COMMUNICATION

Sequencing of Uniform Multifunctional Oligoesters via Random Chain Cleavages

Matthieu Soete[a] and Filip E. Du Prez*[a]

[a] Polymer Chemistry Research group, Centre of Macromolecular Chemistry (CMaC), Department of organic and Macromolecular Chemistry, Faculty of Sciences, Ghent University, Krijgslaan 281 S4-bis, B-9000 Ghent, Belgium * Corresponding author: E-mail: Filip.DuPrez@UGent.be

Supporting information for this article is given via a link at the end of the document.

Abstract: Sequence-defined polymers have been the object of many fascinating studies that focus on their implementation in both material and life science applications. In parallel, iterative synthetic methodologies have become more efficient, whereas the structure elucidation of these molecules is generally dependent on MS/MS analysis. Here, we report an alternative, simple strategy for the determination of the monomer order of uniform oligo(thioether ester)s. This approach, which relies on random cleavages of ester units within the macromolecular backbone *via* a basic treatment, enables the swift characterization of these macromolecules without the need for MS/MS. Consequently, this method can be used for decoding any information stored within the primary structure of oligoesters by means of ESI- or LC-MS. Finally, we speculate that a range of structurally diverse backbones could be susceptible towards this approach, which could promptly expand the library of chemically sequenceable macromolecules.

The sequencing of biomacromolecules has been a topic of interest ever since the elucidation of their primary structure. For instance, Sanger sequencing dominated the early days of DNA sequencing, while the order of nucleobases comprising a DNA strand is nowadays conveniently determined by means of next-generation sequencing techniques.[1,2] In addition, Edman degradation, a process that consists of the stepwise chemical degradation of peptides, has aided the sequence determination of numerous proteins.[3] Ladder sequencing further improved upon this method by reducing the overall time required for the process, while at the same time limiting the analysis to a single mass spectrum after the final step.[4] Those developments resulted in the successful identification of a wide range of peptides and simultaneously led to the formation of huge databases. The latter play a crucial role in, for example, peptide mass fingerprinting, whereby tryptic peptides of a protein are matched with fragments from a database.[5–7]

On the other hand, the structure determination of monodisperse synthetic macromolecules has only more recently attracted widespread attention, mainly as a result of the advancements made within the area of sequence-defined polymers.[8] These macromolecules are, amongst other applications, currently being investigated as a synthetic data storage platform, where digital information can be stored or encoded using the monomers' order.[9–18] Sequencing information-containing macromolecules is generally achieved by using MS/MS techniques, and different research groups have developed algorithms to facilitate the otherwise time-consuming read-out process.[19–22]

Even though MS/MS is an efficient method for the determination of the monomer order of synthetic macromolecules, sequencing the latter can in principle also be achieved by making use of more conventional MS^1 techniques, provided that the fragmentation of the macromolecules is performed in a chemical manner. However, in contrast to biomacromolecules, very few molecular sequencing strategies have been developed for synthetic sequence-defined macromolecules. One of the rare examples was introduced by Zuckermann and co-workers, who described the on-resin stepwise chemical degradation of peptoids.[23] Another strategy for generating a series of truncated structures consists of the depolymerization of synthetic macromolecules, an approach that was recently adopted by Van Anslyn and co-workers.[24,25] They reported the elegant sequencing of sequence-defined oligourethanes *via* a controlled self-immolation process. In this case, the terminal hydroxyl moiety was capable of participating in an intramolecular cyclization process, leading to the partial depolymerization of the oligourethanes. Interestingly, a single LC-MS analysis was sufficient for a successful determination of the monomer order since every crucial fragment was present at a specific point in time. Using a similar approach, Lutz and co-workers reported the self-immolation of N-substituted oligourethanes, whereby the depolymerization could be inhibited by increasing the spacer length between different carbamate moieties.[26] Whilst both examples provide interesting insights - sequencing by means of conventional MS^1 techniques remains a highly desirable feature for industrial applications - the majority of synthetic macromolecules is not able to spontaneously depolymerize upon the application of a certain stimulus.

Inspired by the beneficial characteristics associated with a molecular sequencing strategy that requires solely MS^1 analysis, we herein report the sequencing of discrete oligo(thioether ester)s by chemically inducing random chain scissions within the macromolecular backbone. We have chosen those ester-containing macromolecules as model compounds and we anticipate that this strategy can be readily applicable to any backbone that contains degradable functionalities. Specifically, fragmentation of the macromolecules will be demonstrated *via* a simple basic treatment that delivers all fragments required for the sequencing of the macromolecules by means of ESI- or LC-MS. While this proof of concept might rapidly expand the amount of chemically sequenceable abiotic macromolecules, it could also increase the valorisation potential of sequence-defined macromolecules in different areas of research (*e.g.* anti-counterfeiting, macromolecular data storage).[27,28]

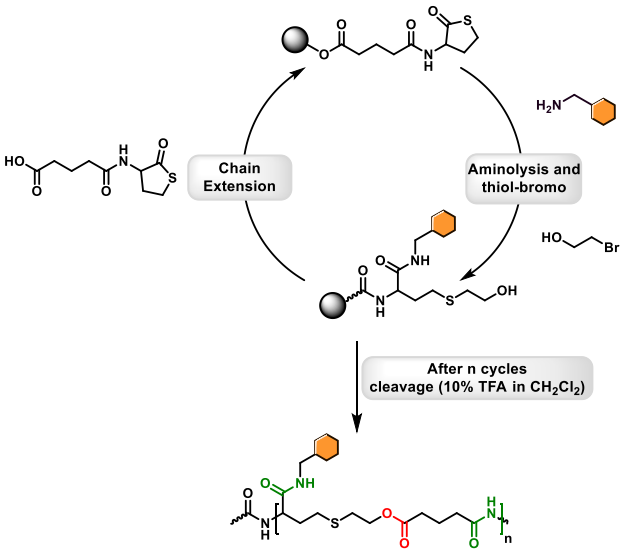


Figure 1: Overview of the protocol used for the synthesis of oligo(thioether ester)s

A chemical platform that enabled the synthesis of sequence-encoded macromolecules susceptible towards on-demand chemical degradation was developed by introducing esters in the backbone of the macromolecules. While esters are generally regarded as being bench-stable, they can be readily hydrolysed upon exposure to specific conditions (*e.g.* acidic/basic environment).[29] For this purpose, we modified a previously described procedure based on an amine-thiolactone-bromo conjugation.[30] Here, the thiolactone moiety can be ring-opened by means of any primary amine, followed by a thiol-bromo reaction with 2-bromoethanol to install an alcohol group (see **Figure 1**). The latter is an important prerequisite, as the alcohol is utilized for an esterification in the subsequent chain extension step to reinstate the thiolactone at the end of the oligomer. By using this protocol, every repeating unit is composed of two amides and one ester moiety, whereby the latter was used for the selective chemical cleavage of the macromolecules (*vide infra*). In addition, the wide range of commercially available primary amines enables the synthesis of macromolecules characterized by a high data storage density as a result of the expanded molecular alphabet.[20]

Initial experiments indicated that the previously used conditions (15 eq. amine and 20 eq. bromide in DMF) could also be used in this protocol, although a decrease in the reaction rate of the thiolactone aminolysis was observed when using aliphatic amines compared to ethanolamine, which was always used in previous studies. This observation can most likely be attributed to the absence of the neighbouring alcohol group, which is expected to facilitate the necessary proton transfer following addition of the amine to the thioester carbonyl.[31] Therefore, the time for each amine-thiol-bromo step was extended from one to two hours in order to reach near-quantitative conversions. By using these conditions, we attempted the synthesis of a sequence-defined tetramer (**M1**, see **Figure 2**). Following its cleavage from the solid-support, LC-MS and ESI-MS indicated that the macromolecule was obtained with high purity (see **Figure S4**-**S5**). In addition, we also investigated the stability of these macromolecules as a function of time, since this is an important prerequisite for applications such as anti-counterfeiting or macromolecular data storage. Therefore, a sample of **M1** was stored on the bench, open to air and ambient conditions, for six months. Aliquots were taken at different time intervals and analysed by LC-MS (see **Figure S13**). Those results did not indicate any sign of degradation within the measured timeframe, suggesting a good long-term stability under ambient conditions of the herein described oligo(thioether ester)s.

Having established the synthetic protocol, we started to explore whether the macromolecules could be efficiently sequenced by relying on chemically-induced random chain cleavages. Even though the nature of this process is uncontrolled, as opposed to the controlled unzipping of self-immolative oligomers, we postulated that all ester moieties within the macromolecules are equally prone towards hydrolysis. In this context, an important prerequisite is that all fragments required for the structure elucidation are present at a single point in time. In other words, either complete degradation or a non-effective cleavage of the ester bonds would result in the absence of certain fragments and thus prohibit the successful sequencing.

Initially, **M1** was treated with 2 eq. lithium hydroxide (LiOH) in an acetonitrile/water (5:1) mixture for five hours at room temperature to stimulate ester hydrolysis. Next, the reaction was quenched by the addition of acetic acid and analysed with ESI-MS. Interestingly, this MS^1 spectrum showed substantial resemblances to an MS/MS spectrum that is generally obtained (see **Figure S15**), and hence enabled a successful determination of the monomer order of **M1**. In this context, it should be noted that the difference of the monoisotopic mass between the starting codon (α-end repeating unit) that contains a primary amide, and a hydrolysed monomer that contains the same coding unit, is around one Dalton. Even though this difference is in principle sufficient to differentiate between an internal or end-group fragment, this issue was readily circumvented by developing an encoding scheme where a specific amine (butylamine in this case) is only used as the starting codon. Therefore, these results confirmed that the

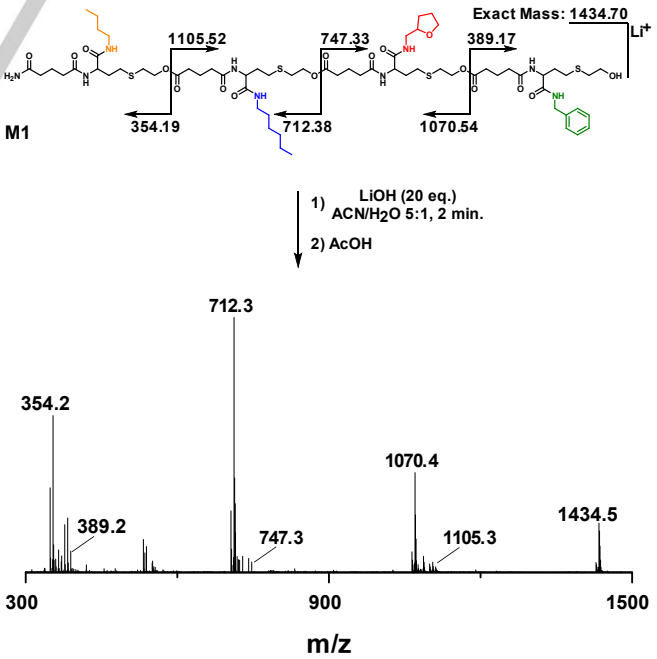


Figure 2: Hydrolysis reaction of **M1** and ESI-MS analysis of the reaction mixture. Structures were observed as their corresponding lithium adduct $[M+Li]^+$ during MS-analysis. Note that acid-containing fragments were observed in lower intensities, though their presence is not required for structure determination.

monomer order of oligo(thioether ester)s can be readily determined *via* a simple chemical treatment and conventional MS^1 analysis.

Encouraged by this proof of concept, we aimed to further increase the applicability of this approach by reducing the time required for fragmentation to occur. Therefore, **M1** was treated with 20 eq. LiOH in the same solvent system, and the reaction was quenched after two minutes. Intriguingly, ESI-MS analysis already contained all fragments required for an accurate sequence determination (see **Figure 2**). In addition, no significant amount of internal fragments, ascribed to ester hydrolysis of already fragmented molecules, could be observed. This suggests that the use of this short timeframe promotes hydrolysis of **M1** and simultaneously limits degradation of the generated fragmentation products, thereby resulting in an MS^1 spectrum that contains even more similarities to a regular MS/MS spectrum. It should also be noted that, in this case, only fragments containing the α-amide and a ω-hydroxyl moiety were used for the determination of the monomer order. This results from the observation that, in positive mode ESI-MS, truncated structures that contain an α-carboxylic acid and a ω-hydroxyl were generally observed in lower intensities (see **Figure 2**). Even though those fragments are not vital for a successful sequence determination, they could in principle provide an extra layer of redundancy.

Following the optimization of the sequencing conditions, a hexamer **M2** was synthesized to investigate whether the protocol would also be compatible with longer structures (see **Figure 3**). Therefore, **M2** was treated with 20 eq. sodium hydroxide (NaOH) in an acetonitrile/water (5:1) mixture for two minutes, after which the reaction was quenched and analyzed by ESI-MS. In this experiment, NaOH instead of LiOH was used to promote the ionization of the sodium adduct of all structures. Similarly, all fragments required for structure determination could be readily identified (see **Figure 3a**).

Simultaneously, a kinetic study was performed to gain insights in the formation of different fragments as a function of time. Because we observed near-complete degradation after five minutes (see **Figure S16**), a solution of **M2** in acetonitrile was cooled to 5°C to observe a more gradual increase of the expected fragments, after which it was treated with an aqueous NaOH solution (20 eq.). Next, an aliquot of the reaction mixture was taken every minute and analysed by ESI-MS (see **Figure 3b**). Even though peak intensities should be interpreted with caution in mass spectrometry, the analysis indicated that **M2** decreased as a function of time, combined with a gradual increase of all fragments. Interestingly, a full structure determination is already possible after one minute at 5°C since the corresponding MS^1 spectrum already contained all fragments, yet in a lower intensity.

To this point, all sequenced structures have been analysed by positive-mode ESI-MS, whereby fragments that contain the α-amide and a ω-hydroxyl were used for the determination of the monomer order. In addition, ester hydrolysis also results in truncated structures that contain an α-carboxylic acid and a ω-hydroxyl, which were generally observed in lower intensities in positive-mode MS-spectra (*vide supra*). To obtain a single series of fragments, we postulated that the sequence analysis of the

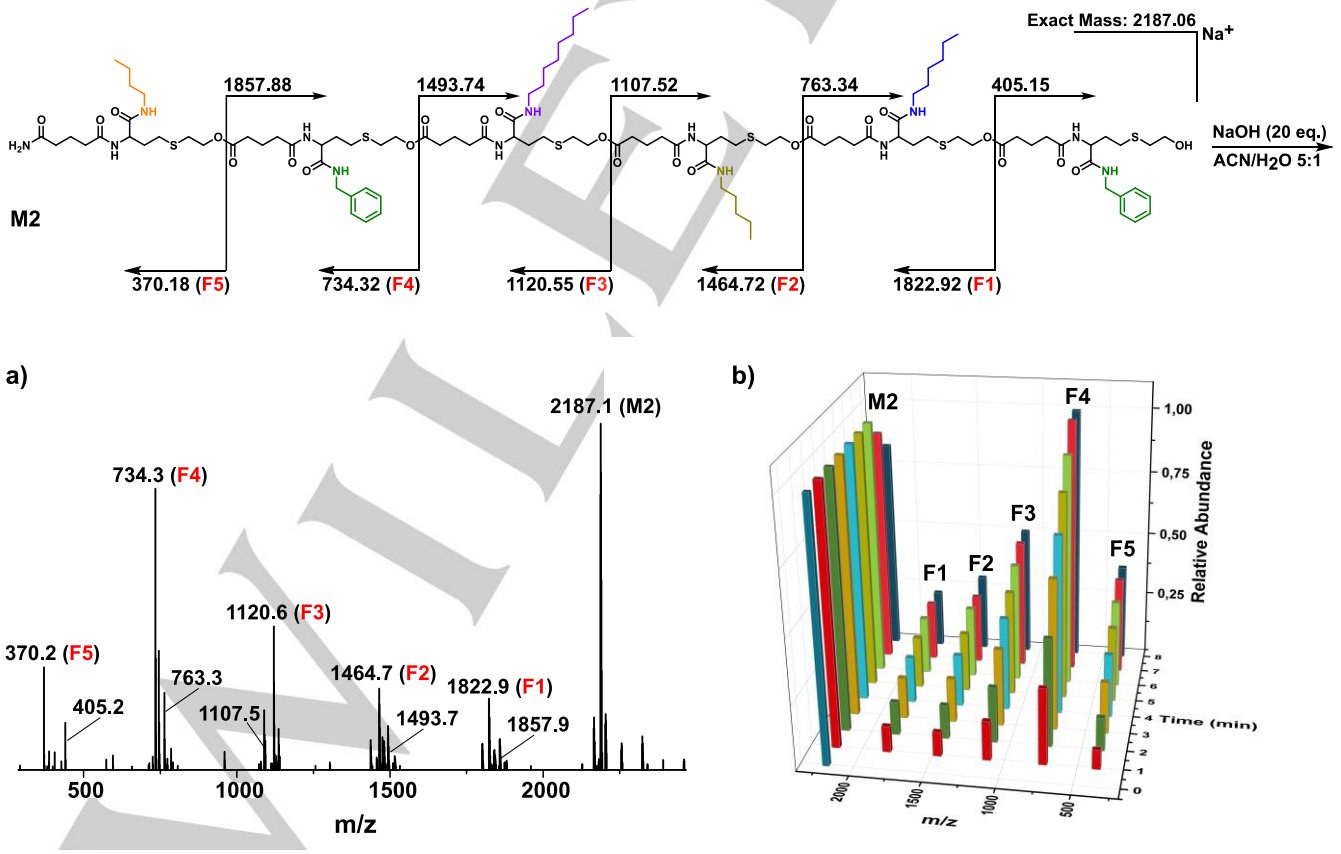


Figure 3: Hexamer **M2**. Fragments used for sequence determination are depicted in red. Acid-containing fragments provide a layer of redundancy but are not necessary for efficient structure elucidation. Structures were observed as their corresponding sodium adduct $[M+Na]^+$ during MS-analysis. a) Sequencing of **M2** and ESI-MS analysis after two minutes. b) Sequencing of **M2** and ESI-MS analysis as a function of time

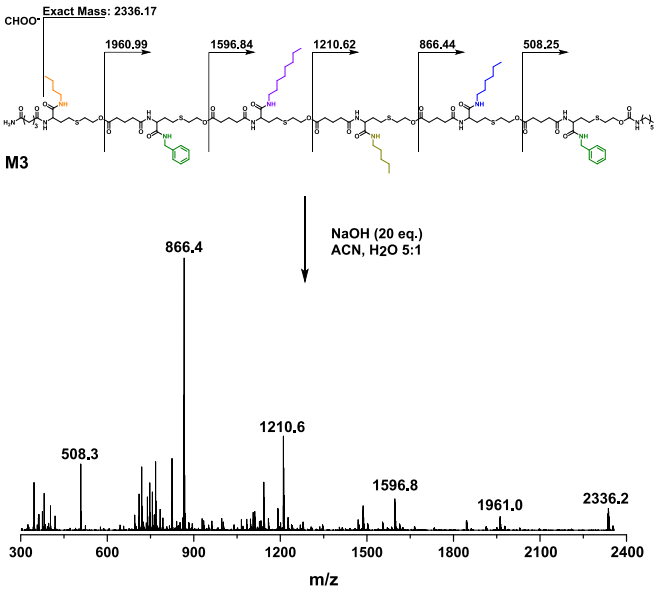


Figure 4: Sequencing and negative mode ESI-MS analysis of hexamer **M3**. Truncated structures were observed as their corresponding $[M-H]^-$ species, while **M3** was observed as the corresponding formic acid adduct $[\mathbf{M3}+CHOO]^-$.

macromolecules by negative-mode ESI-MS could promote the ionization of acid-containing fragments, while at the same time suppress the ionization of other fragments. Prior to sequencing, the ω-hydroxyl moiety was modified with hexyl isocyanate in order to differentiate the end-group from any other repeating unit in **M3**. Next, random chain scissions were induced by subjecting the macromolecule to the previously used hydrolysis conditions, followed by negative-mode ESI-MS analysis (see **Figure 4**).

All acid-containing fragments could be readily identified as their $[M-H]^-$ species, together with the formic acid adduct of the starting macromolecule $[\mathbf{M3}+CHOO]^-$. Despite the presence of other fragments, this experiment further highlighted the effectiveness and simplicity of this approach, as the monomer order of the oligomer can be read from the ω- to α-chain end in positive-mode, and *vice versa* in negative-mode.

Even though ESI-MS measurements provided a straightforward method for the structure determination of oligo(thioether ester)s, it was not possible to obtain a single series of fragments (*i.e.* only ω- to α-chain end fragments or *vice versa*), which is a consequence of the uncontrolled hydrolysis reactions. Therefore, we turned our attention to LC-MS to further facilitate the read-out of the monomer order. We postulated that the incorporation of a chromophore at the end of the macromolecules would result in an LC-trace that solely consists of chromophore-containing fragments. Indeed, only the latter will absorb at a specific wavelength, while the structures that do not contain the chromophore will not be observed. For this, **M1** was chain extended with a thiolactone moiety, followed by the introduction of 2,4-dinitrobenzene as a long-wavelength chromophore (**M4**). Following the cleavage from the solid-support, **M4** was treated with the previously developed sequencing conditions and analysed by LC-MS. Interestingly, the LC-trace at 360 nm contained a single series of crucial fragments, without the interference of a significant number of other peaks (see **Figure 5**). By calculating the mass difference between the different peaks in the LC-chromatogram, the monomer order of **M4** could be readily determined. Even though one would expect such an LC-trace in the case of a controlled depolymerization, as is the case for oligourethanes,[24,25] the installation of a long-wavelength chromophore conveniently delivered all truncated structures upon the introduction of random chain cleavages. In addition, the uncontrolled nature of the process enables the use of very short reaction times as the formation of all iterations occurs at a comparable rate.

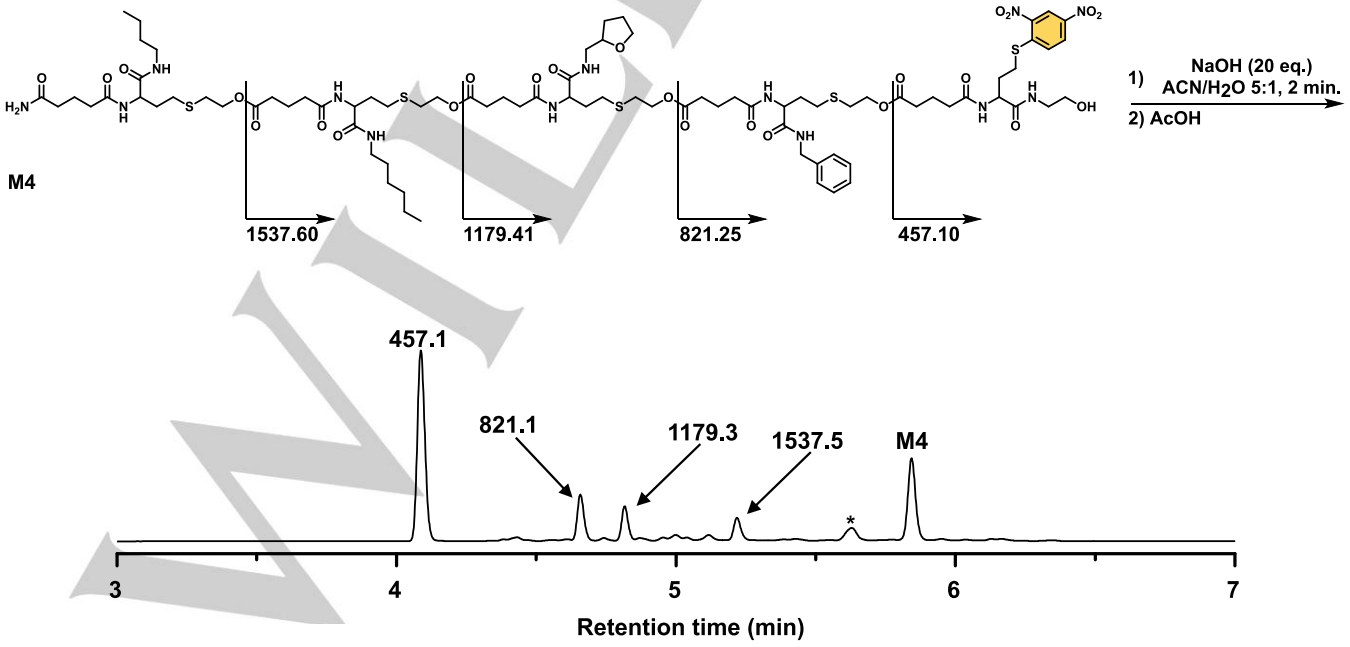


Figure 5: LC-trace (λ = 360 nm) obtained after two minutes of sequencing pentamer **M4**. A single series of fragments can be observed as a result of the presence of the chromophore. *: thioether oxidation of **M4**.

In summary, oligo(thioether ester)s can be readily sequenced by chemically-inducing random chain cleavages within the macromolecular backbone, followed by a single ESI-MS analysis. When using the optimized conditions, a hexamer could efficiently be sequenced by exposing the latter for just two minutes to strongly hydrolytic conditions and subsequent MS^1 analysis. In addition, the introduction of a long-wavelength chromophore conveniently delivers a single series of truncated structures upon LC-MS analysis, a feature that is generally only observed for self-immolative oligomers.

Consequently, this conceptually novel approach provides a simple and economical alternative to MS/MS techniques for the structure determination of sequence-defined macromolecules. Moreover, we believe that this strategy can be used for the sequencing of almost all macromolecular backbones that have the intrinsic ability to be (selectively) degraded upon the use of a specific stimulus. We hope that this approach will rapidly expand the number of macromolecules that can be sequenced by MS^1 analysis, thereby increasing the applicability of digital macromolecules. For example, the area of macromolecular anti-counterfeiting could greatly benefit from this method, since in this case the developer merely wants to investigate whether a (known) barcode is present in the labelled goods without the necessity for in-depth MS/MS analysis.[32]

Acknowledgements

The authors would like to thank Prof. F. Lynen and Pieter Surmont for providing access to the ESI-MS. The authors would also like to thank Dr. Kevin De Bruycker, Dr. Cody Gale and Dr. Nezha Badi for fruitful discussions. M.S. would like to acknowledge the Research Foundation Flanders (FWO) for the funding of his Ph.D. scholarship under application number 1SA3820N. F.D.P. would like to acknowledge financial support from the Research Foundation Flanders (FWO) under EOS-project 30650939. Also, this project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (CiMaC project - grant agreement No 101021081).

Keywords: Sequence-defined macromolecules • Sequence-controlled polymers • Sequencing • Polymer degradation • Random chain cleavages

[1] F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci.* **1977**, *74*, 5463–5467.

[2] A. Magi, R. Semeraro, A. Mingrino, B. Giusti, R. D'Aurizio, *Brief. Bioinform.* **2017**, *19*, 1256–1272.

[3] P. Edman, E. Högfeldt, L. G. Sillén, P.-O. Kinell, *Acta Chem. Scand.* **1950**, *4*, 283–293.

[4] B. T. Chait, R. Wang, R. C. Beavis, S. B. H. Kent, *Science (80-.).* **1993**, *262*, 89–92.

[5] D. J. C. Pappin, P. Hojrup, A. J. Bleasby, *Curr. Biol.* **1993**, *3*, 327–332.

[6] H. Steen, M. Mann, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 699–711.

[7] C. H. Becker, M. Bern, *Mutat. Res. Toxicol. Environ. Mutagen.* **2011**, *722*, 171–182.

[8] H. Mutlu, J.-F. Lutz, *Angew. Chemie Int. Ed.* **2014**, *53*, 13010–13019.

[9] A. Al Ouahabi, M. Kotera, L. Charles, J.-F. Lutz, *ACS Macro Lett.* **2015**, *4*, 1077–1080.

[10] K. S. Wetzel, M. Frölich, S. C. Solleder, R. Nickisch, P. Treu, M. A. R. Meier, *Commun. Chem.* **2020**, *3*, 63.

[11] N. Zydziak, W. Konrad, F. Feist, S. Afonin, S. Weidner, C. Barner-Kowollik, *Nat. Commun.* **2016**, *7*, 13672.

[12] X. Zhang, F. Gou, X. Wang, Y. Wang, S. Ding, *ACS Macro Lett.* **2021**, *10*, 551–557.

[13] Z. Huang, Q. Shi, J. Guo, F. Meng, Y. Zhang, Y. Lu, Z. Qian, X. Li, N. Zhou, Z. Zhang, X. Zhu, *Nat. Commun.* **2019**, *10*, 1918.

[14] B. Liu, Q. Shi, L. Hu, Z. Huang, X. Zhu, Z. Zhang, *Polym. Chem.* **2020**, *11*, 1702–1707.

[15] M. Soete, C. Mertens, R. Aksakal, N. Badi, F. Du Prez, *ACS Macro Lett.* **2021**, *10*, 616–622.

[16] J. M. Lee, M. B. Koo, S. W. Lee, H. Lee, J. Kwon, Y. H. Shim, S. Y. Kim, K. T. Kim, *Nat. Commun.* **2020**, *11*, 56.

[17] T. Mondal, V. Greff, B. É. Petit, L. Charles, J. F. Lutz, *ACS Macro Lett.* **2019**, *8*, 1002–1005.

[18] A. C. Boukis, M. A. R. Meier, *Eur. Polym. J.* **2018**, *104*, 32–38.

[19] S. Martens, A. Landuyt, P. Espeel, B. Devreese, P. Dawyndt, F. Du Prez, *Nat. Commun.* **2018**, *9*, 4451.

[20] E. Laurent, J.-A. Amalian, M. Parmentier, L. Oswald, A. Al Ouahabi, F. Dufour, K. Launay, J.-L. Clément, D. Gigmes, M.-A. Delsuc, L. Charles, J.-F. Lutz, *Macromolecules* **2020**, *53*, 4022–4029.

[21] R. Aksakal, C. Mertens, M. Soete, N. Badi, F. Du Prez, *Adv. Sci.* **2021**, *8*, 2004038.

[22] M. Soete, K. De Bruycker, F. Du Prez, *Angew. Chemie Int. Ed.* **2022**, *61*, DOI 10.1002/anie.202116718.

[23] C. Proulx, F. Noë, S. Yoo, M. D. Connolly, R. N. Zuckermann, *Pept. Sci.* **2016**, *106*, 726–736.

[24] S. D. Dahlhauser, P. R. Escamilla, A. N. VandeWalle, J. T. York, R. M. Rapagnani, J. S. Shei, S. A. Glass, J. N. Coronado, S. R. Moor, D. P. Saunders, E. V. Anslyn, *J. Am. Chem. Soc.* **2020**, *142*, 2744–2749.

[25] S. D. Dahlhauser, S. R. Moor, M. S. Vera, J. T. York, P. Ngo, A. J. Boley, J. N. Coronado, Z. B. Simpson, E. V. Anslyn, *Cell Reports Phys. Sci.* **2021**, *2*, 100393.

[26] T. Mondal, L. Charles, J. Lutz, *Angew. Chemie* **2020**, *132*, 20570–20573.

[27] J. O. Holloway, F. Van Lijsebetten, N. Badi, H. A. Houck, F. E. Du Prez, *Adv. Sci.* **2020**, *7*, 1903698.

[28] U. S. Gunay, B. E. Petit, D. Karamessini, A. Al Ouahabi, J. A. Amalian, C. Chendo, M. Bouquey, D. Gigmes, L. Charles, J. F. Lutz, *Chem* **2016**, *1*, 114–126.

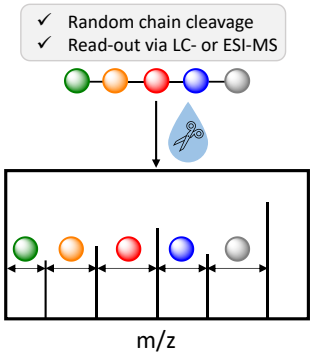
[29] S. L. Johnson, in *Adv. Phys. Org. Chem.*, **1967**, pp. 237–330.

[30] C. Mertens, M. Soete, M. L. Ślęczkowski, A. R. A. Palmans, E. W. Meijer, N. Badi, F. E. Du Prez, *Polym. Chem.* **2020**, *11*, 4271–4280.

[31] V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoët, J.-C. M. Monbaliu, O. Melnyk, *Chem. Rev.* **2019**, *119*, 7328–7443.

[32] H. Zhang, D. Hua, C. Huang, S. K. Samal, R. Xiong, F. Sauvage, K. Braeckmans, K. Remaut, S. C. De Smedt, *Adv. Mater.* **2020**, *32*, 1905486.

Entry for the Table of Contents



A strategy for the determination of the monomer order of sequence-defined macromolecules by making use of ESI- or LC-MS is described. This approach, which relies on the introduction of random chain cleavages within the macromolecular backbone, enabled the rapid chemical sequencing of abiotic macromolecules.

Institute and/or researcher Twitter usernames: @Filip_Du_Prez