RESEARCH ARTICLE



Sequence distribution determination by SWAMP-MS a systematic way of analyzing multiple fragmented polymers with mass spectrometry

Ynze Mengerink^{1,2} | Harry Philipsen^{1,3} | Jan Jordens⁴ | | Josh Mengerink⁵ | Rob van der Hoeven⁶ | Ron A. H. Peters^{7,8} |

The Netherlands

³Engineering Materials, DSM, Geleen, The Netherlands

⁴VITO, Mol, Belgium

⁵Codenext21, Eindhoven, The Netherlands

⁶Food Specialties, DSM, Delft,

The Netherlands

⁷Group Innovation, Covestro, Waalwijk, The Netherlands

⁸Van't Hoff Institute for Molecular Science (HIMS), University of Amsterdam, Amsterdam, The Netherlands

Correspondence

Ynze Mengerink, ACC, DSM, PO Box 1066, 6160 BB Geleen, The Netherlands. Email: ynze.mengerink@dsm.com

Funding information

DSM, Grant/Award Number: n/a

Abstract

A new method to determine the blockiness and especially the block length distribution (BLD) of copolymers is described. This Systematic Workflow to Analyze Multi-fragmented Polymers with Mass Spectrometry (SWAMP-MS) is developed to characterize the sequence distribution of synthetic polymers. Copolymerization of polyamide 4,6 and polyamide 4,10 is used as a model system. Supercharged polymer ions, generated by electrospray ionization, swamp the mass spectrum due to <1> different chemical related distributions (chain composition, end groups), <2> the number of repeating units, <3> the charge and adduct distribution, and <4> isotopes distribution. Without selecting specific substrate ions, MSMS will transform this total ion-SWAMP into chemical rich information of the polymer backbone. The generated fragments contain up to 20 monomers. The reduction of blockiness is evaluated by the decreasing abundancy of specific ions originating from the starting polymers. Moreover, evaluation of all significant fragments and applying Monte Carlo simulations can rebuild the polymer backbone and generate sequence distributions of the copolymerized samples, which is currently still a holy grail in co-chemical polymer analysis.

KEYWORDS

block length distribution, characterization, mass spectrometry, Monte carlo, polymer, simulation

1 | INTRODUCTION

The sequence of building blocks in a copolymer is important for understanding the properties of macromolecules. One of the most famous sequences ever elucidated was the mapping of the human genome in 2001 by J. Craig Venter et al. Besides DNA, protein sequencing also became important in understanding the human genome. Already in 1962, Biemann determined the amino acid sequence of a peptide with Mass Spectrometry. McLafferty et al. applied MS–MS

to characterize a 29 kDa intact protein with the so-called top-down approach.³ To this day, Mass Spectrometry (MS) is still a strong tool in a wide range of protein applications in both top-down and bottom-up (MS and MS–MS on enzymatic digested proteins) approaches. Typical proteomic applications are collagen, ophthalmology, osteoarthritis, and both applications and theory are still evolving.^{4–7}

In contrast to the biological macromolecules mentioned above, which most often have a well-defined backbone sequence, synthetic polymers have lesser-defined

¹ACC, DSM, Geleen, The Netherlands

²Biomedical, DSM, Geleen,

sequences. An AB-copolymer chain with 100 monomeric units can lead to nearly 2¹⁰⁰ different compositions, making full analysis of all possible structures impossible, let alone correlate this to polymer properties.

Crystallization capability is an important materials property, which depends strongly on the distribution of repeating monomeric units. There are several different techniques available to determine such sequences, as described by Multe et al. Recently, pyrolysis gas chromatography (GC) coupled to MS combined with nuclear magnetic resonance (NMR) is described to better elucidate complex copolymers.

Amongst these techniques, NMR is known to be the quantitative golden standard to analyze average randomness of copolymers using diads and triads (two and three consecutive monomeric units),^{11–13} Randomness is calculated using Equation (1).

$$R = \left[\frac{BA}{(BB + BA)}\right] + \left[\frac{AB}{(AB + AA)}\right] \tag{1}$$

Equation (1) to calculate Randomness (*R*) using areas obtained by NMR signals, where AB,BA,AA, BB are the subsequent monomer-monomer couplings.

Matsuda et al. used this to calculate the average sequence length (NASL (Number Average Sequence Length)). As an example, Berti et al. used NMR to find the degree of randomness of a complex system of polyamides, containing hexamethylene diamine and isophthalic and terephthalic acid and polyamides containing m-xylene diamine and adipic acid. 15

However, NMR is limited to these triads. ¹⁶ The main drawback is the relatively short 1.5 repeating units for polyamides or 8–10 Carbon bonds in general, making modulation of long repeating units with a high number of degrees of freedom difficult. ^{17,18} It is known that interaction chromatography is more sensitive to the early stages of copolymerization compared to NMR. ¹⁹ However, it is still impossible to determine the sequence distribution of repeating units.

MS is a powerful technique for the chemical characterization of polymers and excellent reviews have been written. ^{20–24} Matrix assisted laser desorption ionization mass spectrometry (MALDI-MS) and electrospray ionization mass spectrometry (ESI-MS) are the most prominent ionization techniques used for MS for polymer research. ²⁵ However, so far, all applied MS-MS methods to determine polymeric sequences are limited to the selection of a specific ion at rather low molecular weight values (5–10 kDa).

MALDI-MS is known to be able to characterize polymers but is struggling with broadly dispersed polymers. ^{26,27} This so-called mass discrimination effect can be

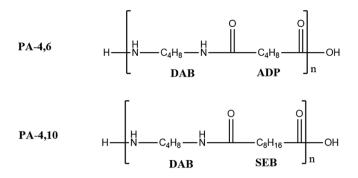


FIGURE 1 Chemical structures of PA-4,6 and PA-4,10 and a possible part of a copolymerization product PA 4,6/10.

reduced using SEC (Size Exclusion Chromatography) fractionation preceding MALDI-MS analysis. ²⁸ Although SEC-MALDI-MS is applicable for homopolymers, complex polymeric systems will be more problematic. ²⁹ Each addition of a monomeric species will strongly increase spectral complexity. Town et al. used MALDI-MS to analyze diblock polyacrylate copolymers and were successful to characterize poly (methyl acrylate-b-ethyl acrylate) copolymers within the 1–3 kDa range with additional laser induced dissociation (LID) fragmentation. ³⁰ Willemse et al. studied the growth of polystyrene-block-polyisoprene copolymer with MALDI-MS and were able to distinguish specific macromolecular species around 4–5 kDa. ³¹ Both examples show the mass-range limitations above 5 kDa.

ESI-MSMS is often applied in combination with liquid chromatography (LC) but suffers similar shortcomings. Gruendlich et al. optimized MS source conditions for molecular ions originating from poly- (methyl methacrylate). Macromolecules (Mw = 5-10 kDa) were visualized using charge states of z = 1 and z = 4. Voeten et al. performed direct infusion and included trapped ion mobility spectrometry (TIMS)-MS to elucidate complex sequences of branched, low molecular weight polyesters.³³ Recently, Nitsche et al. investigated the ionization of carboxylic acid containing polystyrene generating [M-H] and [M-H]²⁻ ions with m/z values of 8 kDa. 34 Steinkoenig et al. used supercharging with multiple chloride ions to determine amongst others the low Mw part of polystyrene (Mw up to 18 kDa), poly(2-vinylpyridine) (Mw = 5 kDa), and polybutadiene.³⁵ Specific fragments originating from multicharged ($z_{max} = 4$) peaks could be distinguished. However, all were hampered by the selection of specific ions and are therefore limited to m/z range of 5-10 kDa. For homo polymers this range might be extended. As an example, Ridgeway et al. including TIMS (Trapped Ion Mobility Spectrometry) and separated n = 100 and n = 101 for polyamide-6 but obviously, sequence distributions are of no interest for these homopolymers.³⁶

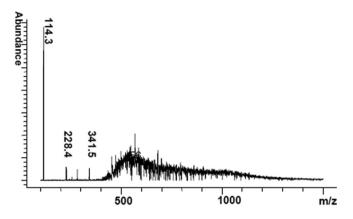


FIGURE 2 ESI (electrospray ionization) (+)-MS of the linear structures of polyamide-6 with a broad hump above 400 Da. (reprinted with permission from Elsevier)

In this present work, a copolymerization model system (see Figure 1) was studied containing polyamide 4,6 (also named A: PA-4,6, consisting of diamino butane (DAB) and adipic acid (ADP)) and polyamide 4,10 (also named B: PA-4,10 consisting of DAB and sebacic acid (SEB)).

In other studies, NMR and MALDI-MS results showed an essential role of the chain ends in copolymerization reactions for melt mixing of polyamide-6 and polyamide-6,6. The main problem with mass spectrometric fragmentation of poly-condensates in general is that the molecular weight distribution is too wide to find discrete m/z values for fragmentation. Although low oligomers are present, they are typically cyclic or very low abundant linear oligomers. ESI-MS fragmentation patterns of low molecular weight polyamide is relative straight forward as typical so called b_n and y_n fragmentions, which means that cleavages take places at the amide functions. The higher molecular weight part of a polyamide-6 was first made visible by supercharging MS as an unresolved broad spectrum (Figure 2). 40,41

This broadly distributed linear homopolymer resulted in a very undefined mass spectrum. All signals originating from <1> macromolecules with different number of the backbone units (typically n=1 to 1000), <2> its isotopic pattern (typically 0–50 13 C atoms but include also 18 O and 15 N) and <3> its charge distributions including adducts (typically z=1–100). This enormous variety of ions swamp the MS-spectrum. 42,43 High-end instruments like Fourier-transform ion cyclotron resonance-MS (FT-ICR-MS), coupled with ion mobility did not reveal better resolutions of the MS-spectrum of this rather simple polymer. 44

In this work, we introduce a new Systematic Way of Analyzing Multi fragmented Polymers (SWAMP) to determine sequence distributions in more complex copolymers. The method is based on MS-MS fragmentation of the supercharged copolymer after ESI without selecting specific ions. Highly informative sequence data can be revealed as the MS-MS spectra contains large fragments of up to 3 kDa. Monte Carlo simulations can rebuild the polymer and sequence distributions can be determined.

The first part of this procedure is related to SWATH-MS (Sequential Window Acquisition of All Theoretical Mass Spectra), which was introduced in 2011 to analyze all chromatographically eluting proteins with a generic MS-MS setup. 45,46 Unlike SWATH-MS, the SWAMP-MS workflow is not using a ramping mass-window to select specific ions from a complex sample, but the fragmentation will be an average of precursors ions of different origins: (multiple) protonated, cationized and different chemically polymeric precursors with different molecular masses. This is the first time that the polymeric sequence distribution is elucidated by applying an MS-MS workflow.

2 | EXPERIMENTAL

An UHPLC-MS (Agilent 1290 QTOF 6540, Wallborn, Germany) was used. Polyamide samples were dissolved in HFIP (hexafluoro isopropanol) (5 mg/ml), and 1 μ l sample was injected on a 2.1x50 mm Acquity UPLC (Ultra Performance Liquid Chromatography) BEH C18 1.7 μ m column (Waters, Milford, MA, USA), which was kept at $T=40^{\circ}$ C. Mobile phase A was aqueous; 1%(v/v) formic acid in water, while mobile phase B was hexafluoro isopropanol (HFIP). A solvent gradient was applied of $40\%B \rightarrow 95\%B$ in 15 min. with an additional 5 min. of 95% HFIP. The flow rate was 0.35 ml/min.

Electrospray ionization in the positive mode was performed with 8 l/min (35 psig) N2 gas at 300°C. Nitrogen is used as the sheath gas (11 ml/min) and the temperature was 350°C. Source conditions were set at 3500 VCap, 1000 Volt nozzle, 175 V fragmentor voltage and 65 V skimmer voltage. Collision Induced Dissociation was applied at 100 eV, scan range was 50-3000 Da.

NMR conditions: approximately 15 mg sample was dissolved in 600 μ l HFIP. After dissolution, a 13 C NMR spectrum was acquired on a 500 MHz NMR spectrometer (BrukerDaltonics, Bremen Germany) equipped with a cryogenically cooled probe head, operating at 40° C.

Homopolymers PA4,6 and PA4,10 resulted from plant scale bulk melt polymerization as regularly used for production of these base polymers. Copolymers were obtained via a transamidation procedure carried out in a lab scale mini extruder at 340 C whereby residence time was varied by varying the rotation speed of the extruder screw.

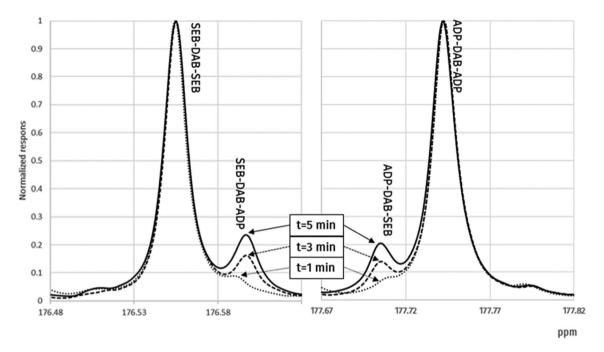


FIGURE 3 Carbonyl region in the ¹³C NMR spectra for PA-4,6/PA-4,10 copolymerization in HFIP with clear triads representing the copolymerization and its progress during copolymerization.

Samples were taken for further analysis without any further purification. These copolyamide samples used are experimental research samples specially prepared by DSM to develop this (and other) methodologies: a 50/50 (%w/w) PA-4,6 – PA-4,10 blend was extruded at 340° C at 200 rpm for t=1, t=2, t=3, t=4 and t=5 min (t1-t5). Beside these five samples, also the pure PA-4,6 and PA-4,10 samples were analyzed as such or as a mixture. Molar mass distributions as determined by Size Exclusion Chromatography (SEC) were found to obey the Flory distribution as may be anticipated for polycondensates including polyamides. (Absolute) weight average molar mass for all samples lies between 20 kDa and 40 kDa and polydispersity Mw/Mn) was found to be 2.0 in all cases.

As input for the Monte Carlo simulation, a probability distribution is provided, that assigns a probability [0,1] to each fragment specification. Such a fragment specification dictates how many A's (DAB-ADP), and B's (DAB-SEB) should occur in a fragment once generated. Each throughput time has a dedicated probability distribution (PT1, PT2, PT3, PT4, and PT5) and each includes probabilities for fragments of size (size = #A + #B) 2 through 10. Above 10 still fragments are visible, but the s/n becomes too low.

When running the Monte Carlo procedure for a single throughput time (e.g., $t=1\,$ min), 100 numeric sequences are constructed, each by iteratively appending 100,000 (randomly) chosen fragments, based on abundance in the MS-MS spectra. To construct an individual fragment a fragment-specification is selected

non-uniformly according to the proper distribution (e.g., PT1).

The source code for this simulation is made available online: https://gitlab.com/swamp-ms/monte-carlo

3 | RESULTS AND DISCUSSION

3.1 | Results

NMR is the golden standard to determine the randomness (NASL) as an average of the sequence distribution. Typical results for the various copolymerized samples are given in Figure 3. The increase of the ADP-DAB-SEB triads can be visualized and quantified. Randomness results are compared with MSMS at the end of this document (Table 5), where according to Equation (1) AA is the area of the signal of ADP-DAB-ADP; AB is the area of the signal of ADP-DAB-SEB, BB=SEB-ADP-SEB and BA = DEB-DAB-ADP.

As with MSMS fragmentations larger sequences can be made visible, a new workflow was developed. This workflow named SWAMP-MS is outlined in Figure 4. The different steps are described in detail in the remainder of this section. For understanding of the fragments named in the probability list, C:3:3:0 stand for a specific fragment with 3 monomers, 3 of them being A and 0 being B.

<u>Step 1: Supercharging</u>: An aqueous mobile phase with HFIP containing 1% formic acid and subsequent ESI will supercharge the polymeric fraction of the polyamide

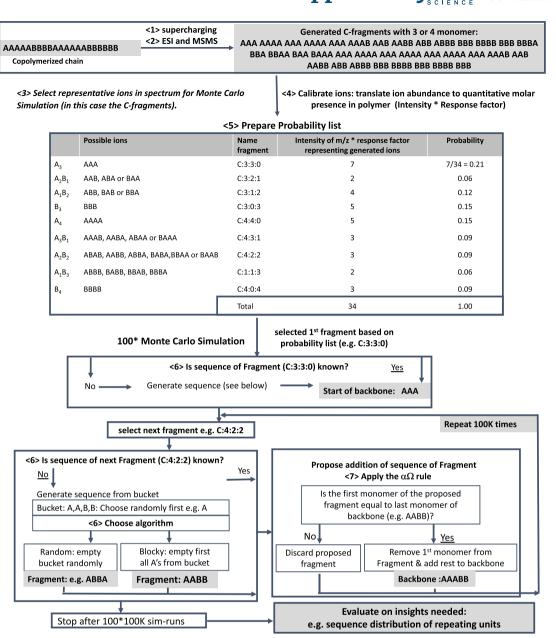


FIGURE 4 SWAMP: top: from ion generation to probability list, bottom: Monte Carlo simulation steps. [Color figure can be viewed at wileyonlinelibrary.com]

samples directly after elution from the LC-column. All generated ions of PA-4,6, PA-4,10, and the copolymerized PA-4,6/10 swamp the mass spectrum (Figure 5 top). The narrow mass spectrometric peaks on the broad unresolved mass spectrum originate largely from gradient contaminations. For illustrative purposes, a non-background corrected chromatogram has been shown to circumvent negative peaks as seen in Figure 2.

<u>Step 2: MSMS fragmentation</u>: Information can be extracted from this broad unresolved MS spectrum by applying fragmentation energy. Figure 5 (bottom) shows the results when a high collision-induced dissociation

(CID) energy is applied on the entire polymeric swamp (all ions are fragmented, no specific pre-selection of ions).

The positive ions generated by ESI are allowed to pass through the first quadrupole, are fragmented, and thereafter separated by a time-of-flight (TOF) mass selector. The complexity of the compositions of the polymeric structure will result in an optimal collision energy of about 100 eV. Using higher voltages will result in more severe fragmentation and thereby less structural information.

The resulting MSMS spectra are already relative complex for this three-monomer system, with an AB sequence.

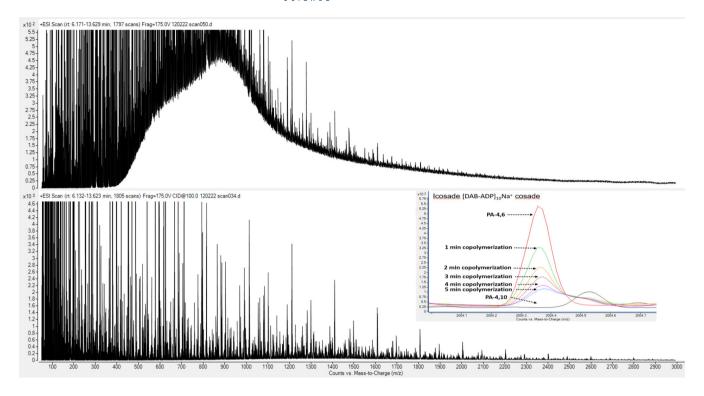


FIGURE 5 Influence of different CID energies applied. Upper-trace (CID = 0 eV) with a zoom in of the polymeric swamp of PA-46, 2nd trace (100 eV) including PA-4,6, PA-4,10 and the copolymerized sample after t = 5 min, Insert is the decrease of the [DAP-ADP]₁₀ C-fragment due to copolymerization. [Color figure can be viewed at wileyonlinelibrary.com]

<u>Step 3: Select representative ions</u>: Table 1 shows all the main fragments observed. Typically mono-charged fragments are observed. Figure 6 shows the most abundant fragment, which is a linear block with an imine and aldehyde end group. This fragment is further indicated as C-fragment, which is the most abundant fragment and therefore used for further evaluation. The C-fragment is annotated as $C:X_1:X_2:X_3$. X_1 stands for the number of DAB units, X_2 represents the number of ADP units and X_3 the number of SEB units. For C-fragments $X_1 = X_2 + X_3$.

Lower Mw fragments tend to be protonated, while higher Mw fragments are often cationized with sodium (and in a lesser extent with potassium). As an example, all C-fragmented 20-mers (including [DAP-ADP]₅-[DAP-SEB]₅ (C:10:5:5), [DAP-ADP]₁₀ (C:10:10:0) and [DAP-SEB]₁₀ (C:10:0:10)) are sodium cationized and C-fragment inserted in the 2nd trace of Figure 5 the MNa⁺ of C:10:10:0.

Studying the fragmentation results in more detail, in general, amide cleavage (b / y cleavage) is observed. The nomenclature of a,b,c versus x,y,z-ions, as used in protein analysis, does not hold for these fragmentations as there is no clear amine or acid termination in the chain.

Fragmentation of a five-times charged linear diacid functionalized oligomer n = 16, ADP-[DAP-ADP]₁₆, originating from an oligomeric PA-4,6 sample, revealed,

TABLE 1 Typical fragments observed for the PA-4,6 and PA 4,10 and copolymerized samples

Main fragment observed both as $[M + H]^+$ and/or $[M + Na]^+$
Block – H ₂ O
Block – $H_2O + NH_3$
Block (see Figure 6)
$Block + H_2O - NH_3$
$Block + H_2O$
Block + DAB
$Block + DAB + H_2O - NH_3$
$Block + DAB + H_2O$

amongst others, di-imino (-CH=NH) terminated fragments indicating at least two cleavages.

The loss of water from the polymer fragments is on the terminus of the fragments, as there is no water loss from di-imino fragments (Block + DAB). Also, there is no addition of NH $_3$ from di-imino terminated fragments, but they are visible if there is an aldehyde terminated fragment, indicating rearrangements at the cleaved amide.

Diacids are not visible, suggesting that cationization also takes place on the imine function.

FIGURE 6 Main block structure of the fragmentations observed of polyamides samples.

The obtained C-fragments are best suitable for further evaluation to generate sequence distributions.

<u>Step 4: Calibrating ions</u>: For the intended Monte Carlo simulations, the abundance of an ion in the MSMS spectrum should represent the probability of its present in the backbone of the polymer.

Several factors contribute to the number of potential fragments and their abundance after MSMS. Each obtained fragment should contribute to the Monte Carlo simulations according its true, relative presence in the backbone. For this, the MSMS abundances of the fragments of interest need to be corrected for differences in sensitivity between PA-4,6 and PA-4,10 and between differences in sensitivity originating from the size of the fragments.

As an over-simplified basic example (without further considerations like the $\alpha\Omega$ -rule explained further), the sequence ABBB would give the fragments A, B, AB and BB. If the abundance of A would be 100, B would be 10, AB would be 20 and BB would be 1, the probability of building these fragments in the final polymer should adjusted by these differences in abundances using calibration with pure standards.

Calibration of the abundances starts by selecting the homo-fragments (which only contain one specific monomer e.g., AAAA). The weight response factors (*Rf) of these homo-fragments can be calculated easily from pure reference PA-4,6 and PA-4,10 as the reciprocal intensity of the sum of the most abundant isotope of the H+ and Na + ions of the C-fragment (*Rf = 1/ Intensity). If this intensity of the C-fragment of the pure PA-4,6 and pure PA-4,10 were now multiplied by their specific *Rf, their concentration would be set to unity for these reference materials and would contribute equally. The weight response factor (*Rf) is transformed to molar response factors (Rf), by dividing the (*Rf) by the average Mw of a monomeric backbone unit of the fragment (198 for DAB-ADP and 254 for DAB-SEB).

An example for the pure C:3:3:0, C:3:0:3, C:4:4:0 and C:4:0:4 fragments is given in Table 2 below.

The measured Rf of all the homo C-fragments (C:X1: X2:0 and C:X1:0:X3) are given in gray in Table 3. The

adjustment factors of the copolymerization C-fragments (C:X1:X2:X3 with $X2 \neq 0$ and $X3 \neq 0$)) cannot be separately determined, as they are not present from pure polymers. Their values are estimated by linear interpolations of the pure C-fragments.

<u>Step 5: Probability list:</u> The product of the intensities of a C-fragment and its corresponding response factors from Table 3 (i.e., Rf * Intensity) is a measure of the probability of this specific sequence in the specific copolymerized polymer. These relative amounts were used to assess the probability of this fragment to be present in the backbone, i.e., the ratio of corrected intensity of the C-fragment and the total intensity of the C-fragments (see Table 4).

The results in Table 4 are highly interesting. The probability of fragments originating from the starting PA-4,6 (C:X₁:X₂:0) and PA-4,10 (C:X₁:0:X₃) decrease as a function of the copolymerization time, as an obvious result of the polymer becoming less blocky. Also, the probability of all other fragments (C:X₁:X₂:X₃ with $X_2 \neq 0$ and $X_3 \neq 0$) to be present increase as a function of the polymerization time for the exact same reason as they are generated during polymerization.

Mole fractions of PA-4,6 and P-4,10 can be calculated easily from Table 3. The results are completely in line with NMR results and the expected loading of the samples (see Table 5). This finding supports the correctness of this calibration procedure.

<u>Step 6: Sequence generation of fragments</u>: The exact sequence of longer fragments containing both A and B (C:X1:X2:X3 with $X2 \neq 0$ and $X3 \neq 0$) cannot be revealed from the m/z value of the fragment ion. There are different approaches to cope with this. It starts with a virtual bucket containing the number of monomers as defined in the C:X1:X2:X3 equation and a fragment sequence generator. The fragment sequence generator will always choose randomly the first monomer (i.e., A or B), the rest of the fragment can be built by different algorithms, which can be chosen or defined freely:

<1> The fragment sequence is blocky.

TABLE 2 Example to calculate Molar Response factors

		Sum inter	nsity		Molar correct	Rf=*Rf/mw (*1E6)
Fragment		PA-4,10	PA-4,6	*Rf=1/intensity (*1E6)	(mw of 1 block)	molar correction
C:3:0:3 H+ & Na+	BBB	4889		205	254	0.81
C:3:3:0 H+ & Na+	AAA		10377	96	198	0.49
C:4:0:4 H+ & Na+	BBBB	9867		101	254	0.40
C:4:4:0 H+ & Na+	AAAA		7305	137	198	0.69

TABLE 3 Response factors for all C: X_1 : X_2 : X_3 fragments. X_2 is not given in the table, but $X_2 = X_1$ - X_3 . Adjustment factors for the X_2 or X_3 fragments (in gray) are derived directly from the pure PA-46 and PA 410 Mass Spectrum. The rest are interpolated values (e.g., C:2:1:1 = 0.82 = (0.5* 1.03) + (0.5*0.61)

	C:X1:X	₂ :X ₃ (X ₂ =X	X ₁ -X ₃): Res	sponse fact	ors * 10e6						
$X_1 \downarrow X_3 \to$	0	1	2	3	4	5	6	7	8	9	10
2	0.49	0.56	0.63								
3	0.49	0.59	0.70	0.81							
4	0.69	0.62	0.55	0.47	0.40						
5	0.87	0.80	0.73	0.66	0.60	0.53					
6	1.16	1.14	1.12	1.09	1.07	1.05	1.03				
7	1.64	1.60	1.57	1.54	1.50	1.47	1.43	1.40			
8	2.71	2.67	2.64	2.60	2.56	2.52	2.48	2.45	2.41		
9	5.11	5.47	5.83	6.18	6.54	6.90	7.26	7.61	7.97	8.33	
10	7.61	8.61	9.60	10.60	11.59	12.59	13.58	14.58	15.58	16.57	17.57

- <2> The fragment sequence is randomly defined.
- <3> The fragment sequence is steered.

<1> Blocky sequence generator: The sequence of a mixed fragment is defined as just two blocks. As an example: If the simulation chooses the fragment C:4:2:2 from the probability list, the generator adds 2 A's and 2 B's in an imaginary bucket. If the first randomly chosen monomer is an A, the generator will first take all remaining A's from the bucket and after that the sequence is completed with the remaining B's. The result will be AABB.

<2> Random sequence generator: The sequence is built completely randomly. As an example: If the simulation chooses, as above, fragment C:4:2:2 and the generator adds 2 A's and 2 B's into the imaginary bucket, the exact structure is chosen randomly. If the first randomly chosen monomer is an A, the bucket content remains with 1 A and 2 B's. Again, in each step a random monomer is chosen until the bucket is empty. As a result, the sequence of the fragment will be AABB, ABAB or ABBA.

<3> Steered: as <2> random, but steered to meet other measured results, e.g., randomness data originating from

NMR, e.g., to reach a certain blockiness. This possibility is not further explored, as the blocky sequence generator delivered already a reasonably good match. As an example: If the simulation chooses to add fragment C:4:2:2 and A is selected as the first monomer, the chance in <2> of selecting the second monomer is ½ for the remaining A and ½ for the two remaining B's. If the specific sample has a high randomness (>1) the chance of selecting a B is increased and the chance of an A is reduced. For a low randomness (<1) the chance of selecting an A would be increased and the chance of B would be decreased.

TABLE 4 Probability of the presence of a specific fragment into the backbone. The sum for each sample equals unity

Probability list	lity list												
		t=1 min	t=2 min	t=3 min	t=4 min	t=5 min			t=1 min	t=2 min	t=3 min	t=4 min	t= 5 min
C:2:0:2	BB	0.070	0.070	0.069	0.065	0.062	C:8:0:8	BBBBBBBB	0.019	0.017	0.012	0.010	0.008
C:2:1:1	AB / BA	0.004	0.008	0.010	0.013	0.016	C:8:1:7		0.001	0.003	0.003	0.004	0.003
C:2:2:0	AA	0.059	0.049	0.047	0.042	0.041	C:8:2:6		0.002	0.003	0.004	0.005	0.005
C:3:0:3	BBB	0.058	0.056	0.053	0.049	0.045	C:8:3:5		0.002	0.004	0.005	0.007	0.007
C:3:1:2	BBA, BAB, ABB	0.005	0.009	0.011	0.016	0.018	C:8:4:4		0.003	0.004	0.007	0.008	0.009
C:3:2:1	AAB, ABA,BAA	0.004	0.007	0.009	0.012	0.014	C:8:5:3		0.003	0.007	0.009	0.011	0.012
C:3:3:0	AAA	0.052	0.049	0.041	0.041	0.039	C:8:6:2		0.004	0.007	0.010	0.011	0.012
C:4:0:4	BBBB	0.045	0.047	0.046	0.043	0.038	C:8:7:1		0.004	0.007	0.008	0.010	0.011
C:4:1:3		0.003	0.004	900.0	0.009	0.010	C:8:8:0	AAAAAAA	0.042	0.035	0.027	0.022	0.020
C:4:2:2		0.003	0.005	0.007	0.009	0.010	C:9:0:9	BBBBBBBBB	0.033	0.025	0.019	0.013	0.010
C:4:3:1		0.003	900.0	0.008	0.011	0.012	C:9:1:8		0.004	0.004	0.005	0.005	0.005
C:4:4:0	AAAA	0.059	0.049	0.045	0.038	0.036	C:9:2:7		0.003	0.004	900'0	90000	0.007
C:5:0:5	BBBBBB	0.035	0.034	0.032	0.027	0.022	C:9:3:6		0.004	0.005	0.008	0.011	0.012
C:5:1:4		0.003	0.005	0.007	0.009	0.010	C:9:4:5		0.004	0.007	0.010	0.012	0.015
C:5:2:3		0.003	0.005	0.007	0.008	0.009	C:9:5:4		0.005	0.009	0.012	0.015	0.017
C:5:3:2		0.004	900.0	0.008	0.010	0.012	C:9:6:3		9000	0.011	0.014	0.017	0.021
C:5:4:1		0.004	0.007	0.009	0.012	0.013	C:9:7:2		90000	0.010	0.013	0.016	0.016
C:5:5:0	AAAAA	0.058	0.047	0.042	0.035	0.033	C:9:8:1		9000	0.009	0.011	0.011	0.012
C:6:0:6	BBBBBB	0.033	0.029	0.025	0.020	0.017	C:9:9:0	AAAAAAAA	0.050	0.037	0.031	0.023	0.021
C:6:1:5		0.003	0.004	900.0	0.007	0.007	C:10:0:10	BBBBBBBBB	0.034	0.025	0.018	0.013	600.0
C:6:2:4		0.002	0.004	900.0	0.007	0.007	C:10:1:9		0.002	0.005	0.004	0.005	0.005
C:6:3:3		0.003	0.005	0.007	0.009	0.011	C:10:2:8		0.003	0.004	0.005	0.005	0.005
C:6:4:2		0.004	0.007	0.009	0.011	0.013	C:10:3:7		0.005	0.005	0.007	0.010	0.011
C:6:5:1		0.004	0.008	0.010	0.012	0.014	C:10:4:6		0.005	0.007	0.010	0.013	0.014
C:6:6:0	AAAAA	0.056	0.045	0.040	0.033	0.030	C:10:5:5		0.005	0.009	0.011	0.014	0.017
C:7:0:7	BBBBBBB	0.023	0.018	0.015	0.012	0.009	C:10:6:4		9000	0.011	0.014	0.018	0.020
C:7:1:6		0.002	0.002	0.003	0.003	0.003	C:10:7:3		9000	0.011	0.015	0.018	0.020
C:7:2:5		0.002	0.003	0.004	0.005	0.005	C:10:8:2		9000	0.010	0.012	0.015	0.017
													(Continues)

(Continued)
TABLE 4

Probability list	ity list													
		t = 1	t=2	t=3	<i>t</i> = 4	t = 5			t=1	t=2	t=3	<i>t</i> = 4	t= 5	
		min	min	min	min	min			min	min	min	min	min	
C:7:3:4		0.002	0.004	0.005	0.007	0.008	C:10:9:1		90000	0.009	0.010	0.010	0.011	
C:7:4:3		0.003	900.0	0.008	0.010	0.011	C:10:10:0	C:10:10:0 AAAAAAAAA 0.051	0.051	0.037	0.030	0.022	0.020	
C:7:5:2		0.005	0.007	0.009	0.011	0.013								
C:7:6:1		0.004	0.008	0.010	0.012	0.013								
C:7:7:0	AAAAAA	0.054	0.041	0.035	0.028	0.025								

be a little less randomized and therefore the α monomer is not included in building the backbone.

As an example: The last part of the backbone generated by the Monte Carlo simulation could be -AAB and the next fragment selected using the probability table (Table 3) could be C:4:2:2. The fragment sequence generator could define this as ABBA. As the Ω monomer (B) of the last part of the simulated backbone is not identical to α monomer (A) of the newly selected fragment, the fragment will not be used in the building of the backbone. If the next selected fragment would again be C:4:2:2 and the random sequence generator could come with BABA, the new backbone would be defined as -AABABA, as the α and Ω monomers (both B) are identical. The α monomer (B) is not included in the new backbone.

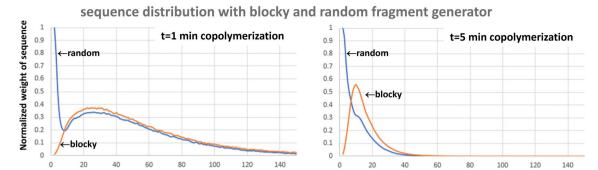
From this point on, the algorithm will generate a backbone by repeating step 6 and 7. Fragments chosen from the probability list are added to the backbone if the $\alpha\Omega$ rule is obeyed. From the final sequence, distributions of repeating units can be easily retrieved.

DISCUSSION

The results of this Monte Carlo algorithm are given in Figure 7. The differences between the random and blocky fragment generator are clearly visible in Figure 7-top. The random fragment generator produces smaller sequences with identical monomers, while the blocky fragment generator produces longer sequences with identical monomers. However, differences are only significant at the low end of the distributions and the maximum of the distribution is for all different copolymerization times almost identical (t = 1 min and t = 5 min given in Figure 7-top). The blocky fragment generator gives a good match in randomness compared to NMR (see Table 6). The random fragment generator clearly overestimates randomness as it generates small random sequences. Therefore, the distributions move to higher values for smaller sequences of identical monomers. For more randomized polymers, the random or steered generator could give better matches.

The randomness obtained from the NMR data is probably a little underestimated, as the small copolymerization peaks AB and BA are measured from its flanks of the main AA and BB peaks.

Figure 7-bottom nicely shows the influence of copolymerization on the sequence distribution. With short copolymerization times, the broader distribution of identical monomers is clearly visible. Moving to longer copolymerization times, this distribution decreases to shorter chain-lengths.



Sequence of consecutive identical monomers

Sequence distribution with different degree of copolymerization

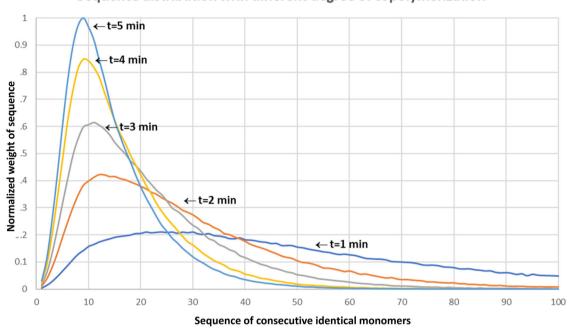


FIGURE 7 Top: comparing blocky and random model (see text) for (t = 1 min and t = 5 min) and bottom sequence distribution as a function of the co-polymerization time, as determined by SWAMP MS. [Color figure can be viewed at wileyonlinelibrary.com]

Although the MS-MS results are somewhat biased at the low end of the distribution (n < 10), of the distribution is not influenced by an integrated steered approach, i.e., using the NMR data for the generation of the sequence in the fragments by including a factor that would steer the total randomness values towards the received results from the NMR experiments. As a downside, there would not be additional control data available.

5 | CONCLUSIONS AND LOOKOUT

"SWAMP-MS" has been developed to analyze the sequence distribution of repeating monomers of a copolymer. This new workflow was applied towards the copolymerization of polyamide 4,6 and polyamide 4.10. In an acidified environment after electrospray ionization, these polymers became supercharged. Using CID at 100 eV, fragments containing up to and including 20 monomeric units were generated. These informative-rich fragments could be used to monitor copolymerization. Monte Carlo-based algorithms were developed to reconstruct the polymer backbone. The resulting sequence distributions of repeating monomeric units clearly distinguished between the different copolymerization times. Overall quantitative numbers are in fair agreement with obtained NMR data. The use of different algorithms to estimate the sequence of fragments delivers identical distributions with respect to sequences longer than 10 identical repeating units. A bias occurs for the shorter sequences and matching with e.g., NMR randomness data can cope for that difference.

Mole fraction in %				
Copolymerization	PA-4,6 (loa	ding 56%)	PA-4,10 (lo	oading 44%)
	NMR	MSMS	NMR	MSMS
$t = 1 \min$	57	58	43	42
$t = 2 \min$	56	56	44	44
$t = 3 \min$	57	57	43	43
$t = 4 \min$	56	56	44	44
$t = 5 \min$	56	58	44	42
Average	56	57	44	43

TABLE 5 Comparison between the total mol fraction of the samples by NMR and MSMS versus the loading of the polymers in the original samples

TABLE 6 Comparison of randomness by NMR and MSMS of different copolymerized samples Randomness (0 = no copolymerization only AA and BB, no AB or BA, 1 = equal amount of (AA+BB) and (AB+BA) and 2 = no AA and no BB, only AB and BA). For MSMS the sequence of the fragments is determined by a blocky and random approach (see text).

Randomness			
Copolymerizati time	on NMR	MSMS blocky	MSMS random
1 min	0.04	0.07	0.19
2 min	0.06	0.12	0.33
3 min	0.11	0.15	0.42
4 min	0.12	0.18	0.51
5 min	0.16	0.21	0.57

This SWAMP-MS method could be applied in a broader context. Nitrogen-containing poly-condensates are a good candidate for being supercharged. Moreover, alternative supercharging procedures could be used to extend the field of application. There is still a lot to discover with this method; Optimization of the procedure, supercharging of other polymers, new algorithms, better estimation of the sequence of the fragments, better calibration procedures and easier visualization of the obtained information to name a few.

AUTHOR CONTRIBUTIONS

Ynze Mengerink: Conceptualization (lead); formal analysis (lead); investigation (lead); project administration (equal); visualization (lead); writing – original draft (lead). Harry Philipsen: Funding acquisition (lead); methodology (equal); project administration (lead); writing – review and editing (equal). Jan Jordens: Conceptualization (equal); investigation (equal); methodology (equal); writing – review and editing (equal). Josh

Mengerink: Methodology (equal); software (lead); visualization (equal). **Rob van der Hoeven:** Conceptualization (equal); methodology (equal); writing – review and editing (equal). **Ron A. H. Peters:** Conceptualization (lead); investigation (equal); writing – review and editing (equal).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Ynze Mengerink https://orcid.org/0000-0001-8521-2149

Jan Jordens https://orcid.org/0000-0001-8711-7858

Josh Mengerink https://orcid.org/0000-0002-1199-2311

Ron A. H. Peters https://orcid.org/0000-0003-0667-1465

REFERENCES

[1] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. D. Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Y. Wang, A. Wang, X. Wang, J. Wang, M. H. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. C. Zhu, S. Zhao, D. Gilbert, S.

Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigó, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. D. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, X. Zhu, Science 2001, 291, 1304.

- [2] K. Biemann, Mass Spectrometry. Organic Chemical Applications, McGraw-Hill, New York 1962, p. 260.
- [3] S. K. Sze, Y. Ge, H. B. Oh, F. W. McLafferty, PNAS 2002, 99, 1774.
- [4] N. A. van Huizen, J. N. M. IJzermans, P. C. Burgers, T. M. Luider, Mass Spectrom. Rev. 2020, 39, 309.
- [5] B. N. A. Coronado, F. B. S. da Cunha, O. de Toledo Nobrega, A. M. A. Martins, *Int. Ophthalmol.* 2021, 41, 2619.
- [6] M. J. J. Haartmans, K. S. Emanuel, G. J. M. Tuijthof, R. M. A. Heeren, P. J. Emans, *Expert Rev. Proteomic.* **2021**, *18*, 693.
- [7] P. Zhang, I. L. Ang, R. Wei, K. M. K. Lei, T. C. W. Poon, W. Chan, M. M. T. Lam, Sci. Rep. 2019, 9, 6453.
- [8] C. De Rosa, F. Auriemma, Crystal Structures of Polymers, Handbook of Crystallization, Piorkowska, E; Rutledge, G.C., Wiley, Hoboken, New Jersey 2013, p. 31.
- [9] H. Mutlu, J. F. Lutz, Angew. Chem., Int. Ed. 2014, 53, 13010.
- [10] W. C. Knol, T. Gruendling, P. J. Schoenmakers, B. W. J. Pirok, R. A. H. Peters, J. Chromatogr. A 2020, 1670, 462973.
- [11] J. C. Randall, Polymer Sequence Determination: Carbon-13 NMR Method, Academic Press, New York 1977.
- [12] H. N. Cheng, A. D. English, NMR Spectroscopy of Polymers in Solution and in the Solid State, ACS symposium series, ACS 2012, 834. https://doi.org/10.1021/bk-2003-0834
- [13] J. Devaux, P. Godard, J. P. Mercier, J. Polym. Sci., Polym. Phys. Ed. 1982, 20, 1875.
- [14] H. Matsuda, T. Asakura, T. Miki, Macromolecules 2002, 35, 4664.
- [15] C. Berti, A. Celli, P. Marchese, S. Sullalti, M. Vannini, C. Lorenzetti, Eur. Polym. J. 2012, 48, 1923.
- [16] H. N. Cheng, M. J. Miri, ACS Symp. Ser. 2011, 1077, 371.

- [17] K. L. L. Eersels, G. Groeninckx, Y. Mengerink, V. der Wal, Sj., Macromolecules 1996, 29, 6744.
- [18] K. Hatada, T. Kitayama, NMR Spectroscopy of Polymers, Springer, Verlag Berlin Heidelberg 2004. https://doi.org/10. 1007/978-3-662-08982-8
- [19] K. Eersels, PhD thesis, KU Leuven (Leuven), 1996, PhD thesis KU Leuven
- [20] M.S. Montaudo, Mathematical Approaches to Polymer Sequence Analysis and Related Problems (Ed: R. Bruni), Springer, New York 2010, pp. 227–246. https://doi.org/10. 1007/978-1-4419-6800-5
- [21] T. Gruendling, S. Weidner, J. Falkenhagen, C. Barner-Kowollik, Polym. Chem 2010, 1, 599.
- [22] M. R. L. Painea, P. J. Barker, S. J. Blanksbya, *Analytica Chemica Acta* 2014, 808, 70.
- [23] E. Altuntas, U. S. Schubert, Analytica Chemica Acta 2014, 808, 56.
- [24] S. Crotty, S. Gerişlioğlu, K. Endres, C. Wesdemiotis, U. S. Schubert, Anal. Chim. Acta 2016, 932, 1.
- [25] e. Altuntaş, a. Krieg, A. Baumgaertel, A. C. Crecelius, A. C. Schubert, J. Polym. Sci. Part A: Polym. Chem. 2013, 51, 1595.
- [26] B. Staal, Ph.D. Thesis, TU/e (Eindhoven), 2005. https://doi.org/ 10.6100/IR583506
- [27] H. C. M. Byrd, C. N. McEwen, Anal. Chem. 2000, 72, 4568.
- [28] S. D. Hanton, X. M. Liu, Anal. Chem. 2000, 72, 4550.
- [29] M. W. F. Nielen, Mass Spectrom. Rev. 1999, 18, 309.
- [30] J. S. Town, G. R. Jones, D. M. Haddleton, *Polym. Chem* 2018, 9, 4631.
- [31] R. X. E. Willemse, B. B. P. Staal, E. H. D. Donkers, A. M. Van Herk, *Macromolecules* 2004, 37, 5717.
- [32] T. Gruendling, M. Guilhaus, C. Barner-Kowollik, *Macromol. Rapid Commun.* **2009**, *30*, 589.
- [33] R. L. C. Voeten, B. Van de Put, J. Jordens, Y. Mengerink, R. A. H. Peters, R. Haselberg, G. W. Somsen, J. Am. Soc. Mass Spectrom. 2021, 32, 1498.
- [34] T. Nitsche, M. M. Sheil, J. P. Blinco, C. Barner-Kowollik, S. J. Blanksby, J. Am. Soc. Mass Spectrom. 2021, 32, 2123.
- [35] J. Steinkoenig, M. M. Cecchini, S. Reale, A. S. Goldmann, C. Barner-Kowollik, *Macromolecules* 2017, 50, 8033.
- [36] M. Ridgeway, M. Lubeck, J. Jordens, M. Mann, M. A. Park, Int. J. Mass Spectrom. 2018, 425, 22.
- [37] F. Samperi, C. Puglisi, R. Alicata, G. Montando, J. Polym. Sci., Part A: Polym. Chem. 2003, 41, 2778.
- [38] C. Barrère, M. Hubert-Roux, C. Afonso, M. Rejaibi, N. Kebir, N. Désilles, L. Lecamp, F. Burel, C. Loutelier-Bourhis, *Anal. Chim. Acta* 2014, 808, 3.
- [39] I. A. Papayannopoulos, Mass Spectrom. Rev. 1995, 14, 49.
- [40] Y. Mengerink, PhD thesis, TU/e (Eindhoven), 2001. https://doi.org/10.6100/ir549993
- [41] Y. Mengerink, R. A. H. Peters, C. G. De Koster, S. van der Wal, H. A. Claessens, C. A. Cramers, J. Chromatogr.: A 2001, 914, 131.
- [42] A. Ghaffar, Degradation and analysis of synthetic polymeric materials for biomedical applications, *PhD thesis*, 2011 isbn: 978-90-5776-231-4
- [43] A. Ghaffar, G. J. J. Draaisma, G. Mihov, A. A. Dias, P. J. Schoenmakers, V. der Wal, Sj., Biomacromolecules 2011, 12, 3243.
- [44] J. Jordens, Y. Mengerink, M. Park, M. Ridgeway, M. Honing, Resolving ionization processes of polyamides by ESI-IMS-MS, presented at ISMS, Geneva, 2014, copy on request

- [45] P. Navarro, L. Gillet, C. Carapito, H. Röst, L. Reiter, O. Rinner, S. Tate, R. Bonner, L. Malmström, R. Aebersold, SWATH-MS: A new data independent acquisition LC-MS methodology for quantitative complete proteome analysis, proteómica., 2011
- [46] L. C. Gillet, P. Navarro, S. Tate, H. Roest, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Mol. Cell. Proteomics 2012, 11, 0111.016717.

How to cite this article: Y. Mengerink, H. Philipsen, J. Jordens, J. Mengerink, R. van der Hoeven, R. A. H. Peters, *J. Appl. Polym. Sci.* **2023**, e53683. https://doi.org/10.1002/app.53683