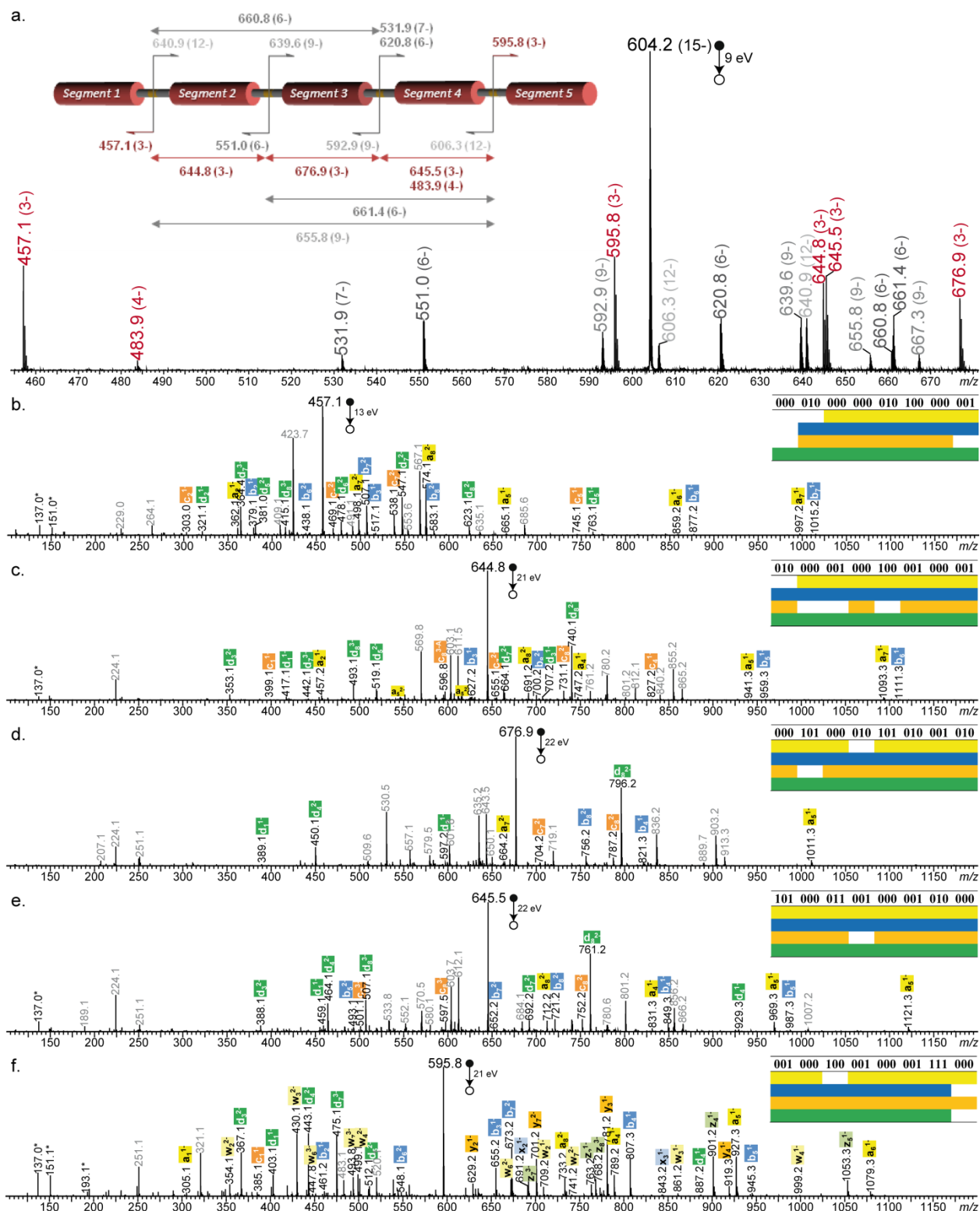


Figure S18. Sequencing of P5. (a) ESI-MS² spectrum of the [M-15H]¹⁵⁻ precursor ion at m/z 604.2, where m/z values measured for triply charged byte-fragments (in red) reveal their initial location in the polymeric chain (see Table S2). Other fragment assignment is indicated in the dissociation scheme shown in inset. Pseudo-MS³ spectra for (b) the first segment at m/z 457.1, (c) the second segment at m/z 644.8, (d) the third segment at m/z 676.9, (e) the fourth segment at m/z 645.5, and (f) the fifth segment at m/z 595.8. Fragments d_i^{2-} (green), c_i^{2-} (orange), b_i^{2-} (blue) and a_i^{2-} (yellow) used for sequencing are not all annotated for the sake of clarity but their detection is summarized in the associated sequence coverage table. Deprotonated monomers are designated by an asterisk and peaks annotated in grey are secondary products. All collision energies are given in the laboratory frame.

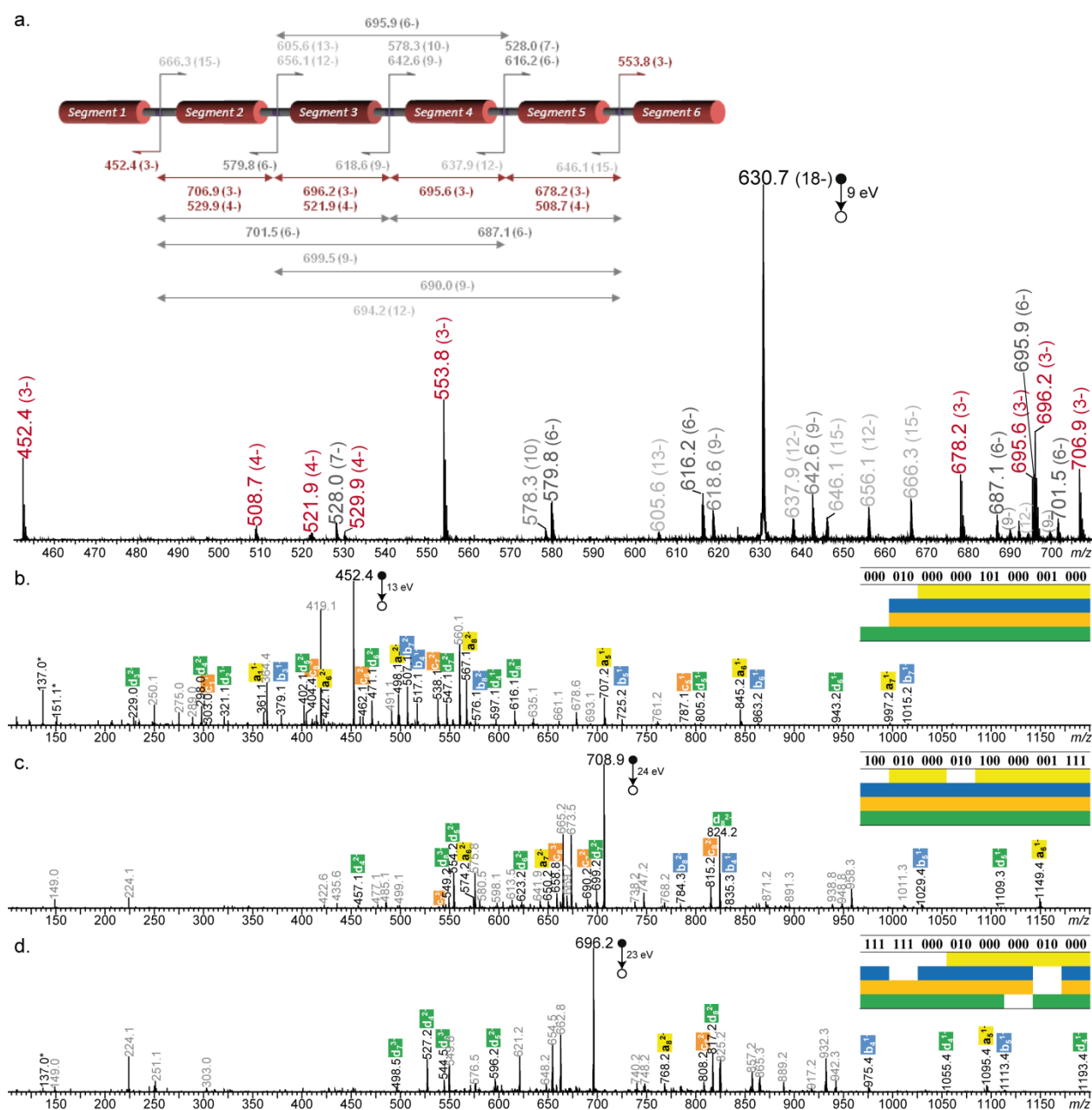


Figure S19. Sequencing of **P6**. (a) ESI-MS² spectrum of the [M-18H]¹⁸⁻ precursor ion at m/z 630.7, where m/z values measured for triply charged byte-fragments (in red) reveal their initial location in the polymeric chain. Other fragment assignment is indicated in the dissociation scheme shown in inset. Pseudo-MS³ spectra for (b) the first segment at m/z 452.4, (c) the second segment at m/z 706.9, and (d) the third segment at m/z 696.2. Fragments d_i^{z-} (green), c_i^{z-} (orange), b_i^{z-} (blue) and a_i^{z-} (yellow) used for sequencing are not all annotated for the sake of clarity but their detection is summarized in the associated sequence coverage table. Deprotonated monomers are designated by an asterisk and peaks annotated in grey are secondary products. All collision energies are given in the laboratory frame.

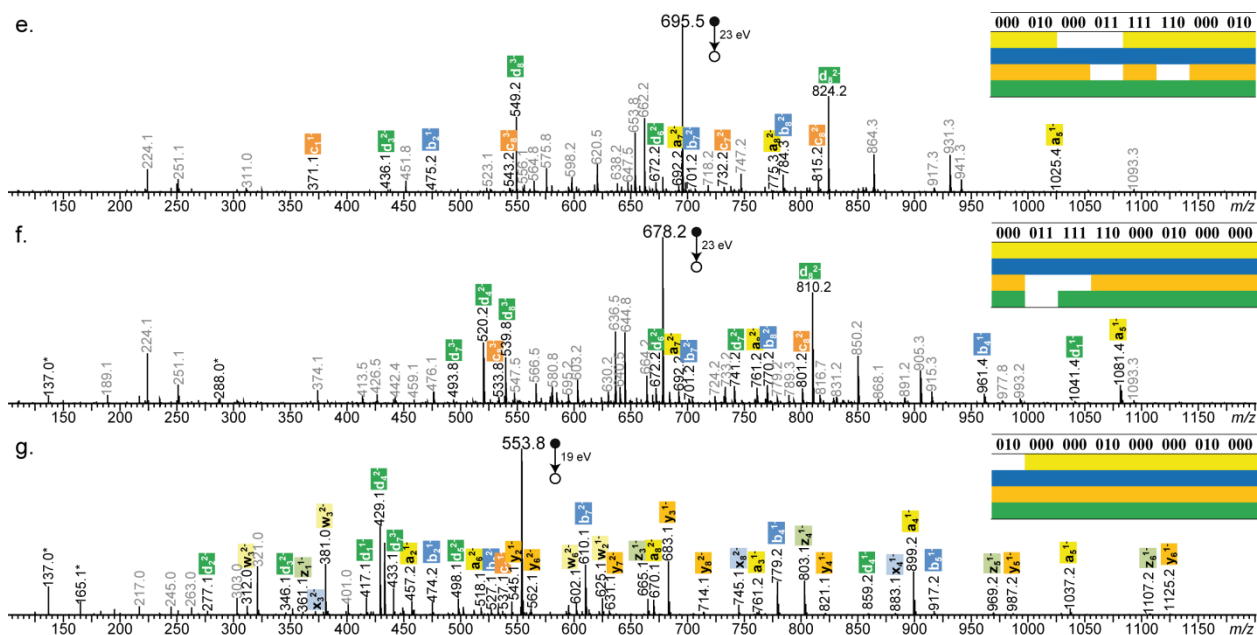


Figure S19. Sequencing of P6 (Cont'd). Pseudo-MS³ spectra for (e) the fourth segment at m/z 695.6, (f) the fifth segment at m/z 678.2, and (g) the sixth segment at m/z 553.8. Fragments d_1^{z-} (green), c_1^{z-} (orange), b_1^{z-} (blue) and a_1^{z-} (yellow) used for sequencing are not all annotated for the sake of clarity but their detection is summarized in the associated sequence coverage table. Deprotonated monomers are designated by an asterisk and peaks annotated in grey are secondary products. All collision energies are given in the laboratory frame.

E. References

- (1) 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.
- (2) Al Ouahabi, A.; Amalian, J.-A.; Charles, L.; Lutz, J.-F. Mass spectrometry sequencing of long digital polymers facilitated by programmed inter-byte fragmentation; *Nat Commun* **2017**, *8*, 967.
- (3) 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.