Optimal conditions for tandem mass spectrometric sequencing of information-containing nitrogen-substituted polyurethanes

Laurence Charles, Tathagata Mondal, Vincent Greff, Mattia Razzini, Valérie Monnier, Alexandre Burel, Christine Carapito, Jean-François Lutz

Optimal conditions for MS/MS sequencing of information-containing N-substituted polyurethanes

Laurence Charles,[1]* Tathagata Mondal,[2] Vincent Greff,[2] Mattia Razzini,[2] Valérie Monnier,[3] Alexandre Burel,[4] Christine Carapito,[4] and Jean-François Lutz[2]*

[1] Aix Marseille Univ, CNRS, ICR, Institut de Chimie Radicalaire, Marseille – France

[2] Université de Strasbourg, CNRS, Institut Charles Sadron UPR22, Strasbourg – France

[3] Aix Marseille Univ, CNRS, Fédération des Sciences Chimiques de Marseille, Marseille – France

[4] Université de Strasbourg, CNRS, Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC, Strasbourg – France

* To whom correspondence should be addressed. E-mail: laurence.charles@univ-amu.fr Phone: +33 491 28 8678. Fax: +33 491 28 2897. Email: jflutz@unistra.fr. Phone: +33 388 41 4016. Fax: +33 388 41 4099.

Published in *Rapid Communications in Mass Spectrometry, 34*, e8815 **(2020)**
https://doi.org/10.1002/rcm.8815

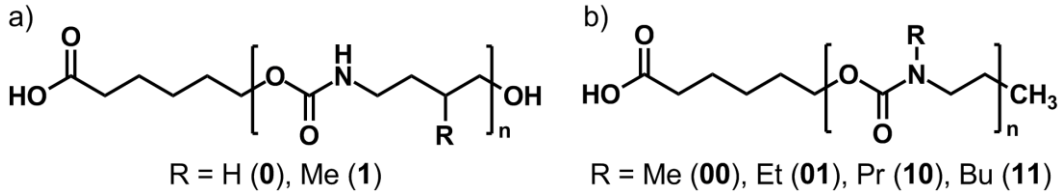
1. INTRODUCTION

Owing to its sequencing capability, tandem mass spectrometry (MS/MS) has rapidly become the key technique to retrieve digital information stored in sequence-defined non-natural polymers. In these monodisperse macromolecules, co-monomers are defined as letters of an alphabet (*e.g.*, 0- and 1-bits of the ASCII code) and their controlled arrangement throughout the chain allows chemical storage of binary messages.[1-3] As long as co-monomers have different masses, reading such messages is a sequencing task as typically achieved in MS/MS, where the nature and relative location of building units can be retrieved by analyzing data with regard to specific fragmentation rules. This was reported by most groups acting in the field for a variety of polymers such as poly(alkoxyamine amide)s,[4] poly(triazole amide)s,[5] polyurethanes,[6] poly(alkoxyamine phosphodiester)s,[7] poly(ester amide)s,[8] poy(urethane amide)s,[9] poly(succinimide thioester)s,[10, 11] and polyesters.[12] As shown in all these studies using collisional activation, one primary requirement for reliable sequencing is occurrence of backbone bond cleavages, independently of the sequence, in order to generate product ions that contain either one or the other original chain end. Yet, the type of functional moieties in these polymers results from the synthesis route adopted to produce, at high yield, chains with length monodispersity and perfect control over their sequence. Because fragmentation patterns of synthetic species strongly depend on their chemistry, MS/MS data of digital polymers are sometimes complex and, although sequence coverage can be achieved, formation of numerous product ion series favors signal dilution and jeopardizes detection of useful product ions as the precursor size increases.

Accordingly, our group early included MS/MS readability as part of the design strategy to develop any new sequence-defined polymers. For example, changing the tri(ethylene glycol) spacer to an alkyl group in poly(triazole amide)s permitted to reduce the number of dissociation routes to the sole amide bond cleavage in each monomer.[13] In contrast, an additional MS-active group was introduced in the end-group of sequence-controlled polyurethanes to improve their ionization yield, another key requirement for reliable sequencing by MS/MS.[14] In order to capitalize on the robustness of phosphoramidite used in the automated synthesis of long poly(phosphodiester)s, specific reagents were prepared to include alkoxyamine linkages in the final chains, the low bond dissociation energy of which makes all phosphate bonds MS/MS silent, lowering product ion series from eight to two and hence ensuring full sequence coverage of long chains.[15] Similarly, Zhang and co-workers recently showed that controlling the oxidation degree of sulfur atoms in the backbone of poly(succinimide thioester)s drastically enhances dissociation selectivity.[16] Beside re-engineering structures, fragmentation of synthetic

polymers can also be manipulated by appropriate adjustments of experimental parameters in the ionization step. This was recently shown for polyurethanes, with sodium salt supplemented to electrosprayed solution permitting to change their acidic α termination into a carboxylate moiety and hence simplify their dissociation pattern in the positive ion mode.[17]

The same strategy is followed here to optimize collision induced dissociation (CID) of poly(N-substituted urethane)s, further named N–R PUs to distinguish them from regular N–H PUs after the nomenclature proposed by Montaudo *et al*.[18] As compared to N–H PUs with the alkyl coding moiety in the backbone (Scheme 1A), N–R PUs have their code implemented in the alkyl N-substituent (Scheme 1B), which precludes hydrogen bonding between urethane units, hence preventing crystallization and solubility issues.[19] N–R PUs were prepared with amino alcohol building blocks containing secondary amine functions[20] and the use of four monomers holding R chains of different mass permits definition of a dyad-based alphabet (Scheme 1B), hence also enhancing storage density. Another structural improvement of N–R PUs over N–H PUs is the methyl ω end-group instead of OH, which permits to avoid esterification of the COOH-,OH-terminated oligomers sometimes observed during chain release from the synthesis solid support.[21] Although minor, the latter structural change also contributes to major changes in CID of N–R PUs compared to N–H PUs. In this context, the present study explores mechanisms underlying dissociation reactions observed for N–R PUs prior to optimizing experimental conditions that ensure simple and robust sequencing rules for reliable decoding of long chains.



Scheme 1. Structural features of sequence-defined A) regular polyurethanes (N–H PUs) and B) poly(N-substituted urethane)s (N–R PUs).

2. EXPERIMENTAL

2.1. Chemicals

Sequence-defined N-substituted polyurethanes studied in this work (Table S1, supporting information) were synthesized using an iterative solid-phase approach described in a recent publication.[20] All oligomer samples (a few mg) were first solubilized in methanol (SDS, Peypin, France), then diluted ($1/10^3$ to $1/10^4$, v/v) in a methanolic solution of ammonium acetate (3

mM) or sodium chloride (0.1 mM), both purchased from Sigma Aldrich (Saint Louis, MO), prior being subjected to ESI. Poly(methylmethacrylate) (PMMA) used as internal standards for accurate mass measurements were from Sigma Aldrich.

2.3. Mass Spectrometry

High resolution MS and MS/MS experiments were performed using a QStar Elite mass spectrometer (Applied Biosystems SCIEX, Concord, ON, Canada), a QTOF instrument enabling high resolution mass measurements of ions in an orthogonal acceleration time-of-flight (oa-TOF) mass analyzer, working in tandem with a quadrupole in the MS/MS mode. This instrument is equipped with an electrospray ionization (ESI) source operated in the negative (capillary voltage: -4200 V; cone voltage: -75 V) or the positive ion mode (capillary voltage: +5500 V; cone voltage: +75 V). Air was used as the nebulizing gas (10 psi) and nitrogen as the curtain gas (20 psi). Sample solutions were introduced in the ESI source with a syringe pump at a 10 μL min^{-1} flow rate. Ions were accurately mass measured after internal calibration of the oa-TOF mass analyzer, using two cationic adducts of PMMA to bracket the targeted analyte *m/z* value in the MS mode[22] or the precursor ion signal in the MS/MS mode. In CID experiments, precursor ions were selected by the quadrupole mass analyzer, which resolving power was intentionally decreased (selection window ≈ 2 *m/z*) to also sample ^{13}C isotopes so that product ion charge state can be determined from their partial isotopic pattern.

Selected precursor ions were then injected into the collision cell filled with nitrogen, and product ions were measured in the oa-TOF. Different collision energies were used as a function of ions to be fragmented (see text). Instrument control, data acquisition and data processing of all experiments were achieved using Analyst software (QS 2.0) provided by Applied Biosystems SCIEX.

2.4. MS-DECODER for automated sequencing of N-coded PUs

MS-DECODER is a platform-independent in-house developed software that can run on any computer with Java8 installed (Windows, macOS, Linux derivatives …).[23] A new algorithm supporting dyad detection in sequence-defined N-substituted PUs was recently implemented and allowed successful sequencing of these coded polymers in some tens of milliseconds. Spectra can be treated in batch mode according to predefined user settings including *(i)* the *m/z* tolerance window to include matching ions and *(ii)* absolute intensity threshold for a peak to be considered as a real non-background signal. Sequencing outcome can be clipped to the clipboard or exported as CSV (Coma Separated Value) or XLSX (Microsoft excel) files.

3. RESULTS AND DISCUSSION

3.1. CID in the negative ion mode

Sequence-defined N–H PUs are certainly the most easy-to-read encoded polymers when analyzed by MS/MS in the negative ion mode.[6] Indeed, when activated as deprotonated molecules, dissociation of N–H PUs proceeds *via* O–(CO) bond cleavage in each repeating unit, leading to one single and complete a_i^- ion series (Figure S1, supporting information). This allows their full sequence to be readily recovered based on peak-to-peak distances being equal to the mass of one or the other coding unit. Cleavage of O–(CO) bonds was shown to proceed according to two main pathways, involving H-atom transfer either from the NH group (as also proposed by Gies and Hercules for related species[24]) or from the terminal OH group onto the ether oxygen of the carbamate group.[14] These reactions can obviously not occur from N–R PUs, with an alkyl group on the carbamate nitrogen and CH_3 as the ω end-group. Instead, CID of deprotonated N–R PUs gives rise to two main product ion series of low abundance, as shown in Figure 1 for the P1 oligomer of sequence 00·01·10·11·00·01·10·11. Both series are formed upon cleavage of the CH_2–O bond between repeating units so, according to Wesdemiotis' nomenclature for polymer product ions,[25] these ions are named α_i^- and w_j^-, respectively (Scheme S1, supporting information).

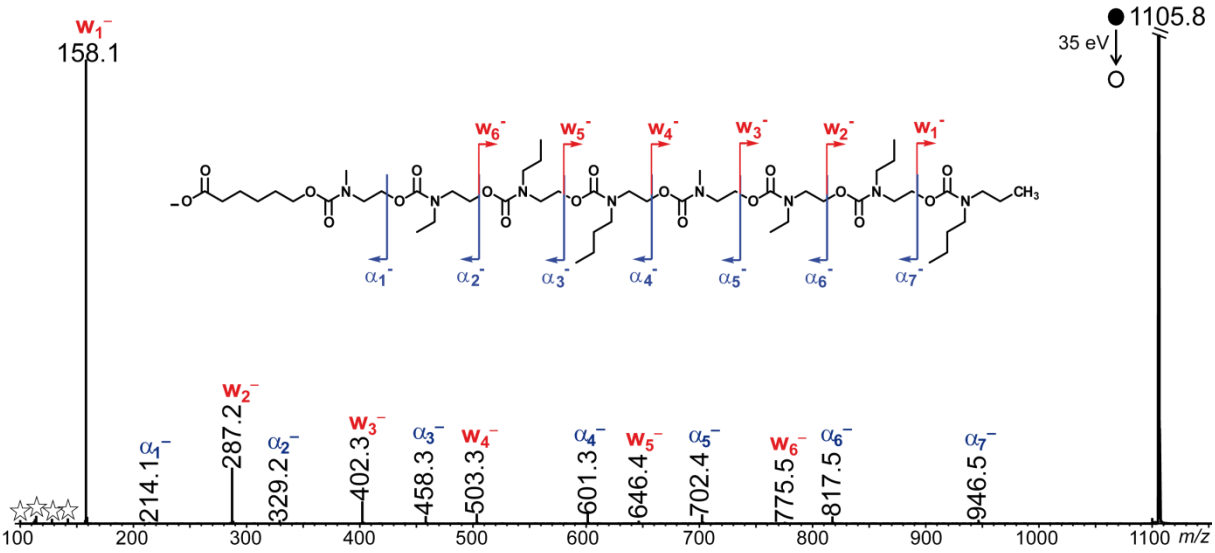
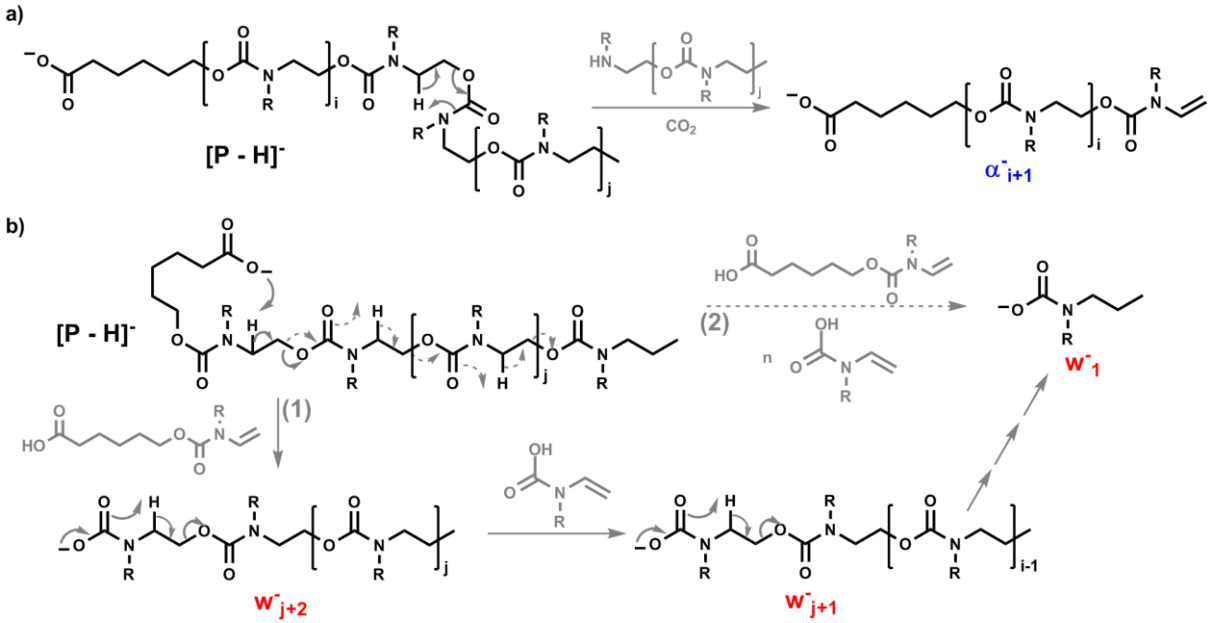


Figure 1. Negative mode ESI-MS/MS of $[P1 - H]^-$ at *m/z* 1105.8 (with a 3.0 magnification factor of the y-axis). Product ion assignments (in inset) are supported by accurate mass measurements (Table S2, supporting information). Peaks designated by white stars correspond to deprotonated monomers: $[00 - H]^-$ at *m/z* 100.1, $[01 - H]^-$ at *m/z* 114.1, $[10 - H]^-$ at *m/z* 128.1, $[11 - H]^-$ at *m/z* 142.1. Collision energy is given in the laboratory frame.

Formation of α_i^- product ions is proposed to occur in a competitive manner, after 1,5-H transfer from a methylene group onto the carbamate nitrogen atom of the next repeating unit (Scheme 2A). Although detected with low abundance, these product ions allow full coverage of the sequence (see inset of Figure 1). Detection of w_i^- ions containing the original ω end-group when starting from precursor ions having the negative charge initially located in the α termination implies that the charge has moved from the left- to the right-hand side. Accordingly, mechanisms proposed for formation of w_j^- ions rely on high mobility of the negative charge which would propagate through the entire chain. As depicted in Scheme 2B, w_j^- product ions could be formed either *via* a multiple step pathway (plain arrows, pathway 1) or after rapid propagation of the negative charge (dotted arrows, pathway 2). The latter process would be more efficiently related to the smallest w_j^- ions always observed with the highest abundance. Yet, this ion series is not complete and can only be used to confirm the right-end part of the sequence established with the α_i^- product ions. However, full sequence coverage is no longer achieved as the size of the precursor ion exceeds eight monomer units. This is because the most useful α_i^- ions are formed with very low abundance, so signal dilution becomes critical as the chain length increases. Sequence coverage became worse upon enhancement of the collision energy, which mainly favored production of the smallest w_i^- and more particularly w_1^-. Efficiency of MS/MS sequencing was hence explored in the positive ion mode.



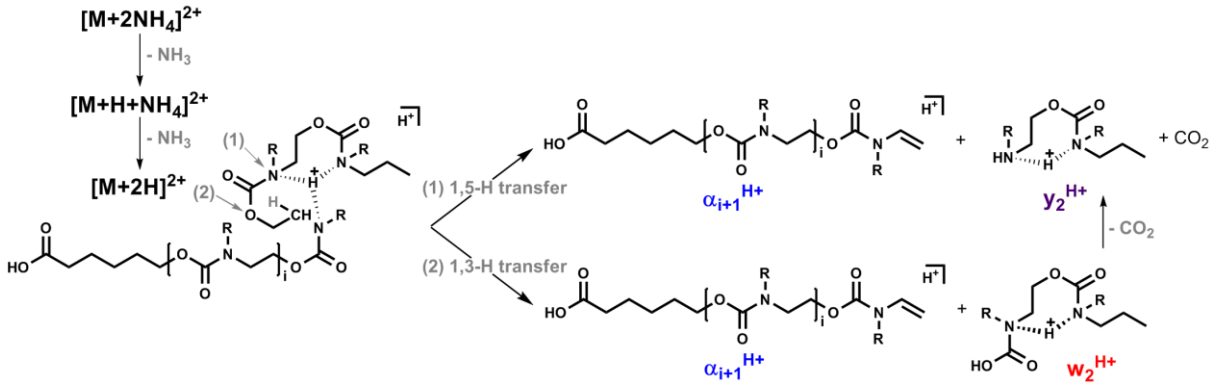
Scheme 2. Mechanisms proposed for dissociation of deprotonated N–PUs (with R = Me for 00, Et for 01, Pr for 10 or Bu for 11) conducting to the formation of A) α_i^- ions and B) w_j^- ions.

3.2. CID in the positive ion mode

Sequence-defined N–H PUs were easily ionized as cationic adducts but exhibited complex MS/MS spectra, unless activated as $[M - H + (z+1)Na]^{z+}$ precursor ions, where the acidic proton of the α end-group has been exchanged with an alkali.[17] The high mobility of this proton was indeed demonstrated to assist bond cleavages, yielding up to five series of product ions. In contrast, as also reported in the case of unsaturated fatty acids,[26, 27] this reactivity can be inhibited by changing the acidic to a carboxylate termination, as shown for CID of N–H PUs leading to two ion series only, upon cleavage of O–(CO) bonds (Figure S2, supporting information). A quite similar scenario was observed for N–R PUs, although they exhibit some specificities, starting from the MS mode where they readily adopt multiple charges. Even for short oligomers, doubly charged adducts were produced with higher abundance than their singly charged counterparts (Figure S3, supporting information). This offers the opportunity to study CID behavior of multiply charged species, which is more relevant in the perspective of long chain sequencing while keeping MS/MS data simple, as shown hereafter for ammonium and sodium adducts of P1.

Collisional activation of $[P1+2NH_4]^{2+}$ at *m/z* 571.4 first leads to two successive losses of NH_3, respectively yielding $[P1+H+NH_4]^{2+}$ at *m/z* 562.9 and $[P1+2H]^{2+}$ at *m/z* 554.3 (Figure 2A). Since ammonium was not found in any other product ions (Table S3, supporting information) and because the same product ions were generated from $[P1+H+NH_4]^{2+}$ and $[P1+2H]^{2+}$ (Figure S4, supporting information), the doubly protonated molecule was considered here as the actual precursor ion for the sake of simplicity. Cleavage of the CH_2–O bond between two repeating units (and subsequent small neutral release) can explain the formation of the three types of product ions observed from $[M+2H]^{2+}$. As depicted in Scheme 3, one adducted proton would be solvated by N atoms from at least three consecutive units. Then, one H atom from the methylene group of the third unit (in grey) could be transferred to the nitrogen atom of the preceding unit (1,5-transfer), yielding α_i^{H+} (in blue) and y_j^{H+} (in purple) after elimination of a CO_2 molecule. The alternative 1,3-transfer of the same atom onto the nearby O atom would result in the formation of α_i^{H+} and w_j^{H+} (in red) product ions. 1,5 H-atom transfer involving the carbonyl group was also envisaged to explain the formation of α_i^{H+} and w_j^{H+} (with the same structure as shown in Scheme 3). However, this reaction was not considered since it is expected to cleave any CH_2–O bond, including the last one to produce w_1^{H+} which is never detected. Instead, the need for the precursor ion to adopt the proton-bound conformation depicted in Scheme 3 would explain why the last CH_2–O bond in the chain is never observed to cleave. This prevents the two last repeat units from being identified and hence full sequence coverage

to be achieved. Additional low abundance internal product ions designated by pale grey stars in Figure 2A correspond to protonated units (Scheme S2, supporting information).



Scheme 3. Mechanisms proposed for the cleavage of CH_2–O bonds in protonated N–R PUs, leading either to α_i^{H+} and y_j^{H+} after loss of CO_2, or α_i^{H+} and w_j^{H+}.

As compared to $[P1+2NH_4]^{2+}$, MS/MS data recorded for the doubly sodiated oligomer show the same product ion series (but designated as α_i^{Na+}, y_j^{Na+} and w_j^{Na+}) as well as an additional $w'_j{}^{Na+}$ ion series (in green, Figure 2B). Another feature is the release of a naked Na^+, as evidenced by detection of $[P1+Na]^+$. Two distinct forms of the $[P1+2Na]^{2+}$ precursor ion were envisaged to explain the observed dissociation pattern. One form would be $[(P1–H+Na)+H+Na]^{2+}$ obtained after H/Na exchange in the α termination. This makes one proton available to assist dissociation according to the process illustrated in Scheme 4A, leading to α_i^{Na+} and y_j^{Na+} ions (except y_1 which lacks oxygen atoms to interact with sodium) after loss of CO_2. Alternatively, precursors having one carbamate N atom protonated could dissociate according to a Na^+-assisted process, to yield α_i^{Na+} together with w_j^{Na+} (Scheme 4B). The α_i^{Na+} ions would experience a depolymerization process, that is, monomer elimination from large ions to form smaller congeners, ultimately yielding α_1^{Na+} which would further eliminate the α moiety as a neutral to produce the I_1^+ internal product ion (Scheme 4B). The latter pathway is proposed to account for the abundance of I_1^+, always detected as the most prominent product ion. Consistent with its structure, its *m/z* value depends on the first repeating unit in the chain: from *m/z* 102.1 for P1 starting with the 00 unit of mass 101.0 Da (Figure 2B), I_1^+ is observed to shift to *m/z* 116.1 for species like P2 which first unit is 01 of mass 115.1 Da (Figure S5A, supporting information) and to *m/z* 144.1 for sequences starting with 11 of mass 143.1 Da as in P3 (Figure S5B, supporting information). In contrast, all other repeating units of the chain are released as minor sodiated molecules (Scheme S2, supporting information) at *m/z* 124.1, *m/z* 138.1, *m/z* 152.1, and *m/z* 166.1 (dark grey stars in Figure 2B).

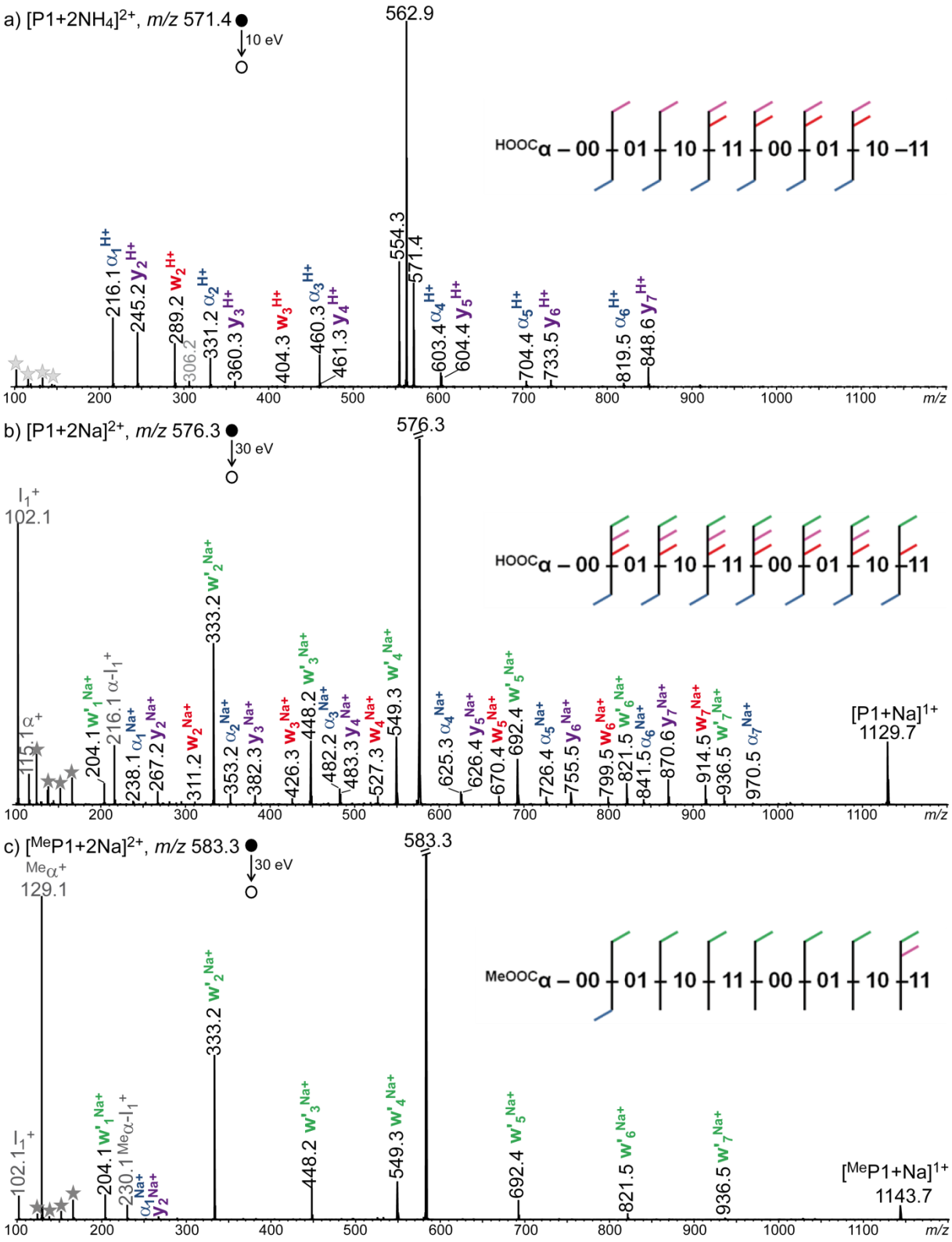
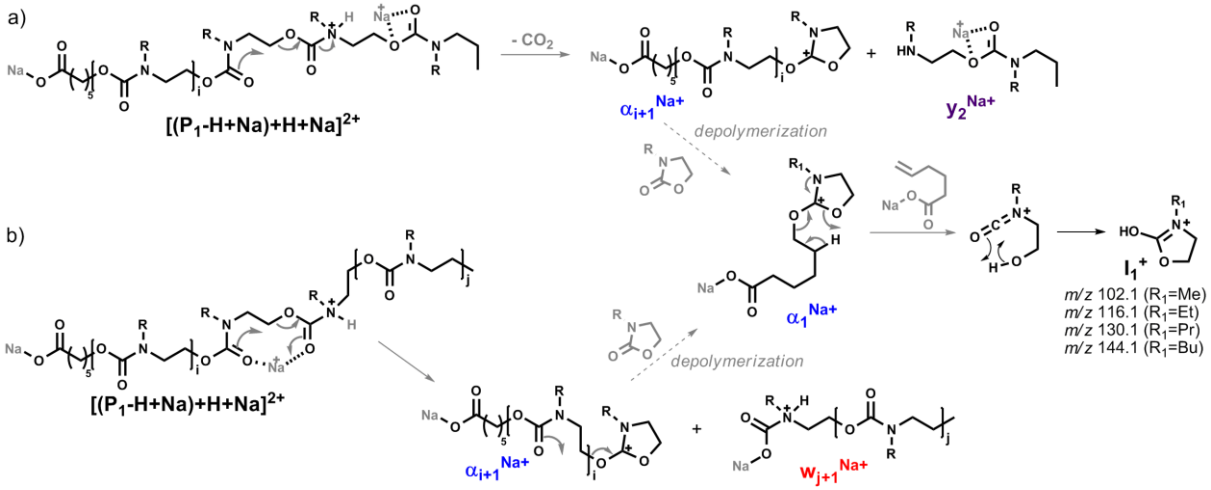
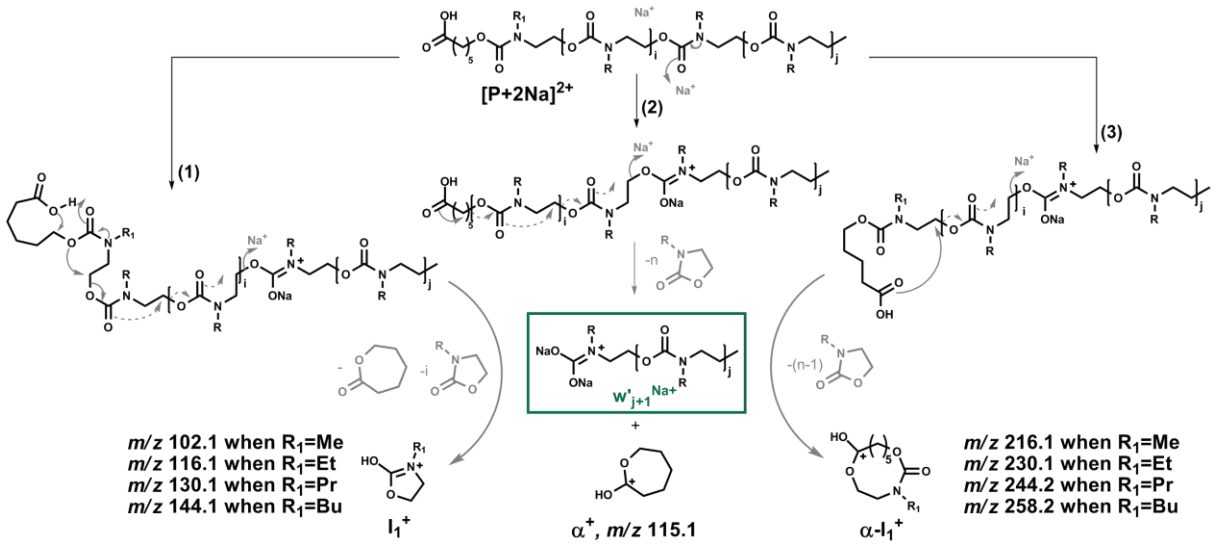


Figure 2. Positive mode ESI-MS/MS of A) $[P1+2NH_4]^{2+}$ at *m/z* 571.4, B) $[P1+2Na]^{2+}$ at *m/z* 576.3, and C) $[^{Me}P1+2Na]^{2+}$ at *m/z* 583.3 (with ^{Me}P1, the oligomer with methylated α group). The same color code is employed for peak annotation and for patterns designating detected product ions in the dissociating scheme shown in inset: blue for α_i, red for w_j, purple for y_j, and green for w'_j. Secondary product ions are in grey, including protonated (pale grey) or sodiated (dark grey) monomers designated by stars. Collision energy is given in the laboratory frame. Magnification factor of the y-axis is 2.8 in B) and 1.5 in C).



Scheme 4. Dissociation of $[P1+2Na]^{2+}$ after H/Na exchange in the acidic α end-group: mechanisms proposed for the cleavage of CH_2–O bonds lead to A) α_i^{Na+} and y_j^{Na+} after loss of CO_2 according to a proton-assisted reaction, or B) α_i^{Na+} and w_j^{Na+} in a process assisted by both H^+ and Na^+, with α_i^{Na+} further experiencing a depolymerization process to yield I_1^+.

The second form of the precursor, *i.e.*, P1 with an acidic α termination and the two Na^+ cations distributed along the backbone, has to be considered to account for the formation of ions of the fourth $w'_j{}^{Na+}$ series that contain two sodium atoms (Table S4, supporting information). Moreover, since no complementary product was ever observed, formation of $w'_j{}^{Na+}$ is proposed to occur together with a depolymerization process leading to low *m/z* ions. Three pathways were actually considered, starting from the same ion in which delocalization of the electron pair of one carbamate N atom has assisted the formation of one O–Na bond (Scheme 5). Then, adduction of the second Na^+ onto the other oxygen of the same carbamate group would permit formation of $w'_j{}^{Na+}$ ion with the positive charge held by the nitrogen atom. The latter reaction could be promoted by the attack of the carbonyl oxygen of the carboxylic acid onto the last methylene moiety i) of the α end-group to produce the cyclic *m/z* 115.1 ion named α^+ (pathway 2 in Scheme 5) or ii) of the first repeating unit to yield α-I_1^+ (pathway 3 in Scheme 5), after release of neutral repeating units. Alternatively, transfer of the α acidic proton onto the carbonyl O atom of the first unit would induce the concerted process depicted in pathway 1 of Scheme 5, leading to i) elimination of the end-group as a cyclic neutral, ii) formation of I_1^+, iii) depolymerization of the chain up to iv) formation of $w'_j{}^{Na+}$, the size of which depends on which carbamate group is adducted to the second Na^+ (not necessarily the same one that contains the first O–Na bond as shown in Scheme 5).



Scheme 5. Dissociation of $[P+2Na]^{2+}$ precursors still having their acidic α termination to account for the formation of $w'_j{}^{Na+}$ product ion together with the internal $I_1{}^+$ ion (pathway 1), α^+ at *m/z* 115.1 (pathway 2), or α-$I_1{}^+$ detected at *m/z* 216.1, *m/z* 230.1, *m/z* 244.1 or *m/z* 258.1 depending on the R_1 substituent being Me (00), Et (01), Pr (10) or Bu (11), respectively.

Dissociation reactions proposed in Schemes 4-5 are supported by MS/MS data recorded for P1 having a methyl carboxylate as the α termination, found to form quite readily during storage of N–R PUs in methanol (*vide infra*). Indeed, when activated as a doubly sodiated adduct, dissociation of the so-called ^{Me}P1 mostly corresponds to a single series of $w'_j{}^{Na+}$ ions (Figure 2C and Table S5, supporting information). This shows that, once the reactive acidic α group is removed, reactions leading to $\alpha_i{}^{Na+}$, $y_j{}^{Na+}$ and $w_j{}^{Na+}$ (Scheme 4) no longer occur. Similarly, pathway 1 in Scheme 5 can no longer proceed, accounting for the major decrease of $I_1{}^+$ abundance. Instead, the most abundant product ion becomes $^{Me}\alpha^+$ at *m/z* 129.1 (the methylated homologue of α^+), proposed to form together with the main $w'_j{}^{Na+}$ ion series according to pathway 2 in Scheme 5. In contrast, and as expected from reactions shown in Scheme 4, ^{Me}P1 adducted with ammonium exhibits the same CID pattern compared to P1 (Figure S6 and Table S6, supporting information). In summary, both Na adduction and the carboxylate form of α are required to simplify MS/MS data of N–R PUs.

In the case of N–H PUs, the carboxylate form of α was promoted by H/Na exchange in solution (Figure S2C, supporting information), but requires the use of NaOH for optimal exchange.[17] While this approach also works for N–R PUs, the exchanged Na^+ does not always remain in the α carboxylate function, *i.e.*, in the form $[(P - H + Na) + 3Na]^{3+}$, but exhibits a certain mobility to generate, at least in the gas phase, the distonic species $[(P - H) + 4Na]^{3+}$ in which the deprotonated oligomer is adducted with four Na^+. This was evidenced for long chains

such as P5 with 28 units, which dissociation leads to formation of $w'_j{}^{3+}$ product ions carrying four sodium (Figure S7, supporting information). In contrast, esterification of the terminal carboxylic acid was observed to occur within a few hours of storage of N–R PUs in methanol and remains irreversible. This reaction was never observed for N–H PUs, even after months of storage, and would be favored by the increased basicity of the N atom in N–R PUs (due to inductive effect of the alkyl substituent) promoting nucleophilic attack of the α end-group by CH_3O^-. This particular reactivity offers a straightforward process to make the α termination "MS-MS silent" for best sequencing N–R PUs using $w'_j{}^{z+}$ product ions.

3.3. MS/MS sequencing of N-coded PUs

In order to illustrate the sequencing methodology established for N–R PUs, the 16-mer P4 is considered hereafter as an unknown and its sequence reconstructed based on CID data recorded for its ^{Me}P4 homologue adducted with two Na^+ (Figure 3). The first member of the series, $w'_1{}^{Na+}$, has its *m/z* value depending on the last unit, *i.e.*, *m/z* 162.1 if 00 (R=Me), *m/z* 176.1 if 01 (R=Et), *m/z* 190.1 if 10 (R=Pr), or *m/z* 204.1 if 11 (R=Bu). As can be seen in the inset of Figure 3, only *m/z* 176.1 was detected, indicating that the last coding unit of ^{Me}P4 is 01. Owing to the absence of any other ions in this MS/MS spectrum, measuring the peak-to-peak distance permits determination of the retro-sequence (reading from the right- to the left-hand side) of the polymer based on the mass of the four possible units: 101.0 Da (00), 115.1 Da (01), 129.1 Da (10) or 143.1 Da (11). Alternatively, the methodology adopted in the algorithm implemented in the MS-DECODER software[23] for automated sequencing consists of iterative searching of each member of the $w'_j{}^{Na+}$ series by adding to the previously detected congener the mass of one or the other four units (as illustrated in the top table of Figure 3). Amongst the four values obtained when starting from $w'_1{}^{Na+}$ at *m/z* 176.1 (containing 01), a signal is observed only at *m/z* 319.1 (= 176.1 + 143.1 → 11). This $w'_2{}^{Na+}$ *m/z* value becomes the new starting value for monomer mass addition, allowing $w'_3{}^{Na+}$ to be found at *m/z* 448.2 (= 319.1 + 129.1 → 10). Search for $w'_j{}^{Na+}$ ions is continued until no signal is detected for any predicted *m/z* value. This approach allows all coding units but the first one to be determined. Accordingly, the following partial retro-sequence is reconstructed for P4: 01·11·10·01·01·01·10·01·00·10·10·01·11·00·00. To determine the first coding unit linked to ^{Me}α and hence complete the sequence, one can subtract the mass of last $w'_j{}^{Na+}$ ion to that of the precursor ion, which corresponds to the mass of the ^{Me}α-XX moiety, *i.e.*, 230.1 Da, 244.2 Da, 258.2 Da or 272.2 Da depending on the first XX coding unit being 00, 01, 10 or 11, respectively. Applying such a calculation to data of Figure 3 indicates that the first coding unit of P4 is 01 ((1050.6 x 2) – 1857.0 = 244.2). The

same result can be found in a more straightforward manner from *m/z* values experimentally measured for I_1^+ and $^{Me}\alpha\text{-}I_1^+$ at 116.1 and 244.2, respectively. Accordingly, full coverage is achieved for the P4 sequence, determined to be 01·00·00·11·01·10·10·00·01·10·01·01·01·10·11·01. This sequencing task was successfully performed in 100 ms when running the MS-DECODER software on a computer with 4 cores with a spinning disk drive, using 2GB of memory.

	$w'_1{}^{Na+}$	$w'_2{}^{Na+}$	$w'_3{}^{Na+}$	$w'_4{}^{Na+}$	$w'_5{}^{Na+}$	$w'_6{}^{Na+}$	$w'_7{}^{Na+}$	$w'_8{}^{Na+}$	$w'_9{}^{Na+}$	$w'_{10}{}^{Na+}$	$w'_{11}{}^{Na+}$	$w'_{12}{}^{Na+}$	$w'_{13}{}^{Na+}$	$w'_{14}{}^{Na+}$	$w'_{15}{}^{Na+}$
00	~~162.1~~	~~277.1~~	~~420.2~~	~~549.3~~	~~664.4~~	~~779.4~~	~~894.5~~	~~1023.6~~	**1138.6**	~~1239.7~~	~~1368.7~~	~~1497.8~~	~~1612.9~~	**1756.0**	**1857.0**
01	**176.1**	~~291.1~~	~~434.2~~	**563.3**	**678.4**	**793.4**	~~908.5~~	**1037.6**	~~1152.6~~	~~1253.7~~	~~1382.8~~	**1511.8**	~~1626.9~~	~~1770.0~~	~~1871.0~~
10	~~190.1~~	~~305.1~~	**448.2**	~~577.3~~	~~692.4~~	~~807.4~~	**922.5**	~~1051.6~~	~~1166.7~~	**1267.7**	**1396.8**	~~1525.9~~	~~1640.9~~	~~1784.0~~	~~1885.1~~
11	~~204.1~~	**319.2**	~~462.3~~	~~591.3~~	~~706.4~~	~~821.5~~	~~936.5~~	~~1065.6~~	~~1180.7~~	~~1281.7~~	~~1410.8~~	~~1539.9~~	**1654.9**	~~1798.0~~	~~1899.1~~

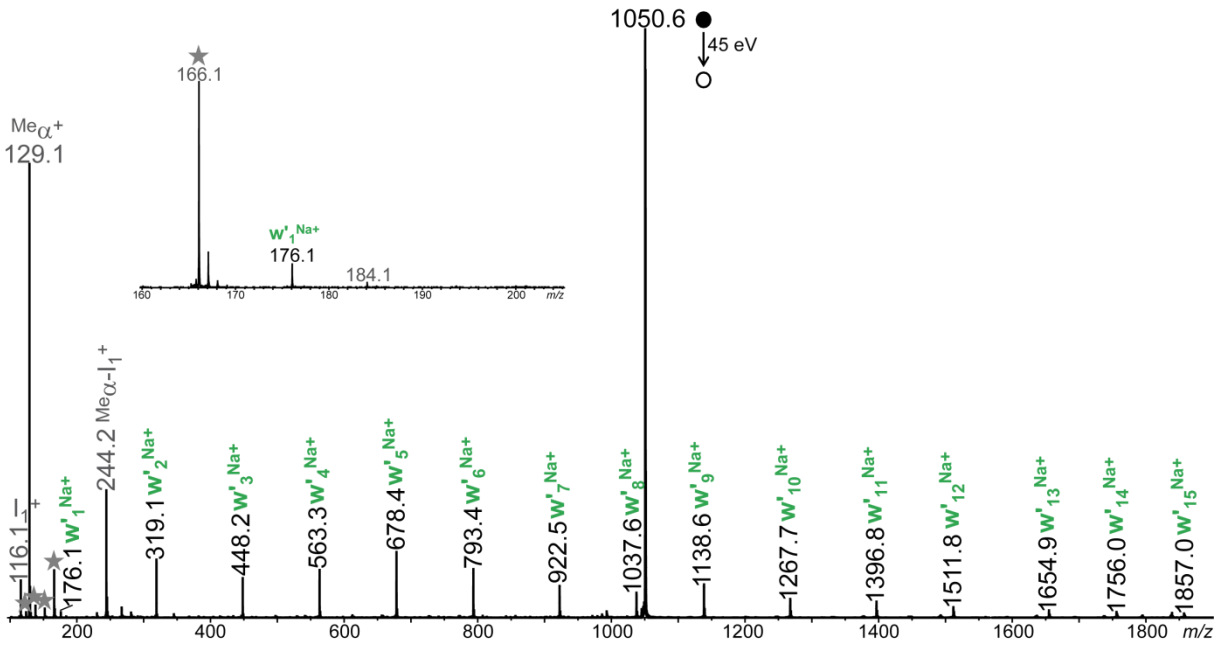


Figure 3. Positive mode ESI-MS/MS of $[^{Me}P4+2Na]^{2+}$ at *m/z* 1050.6 (collision energy in the laboratory frame). Inset: zoom on the 160–205 *m/z* range for identification of the $w'_1{}^{Na+}$ product ion. Top table: calculated *m/z* values for $w'_j{}^{Na+}$ ions by adding the mass of 00 (101.0 Da), 01 (115.1 Da), 10 (129.1 Da) or 11 (143.1 Da) to the *m/z* value of the detected previous congener (in green). Secondary product ions are in grey, including sodiated monomers designated by grey stars.

4. CONCLUSIONS

Unlike N–H PUs dissociating *via* O–(CO) bond cleavage in all carbamate groups, CID of N–R PUs proceeds by cleavage of the CH_2–O linkage between consecutive units, in both the negative and positive ion modes. The number of dissociating routes experienced by cationized N–R PUs

could be lowered to a single one when removing all sources of proton in the fragmenting precursor, *i.e.*, changing the original acidic α termination to a carboxylate moiety and avoiding H^+ or NH_4^+ as adducted charges. A similar effect was previously shown for N–H PUs but required the use of NaOH for efficient H/Na exchange in the α end-group. In contrast, methylation of α was found to occur spontaneously during storage of N–R PUs in methanol, allowing less harsh conditions to be used in ESI. When adducted to sodium, dissociation of methylated N–R PUs yields a single series of product ions which enables full coverage of their sequence. Combined with high ionization yield in positive mode ESI, these optimal MS/MS conditions have the possibility of reliable decoding of very large amount of information owing to the fact that each unit of N–R PUs actually codes for a dyad.

ACKNOWLEDGMENTS

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). This work was supported by CNRS, the University of Strasbourg and the LabEx CSC.

REFERENCES

1. Lutz J-F, Ouchi M, Liu DR, Sawamoto M. Sequence-Controlled Polymers. *Science.* 2013; 341(6146):628-+. https://doi.org/10.1126/science.1238149
2. Colquhoun H, Lutz J-F. Information-containing macromolecules. *Nat Chem.* 2014;6(6):455-456. https://doi.org/10.1038/nchem.1958
3. Lutz J-F. Coding Macromolecules: Inputting Information in Polymers Using Monomer-Based Alphabets. *Macromolecules.* 2015;48(14):4759-4767. https://doi.org/10.1021/acs.macromol.5b00890
4. Roy RK, 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;6:7287. https://doi.org/10.1038/ncomms8237
5. Amalian J-A, Trinh TT, Lutz J-F, Charles L. MS/MS digital readout: analysis of binary information encoded in the monomer sequences of poly(triazole amide)s. *Anal Chem.* 2016;88(7):3715-3722. https://doi.org/10.1021/acs.analchem.5b04537
6. Gunay US, Petit BE, Karamessini D, et al. Chemoselective synthesis of uniform sequence-coded polyurethanes and their use as molecular tags. *Chem.* 2016;1(1):114-126. https://doi.org/10.1016/j.chempr.2016.06.006

7. 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(30): 9417-9420. https://doi.org/10.1021/jacs.6b06222
8. Boukis AC, Meier MAR. Data storage in sequence-defined macromolecules via multicomponent reactions. *Eur Polym J.* 2018;104:32-38. https://doi.org/10.1016/j.eurpolymj.2018.04.038
9. 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. https://doi.org/10.1038/s41467-018-06926-3
10. Ding KS, Zhang YJ, Huang ZH, et al. Easily encodable/decodable digital polymers linked by dithiosuccinimide motif. *Eur Polym J.* 2019;119:421-425. https://doi.org/10.1016/j.eurpolymj.2019.08.017
11. Huang ZH, Shi QN, Guo J, et al. Binary tree-inspired digital dendrimer. *Nat Commun.* 2019;10:1918. https://doi.org/10.1038/s41467-019-09957-6
12. Lee JM, Koo MB, Lee SW, et al. High-density information storage in an absolutely defined aperiodic sequence of monodisperse copolyester. *Nat Commun.* 2020;11:56. https://doi.org/10.1038/s41467-019-13952-2
13. Charles L, Cavallo G, Monnier V, Oswald L, Szweda R, Lutz J-F. MS/MS-assisted design of sequence-controlled synthetic polymers for improved reading of encoded information. *J Am Soc Mass Spectrom.* 2017;28(6):1149-1159. https://doi.org/10.1007/s13361-016-1543-5
14. Amalian J-A, Poyer S, Petit BE, et al. Negative mode MS/MS to read digital information encoded in sequence-defined oligo(urethane)s: A mechanistic study. *Int J Mass Spectrom.* 2017;421:271-278. https://doi.org/10.1016/j.ijms.2017.07.006
15. Amalian J-A, Al Ouahabi A, Cavallo G, et al. Controlling the structure of sequence-defined poly(phosphodiester)s for optimal MS/MS reading of digital information. *J Mass Spectrom.* 2017;52(11):788-798. https://doi.org/10.1002/jms.3947
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. https://doi.org/10.1039/c9py01939a
17. Poyer S, Petit BE, Telitel S, Karamessini D, Lutz J-F, Charles L. Promoting carboxylate salts in the ESI source to simplify positive mode MS/MS sequencing of acid-terminated encoded polyurethanes. *Int J Mass Spectrom.* 2020;448:116271. https://doi.org/10.1016/j.ijms.2019.116271

18. Montaudo G, Puglisi C, Scamporrino E, Vitalini D. Mechanism of thermal-degradation of polyurethanes - Effect of ammonium phosphate. *Macromolecules.* 1984;17(8):1605-1614. https://doi.org/10.1021/ma00138a032
19. Petit BE, Lotz B, Lutz J-F. About the Crystallization of Abiotic Coded Matter. *ACS Macro Lett.* 2019;8(7):779-782. https://doi.org/10.1021/acsmacrolett.9b00307
20. Mondal T, Greff V, Petit BE, Charles L, Lutz J-F. Efficient Protocol for the Synthesis of "N-Coded" Oligo- and Poly(N-Substituted Urethanes). *ACS Macro Lett.* 2019;8(8):1002-1005. https://doi.org/10.1021/acsmacrolett.9b00446
21. Karamessini D, Poyer S, Charles L, Lutz J-F. 2D Sequence-Coded Oligourethane Barcodes for Plastic Materials Labeling. *Macromol Rapid Commun.* 2017;38(24):1700426. https://doi.org/10.1002/marc.201700426
22. Charles L. Influence of internal standard charge state on the accuracy of mass measurements in orthogonal acceleration time-of-flight mass spectrometers. *Rapid Commun Mass Spectrom.* 2008.22(2)/151-155. https://doi.org/10.1002/rcm.3347
23. Burel A, Carapito C, Lutz J-F, Charles L. MS-DECODER: Milliseconds Sequencing of Coded Polymers. *Macromolecules.* 2017;50(20):8290-8296. https://doi.org/10.1021/acs.macromol.7b01737
24. Gies AP, Hercules DM. Collision induced dissociation study of ester-based polyurethane fragmentation reactions. *Anal Chim Acta.* 2014;808:199-219. https://doi.org/10.1016/j.aca.2013.09.035
25. Wesdemiotis C, Solak N, Polce MJ, Dabney DE, Chaicharoen K, Katzenmeyer BC. Fragmentation pathways of polymer ions. *Mass Spectrom Rev.* 2011;30(4):523-559. https://doi.org/10.1002/mas.20282
26. Hsu FF, Turk J. Distinction among isomeric unsaturated fatty acids as lithiated adducts by electrospray ionization mass spectrometry using low energy collisionally activated dissociation on a triple stage quadrupole instrument. *J Am Soc Mass Spectrom.* 1999;10(7):600-612. https://doi.org/10.1016/S1044-0305(99)00041-0
27. Zehethofer N, Pinto DM, Volmer DA. Plasma free fatty acid profiling in a fish oil human intervention study using ultra-performance liquid chromatography/electrospray ionization tandem mass spectrometry. *Rapid Commun Mass Spectrom.* 2008;22(13):2125-2133. https://doi.org/10.1002/rcm.3597

Supporting information

Optimal conditions for MS/MS sequencing of information-containing N-substituted polyurethanes

Laurence Charles, Tathagata Mondal, Vincent Greff, Mattia Razzini, Valérie Monnier, Alexandre Burel, Christine Carapito, and Jean-François Lutz

name (DP)	sequence	composition	mass (Da)*
P1 (8)	00·01·10·11·00·01·10·11	$C_{51}H_{94}N_8O_{18}$	1106.6686
P2 (8)	01·00·11·10·01·10·10·01	$C_{51}H_{94}N_8O_{18}$	1106.6686
P3 (8)	11·00·01·11·01·10·00·01	$C_{50}H_{92}N_8O_{18}$	1092.6530
P4 (16)	01·00·00·11·01·10·10·00·01·10·01·01·01·10·11·01	$C_{92}H_{168}N_{16}O_{34}$	2041.1909
P5 (28)	01·01·00·00·01·10·11·11·01·10·11·00·01·11 ·10·01·01·10·11·01·01·10·01·01·01·11·00·10	$C_{161}H_{294}N_{28}O_{58}$	3548.0917

Table S1. Description of N–R PUs used in this study. *monoisotopic mass value.

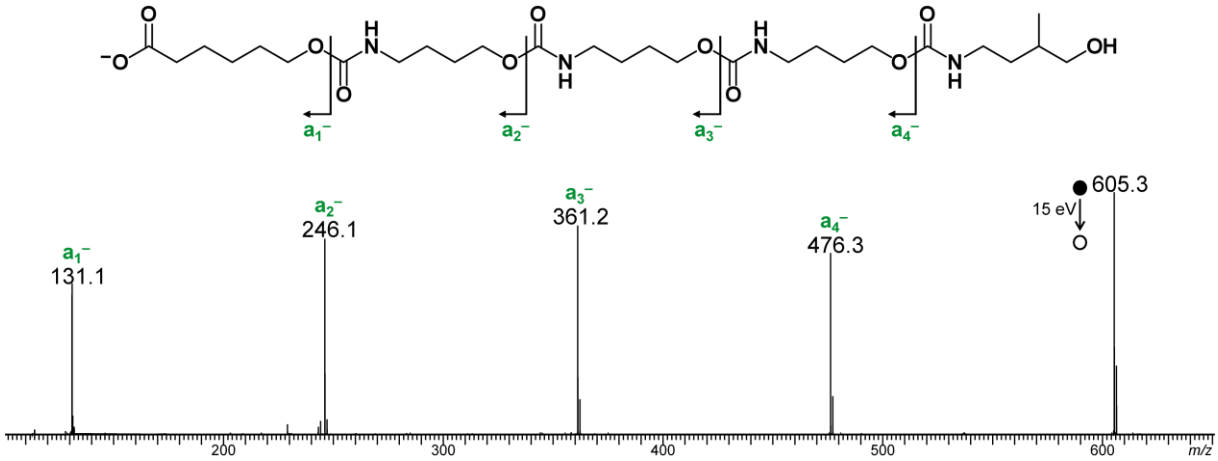
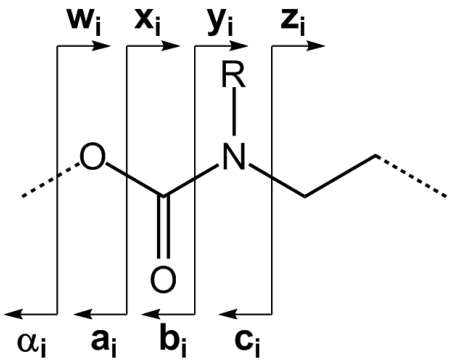


Figure S1. Negative mode ESI-MS/MS of the deprotonated N–H PU 4-mer of sequence 0·0·0·1, with dissociation scheme on top for ion assignment. Collision energy is given in the laboratory frame.



Scheme S1. Nomenclature used for product ions of N-PUs, adapted from Wesdemiotis *et al.*[1] and focused on the cleavable bonds of the carbamate moiety.

1. Wesdemiotis, C.; Solak, N.; Polce, M. J.; Dabney, D. E.; Chaicharoen, K.; Katzenmeyer, B. C., Fragmentation pathways of polymer ions. *Mass Spectrometry Reviews* **2011,** *30* (4), 523-559.

$[P1 - H]^-$: $C_{51}H_{93}N_8O_{18}^-$, *m/z* 1105.6613 (I.S.)							
α-containing product ions				ω-containing product ions			
	composition	m/z_{th}	m/z_{exp}		composition	m/z_{th}	m/z_{exp}
α_1^-	$C_{10}H_{16}NO_4^-$	214.1085	214.1088	w_1^-	$C_8H_{16}NO_2^-$	158.1187	158.1184
α_2^-	$C_{15}H_{25}N_2O_6^-$	329.1718	329.1709	w_2^-	$C_{14}H_{27}N_2O_4^-$	287.1976	287.1969
α_3^-	$C_{21}H_{36}N_3O_8^-$	458.2508	458.2480	w_3^-	$C_{19}H_{36}N_3O_6^-$	402.2610	402.2586
α_4^-	$C_{28}H_{49}N_4O_{10}^-$	601.3454	601.3441	w_4^-	$C_{23}H_{43}N_4O_8^-$	503.3086	503.3052
α_5^-	$C_{32}H_{56}N_5O_{12}^-$	702.3931	702.3927	w_5^-	$C_{30}H_{56}N_5O_{10}^-$	646.4033	646.3992
α_6^-	$C_{37}H_{65}N_6O_{14}^-$	817.4564	817.4525	w_6^-	$C_{36}H_{67}N_6O_{12}^-$	775.4822	775.4790
α_7^-	$C_{43}H_{76}N_7O_{16}^-$	946.5354	946.5356	w_7^-	$C_{41}H_{76}N_7O_{14}^-$	890.5456	n.d.

Table S2. Accurate measurements of product ions generated upon CID of $[P1 - H]^-$ (Figure 1), of sequence 00·01·10·11·00·01·10·11 (Figure 1), with 00: $C_4H_7NO_2$ (101.0477), 01: $C_5H_9NO_2$ (115.0633), 10: $C_6H_{11}NO_2$ (129.0790), and 11: $C_7H_{13}NO_2$ (143.0946). n.d.: not detected. I.S.: used as an internal standard for calibration.

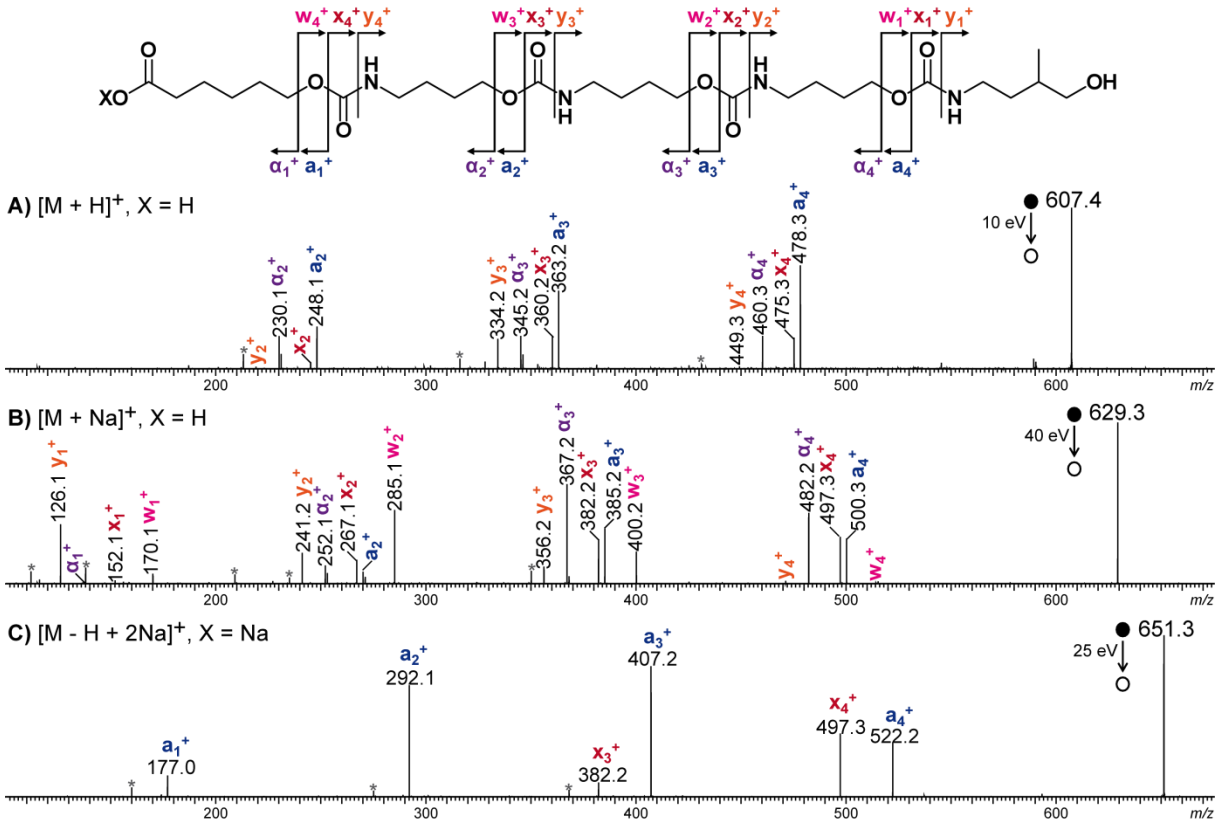


Figure S2. Positive mode ESI-MS/MS spectra of the N–H PU 4-mer (0·0·0·1) when activated as A) $[M + H]^+$ at *m/z* 607.4 (X=H), B) $[M + Na]^+$ at *m/z* 629.3 (X=H) and C) $[M - H + 2Na]^+$ at *m/z* 651.3 (X=Na), with dissociation scheme on top for ion assignment. Secondary product ions are designated by asterisks. Collision energy is given in the laboratory frame.

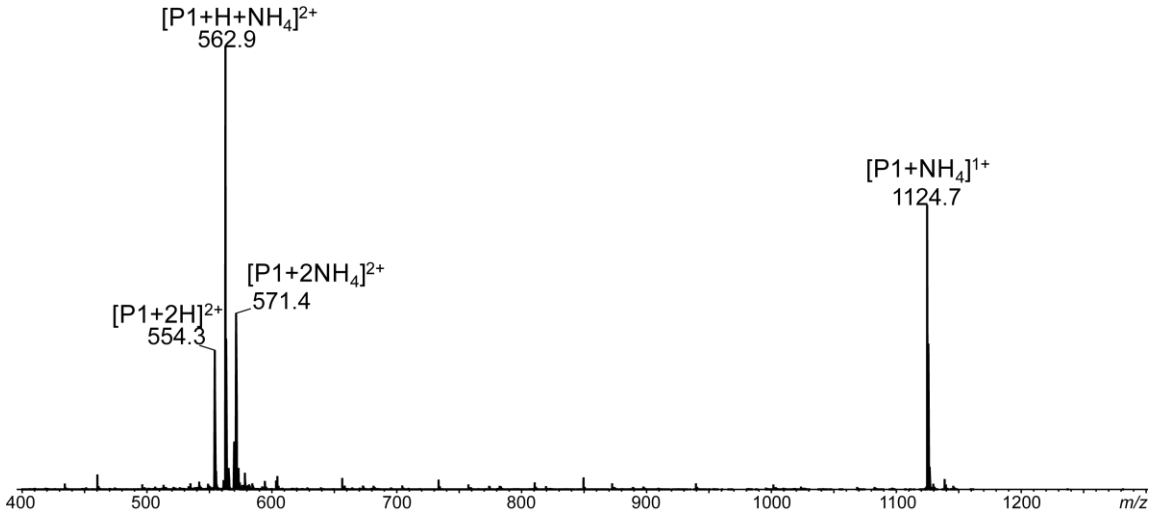


Figure S3. Positive mode ESI mass spectrum of the N–R PU P1 of sequence 01·00·11·11·01·10·11·10 in a methanolic solution containing ammonium acetate (3 mM).

[P1 + 2NH4]$^{2+}$: $C_{51}H_{102}N_{10}O_{18}^{2+}$, *m/z* 571.3681 (I.S.)							
α-containing product ions				internal product ions			
	composition	*m/z*$_{th}$	*m/z*$_{exp}$		composition	*m/z*$_{th}$	*m/z*$_{exp}$
α_1^{H+}	$C_{10}H_{18}NO_4^+$	216.1230	216.1233	00^{H+}	$C_4H_8NO_2^+$	102.0550	102.0572
α_2^{H+}	$C_{15}H_{27}N_2O_6^+$	331.1864	331.1858	01^{H+}	$C_5H_{10}NO_2^+$	116.0706	116.0731
α_3^{H+}	$C_{21}H_{38}N_3O_8^+$	460.2653	460.2624	10^{H+}	$C_6H_{12}NO_2^+$	130.0863	130.0887
α_4^{H+}	$C_{28}H_{51}N_4O_{10}^+$	603.3600	603.3574	11^{H+}	$C_7H_{14}NO_2^+$	144.1019	144.1047
α_5^{H+}	$C_{32}H_{58}N_5O_{12}^+$	704.4076	704.4040	α^+	$C_6H_{11}O_2^+$	115.0754	115.0776
α_6^{H+}	$C_{37}H_{67}N_6O_{14}^+$	819.4710	819.4729				
α_7^{H+}	$C_{43}H_{78}N_7O_{16}^+$	948.5500	n.d.				
ω-containing product ions							
	composition	*m/z*$_{th}$	*m/z*$_{exp}$		composition	*m/z*$_{th}$	*m/z*$_{exp}$
w_1^{H+}	$C_8H_{18}NO_2^+$	160.1332	n.d.	y_1^{H+}	$C_7H_{18}N^+$	116.1434	n.d.
w_2^{H+}	$C_{14}H_{29}N_2O_4^+$	289.2122	289.2122	y_2^{H+}	$C_{13}H_{29}N_2O_2^+$	245.2224	245.2218
w_3^{H+}	$C_{19}H_{38}N_3O_6^+$	404.2755	404.2753	y_3^{H+}	$C_{18}H_{38}N_3O_4^+$	360.2857	360.2840
w_4^{H+}	$C_{23}H_{45}N_4O_8^+$	505.3232	505.3189	y_4^{H+}	$C_{22}H_{45}N_4O_6^+$	461.3334	*
w_5^{H+}	$C_{30}H_{58}N_5O_{10}^+$	648.4178	648.4142	y_5^{H+}	$C_{29}H_{58}N_5O_8^+$	604.4280	*
w_6^{H+}	$C_{36}H_{69}N_6O_{12}^+$	777.4968	n.d.	y_6^{H+}	$C_{35}H_{69}N_6O_{10}^+$	733.5070	733.5042
w_7^{H+}	$C_{41}H_{78}N_7O_{14}^+$	892.5601	n.d.	y_7^{H+}	$C_{40}H_{78}N_7O_{12}^+$	848.5703	848.5708

Table S3. Accurate mass measurement of product ions generated upon CID of $[P1+2NH_4]^{2+}$ (Fig 2A). Elemental compositions show that all product ions contain a number of N atoms equal to their polymerization degree, hence indicating that they are not ammonium adducts. Structure of P1 is α-00·01·10·11·00·01·10·11-CH_3, with α: $C_6H_{11}O_2$ (115.0759 Da), 00: $C_4H_7NO_2$ (101.0477), 01: $C_5H_9NO_2$ (115.0633), 10: $C_6H_{11}NO_2$ (129.0790), and 11: $C_7H_{13}NO_2$ (143.0946). *: these two ions could not be accurately mass measured due to interferences of ^{13}C from α_3^{H+} or α_4^{H+}. n.d.: not detected. I.S.: used as an internal standard for calibration.

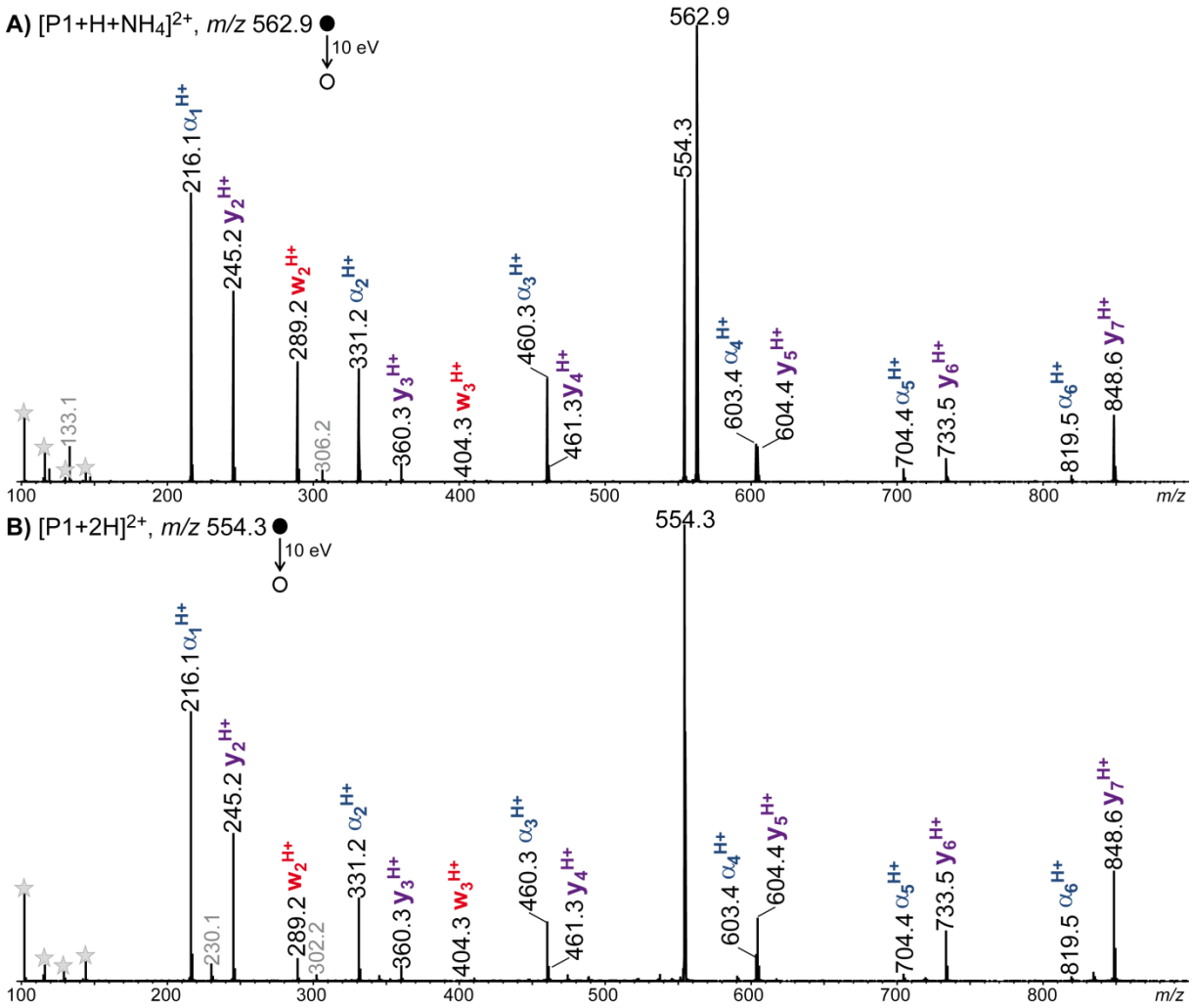
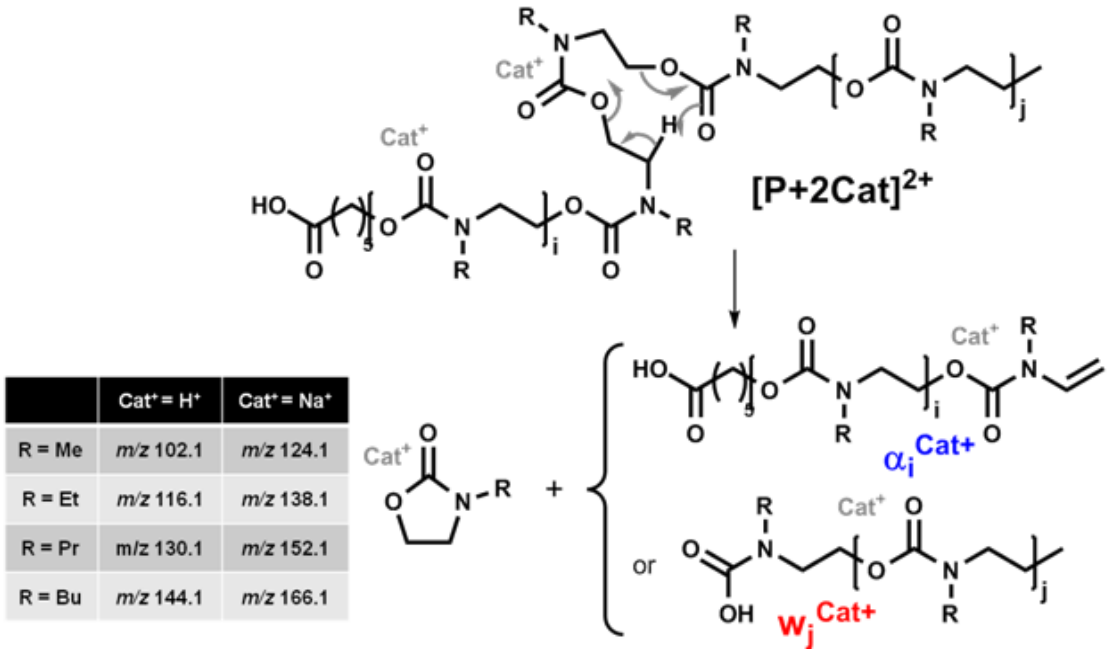


Figure S4. CID spectra of A) $[P1+H+NH_4]^{2+}$ at *m/z* 562.9 and B) $[P1+2H]^{2+}$ at *m/z* 554.3 show the same product ions compared to their $[P1+2NH_4]^{2+}$ precursor at *m/z* 571.4 (Fig 2A).



	Cat⁺ = H⁺	Cat⁺ = Na⁺
R = Me	*m/z* 102.1	*m/z* 124.1
R = Et	*m/z* 116.1	*m/z* 138.1
R = Pr	m/z 130.1	*m/z* 152.1
R = Bu	*m/z* 144.1	*m/z* 166.1

Scheme S2. Formation of cationized monomers (Cat = H or Na) as internal product ions, complementarily to α_i^{Cat+} or w_j^{Cat+} ions from doubly charged N–R PUs.

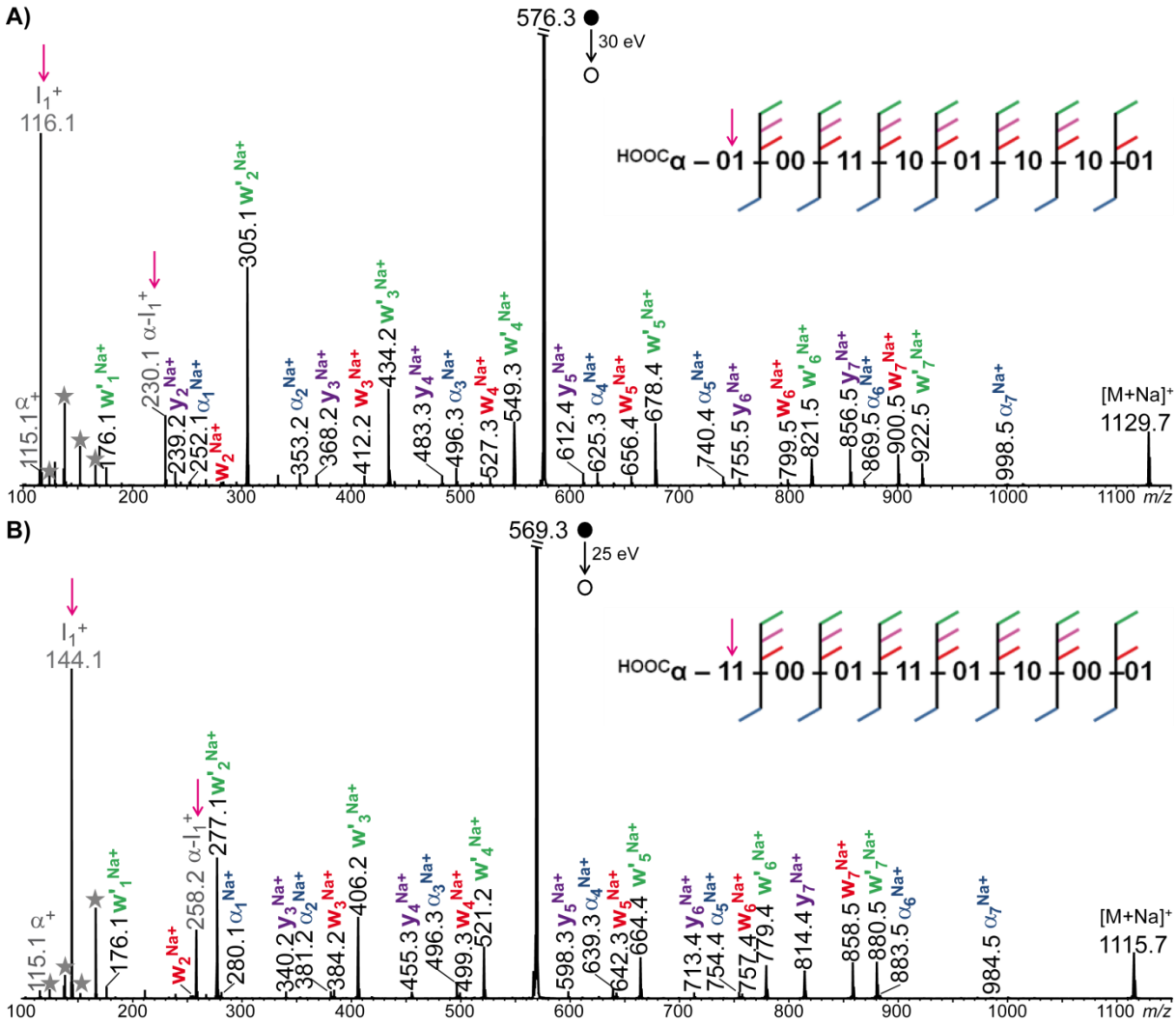


Figure S5. Influence of the first unit on *m/z* values of I_1^+ and α-I_1^+ generated during CID of doubly sodiated oligomers. The I_1^+ and α-I_1^+ internal product ions are respectively expected at A) *m/z* 116.1 and *m/z* 230.1 for chains starting with unit 01 (115.1 Da), as in P2 of sequence 01·00·11·10·01·10·10·01, or B) *m/z* 144.1 and *m/z* 258.2 for chains starting with unit 11 (143.1 Da), as in P3 of sequence 11·00·01·11·01·10·00·01. The same 2.8 magnification factor of the y-axis was applied to the two MS/MS spectra.

[P1 + 2Na]$^{2+}$: $C_{51}H_{94}N_8O_{18}Na_2^{2+}$, *m/z* 576.3235 (I.S.)							
α-containing product ions				internal product ions			
	composition	m/z_{th}	m/z_{exp}		composition	m/z_{th}	m/z_{exp}
α_1^{Na+}	$C_{10}H_{17}NO_4Na^+$	238.1050	238.1030	00^{Na+}	$C_4H_7NO_2Na^+$	124.0369	124.0362
α_2^{Na+}	$C_{15}H_{26}N_2O_6Na^+$	353.1683	353.1669	01^{Na+}	$C_5H_9NO_2Na^+$	138.0525	138.0521
α_3^{Na+}	$C_{21}H_{37}N_3O_8Na^+$	482.2473	482.2457	10^{Na+}	$C_6H_{11}NO_2Na^+$	152.0682	152.0673
α_4^{Na+}	$C_{28}H_{50}N_4O_{10}Na^+$	625.3419	625.3402	11^{Na+}	$C_7H_{13}NO_2Na^+$	166.0839	166.0830
α_5^{Na+}	$C_{32}H_{57}N_5O_{12}Na^+$	726.3896	726.3895	I_1^+	$C_4H_8NO_2^+$	102.0550	102.0544
α_6^{Na+}	$C_{37}H_{66}N_6O_{14}Na^+$	841.4529	841.4527	α^+	$C_6H_{11}O_2^+$	115.0754	115.0750
α_7^{Na+}	$C_{43}H_{77}N_7O_{16}Na^+$	970.5319	970.5319	$\alpha\text{-}I_1^+$	$C_{10}H_{18}NO_4^+$	216.1230	216.1226
ω-containing product ions							
	composition	m/z_{th}	m/z_{exp}		composition	m/z_{th}	m/z_{exp}
w_1^{Na+}	$C_8H_{17}NO_2Na^+$	182.1152	182.1150	$w'_1{}^{Na+}$	$C_8H_{16}NO_2Na_2^+$	204.0971	204.0943
w_2^{Na+}	$C_{14}H_{28}N_2O_4Na^+$	311.1941	311.1916	$w'_2{}^{Na+}$	$C_{14}H_{27}N_2O_4Na_2^+$	333.1761	333.1742
w_3^{Na+}	$C_{19}H_{37}N_3O_6Na^+$	426.2575	426.2555	$w'_3{}^{Na+}$	$C_{19}H_{36}N_3O_6Na_2^+$	448.2394	448.2374
w_4^{Na+}	$C_{23}H_{44}N_4O_8Na^+$	527.3051	527.3035	$w'_4{}^{Na+}$	$C_{23}H_{43}N_4O_8Na_2^+$	549.2871	549.2843
w_5^{Na+}	$C_{30}H_{57}N_5O_{10}Na^+$	670.3998	670.3978	$w'_5{}^{Na+}$	$C_{30}H_{56}N_5O_{10}Na_2^+$	692.3817	692.3801
w_6^{Na+}	$C_{36}H_{68}N_6O_{12}Na^+$	799.4787	799.4762	$w'_6{}^{Na+}$	$C_{36}H_{67}N_6O_{12}Na_2^+$	821.4607	821.4585
w_7^{Na+}	$C_{41}H_{77}N_7O_{14}Na^+$	914.5421	914.5410	$w'_7{}^{Na+}$	$C_{41}H_{76}N_7O_{14}Na_2^+$	936.5240	936.5205
y_1^{Na+}	$C_7H_{17}NNa^+$	138.1253	n.d.				
y_2^{Na+}	$C_{13}H_{28}N_2O_2Na^+$	267.2043	267.2022				
y_3^{Na+}	$C_{18}H_{37}N_3O_4Na^+$	382.2676	382.2654				
y_4^{Na+}	$C_{22}H_{44}N_4O_6Na^+$	483.3153	483.3125				
y_5^{Na+}	$C_{29}H_{57}N_5O_8Na^+$	626.4099	626.4038				
y_6^{Na+}	$C_{35}H_{68}N_6O_{10}Na^+$	755.4889	755.4864				
y_7^{Na+}	$C_{40}H_{77}N_7O_{12}Na^+$	870.5522	870.5501				

Table S4. Accurate mass measurement of product ions generated upon CID of $[P1+2Na]^{2+}$ (Fig 2B). Elemental compositions determined for $w'_j{}^{Na+}$ product ions clearly show that these singly charged species contain two Na atoms. Structure of P1 is α-00·01·10·11·00·01·10·11-CH_3, with α: $C_6H_{11}O_2$ (115.0759 Da), 00: $C_4H_7NO_2$ (101.0477), 01: $C_5H_9NO_2$ (115.0633), 10: $C_6H_{11}NO_2$ (129.0790), and 11: $C_7H_{13}NO_2$ (143.0946). n.d.: not detected. I.S.: used as an internal standard for calibration.

$[^{Me}P1 + 2Na]^{2+}$: $C_{52}H_{96}N_8O_{18}Na_2^{2+}$, *m/z* 583.3314 (I.S.)							
ω-containing product ions				other minor product ions			
	composition	m/z_{th}	m/z_{exp}		composition	m/z_{th}	m/z_{exp}
$w'_1{}^{Na+}$	$C_8H_{16}NO_2Na_2^+$	204.0971	204.0975	$\alpha_1{}^{Na+}$	$C_{11}H_{19}NO_4Na^+$	252.1206	252.1206
$w'_2{}^{Na+}$	$C_{14}H_{27}N_2O_4Na_2^+$	333.1761	333.1756	$y_2{}^{Na+}$	$C_{13}H_{28}N_2O_2Na^+$	267.2043	267.2049
$w'_3{}^{Na+}$	$C_{19}H_{36}N_3O_6Na_2^+$	448.2394	448.2381				
$w'_4{}^{Na+}$	$C_{23}H_{43}N_4O_8Na_2^+$	549.2871	549.2846				
$w'_5{}^{Na+}$	$C_{30}H_{56}N_5O_{10}Na_2^+$	692.3817	692.3792				
$w'_6{}^{Na+}$	$C_{36}H_{67}N_6O_{12}Na_2^+$	821.4607	821.4554				
$w'_7{}^{Na+}$	$C_{41}H_{76}N_7O_{14}Na_2^+$	936.5240	936.5186				
internal product ions							
	composition	m/z_{th}	m/z_{exp}		composition	m/z_{th}	m/z_{exp}
00^{Na+}	$C_4H_7NO_2Na^+$	124.0369	124.0382	$I_1{}^+$	$C_4H_8NO_2^+$	102.0550	102.0563
01^{Na+}	$C_5H_9NO_2Na^+$	138.0525	138.0533	α^+	$C_7H_{13}O_2^+$	129.0910	129.0925
10^{Na+}	$C_6H_{11}NO_2Na^+$	152.0682	152.0687	α-$I_1{}^+$	$C_{11}H_{20}NO_4^+$	230.1387	230.1391
11^{Na+}	$C_7H_{13}NO_2Na^+$	166.0839	166.0845				

Table S5. Accurate mass measurement of product ions generated upon CID of $[^{Me}P1+2Na]^{2+}$ (Fig 2C). Structure of $^{Me}P1$ is $^{Me}\alpha$-00·01·10·11·00·01·10·11-CH_3, with $^{Me}\alpha$: $C_7H_{13}O_2$ (129.0916 Da), 00: $C_4H_7NO_2$ (101.0477), 01: $C_5H_9NO_2$ (115.0633), 10: $C_6H_{11}NO_2$ (129.0790), and 11: $C_7H_{13}NO_2$ (143.0946). I.S.: used as an internal standard for calibration

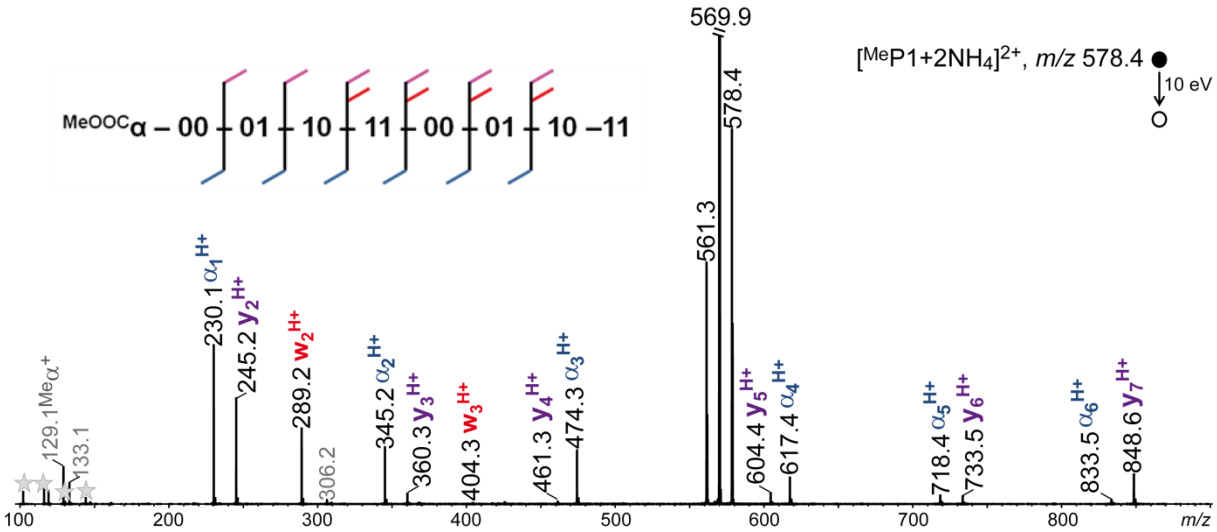
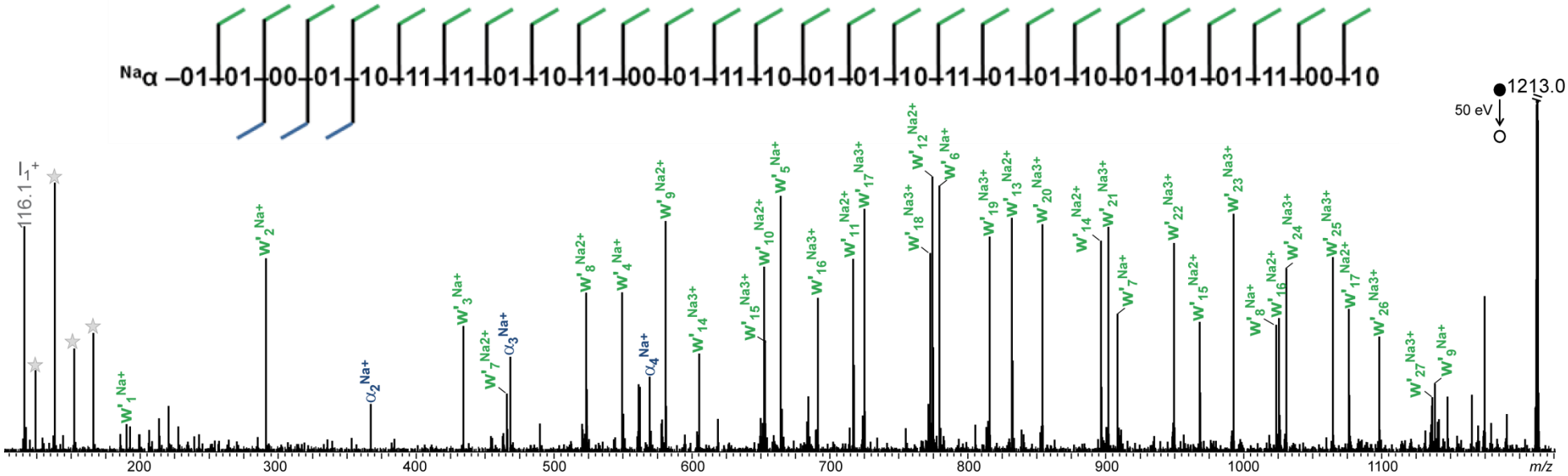


Figure S6. ESI-MS/MS of $[^{Me}P1+2NH_4]^{2+}$ at *m/z* 578.4, with product ion assignments in inset. Secondary product ions are in grey, including protonated monomers designated by stars. Collision energy is given in the laboratory frame. Magnification factor of the y-axis is 3.5.

$[^{Me}P1 + 2NH_4]^{2+}$: $C_{52}H_{104}N_{10}O_{18}^{2+}$, *m/z* 578.3760 (I.S.)							
α-containing product ions				internal product ions			
	composition	m/z_{th}	m/z_{exp}		composition	m/z_{th}	m/z_{exp}
α_1^{H+}	$C_{11}H_{20}NO_4^+$	230.1387	230.1398	00^{H+}	$C_4H_8NO_2^+$	102.0550	102.0558
α_2^{H+}	$C_{16}H_{29}N_2O_6^+$	345.2020	345.2035	01^{H+}	$C_5H_{10}NO_2^+$	116.0706	116.0713
α_3^{H+}	$C_{22}H_{40}N_3O_8^+$	474.2810	474.2817	10^{H+}	$C_6H_{12}NO_2^+$	130.0863	130.0900
α_4^{H+}	$C_{29}H_{53}N_4O_{10}^+$	617.3756	617.3753	11^{H+}	$C_7H_{14}NO_2^+$	144.1019	144.1027
α_5^{H+}	$C_{33}H_{60}N_5O_{12}^+$	718.4233	718.4271	$^{Me}\alpha^+$	$C_7H_{13}O_2^+$	129.0910	129.0921
α_6^{H+}	$C_{38}H_{69}N_6O_{14}^+$	833.4866	833.4919				
α_7^{H+}	$C_{44}H_{80}N_7O_{16}^+$	962.5656	n.d.				
ω-containing product ions							
	composition	m/z_{th}	m/z_{exp}		composition	m/z_{th}	m/z_{exp}
w_1^{H+}	$C_8H_{18}NO_2^+$	160.1332	n.d.	y_1^{H+}	$C_7H_{18}N^+$	116.1434	n.d.
w_2^{H+}	$C_{14}H_{29}N_2O_4^+$	289.2122	289.2136	y_2^{H+}	$C_{13}H_{29}N_2O_2^+$	245.2224	245.2229
w_3^{H+}	$C_{19}H_{38}N_3O_6^+$	404.2755	404.2780	y_3^{H+}	$C_{18}H_{38}N_3O_4^+$	360.2857	360.2877
w_4^{H+}	$C_{23}H_{45}N_4O_8^+$	505.3232	505.3255	y_4^{H+}	$C_{22}H_{45}N_4O_6^+$	461.3334	461.3353
w_5^{H+}	$C_{30}H_{58}N_5O_{10}^+$	648.4178	648.4237	y_5^{H+}	$C_{29}H_{58}N_5O_8^+$	604.4280	604.4276
w_6^{H+}	$C_{36}H_{69}N_6O_{12}^+$	777.4968	n.d.	y_6^{H+}	$C_{35}H_{69}N_6O_{10}^+$	733.5070	733.5120
w_7^{H+}	$C_{41}H_{78}N_7O_{14}^+$	892.5601	n.d.	y_7^{H+}	$C_{40}H_{78}N_7O_{12}^+$	848.5703	848.5715

Table S6. Accurate mass measurement of product ions generated upon CID of $[^{Me}P1+2NH_4]^{2+}$ (Fig S6). Structure of $^{Me}P1$ is $^{Me}\alpha$-00·01·10·11·00·01·10·11-CH_3, with $^{Me}\alpha$: $C_7H_{13}O_2$ (129.0916 Da), 00: $C_4H_7NO_2$ (101.0477), 01: $C_5H_9NO_2$ (115.0633), 10: $C_6H_{11}NO_2$ (129.0790), and 11: $C_7H_{13}NO_2$ (143.0946). n.d.: not detected. I.S.: used as an internal standard for calibration.



j	↓	composition	m/z_{th}	m/z_{exp}	j	↓	composition	m/z_{th}	m/z_{exp}	j	↓	composition	m/z_{th}	m/z_{exp}
1	**10**	$C_7H_{14}NO_2Na_2^+$	190.0814	190.0780	10	**11**	$C_{56}H_{103}N_{10}O_{20}Na_3^{2+}$	652.3516	652.3465	19	**10**	$C_{107}H_{196}N_{19}O_{38}Na_4^{3+}$	815.7854	815.7786
2	**00**	$C_{11}H_{21}N_2O_4Na_2^+$	291.1291	291.1264	11	**10**	$C_{62}H_{114}N_{11}O_{22}Na_3^{2+}$	716.8911	716.8851	20	**01**	$C_{112}H_{205}N_{20}O_{40}Na_4^{3+}$	854.1399	854.1395
3	**11**	$C_{18}H_{34}N_3O_6Na_2^+$	434.2238	434.2232	12	**01**	$C_{67}H_{123}N_{12}O_{24}Na_3^{2+}$	774.4228	774.4199	21	**11**	$C_{119}H_{218}N_{21}O_{42}Na_4^{3+}$	901.8381	901.8340
4	**01**	$C_{23}H_{43}N_4O_8Na_2^+$	549.2871	549.2841	13	**01**	$C_{72}H_{132}N_{13}O_{26}Na_3^{2+}$	831.9544	831.9502	22	**11**	$C_{126}H_{231}N_{22}O_{44}Na_4^{3+}$	949.5363	949.5312
5	**01**	$C_{28}H_{52}N_5O_{10}Na_2^+$	664.3504	664.3478	14	**10**	$C_{78}H_{143}N_{14}O_{28}Na_3^{2+}$	896.4939	896.4903	23	**10**	$C_{132}H_{242}N_{23}O_{46}Na_4^{3+}$	992.5626	992.5581
6	**01**	$C_{33}H_{61}N_6O_{12}Na_2^+$	779.4137	779.4075	15	**11**	$C_{85}H_{156}N_{15}O_{30}Na_4^{3+}$	653.0239	653.0212	24	**01**	$C_{137}H_{251}N_{24}O_{48}Na_4^{3+}$	1030.9171	1030.9155
7	**10**	$C_{39}H_{72}N_7O_{14}Na_2^+$	908.4927	908.4865	16	**01**	$C_{90}H_{165}N_{16}O_{32}Na_4^{3+}$	691.3783	691.3740	25	**00**	$C_{141}H_{258}N_{25}O_{50}Na_4^{3+}$	1064.5996	1064.5981
8	**01**	$C_{44}H_{81}N_8O_{16}Na_3^{2+}$	523.2726	523.2690	17	**00**	$C_{94}H_{172}N_{17}O_{34}Na_4^{3+}$	725.0609	725.0578	26	**00**	$C_{145}H_{265}N_{26}O_{52}Na_4^{3+}$	1098.2822	1098.2831
9	**01**	$C_{49}H_{90}N_9O_{18}Na_3^{2+}$	580.8043	580.8018	18	**11**	$C_{101}H_{185}N_{18}O_{36}Na_4^{3+}$	772.7591	772.7531	27	**01**	$C_{150}H_{274}N_{27}O_{54}Na_4^{3+}$	1136.6366	1136.6373

Figure S7. Top: MS/MS spectrum of $[P5 - H + 4Na]^{3+}$ at *m/z* 1213.0 (collision energy in the laboratory frame; magnification factor of the y-axis: 9). Inset: coverage of the **01·01·00·00·01·10·11·11·01·10·11·00·01·11·10·01·01·10·11·01·01·10·01·01·01·11·00·10** sequence of P5 with $w'_j{}^{Na(z+)}$ ions. Bottom: accurate mass measurements of $w'_j{}^{Na(z+)}$ ions (z = 1-3) using the precursor ion (*m/z* 1213.0138) as an internal standard.