



HAL
open science

Precise Alkoxyamine Design to Enable Automated Tandem Mass Spectrometry Sequencing of Digital Poly(phosphodiester)s

Kévin Launay, Jean-arthur Amalian, Eline Laurent, Laurence Oswald, Abdelaziz Al Ouahabi, Alexandre Burel, Florent Dufour, Christine Carapito, Jean-louis Clément, Jean-François Lutz, et al.

► To cite this version:

Kévin Launay, Jean-arthur Amalian, Eline Laurent, Laurence Oswald, Abdelaziz Al Ouahabi, et al.. Precise Alkoxyamine Design to Enable Automated Tandem Mass Spectrometry Sequencing of Digital Poly(phosphodiester)s. *Angewandte Chemie International Edition*, 2021, 60 (2), pp.917-926. 10.1002/anie.202010171 . hal-03090783

HAL Id: hal-03090783

<https://hal.science/hal-03090783>

Submitted on 30 Nov 2021

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.

Precise alkoxyamine-design enables automated tandem mass spectrometry sequencing of digital poly(phosphodiester)s

Kévin Launay,^[a] Jean-Arthur Amalian,^[a] Eline Laurent,^[b] Laurence Oswald,^[b] Abdelaziz Al Ouahabi,^[b] Alexandre Burel,^[c] Florent Dufour,^[b,c] Christine Carapito,^[c] Jean-Louis Clément,^[a] Jean-François Lutz,^{[b]*} Laurence Charles,^{[a]*} Didier Gigmes^{[a]*}

^[a] Aix Marseille Université, CNRS, Institute for Radical Chemistry, UMR 7273, 23 Av Escadrille Normandie-Niemen, 13397 Marseille Cedex 20, France

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

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

Abstract

Increasing data storage density is one of the major challenges of the information industry. To be widely applicable, writing and reading information have to be dense, convenient and fast. A major step towards reliable reading of information coded in the sequence of long poly(phosphodiester)s was previously achieved by introducing an alkoxyamine spacer between information sub-segments. However, MS/MS decoding had to be performed manually to safely identify useful fragments of low abundance compared to side-products induced by the amide-based alkoxyamine used. Here, alternative alkoxyamines were designed to prevent these side-reactions and enable automated MS/MS sequencing. Different styryl-TEMPO spacers were prepared to increase radical delocalization and stiffness of the structure. Their dissociation behavior was investigated by EPR and best MS/MS data were recorded for polymers with alkoxyamines containing in-chain benzyl ring, avoiding all side-reactions during synthesis or sequencing. As a result, automated decoding of these polymers can be achieved using the MS-DECODER software, which performs interpretation of fragmentation data recorded for each sub-segment prior to align the so-obtained digital sequences in their original order based on the mass tag defining the initial location of each sub-segment.

Key words: alkoxyamine synthesis; digital information; mass spectrometry; radical ions; sequence-defined synthetic polymers; structural design

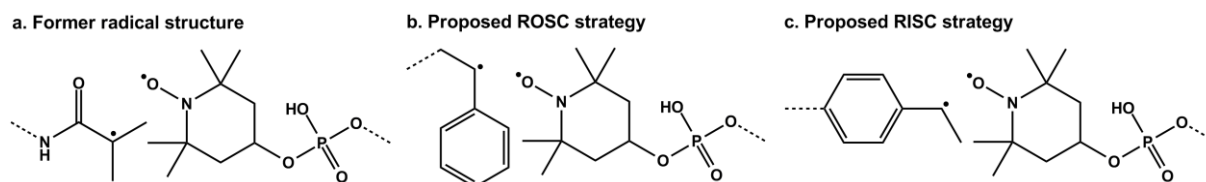
Introduction

As inspired by the information-storage capacity of DNA,^[1] digital polymers are functional macromolecules that store binary information at the molecular level in form of a controlled comonomer sequence.^[2-9] Indeed, a string of information can be implemented in a polymer using two or more comonomers intentionally defined as single bits or groups of bits, allowing encryption of any desired binary sequence in the chains. These polymers are typically prepared by automated solid-phase synthesis, which consists of attaching monomers one-by-one on a solid-support and hence permits to insert coding motifs at precise positions in chains containing up to 100-150 repeating units. Longer chains are not attainable by stepwise synthesis and an efficient strategy to circumvent this limitation is to produce large libraries of medium-sized macromolecules, as demonstrated to store about a megabyte of data in DNA chains.^[10-12] Writing information in the structure of non-natural polymers is also very appealing as they offer many advantages over DNA. Because it can proceed by simpler protecting-group-free methods,^[13] solid-phase synthesis of abiotic polymers is often easier than for DNA. Moreover, the macromolecular structure of non-natural species can be optimized to control their fragmentation for best sequencing by tandem mass spectrometry (MS/MS). Such structural designs always aim at reducing the number of dissociation routes, in order to avoid signal dilution when dealing with long chains, while still enabling full sequence coverage. For this purpose, different strategies have been reported: removing reactive functions,^[14] controlling the oxidation degree of specific atoms,^[15] or introducing linkages of low dissociation energy to make any other bonds MS/MS silent.^[16] The latter approach was successfully employed to decipher large amounts of information stored in sequence-defined poly(phosphodiester)s using a routine mass spectrometer.^[17] Indeed, the structure of long chains produced via automated phosphoramidite chemistry^[18-21] was re-engineered to enable their *de novo* sequencing, in spite of the numerous fragmentation routes experienced by their repeating units when subjected to collision-induced dissociation (CID). To do so, reagents including alkoxyamine linkages were used to introduce weak bonds between short coded sub-segments (for example, between two consecutive bytes). As a result, the reading step is simply based on MS/MS sequencing of sub-segments released upon homolysis of all C–ON bonds in these alkoxyamine-modified poly(phosphodiester)s. Indeed, using a two-step protocol (Supporting Information, Figure S1), intact synthetic chains ionized as polyanions are first subjected to in-source collisional activation that cleaves each inter-byte alkoxyamine linkage, and so-released sub-segments are then individually submitted to a second activation stage for MS/MS sequencing based on polyphosphate dissociation rules.^[22]

Due to the lack of precursor ion selection before the first activation stage, this sequencing protocol based on two CID steps is designated as a pseudo-MS³ experiment. The use of specific tags to label each sub-segment for its initial location in the chains permits to readily reconstruct the information originally encoded in the backbone of these digital poly(phosphodiester)s. Very recently, this simple strategy was used in conjunction with an expanded monomer alphabet (e.g. using 4 or 8 coded monomers instead of two) to afford high-capacity digital poly(phosphodiester)s.^[23] Indeed, 144 bits of data were stored on and extracted from a polymer single chain, which is to our knowledge the longest non-natural decodable sequence ever reported.

Although this analytical strategy allows full sequence coverage of all coded sub-segments, and hence reliable recovery of messages stored in poly(phosphodiester)s, automation of the reading step with the MS-DECODER software^[24] is currently impossible. Indeed, automated identification of fragments in pseudo-MS³ spectra requires ions to be produced with sufficiently high abundance, which is hardly achieved with the cleavable spacer used so far. This spacer is composed of an amide-containing segment, (CH₂)₃NH(CO)C(CH₃)₂, connected to a 2,2,6,6-tetramethylpiperidine-1-oxyl (TEMPO) nitroxide: upon C–ON bond homolysis in MS² experiments, the former moiety becomes the ω termination of all sub-segments but the last one, whereas TEMPO becomes the α termination of all sub-segments but the first one (Supporting Information, Scheme S1). While the TEMPO nitroxide is highly stable, the carbon-centered radical constituting the ω-termination of sub-segments fragments is too reactive during the second activation stage (pseudo MS³ experiments). The observed radical-induced dissociation routes efficiently competes with phosphate bond cleavages, which gives rise to an overall decrease of useful fragment abundance and hence jeopardizes their automated identification.

In this study, phosphoramidite reagents including new TEMPO-based alkoxyamines were developed and employed to prepare automatically-decodable poly(phosphodiester)s. To prevent the above-mentioned side reactions which occurred with the amide moiety (Scheme 1a), herein, we developed original alkoxyamine spacers bearing a highly delocalized rigid linker, namely the 1-ethyl-phenyl moiety. The effect of the linker position, *i.e.* as pendant group (Ring OutSide Chain, namely ROSC, Scheme 1b) or in the main chain (Ring InSide Chain, namely RISC, Scheme 1c), on the alkoxyamine dissociation selectivity was studied by electron paramagnetic resonance (EPR) of the monomeric spacer.



Scheme 1. Comparison between the radicals formed upon C–ON bond homolysis with (a) the former and new alkoxyamine spacers following the (b) ROSC or (c) RISC strategy.

The new structural designs were then scrutinized for their efficiency at limiting undesired radical-induced reactions during pseudo-MS³ of the first sub-segment using poly(phosphodiester)s composed of two bytes, prior to studying longer polymers.

Results and discussion

Reactivity of the terminal carbon-centered radical

In order to rationalize the development of new alkoxyamines, dissociative reactions experienced by sub-segments released upon C–ON bond homolysis of TEMPO-based digital poly(phosphodiester)s were studied in details. Prior to using the expanded phosphoramidite alphabets which have been recently described,^[23] the model digital polymers employed to evaluate alkoxyamine spacers were encrypted using two monomers only. Rules for coding and decoding this binary monomer alphabet are described in details in previous publications.^[17, 20, 21] As previously mentioned, these polymers are prepared with an alkylamide-TEMPO spacer (Scheme 1a) which leaves the $(\text{CH}_2)_3\text{NH}(\text{CO})\text{C}^*(\text{CH}_3)_2$ moiety as the ω termination of the first sub-segment during the first activation stage. This is exemplified with the two-byte containing polymer P0 (Supporting Information, Table S1), which CID leads to two triply charged sub-segments (Supporting Information, Figure S2) when activated as its preferential 6– form (*i.e.*, 3 deprotonated phosphate per sub-segment). Activation of the m/z 433.7 sub-segment containing the first byte conducts to two main types of dissociation pathways (Figure 1). The first one comprises cleavages of backbone phosphate bonds that produce ions useful for sequencing. These are α -containing (*i.e.*, a_i^{z-} , b_i^{z-} , c_i^{z-} , d_i^{z-}) and ω -containing (*i.e.*, w_i^{z-} , x_i^{z-} , y_i^{z-} , z_i^{z-}) product ions typically employed to reconstruct the sequence (annotated with colored square labels in Figure 1).^[22] However, sequence recovery is achieved with α -containing fragments only because most product ions containing the original ω termination are missing, as illustrated by many empty cells in the w/x/y/z rows of the coverage table (inset of Figure 1).

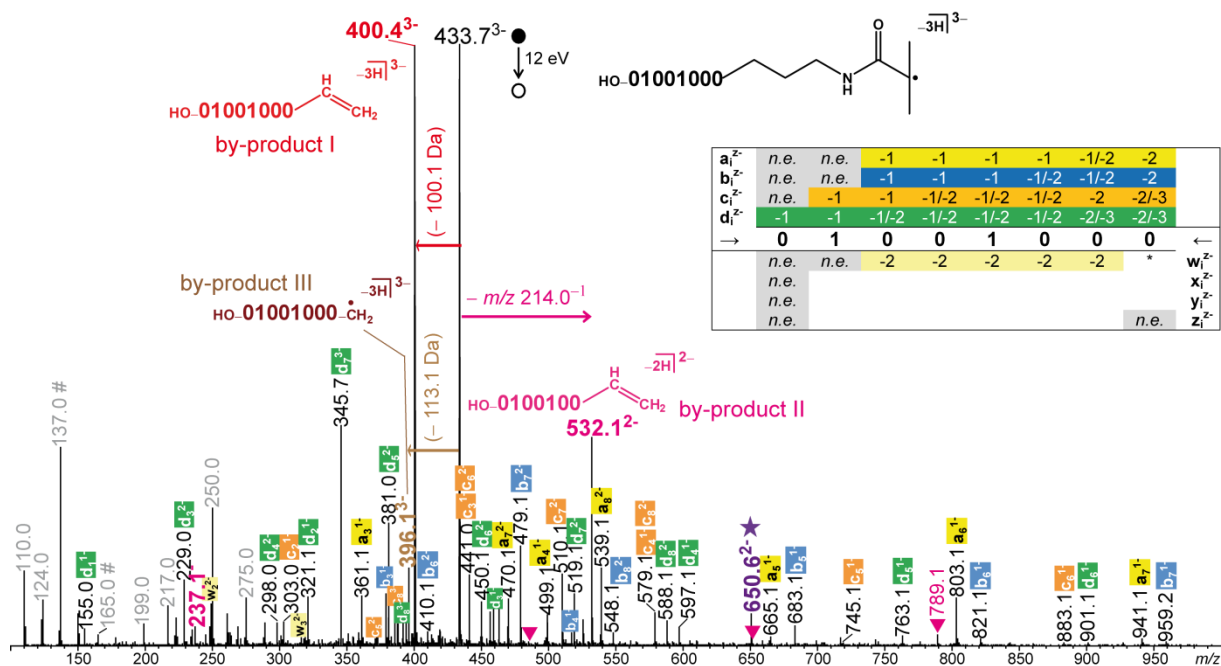
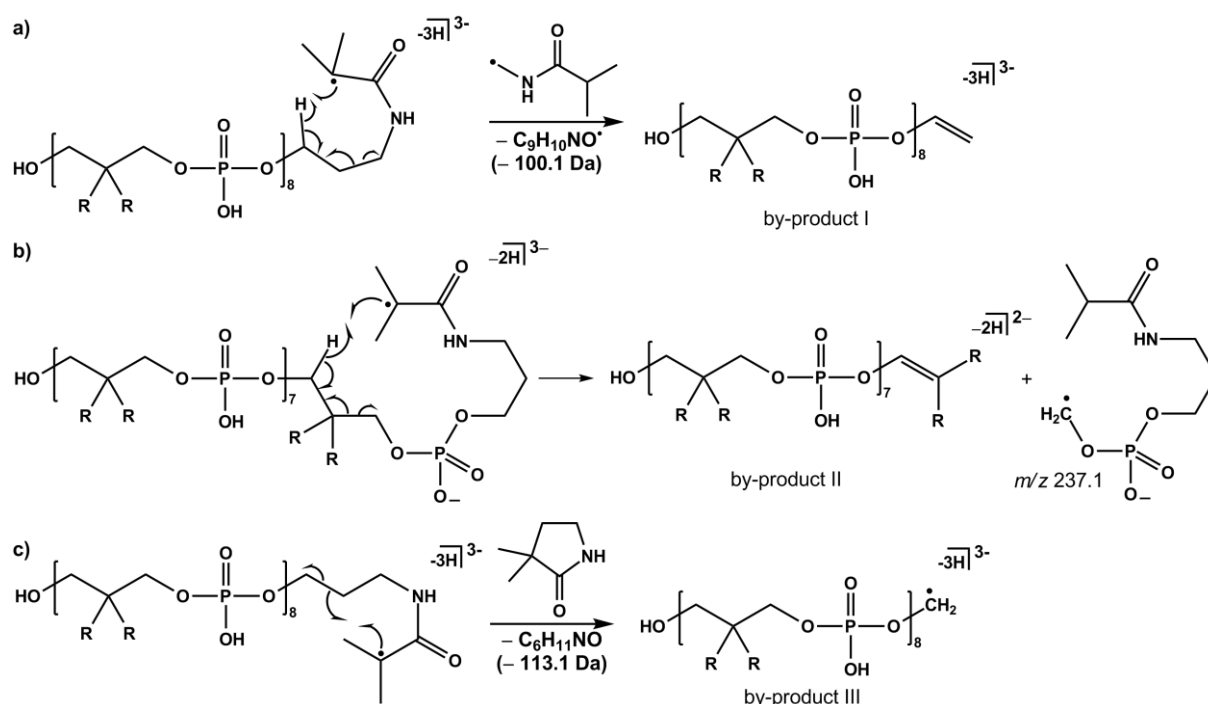


Figure 1. CID spectrum of the m/z 433.7 sub-segment containing the first byte of P0. Besides phosphate bond cleavages leading to fragments annotated with colored square labels, the triply deprotonated precursor (black structure) experiences three main radical-induced side reactions (depicted in Scheme 2) to generate by-products I (m/z 400.4, in red), II (m/z 532.1, in pink), and III (m/z 396.1, in brown). Further dissociation of these byproducts leads to signal annotated with pink triangles. The purple star at m/z 650.6 designates an oxidation product of the precursor ion. Inset: coverage table of the **01001000** byte with each cell reporting the charge states of fragments produced after phosphate bond cleavages (Supporting Information, Table S3). Due to $\alpha = \text{HO}$, some fragments are not expected (*n.e.* in grey cells). Secondary product ions are in grey, including $[\mathbf{0} - \text{H}]^-$ at m/z 137.0 and $[\mathbf{1} - \text{H}]^-$ at m/z 165.0 (annotated by #).

These missing fragments are due to the occurrence of the second type of reactions induced by the reactive carbon-centered radical termination of the dissociating precursor ion. As depicted in Scheme 2a, H^\bullet abstraction by this radical carbon would occur at the first methylene moiety of the ω end-group to generate the triply charged by-product I at m/z 400.4 after elimination of a 100.1 Da radical (in red, Figure 1). Alternatively, H^\bullet could be transferred from the first methylene moiety of the last coding unit (Scheme 2b) to yield the doubly charged by-product II at m/z 532.1 together with the m/z 237.1 anion (both in pink in Figure 1). Finally, the radical-induced elimination of dimethylcyclobutanamide (113.1 Da) depicted in Scheme 2c would generate the triply charged by-product III at m/z 396.1 (in brown, Figure 1). Of note,

collisional activation is also observed to promote oxidation of the triply charged m/z 433.7 precursor, as revealed by the doubly charged fragment at m/z 650.6 (designated by a purple star in Figure 1). Such charge stripping reactions were first reported in high energy (keV) collision regime^[25] but were also observed at lower energy as in the case of oligonucleotides.^[26] It should also be noted that the charge state of the dissociating sub-segment has no influence on the type of observed fragmentation reactions. This last result suggests that long range stabilizing effects reported for some distonic anions^[27, 28] are not significant in the studied species.



Scheme 2. Mechanisms proposed to account for radical-induced pathways experienced by sub-segments containing the first byte of poly(phosphodiester)s built with the alkylamide-TEMPO spacer.

Overall, mechanisms proposed in Scheme 2 for the three main radical-induced side-reactions all rely on a high flexibility of the linear $(\text{CH}_2)_3\text{NH}(\text{CO})\text{C}^\bullet(\text{CH}_3)_2$ terminal moiety. Accordingly, these reactions should no longer proceed with an alternative termination of increased steric hindrance or increased stiffness. In addition, any moiety improving radical stabilization via delocalization of the unpaired electron would help reducing its reactivity. To do so, the ROSC and RISC strategies depicted in Scheme 1 were explored.

Alkoxyamine spacer with increased electron delocalization (ROSC strategy)

Increasing the radical delocalization is correlated to an increase of the C–ON homolytic dissociation rate constant k_d and therefore a diminution of its activation energy, $E_{a,d}$.^[29] This may lead to a poorly stable polymer and the activation energy of the possible ROSC structures was therefore estimated from the k_d calculated with the multiparameter equation (1) established by Bertin *et al.* in 2005 for TEMPO-based alkoxyamines:^[30]

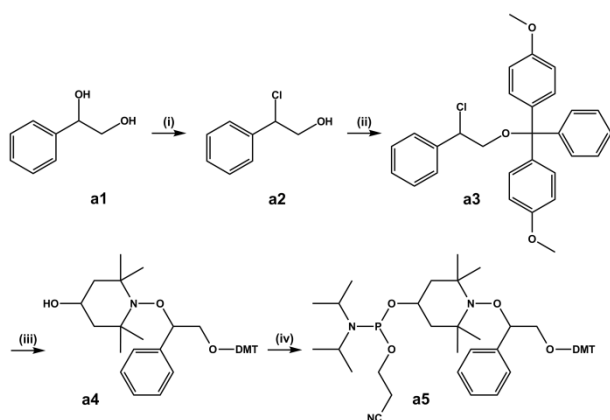
$$\log k_{d,393} = -14.8 (\pm 0.7) + 13.9 (\pm 0.9) \cdot \sigma_{RS} + 13.6 (\pm 3.2) \cdot \sigma_I + 6.6 (\pm 0.7) \cdot \nu \quad (1)$$

This equation is an empirical evaluation of k_d at 120°C from the radical stabilization constant σ_{RS} , the universal electrical Hammett constant σ_I , and the steric constant ν (detailed calculations are given in Supporting Information). Earlier MS/MS experiments have shown that, while CID of poly(phosphodiester)s built with a di-methylated spacer (Table 1, a) only leads to C–ON bond homolysis, phosphate bond cleavages become competitive for polymers constructed with a mono-methylated spacer (Table 1, b).^[17] The dissociation rate constant of the latter alkoxyamine was therefore estimated with equation (1) and considered as upper limit of k_d . From calculations summarized in Table 1, the 1-ethyl-phenyl radical structure without amide moiety (Table 1, f) appears as the structure of choice to combine a low activation energy (high k_d) and a high stability (low k_d) of the final polymer.

| alkyl radical | a) | b) | c) | d) | e) | f) |
|--------------------------------|-------|----------------------|-------|-------|-------------|---|
| σ_{RS} | 0.20 | 0.18 | 0.41 | 0.39 | 0.36 | 0.34 |
| σ_I | 0.07 | 0.09 | 0.10 | 0.12 | 0.07 0.06 | 0.09 0.07 |
| ν | 1.42 | 1.01 | 1.45 | 1.20 | 1.28 | 0.86 |
| $K_{d,393}$ (s ⁻¹) | 0.02 | 1.8.10 ⁻⁵ | 76.39 | 1.56 | 0.41 0.30 | 5.3.10 ⁻⁴ 3.7.10 ⁻⁴ |
| E_a (kJ.mol ⁻¹) | 120.7 | 143.9 | 94.0 | 106.7 | 111.1 112.2 | 132.8 134.0 |

Table 1. Calculated molecular descriptors and activation energy corresponding to the dissociation rate constant calculated with equation (1). An averaged frequency factor $A_d = 2.4.10^{14}$ s⁻¹ was used to estimate E_a from $k_{d,393}$. The σ_I values of structures e) and f) are directly connected to a phosphate in the final polymer (left values) or to a DMT-protecting group in the initial monomer tested in EPR (right values).

Among the various possibilities to prepare alkoxyamine compounds, one of the most robust and convenient consists of performing a coupling reaction between an alkyl radical and a nitroxide under atom transfer radical addition (ATRA) conditions.^[31] The first alkoxyamine candidate includes a pendant phenyl group (hence named Ring OutSide Chain, ROSC) and was prepared according to the 4-step chemical pathway described in Scheme 3. Typically, 2-chloro-2-phenylethan-1-ol (**a2**) obtained by chlorination of phenylethane-1,2-diol (**a1**) with HCl_{aq} was protected with dimethoxytrityl chloride (DMT-Cl) in 85% yield. The resulting halide **a3** was reacted with 4-hydroxy-TEMPO in ATRA conditions (CuBr , Cu^0 , pentamethyldiethylenetriamine (PMDETA) in degassed THF) to produce the alkoxyamine **a4** in 74% yield. Final functionalization with O-2-cyanoethyl-N,N-diisopropylchlorophosphoramidite led to the ROSC spacer **a5** in 75% yield, as confirmed by NMR (Supporting Information, Figure S3) and electrospray ionization high resolution mass spectrometry (ESI-HRMS) performed in both MS and MS/MS modes (Supporting Information, Figure S4). The C–ON bond homolytic dissociation rate constant (k_d) determined by EPR for ROSC was $5.4 \times 10^{-4} \text{ s}^{-1}$ at 128°C , which corresponds to an activation energy of $135.5 \text{ kJ mol}^{-1}$, consistent with literature results associated to 2,2,6,6-tetramethyl-1-(1-phenyl-ethoxy)piperidine.^[32]



Scheme 3. Synthesis of ROSC, **a5**. (i) HCl_{aq} 37%, CHCl_3 , r.t, 1h; (ii) DMT-Cl, anhydrous pyridine, anhydrous THF, r.t, overnight; (iii) 4-hydroxyTEMPO, CuBr , Cu^0 , PMDETA, THF, r.t., overnight; (iv) O-2-cyanoethyl-N,N-diisopropylchlorophosphoramidite, DIPEA, anhydrous DCM, r.t, 1h. See details in supporting information.

The ROSC phosphoramidite reagent was then employed to prepare the P1 polymer composed of two bytes, HO-00000000-ROSC-11111111-T (Supporting Information, Table S1). The negative mode ESI mass spectrum shows a monodisperse sample, with P1

distributed over charge states ranging from -3 to -9 (Supporting Information, Figure S5). During the first activation stage, homolysis of the ROSC spacer cleaves the $[P1 - 6H]^{6-}$ precursor ion in two complementary triply charged sub-segments (Supporting Information, Figure S6). Activation of the m/z 407.4 ion containing the first byte leads to the pseudo-MS³ spectrum shown in Figure 2. The major product ion in this spectrum is a doubly charged species at m/z 504.0 (in red), which formation would require a backbiting process (Supporting Information, Scheme S2) which is obviously not hindered by the bulky benzyl group. The alternative elimination of styrene from the terminal styryl radical yields the triply charged m/z 372.7 species (in pink), which quite low abundance can be explained by a consecutive fast radical-induced depolymerization process (Supporting Information, Scheme S3). Oxidation of the precursor ion is also observed, yielding the doubly charged fragment at m/z 611.1 (designated by a purple star in Figure 2). As a result of these side reactions, many w/x/y/z product ions containing the original ω termination of the m/z 407.4 precursor ion are missing (see coverage table in inset of Figure 2). Yet, full sequence coverage of byte 1 can be achieved with α -containing fragments, all detected but some with very minor abundance (Supporting Information, Table S4). This is in contrast with the sub-segment containing the second byte of P1, which CID yields all expected α - and ω -containing product ions (Supporting Information, Figure S7 and Table S5).

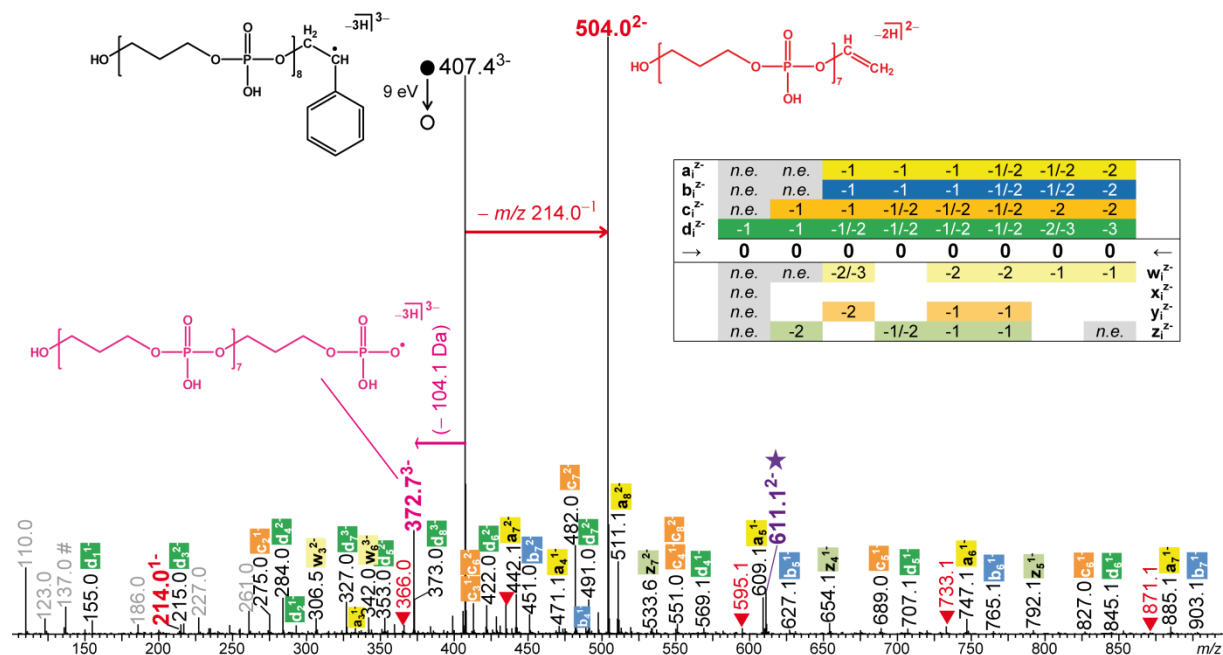
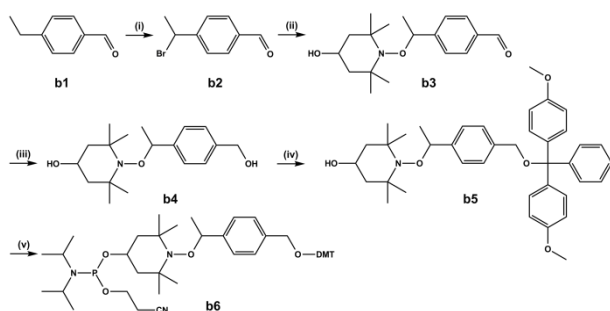


Figure 2. CID spectrum of the m/z 407.4 sub-segment containing the first byte of P1. The triply deprotonated precursor ion (black structure) experiences radical-induced side reactions

leading to i) the m/z 504.0 ion (which further dissociates into products designated with red triangles) and ii) the m/z 327.7 ion (in pink) after elimination of styrene. The purple star at m/z 611.1 designates an oxidation product of the precursor ion. Inset: coverage table of the **00000000** byte with each cell reporting the charge states of fragments produced after phosphate bond cleavages (Supporting Information, Table S4). Due to $\alpha = \text{HO}$, some fragments are not expected (*n.e.* in grey cells). Secondary product ions are in grey, including $[\mathbf{0} - \text{H}]^-$ at m/z 137.0 (#).

Alkoxyamine spacer with increased delocalization and stiffness (RISC strategy)

In order to prevent CID side-reactions still observed for the ROSC-based P1, a second alkoxyamine was designed with a phenyl ring inside the main chain to also increase the stiffness of the ω moiety of the first sub-segment. The so-called RISC (Ring InSide Chain) species was prepared in 5 steps (Scheme 4), starting with Wohl-Ziegler bromination of 4-benzaldehyde **b1** which afforded **b2** in 20% yield, followed by formation of the alkoxyamine **b3** in ATRA conditions (82% yield). The benzaldehyde was then reduced with NaBH_4 to afford diol **b4** in 71% yield, which primary alcohol was selectively protected with DMT-Cl to produce **b5** in 64% yield. Final functionalization with O-2-cyanoethyl-N,N-diisopropylchlorophosphoramidite afforded the targeted RISC **b6** in 42% yield, as confirmed by NMR (Supporting Information, Figure S8) and ESI-HRMS and MS/MS (Supporting Information, Figure S9). EPR study of RISC permitted to determine a dissociation rate constant of $8.7 \cdot 10^{-4} \text{ s}^{-1}$ at 127°C , corresponding to an activation energy of $133.6 \text{ kJ mol}^{-1}$ consistent with values estimated in Table 1 and literature data.^[32]



Scheme 4. Synthesis of RISC, **b6**. (i) BMS, anhydrous THF, rt, overnight; (ii) DMT-Cl, anhydrous pyridine, anhydrous THF, r.t, overnight; (iii) 4-hydroxyTEMPO, CuBr , Cu^0 , PMDETA, degassed THF, r.t., overnight; (iv) O-2-cyanoethyl-N,N-diisopropylchlorophosphoramidite, DIPEA, anhydrous DCM, r.t, 1 h. See details in supporting information.

The byte-truncated polymer P2 (Supporting Information, Table S1) prepared with the RISC phosphoramidite reagent was observed with charge states ranging from -4 to -10 in negative mode ESI-MS (Supporting Information, Figure S10). Activated homolysis of RISC yields the expected complementary sub-segments (Supporting Information, Figure S11) and further CID of the m/z 412.0 ion containing the first byte of P2 gives rise to the pseudo-MS³ data reported in Figure 3. All product ions observed in this spectrum result from cleavages of phosphate bonds, including those holding the original ω termination of the dissociating sub-segment. No side-product resulting from radical-induced reactions could be detected and, interestingly, charge stripping was not observed to occur for the m/z 412.0 precursor. These results clearly indicate a very low MS/MS reactivity of the RISC segment constituting the ω termination of the dissociating sub-segment.

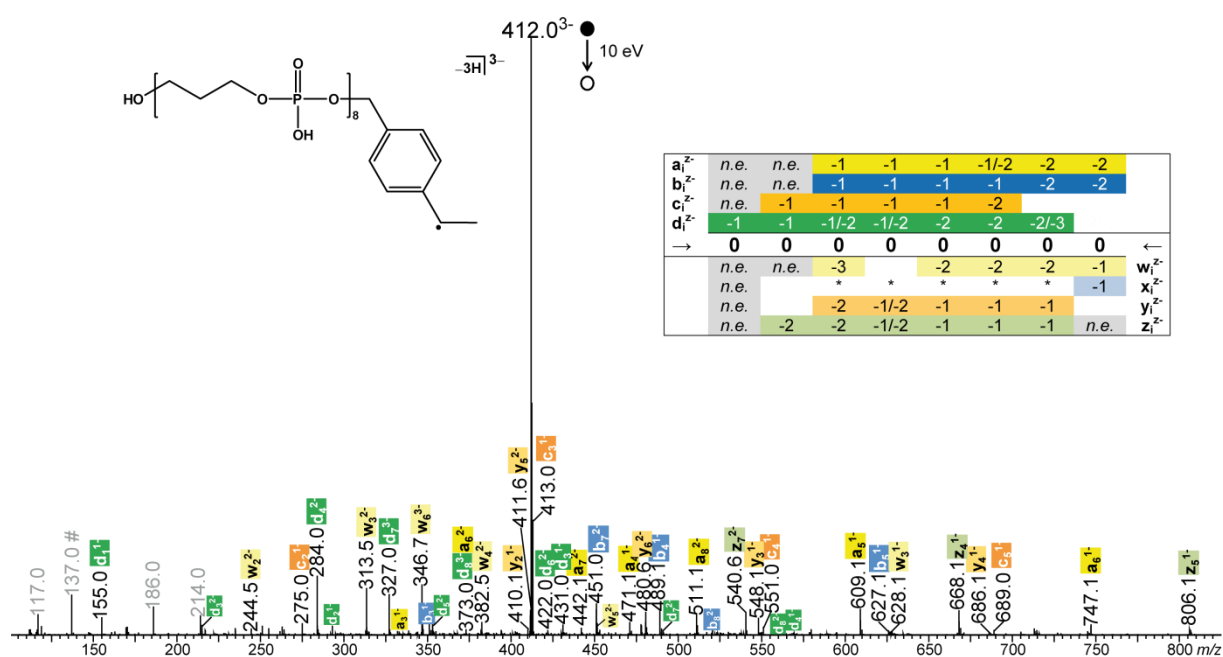
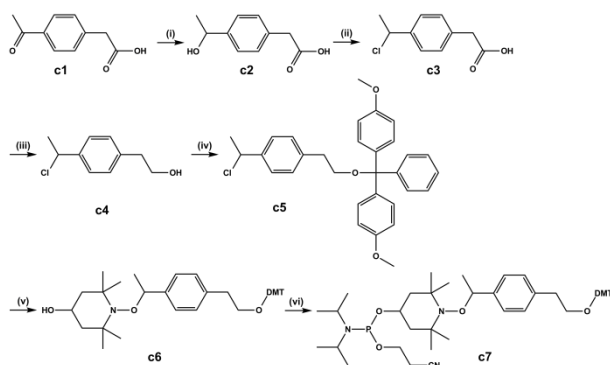


Figure 3. CID spectrum of the m/z 412.0 sub-segment containing the first byte of P2, showing that dissociation only occurs *via* phosphate bond cleavages. Inset: coverage table of the **00000000** byte with each cell reporting the charge states of fragments (Supporting Information, Table S6). Due to $\alpha = \text{HO}$, some fragments are not expected (*n.e.* in grey cells). Most x_1^{z-} ions cannot be determined (hence annotated with an asterisk) because they interfere with ^{13}C isotope of a_1^{z-} ions. Secondary product ions are in grey, including $[\mathbf{0} - \text{H}]^-$ at m/z 137.0 (#).

These CID data allow full coverage of the byte sequence but, as illustrated by the inset table of Figure 3, some fragment series are incomplete. On the one hand, most x_i^{z-} ions cannot be validated due to an isotopic interference with a_i^{z-} ions. On the other hand, some product ions containing the α termination are missing (*e.g.*, the largest c_i^{z-} ions). This is actually a sensitivity issue, due to the quite low abundance of the $[P2 - 6H]^{6-}$ primary precursor yielding small amounts of the m/z 412.0 secondary precursor. Indeed, in contrast to the monodisperse ROSC-based P1 polymer, ESI(-)-MS of the P2 sample shows four main impurities (named A-D) of lower mass but higher intensity than the targeted P2 polymer (Supporting Information, Figure S10). When subjected to CID, all impurities exhibit the same dissociation pattern, with C-ON bond homolysis yielding a single product ion at m/z 600.5 corresponding to the sub-segment containing byte 2 at the right-hand side of the alkoxyamine linkage (Supporting Information, Figure S12). For each species, the defective part at the left-hand side of the C-ON bond has a very low mass (134.1 Da for A, 148.2 Da for B, 196.1 Da for C and 247.1 Da for D) compared to the perfect one (1225.2 Da) in P2. Moreover, this defective part is never detected as a charged fragment, which indicates that it does not contain any phosphate groups. Based on these features and using accurate mass measurements as a support, structural assumptions could be proposed for impurities A (Supporting Information, Figure S12a) and B (Supporting Information, Figure S12b). In both cases, proposed structures suggest a nucleophilic substitution in benzylic position with ammonia (for impurity A) or methylamine (for impurity B), the two reagents used to cleave chains from their solid support at the end of the synthesis. Because such impurities were not formed during the cleavage step of the ROSC-based P1, where the last phosphate group of byte 1 was in a homobenzylic position, the RISC structure was modified accordingly and the RISC2 spacer was investigated.

RISC2 was hence synthesized in 6 steps starting from commercially available 4-acetylphenylacetic acid (Scheme 5) starting with NaBH_4 reduction of 2-(4-acetylphenyl)acetic acid (**c1**) yielding alcohol **c2** in 67% yield followed with chlorination in THF with concentrated HCl in 72% yield. The resulting 2-(4-(1-chloroethyl)phenyl)acetic acid (**c3**) was then selectively reduced with BMS to produce alcohol **c4** in 73% yield, which was then involved in the same synthesis route than ROSC in Scheme 3, with DMT-protection (**c5** in 88% yield), ATRA (**c6** in 86% yield) and phosphoramidite functionalization (**c7** in 48% yield).



Scheme 5. Synthesis of RISC2, **c7**. (i) NaBH₄, EtOH, rt, overnight; (ii) HCl 37%, THF, 0°C, 4h; (iii) BMS, anhydrous THF, rt, overnight; (iv) DMT-Cl, anhydrous pyridine, anhydrous THF, r.t, overnight; (v) 4-hydroxy-TEMPO, CuBr, Cu⁰, PMDETA, degassed THF, r.t., overnight; (vi) O-2-cyanoethyl-N,N-diisopropylchlorophosphoramidite, DIPEA, anhydrous DCM, r.t, 1h. See details in Supporting Information.

RISC2 was fully characterized by NMR (Supporting Information, Figure S13) and mass spectrometry (Supporting Information, Figure S14). This synthesis took advantage of the selectivity of the borane reduction of a benzoic acid in presence of a benzyl chloride.^[33] It should be noted that the molecule is prone to dimerization before DMT protection since the primary alcohol resulting from the borane reduction is a good Lewis base. The **c4** intermediate was hence stored at -18°C before step (ii) and dissolved at low temperature before slow addition to a well stirred solution of DMT-Cl with 1.5 eq of anhydrous pyridine. The dissociation rate constant determined by EPR for RISC2 was $1.1 \cdot 10^{-3} \text{ s}^{-1}$ at 133°C, corresponding to an activation energy of $134.7 \text{ kJ mol}^{-1}$ consistent with values estimated in Table 1 and literature data.^[32]

The byte-truncated poly(phosphodiester) P3 prepared with RISC2 has two identical bytes composed of eight **0** units, which corresponds to a C₇₇H₁₅₆N₃O₇₄P₁₇ elemental composition and a monoisotopic mass of 2833.4076 Da (Supporting Information, Table S1). Negative mode ESI-MS shows a monodisperse sample, with P3 detected with charge state ranging from 3- to 7- (Supporting Information, Figure S15). This result shows that the structure of RISC2 has kept intact during the cleavage step, and hence confirms our assumptions regarding the origin of impurities A and B found in the P2 sample. Collisional activation of [P3 - 6H]⁶⁻ conducts to homolysis of the C-ON bond in RISC2 and subsequent release of the two expected complementary products (Supporting Information, Figure S16). As shown in Figure 4, the CID spectrum of the *m/z* 416.7 sub-segment containing the first byte of P3 only displays fragments of quite high abundance resulting from phosphate bond cleavages. The

reason why all ω -containing fragments are detected except x_i ions remains unclear. Overall, both MS and MS/MS data clearly show that RISC2 has an optimal structure to prevent (i) degradation issues while releasing chains from their solid support and (ii) undesired radical-induced during pseudo-MS³ sequencing experiments.

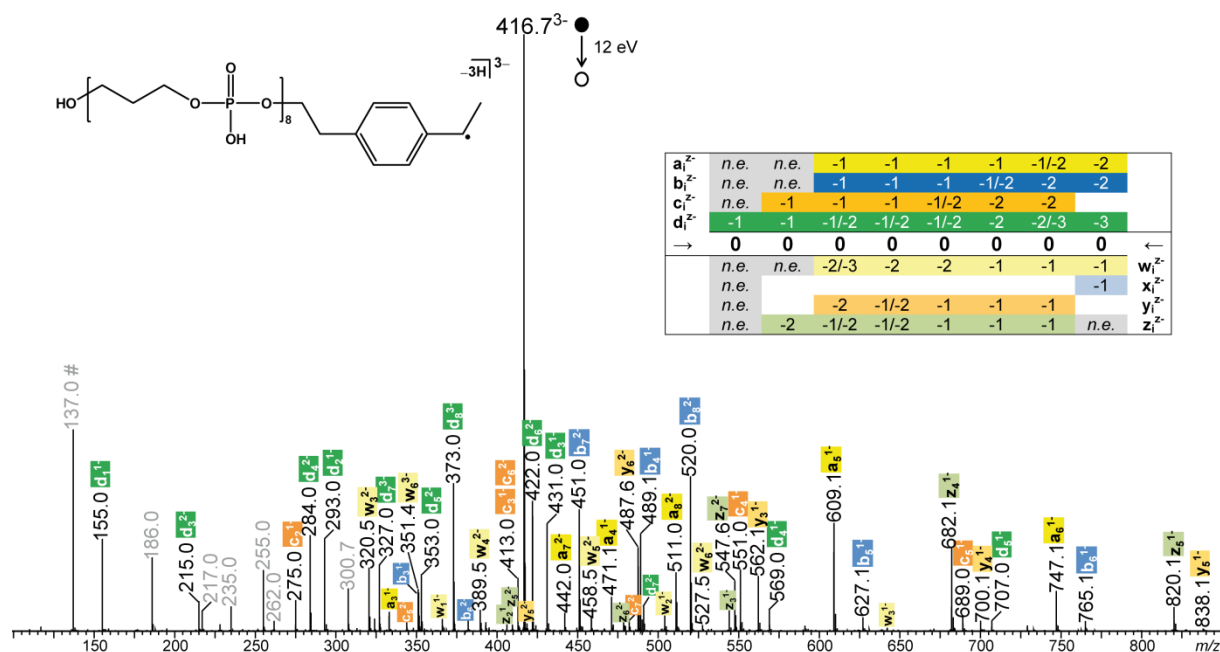


Figure 4. CID spectrum of the m/z 416.7 sub-segment containing the first byte of P3, showing that dissociation only occurs *via* phosphate bond cleavages. Inset: coverage table of the 00000000 byte with each cell reporting the charge states of fragments (Supporting Information, Table S7). Due to $\alpha = \text{HO}$, some ions are not expected (*n.e.* in grey cells). Peaks annotated in grey are secondary fragments including deprotonated 0 units at m/z 137.0 (#).

Sequencing of RISC2-terminated central bytes

The performance of RISC2 to prevent side reactions during MS/MS of first bytes was also tested for central bytes that experience additional undesired dissociations due to the presence of nucleotide tag (used to retrieve initial byte location, Supporting Information, Figure S1) in the vicinity of the terminal radical moiety (Supporting Information, Scheme S1c). When using the amide-containing spacer, these alternative reactions are identical regardless of the nucleotide tag, also rely on a high flexibility of the linear $(\text{CH}_2)_3\text{NH}(\text{CO})\text{C}^*(\text{CH}_3)_2$ terminal moiety (Supporting Information, Scheme S4) and efficiently compete with phosphate bond cleavages useful for sequencing (Supporting Information, Figure S17). The RISC2 spacer was hence used to prepare the four-byte poly(phosphodiester) P4 (Supporting Information, Table

S1) as a monodisperse sample (Figure S18). The $[P4 - 12H]^{12-}$ precursor ion was selected to favor production of sub-segments as triply charged ions during the first activation stage (Supporting Information, Figure S19). CID spectra of the second byte labeled with adenine (Figure 5a) and of the third byte labeled with cytosine (Figure 5b) do not exhibit any side-reaction products but only display ions resulting from phosphate bond cleavages. As illustrated by inset tables of Figure 5, full sequence coverage is readily achieved for both sub-segments. Amongst all expected fragments, only one ion (c_7^{2-}) remains not assignable due to coincidence of its m/z 599.1 value with that of the precursor ion. Mass characteristics of the RISC2 spacer were then implemented in the MS-DECODER algorithm developed for alkoxyamine-containing poly(phosphodiester)s, allowing automated sequence and localization of each byte to be successfully performed in 50 to 100 ms (Supporting Information, Figure S20).

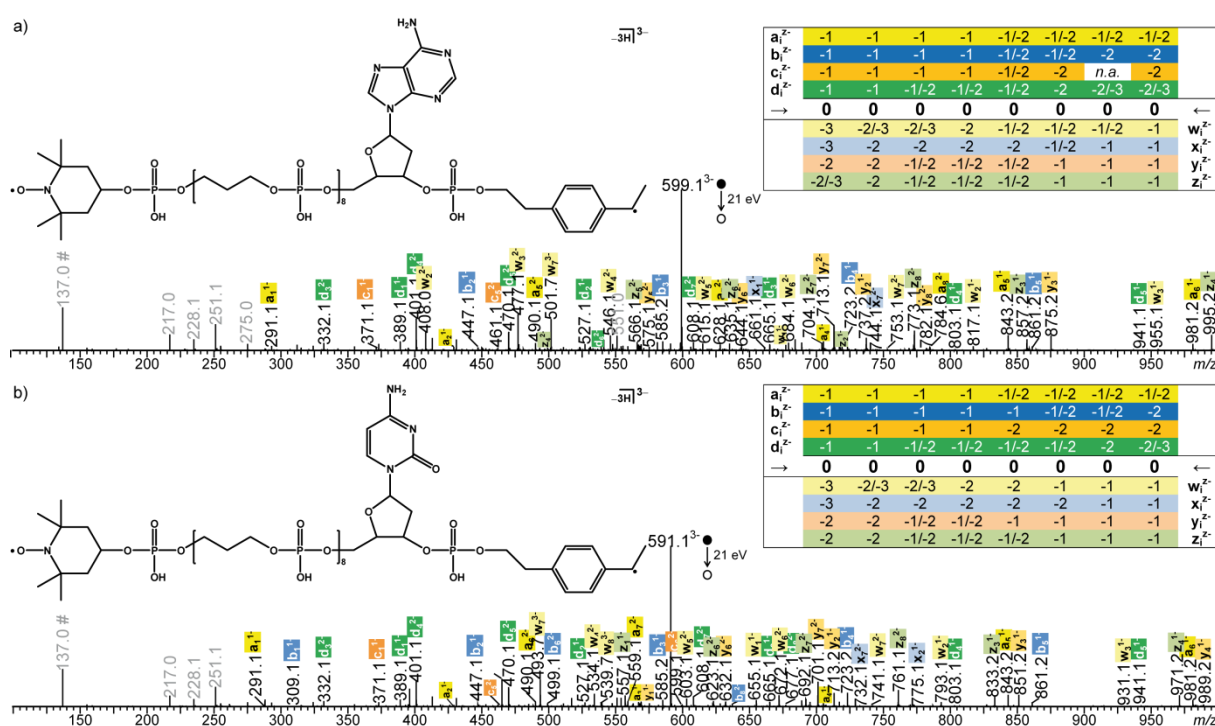


Figure 5. CID spectrum of a) the m/z 599.1 sub-segment labeled with adenine and containing the second byte of P4, and b) the m/z 591.1 sub-segment labeled with cytosine and containing the third byte of P4. Inset: coverage table of the 00000000 byte with each cell reporting the charge states of fragments (Supporting Information, Tables S8-S9). Peaks annotated in grey are secondary fragments including deprotonated 0 units at m/z 137.0 (#).

Automated sequencing of high-capacity digital poly(phosphodiester)s

In order to demonstrate that a complete data file stored in a single molecule can also be read in a fully automated manner, the RISC2 alkoxyamine was used in conjunction with an expanded phosphoramidite alphabet to store a black and white image of 80 pixels in the P5 polymer (Supporting Information, Table S1). Recently developed to enable a storage density of 2 bits per monomer,^[23] this alphabet is based on a library of four binary dyads 00, 01, 10 and 11 (Supporting Information, Scheme S5). The image was created in a grid of 10 columns and 8 rows, converted into a binary sequence using the convention “white = 0, black = 1”, and the binary string was then converted into a coded monomer sequence according to the dyad-language. The so-obtained five-block P5 poly(phosphodiester) was produced as a monodisperse sample, as supported by ESI-MS data (Supporting Information, Figure S21a). The observed charge state distribution ranges from 8– to 19–, with one of the preferential charge states corresponding to 3 negative charges per sub-segment. Activated homolysis of RISC2 in this [P5 – 15H]^{15–} precursor yields the expected complementary sub-segments (Supporting Information, Figure S21b). The *m/z* value measured for each triply charged sub-segment permits to retrieve their initial location thanks to the tag system but is associated to a set of possible dyad combinations (Supporting Information, Table S10), in contrast to the unique 0/1 composition obtained when using the alphabet where one unit codes for one bit. This is a consequence of the mass of dyads used here, with 138.0 Da for 00, 152.0 Da for 01, 166.0 for 10 and 180.1 for 11, leading to mass coincidences when combined within a given sub-segment. For example, the first sub-segment of P5 detected at *m/z* 435.4 as a triply charged ion has a mass of 1309.2 Da which includes the 1160 Da contribution of the coding string that be obtained with 6 “00” + 1 “01” + 1 “11”, 6 “00” + 2 “10”, 5 “00” + 2 “01” + 1 “10”, or 4 “00” + 4 “01”. As a result, it takes more time for the MS-DECODER to decode each sub-segment written with dyads (1 second, on average) compared to the 100 ms requested to retrieve the binary information encoded with monomers “0” and “1”. Yet, all pseudo-MS³ spectra recorded for P5 (Supporting Information, Figure S21c) were successfully processed. Based on the location tag system, the MS-DECODER software enables proper alignment of all sub-segments according to their initial sequence in the chain (Figure 6, left part) and, providing that the user defines the image definition (10 columns x 8 rows), readily reconstruct the black and white image (Figure 6, right part).

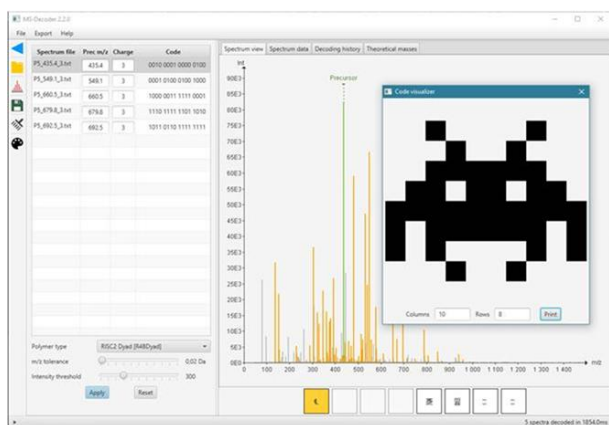


Figure 6. Output of the MS-DECODER software for automated decoding of the 80 pixel image of a space invader stored in P5.

Conclusion

In summary, it was shown herein that precise alkoxyamine-design allows preparation of automatically-decodable digital poly(phosphodiester)s. Among all alkoxyamines studied in this work, RISC2 turned out to be, by far, the most efficient. Indeed, this particular design efficiently suppresses side-peaks that usually prevent MS/MS automated decryption with the software MS-DECODER. As a result, model digital polymers could be decrypted in a fully automated manner: after loading all pseudo-MS³ files recorded for a given polymer, the algorithm was developed to decode each sub-segment and to align all sub-segments according to their initial order in the chain thanks to the location tag system. These results constitute an important step-forward in the field of non-biological digital polymers. Poly(phosphodiester)s being currently the only family of synthetic polymers allowing high capacity storage,^[23] the automatic decryption of these polymers opens up very interesting perspectives for fast and parallel sequencing. Using RISC2 in conjunction with expanded phosphoramidite alphabets was shown to allow fast (5 seconds) interpretation of a black and white image stored in a synthetic poly(phosphodiester) and, when combined with appropriate compression algorithms, this may also allow efficient processing of larger amounts of information as requested to encode texts, videos or secure data. These results further indicate that precise macromolecular-engineering is a key strategy for improving the storage properties of digital macromolecules. Consequently, the performances of non-natural informational polymers may soon equal or even surpass those of biological polymers.^[34]

Experimental section

See Supporting Information

Acknowledgments

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). 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] V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church, W. L. Hughes, *Nature Mater.* **2016**, *15*, 366.
- [2] J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science* **2013**, *341*, 1238149.
- [3] H. Colquhoun, J.-F. Lutz, *Nat. Chem.* **2014**, *6*, 455.
- [4] R. K. Roy, A. Meszynska, C. Laure, L. Charles, C. Verchin, J.-F. Lutz, *Nat. Commun.* **2015**, *6*, 7237.
- [5] M. Rutten, F. W. Vaandrager, J. Elemans, R. J. M. Nolte, *Nat. Rev. Chem.* **2018**, *2*, 365.
- [6] S. Martens, A. Landuyt, P. Espeel, B. Devreese, P. Dawyndt, F. Du Prez, *Nat. Commun.* **2018**, *9*, 4451.
- [7] A. C. Boukis, M. A. R. Meier, *Eur. Polym. J.* **2018**, *104*, 32.
- [8] Z. H. Huang, Q. N. Shi, J. Guo, F. Y. Meng, Y. J. Zhang, Y. T. Lu, Z. F. Qian, X. P. Li, N. C. Zhou, Z. B. Zhang, X. L. Zhu, *Nat. Commun.* **2019**, *10*, 1918.
- [9] J. M. Lee, M. B. Koo, S. W. Lee, H. Lee, J. Kwon, Y. H. Shim, S. Y. Kim, K. T. Kim, *Nat. Commun.* **2020**, *11*, 56.
- [10] G. M. Church, Y. Gao, S. Kosuri, *Science* **2012**, *337*, 1628.
- [11] N. Goldman, P. Bertone, S. Y. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, E. Birney, *Nature* **2013**, *494*, 77.
- [12] Y. Erlich, D. Zielinski, *Science* **2017**, *355*, 950.
- [13] T. T. Trinh, C. Laure, J.-F. Lutz, *Macromol. Chem. Phys.* **2015**, *216*, 1498.
- [14] L. Charles, G. Cavallo, V. Monnier, L. Oswald, R. Szweda, J.-F. Lutz, *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1149.
- [15] B. Liu, Q. Shi, L. Hu, Z. Huang, X. Zhu, Z. Zhang, *Polym. Chem.* **2020**, *11*, 1702.
- [16] J.-A. Amalian, A. Al Ouahabi, G. Cavallo, N. F. Konig, S. Poyer, J.-F. Lutz, L. Charles, *J. Mass Spectrom.* **2017**, *52*, 788.
- [17] A. Al Ouahabi, J.-A. Amalian, L. Charles, J.-F. Lutz, *Nat. Commun.* **2017**, *8*, 967.
- [18] N. Appukutti, C. J. Serpell, *Polym. Chem.* **2018**, *9*, 2210.

- [19] M. Vybornyi, Y. Vyborna, R. Haner, *Chem. Soc. Rev.* **2019**, *48*, 4347.
- [20] A. Al Ouahabi, M. Kotera, L. Charles, J.-F. Lutz, *ACS Macro Letters* **2015**, *4*, 1077.
- [21] A. Al Ouahabi, L. Charles, J.-F. Lutz, *J. Am. Chem. Soc.* **2015**, *137*, 5629.
- [22] S. A. McLuckey, G. J. Vanberkel, G. L. Glish, *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 60.
- [23] E. Laurent, J.-A. Amalian, M. Parmentier, L. Oswald, A. Al Ouahabi, F. Dufour, K. Launay, J.-L. Clement, D. Gigmes, M. A. Delsuc, L. Charles, J.-F. Lutz, *Macromolecules* **2020**, *53*, 4022.
- [24] A. Burel, C. Carapito, J.-F. Lutz, L. Charles, *Macromolecules* **2017**, *50*, 8290.
- [25] R. G. Cooks, J. H. Beynon, T. Ast, *J. Am. Chem. Soc.* **1972**, *94*, 1004.
- [26] D. P. Little, R. A. Chorush, J. P. Speir, M. W. Senko, N. L. Kelleher, F. W. McLafferty, *J. Am. Chem. Soc.* **1994**, *116*, 4893.
- [27] G. Gryn'ova, D. L. Marshall, S. J. Blanksby, M. L. Coote, *Nat. Chem.* **2013**, *5*, 474.
- [28] G. Gryn'ova, M. L. Coote, *J. Am. Chem. Soc.* **2013**, *135*, 15392.
- [29] S. Marque, H. Fischer, E. Baier, A. Studer, *J. Org. Chem.* **2001**, *66*, 1146.
- [30] D. Bertin, D. Gigmes, S. R. A. Marque, P. Tordo, *Macromolecules* **2005**, *38*, 2638.
- [31] K. Matyjaszewski, B. E. Woodworth, X. Zhang, S. G. Gaynor, Z. Metzner, *Macromolecules* **1998**, *31*, 5955.
- [32] S. Marque, C. Le Mercier, P. Tordo, H. Fischer, *Macromolecules* **2000**, *33*, 4403.
- [33] E. R. Burkhardt, K. Matos, *Chem. Rev.* **2006**, *106*, 2617.
- [34] J. F. Lutz, *ACS Macro Letters* **2020**, *9*, 185.

Entry for the Table of Contents



Reliable reading of information coded in the sequence of long poly(phosphodiester)s was previously achieved by introducing an alkoxyamine spacer between information sub-segments. In the present work, we show that the structure of this spacer can be designed in a rational manner to prevent side-reactions during MS/MS sequencing and hence enable automated reading of each byte of information in about 100 ms.