

Rewritable Macromolecular Data Storage with Automated Read-out

Matthieu Soete,^[a] Kevin De Bruycker,^[a] and Filip 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

Abstract: Rewriting data stored on synthetic macromolecules is an interesting feature, even though it is considered as being quite challenging within the area of digital macromolecules. In this context, we initially studied a strategy for modifying the position tag of sequence-encoded macromolecules in a reversible manner. The efficiency of this method, which relies on the orthogonal cleavage of a thioester moiety via aminolysis, was demonstrated by modifying parts of an exemplary sentence. Simultaneously, a novel algorithm was developed to ease the read-out of macromolecular information by means of MS/MS techniques. This program, Oligoreader, can identify potential information-containing macromolecules from a series of MS¹ spectra, analyze the corresponding MS² spectra, and finally decode the data. Consequently, the algorithm simplifies the entire read-out process by avoiding any interference from the operator, which increases the potential for blind sequencing of uniform macromolecules.

Introduction

The storage of digital information on macromolecular chains has recently attracted widespread attention as a result of the many attractive features associated with it.^[1,2] In this context, it has been demonstrated that, by using a defined monomer sequence, both biopolymers (*i.e.* DNA) and synthetic macromolecules can successfully be employed for the storage of information.^[3–11] For instance, the possibility to store more than 200 megabytes of data on DNA strands has already been described, while our and other research groups successfully encoded a QR-code, a figure, or an exemplary sentence on a set of synthetic macromolecules.^[11–15] The focus of these studies is often directed towards the development of an archival data storage medium, whereby rarely-accessed information that has to be saved for an extended period of time, *e.g.* legal documents, is stored on information-containing macromolecules.^[16]

In contrast to writing data on sequence-defined oligomers, little attention has been directed towards editing of information written on these macromolecules. Nevertheless, this could avoid significant (synthetic) efforts when, for example, only small parts of the archived data have to be adjusted. Moreover, since writing of the information is generally regarded as the bottleneck of molecular data storage applications, modifying the information rather than re-synthesizing a complete set of macromolecules is an attractive approach to minimize expensive and time-consuming synthesis steps.

Manipulating data stored on macromolecules can be a complex process that, for example, requires the use of specialized bacterial machinery in the case of nucleic acid sequences.^[17,18] Nevertheless, chemists have the ability to incorporate a wide range of functional groups that can be adapted once the macromolecules have been synthesized. For example, Lutz and co-workers first described the development of oligo(alkoxyamine

amide)s, of which the stored information could be erased by simply heating the macromolecules.^[19] Using an analogous concept, our research group incorporated dynamic covalent bonds in the backbone of information-containing macromolecules.^[20] This feature allowed to scramble the stored data, after which it could only be decoded using the correct pin code. Nonetheless, features such as erasing or scrambling are mainly interesting when targeting anti-counterfeiting applications, yet provide little to no added value for the long-term storage of information.

Rather than editing data by degrading the macromolecular backbone, Lutz and co-workers demonstrated the synthesis of a truly editable information carrier.^[21] More specifically, both light-sensitive and light-inert moieties were incorporated in the side chain of sequence-encoded poly(phosphodiester)s, whereby the sequence could effectively be altered upon irradiation. This was achieved via a well-considered monomer design in which the mass of the light-sensitive moiety was identical to that of the light-inert one (*o*- and *p*-nitrobenzyl, respectively). Irradiating the macromolecule resulted in the loss of the *o*-nitrobenzyl group and thus in the formation of the corresponding free alcohol, hence effectively editing the information. However, only a single coding monomer was changed from a corresponding digital “1” to “0”, which means that the data before and after the irradiation step is both defined by the initial synthesis. Furthermore, no attempts were made to perform additional adjustments to the code by, for example, reverting the mutation and thereby returning to the original state. Even though such a reversible editing cycle introduces some intrinsic difficulties, it would be of undeniable interest within the field of information-containing macromolecules, as (part of) the encoded information could then be changed in a manner similar to conventional storage media.^[22]

In this work, we introduce a straightforward strategy for the reversible post-synthesis modification of the macromolecule's end-group and validate the approach on an exemplary set of information-containing oligomers. More specifically, a series of sequence-encoded macromolecules was equipped with a specific end-group, hereafter referred to as the *position tag*, to discriminate the position of a data fragment within the ensemble of information. The position tag of a macromolecule was easily removed via a simple chemical treatment, while the addition of a new tag effectively resulted in a mutated data fragment. This approach thus allows the reversible insertion and deletion of data within an existing piece of information.

In parallel with the development of a editable position tag, we aimed for the design of an algorithm that enables the automated read-out of data stored on a set of macromolecules, both before and after the editing process. In contrast to other programs developed for the read-out of a single oligomer, this algorithm is capable of scanning a series of MS¹ spectra to identify potential information-bearing macromolecules, to analyze the corresponding MS² spectra, and thereby to decode the

RESEARCH ARTICLE

information stored on a mixture of macromolecules. Making use of this program, a complete sentence, which was stored on a series of different macromolecules, could be automatically decoded within a single measurement run without the interference of an operator.

Results and Discussion

We have recently described the synthesis of oligo(thioether urethane)s, a promising class of dual sequence-encoded macromolecules based on thiolactone chemistry.^[23] These are obtained via aminolysis of the thiolactone, which introduces the first coding monomer, while the subsequent thiol-epoxy reaction installs the second coding monomer, followed by a chain extension reaction (see **Figure 1a**). Since the high storage capacity of these macromolecules facilitates the encoding process of larger pieces of data, we envisioned their use for the introduction of the rewritable position tag. Importantly, the substitution of the end-group has to occur in an orthogonal manner and the used reaction conditions should not affect any other functional group present in the macromolecule (*i.e.* amide, urethane and thioether moieties), as this would alter the encoded information in an uncontrolled manner.

The first investigations were directed towards the introduction of an ester end-group via a previously described amine-thiol-ene conjugation based on thiolactone chemistry.^[24,25] Here, an acrylate introduces the unique mass of the position tag. Removal of the tag was anticipated to occur via ester hydrolysis, yielding the free carboxylic acid, which should be readily alkylated to introduce a position tag with a different mass (see **Figure S1**). At first, a systematic model study was conducted to investigate

whether this envisaged rewriting process would be feasible without significant degradation of the other functional moieties in a typical macromolecule. The model compound, which was obtained via thiolactone ring-opening and an alcohol-isocyanate coupling, was treated with a series of basic hydrolysis conditions (see **Figure S2-S7**). The screening of reaction conditions indicated that the use of 1,5,7-triazabicyclo[4.4.0]dec-5-ene (TBD, 2 eq.) in a 9:1 THF:H₂O mixture for two hours resulted in the selective hydrolysis of the ester moiety. Next, these conditions were applied on an oligo(thioether urethane) that was end-functionalized with an ester group. Despite the successful model studies, partial degradation of the macromolecule was observed as a result of urethane hydrolysis (see **Figure S8**). Decreasing the reaction time or the amount of base could not completely suppress this side reaction. While this strategy might be applicable to other polymer backbones, alternative strategies were investigated that could enable an orthogonal removal of the end-group and allow the re-introduction of a series of different mass tags.

An alternative approach that was simultaneously explored focused on the use of thioesters as the distinctive mass tag. These are readily obtained from a thiolactone-bearing macromolecule through aminolysis and subsequent functionalization of the released thiol. In addition, thioesters are characterized by a good hydrolytic stability but are significantly more prone towards reactions involving amines and thiols.^[26] Consequently, it was hypothesized that the removal of the end-group could occur via a simple aminolysis reaction, followed by the introduction of a novel mass tag via a thioesterification (see **Figure 1b**). The consecutive series of model studies indicated that a methylthioester could be readily cleaved within one hour by using an excess of butylamine (10 eq.) in THF (see **Figure S9**).

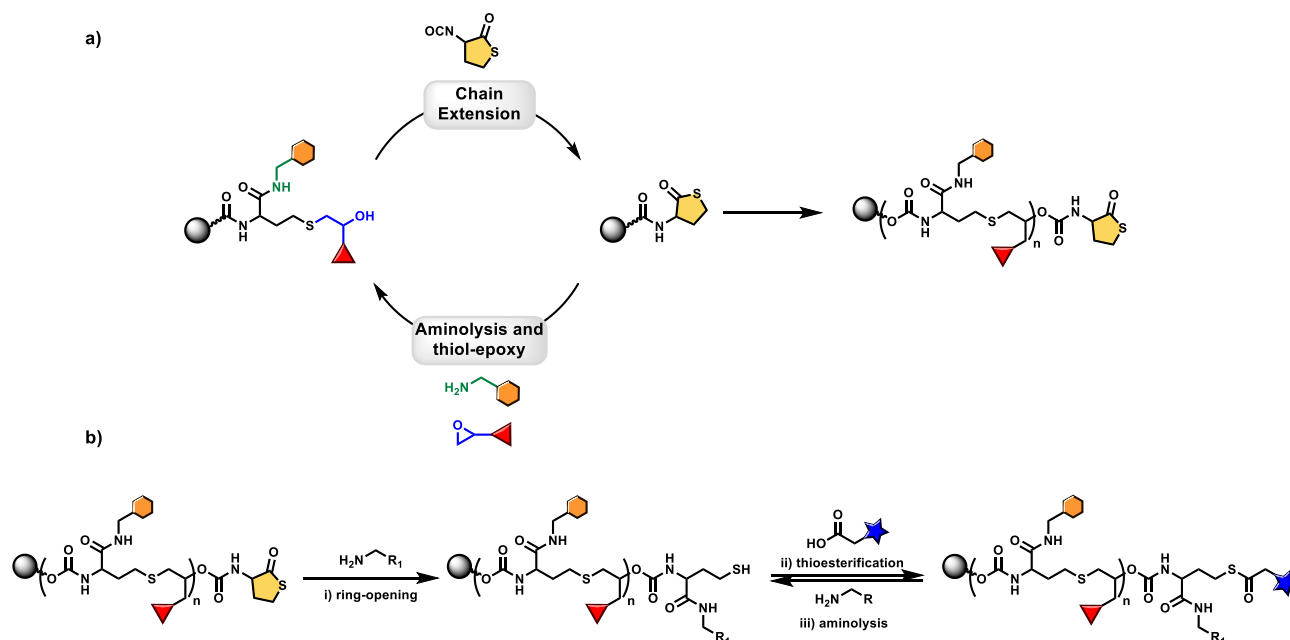


Figure 1: a) Strategy used for the synthesis of sequence-encoded macromolecules. Both the amine and epoxide introduce a distinct moiety, yielding dual-encoded macromolecules. b) Proposed strategy for the reversible modification of the position tag of oligo(thioether urethane). The hexagon represents the functional group introduced by the amine moiety, while the red triangle represents the functional group introduced by the epoxy group during the synthesis of the macromolecules. i) thiolactone ring-opening ii) thioesterification for the introduction of the unique mass tag (represented by a blue star) iii) aminolysis of the thioester for the removal of the tag.

RESEARCH ARTICLE

More specifically, LC-MS analysis showed that, instead of the free thiol, the corresponding disulfide was obtained as the sole product. Dimethylphenylphosphine (DMPP) was thus employed as a reducing agent to obtain the free thiol, followed by a carbodiimide coupling with a different acid (*i.e.* hexanoic acid).^[27] This chemical protocol effectively resulted in the transformation of the end-group thioester, thereby demonstrating its feasibility (see **Figure S10**). As the aim was to develop a reversible rewriting cycle, the removal process was also investigated for the more hydrophobic hexanoic acid-based thioester. Notably, applying the previously used conditions resulted only in a moderate removal of the thioester group, and hence only a minor fraction of the expected disulfide could be detected. Even though full conversion could be reached when extending the reaction time to 14 hours and increasing the amount of amine, we observed that 1,8-diazabicyclo(5.4.0)undec-7-ene (DBU) is an efficient catalyst for the thioester amidation. Indeed, employing a catalytic amount of DBU (0.1 eq.) effectively reduced the reaction time to three hours. Interestingly, replacing butylamine by pyrrolidine whilst keeping all other parameters constant even further decreased the time required for this protocol, since in this case the amidation reaction already reached 92% conversion after 30 minutes, and near-quantitative conversions were obtained after just one hour (see **Figure 2**).

Next, the developed reaction conditions were applied on a macromolecular level. While the conditions used for the previously described ester hydrolysis were ineffective in avoiding degradation of the macromolecular backbone, the thioester-based strategy readily transformed the end-group of the macromolecule without any observable traces that are indicative of undesired side reactions (see **Figure S13**). This observation was in line with our expectations, given the fact that the conditions used for the synthesis of the macromolecules (*i.e.* 10 eq. of amine and catalytic amount of DBU) are identical to the ones used for the removal of the thioester moiety.

Finally, we investigated the stability of the thioester-based (macro)molecules as a function of time. Indeed, for applications in macromolecular data storage, the stability of the molecule is paramount for data retention. To demonstrate the long-term stability of these moieties, a thioester-containing model compound was placed in an open vial on the bench for two months. LC-MS analysis at different time points did not indicate any sign of thioester degradation within the measured time frame

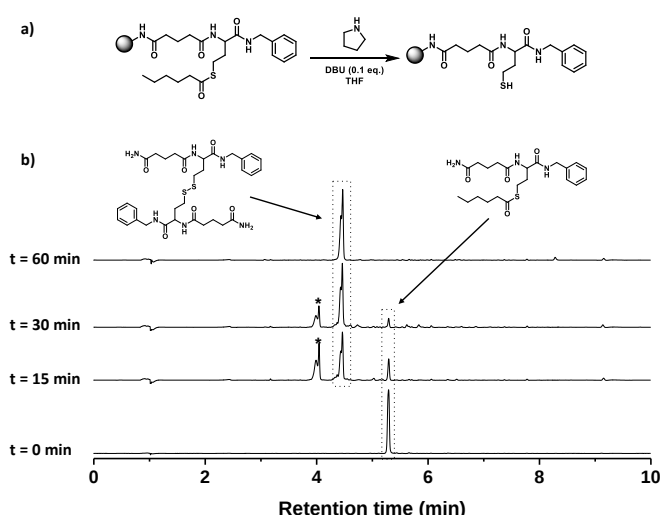


Figure 2: a) overview of the model reaction that was used to determine the time required for complete amidation of the pentylthioester. Reaction conditions: solid-phase bound thioester (1 eq.), pyrrolidine (10 eq.), DBU (0.1 eq.) in THF. (b) Reaction progress followed by LC-MS analysis ($\lambda = 214$ nm) as a function of time. The reaction results in the formation of the free thiol (*), which is readily oxidized to its disulfide analogue in a basic environment (retention time = 4.46 mins).

(see **Figure S14**). Therefore, the presented approach is not only composed of mild writing and rewriting conditions, it also makes use of bench-stable moieties.

Following a series of model studies, the applicability of the proposed rewriting approach was demonstrated by means of an example. In order to do so, a 'molecular alphabet' based on the aforementioned amine-thiolactone-epoxy protocol was developed. More specifically, it was hypothesized that a mass difference of at least 4 Dalton amongst the different repeating units would be sufficient to enable a straightforward read-out of the macromolecules *via* tandem MS techniques. When making use of solely commercially available amines and epoxides, a total of 32 unique combinations were identified that match this criterium. For the specific example discussed in this work, 26 combinations were used so that each amine-epoxide pair directly corresponds to a letter in the (Latin) alphabet (see **Table S1** for an overview of the molecular alphabet). Nevertheless, applications in data storage could also use a binary or a more data-dense hexadecimal encoding scheme.

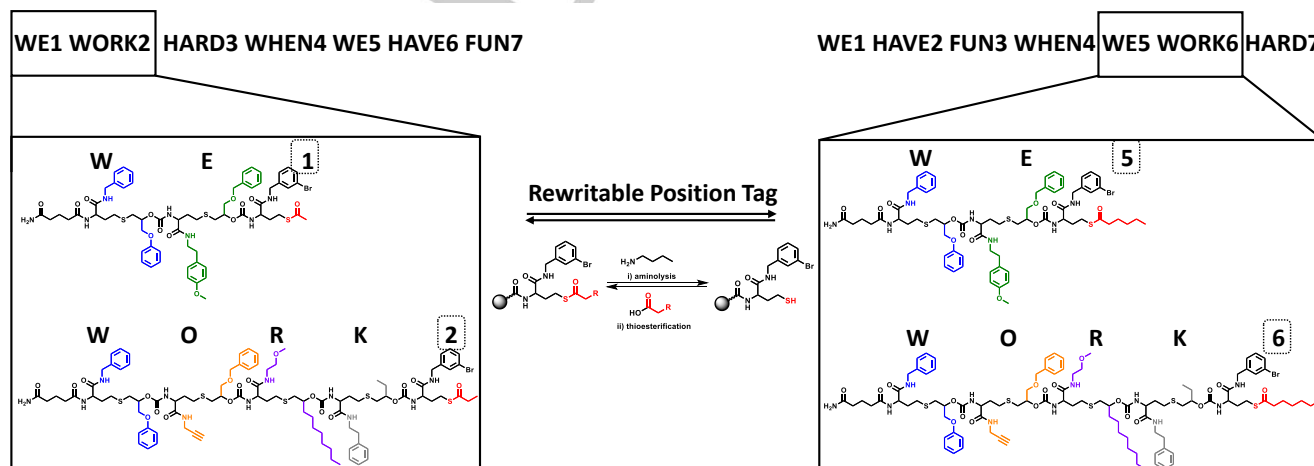


Figure 3: The sentences encoded on the macromolecules, whereby the chemical structure of a selection of oligomers is shown. The strategy for the reversible editing of the position tag is displayed in the middle.

RESEARCH ARTICLE

By making use of the defined molecular alphabet, the sentence "WE WORK HARD WHEN WE HAVE FUN" was successfully encoded on a set of seven different macromolecules, whereby each word of the sentence corresponded to a different oligomer. In this context, it should be emphasized that this specific sentence merely serves as an example to demonstrate the concept of the editable position tags. Indeed, the herein developed approach can also be used to write (and rewrite) information by assembling parts of a pre-synthesized macromolecular library.

Finally, suitable (editable) position tags were added to the macromolecules in order to allow for the correct decoding of the data, resulting in "WE1 WORK2 HARD3 WHEN4 WE5 HAVE6 FUN7". For this, the thiolactone end-groups of the oligomers were first ring-opened by using 3-bromobenzylamine. The introduction of the bromine ensures that, during the read-out *via* MS/MS, potential overlap between the end-group and any of the repeating units can be identified as a result of the unique isotopic pattern associated with the bromine group. Next, the free thiol was coupled with a series of acids to yield the different position tags for each macromolecule. For example, the thiol moiety of the first macromolecule ('WE1') was treated with acetic acid to yield the methylthioester as the unique mass identifier within the series of position tags (see **Figure 3** and **Table S1**).

Given the successful macromolecular encoding of the exemplary sentence, we aimed to read-out the data before attempting the rewriting process. In our previous work, we relied on MALDI-TOF/TOF as the method of choice for the fragmentation and subsequent determination of the monomer order of sequence-defined macromolecules.^[23] Moreover, an algorithm that is capable of translating a set of MS² spectra into the original

data has also been developed.^[13] Nevertheless, one of the drawbacks of using MALDI-TOF/TOF is the time required to collect all individual spectra, combined with a suboptimal resolution in the MS² spectrum. In addition, the algorithm that our group and others previously developed generally requires knowledge regarding the macromolecules that are being decoded, *i.e.* the molar mass of the oligomer has to be known to the operator - both for peak selection in MS¹ and data read-out - limiting the possibility for blind sequencing.

In an attempt to reduce the overall time required for the read-out process, we used ESI-MS/MS for the determination of the monomer order in all macromolecules. More specifically, a full scan MS with data-dependent tandem mass spectrometry (MS/dd-MS²) measurement on an ESI-MS instrument automatically identifies the most abundant ions in the acquired MS¹ spectrum, and subsequently records an MS² spectrum for each of them. In addition, parent ions for which a fragment spectrum has been recorded are excluded in following MS/dd-MS² scans, thereby minimizing the chance of missing the macromolecule of choice. Even though the chromatographic peak of an injected sample is merely 10 seconds, the scan rate is sufficiently high to obtain MS² spectra for all oligomers of the exemplary sentence with a single injection.

To facilitate the analysis of the large number of recorded spectra, we also developed an algorithm, Oligoreader, that fully automates the entire read-out process. Firstly, the algorithm averages out all MS¹ spectra that are recorded within the chromatographic peak. Next, based on a previously described tool,^[28] a library of all potential macromolecules that can be formed within a postulated *m/z* range are determined depending on the given molecular

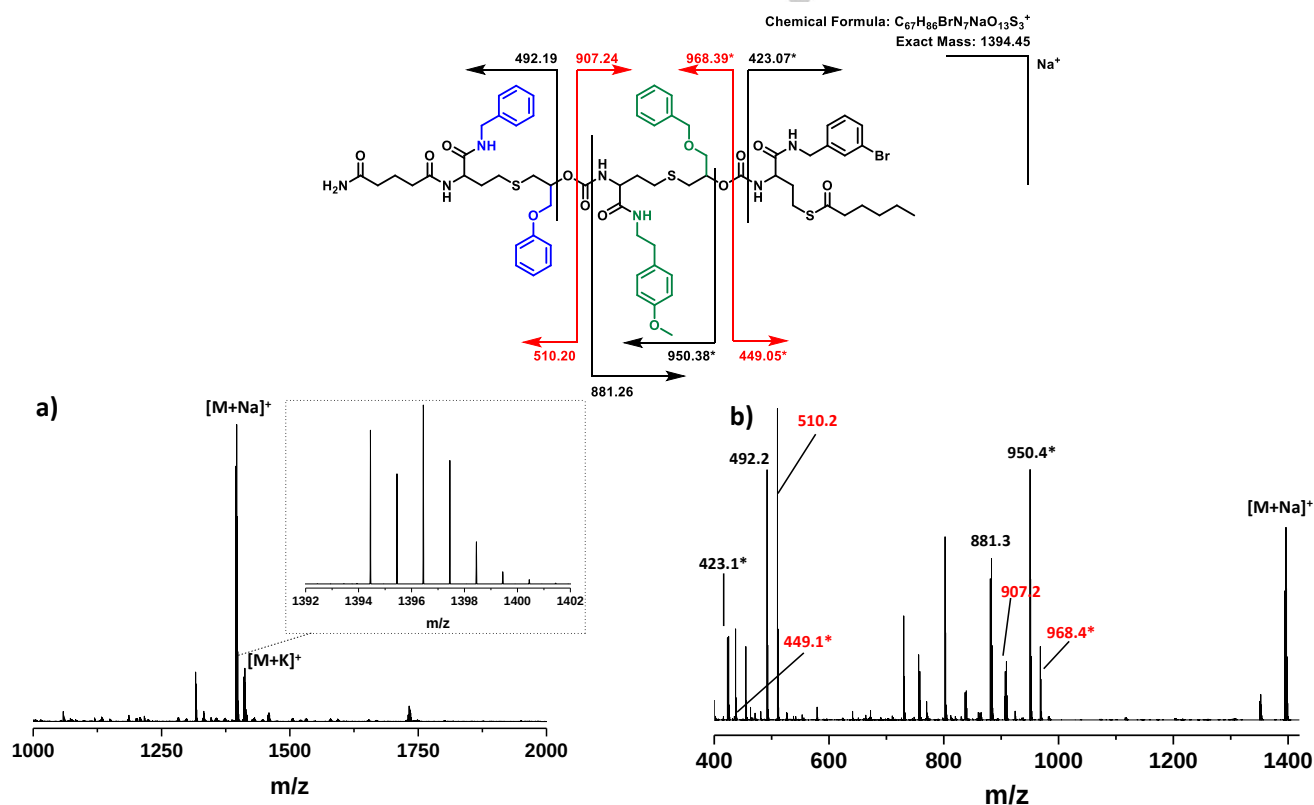


Figure 4: a) ESI-MS analysis of WE5, displaying the sodium and potassium adduct of the product. b) ESI-MS/MS analysis (normalized collision energy (NCE) = 30) of WE5. The peaks originating from fragmentation through urethane decarboxylation are shown in black, and the peaks corresponding to the reversed alcohol-isocyanate reaction are shown in red. All assigned values correspond to the monoisotopic mass of the molecules.* = fragments that originate from the position tag.

RESEARCH ARTICLE

alphabet and position tags. The algorithm then simulates the theoretical isotopic patterns for all of these macromolecules and compares them to the experimentally obtained spectra. If the matched experimental peaks account for at least 50% of the simulated (normalized) isotopic envelope, the theoretical macromolecule is assumed to be present in the MS¹ spectrum.^[28] Interestingly, this means that the analysis of the averaged MS¹ spectrum already indicates which macromolecules can be present, yet without information on the exact order of repeating units. For example, when the macromolecule 'WORK2' is matched, the output of the algorithm after MS¹ analysis is 'KORW2', as the letters are ordered alphabetically when no information about the order of repeating units is available.

Even though Oligoreader is able to identify information-containing macromolecules in an MS¹ spectrum and to deliver the corresponding building blocks, it is not possible to actually decode the information without MS² analysis. In previous work, we observed predominant fragmentation of the urethane bonds into an amine and an olefin, which were the result of a decarboxylation process.^[13,23] However, as we now relied on ESI-MS/MS instead of MALDI-TOF/TOF, we manually injected sequence WE5 and analyzed the obtained MS² spectrum. In doing so, four unique fragmentation sets were readily identified, all of which were the result of urethane bond fragmentation (see **Figure S15** and **Table S2** for an overview of all fragments). Apart from the decarboxylation pathway, the presence of the reversed alcohol-isocyanate addition was also observed (see **Figure 4**). Taking advantage of this observation, all pathways were included in the algorithm to increase the robustness of the read-out process. Indeed, even if some fragments are missing occasionally, the information can still be decoded as a result of the redundancy provided by the fragmentation pathways, avoiding the need for

extensive error correcting schemes. Therefore, when provided with an MS² spectrum and the constituting building blocks for the parent ion, Oligoreader is able to determine the specific monomer sequence of the investigated macromolecule.

It should be noted that we also observed for the first time thioether fragmentation in the MS² spectra. We are currently investigating these fragments as a way for the determination of every individual side chain of the macromolecules, which might open new avenues to consider the amine and epoxide as separate entities in the molecular alphabet, rather than as a pair.

Following the optimization of the algorithm, sequence WE5 was measured again via an automated injection with the mass spectrometer running a full MS/dd-MS² analysis as explained above. This resulted in a total of 564 spectra (94 MS¹ and 470 MS²), whereby the 66 spectra (11 MS¹ and 55 MS²) that were recorded within the chromatographic peak served as the input for Oligoreader. First, the algorithm matched the isotopic pattern of the injected sequence in the MS¹ spectrum to one of the compositions in the library, namely 'EW5'. Next, all MS² spectra resulting from a parent ion within the matched isotopic envelope, together with the identified macromolecular composition, are passed on for sequencing. During the MS² analysis, it should be noted that a distinction can be made between 'backbone' and 'label' fragments. Indeed, the latter fragments only provide information about the ω chain-end, i.e. the label, or more specifically the position tag (see **Figure 4**). Although the position tag is already determined via the MS¹ analysis, i.e. '5' in this example, these fragments in fact provide an error-check of the data passed to the sequencing algorithm and therefore allow to identify false-positive MS¹ matches. The 'backbone' fragments, on the other hand, allow to differentiate between WE5 or EW5, and are thus key for an accurate data decryption. In conclusion,

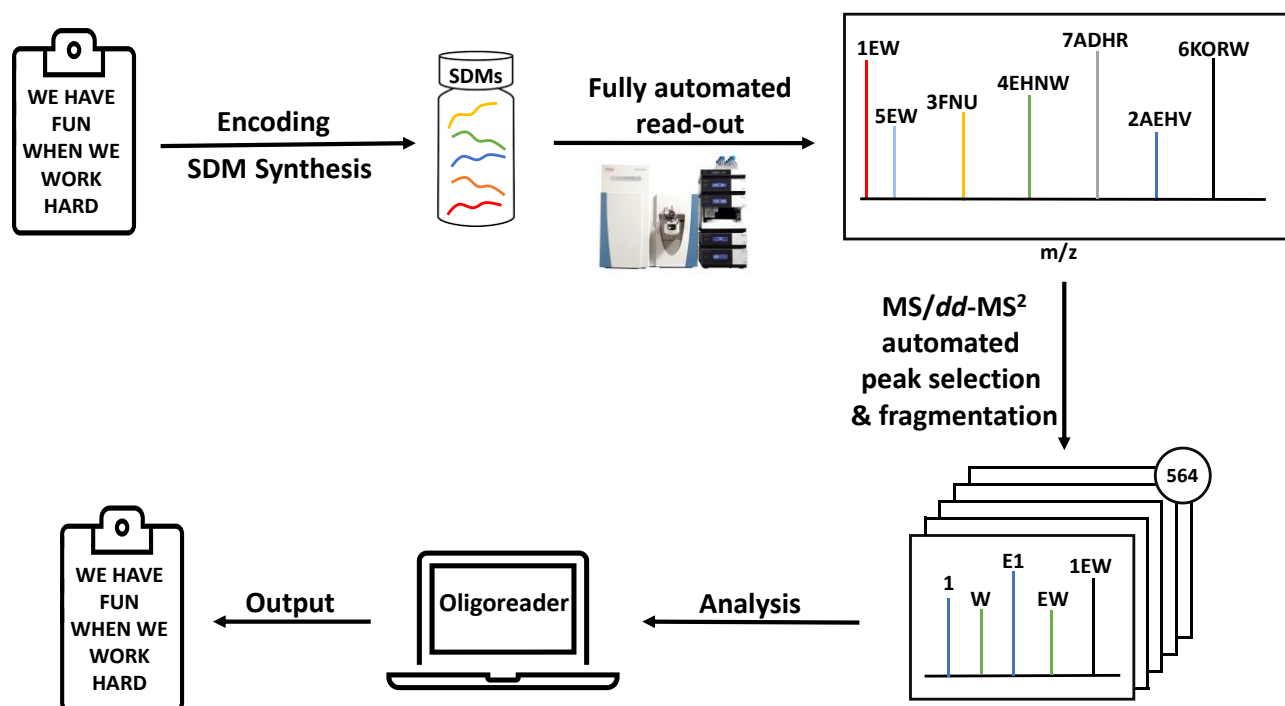


Figure 5: Schematic representation of the read-out process. A mixture of information-containing sequence-defined macromolecules (SDM) is, without any form of purification, analyzed automatically in the full MS/dd-MS² option to yield a library of full MS and MS² spectra. Analysis of this data by Oligoreader then yields the encoded sentence as the output.

RESEARCH ARTICLE

Oligoreader was able to autonomously determine the correct information-containing macromolecule, *i.e.* WE5, from all the recorded spectra (see supporting information for the full output). Since the performance of Oligoreader for WE5 proved to be excellent, the algorithm was further tested with the other macromolecules constituting the exemplary sentence. Nevertheless, to allow for a more controlled environment, we first characterized all individual macromolecules by ESI-MS and manually selected the appropriate parent ion for MS/MS analysis. In all cases, Oligoreader was able to successfully decode the monomer order of the macromolecules. In addition, the scope of the algorithm was further expanded to enable the correct reassembly of the individual data fragments. In the case of our exemplary data, the algorithm thus provided the entire sentence as the output, rather than the individual words encoded on the macromolecules.

Because Oligoreader was able to accurately decipher every word of the sentence, all seven macromolecules were measured via the automated full MS/dd-MS² analysis as well. Even though this resulted in a total of more than 4000 spectra, Oligoreader was still able to effectively translate the macromolecular information to the sentence "WE WORK HARD WHEN WE HAVE FUN". Notably, the entire read-out process consisted of simply placing seven solutions, each of which contained a single macromolecule, in the ESI-MS autosampler, initiating a series of full MS/dd-MS² runs, and finally launching Oligoreader with the collected data. This means that any operator can readily decode pieces of macromolecular information, even without prior knowledge of the analyzed macromolecules, since no precursor ion selection for fragmentation is required.

Lastly, we hypothesized that the overall runtime of the read-out process could be further reduced by analyzing all macromolecules in a single run. Even though we anticipated that the time required for recording an MS/MS spectrum of every single macromolecule could potentially exceed the chromatographic peak time, a stock solution that contained all seven macromolecules was analyzed via the automated full MS/dd-MS² analysis. Interestingly, Oligoreader was able to readily elucidate the encoded sentence without the need for further error-corrections, despite the fact that seven oligomers were involved in a single measurement (see **Figure 5** for a schematic representation). Consequently, Oligoreader can be regarded as an extremely robust read-out algorithm that is able to decode large portions of macromolecular data.

With the development of Oligoreader being successful, the rewriting of macromolecular information was further investigated. For this, all resin-bound macromolecules were treated with a solution of pyrrolidine in THF with DBU as catalyst. During the course of the reaction, an aliquot was taken from HAVE6 after two hours in order to assess whether the thioester removal would be significantly slower on macromolecular level. Even though this transformation of the hexylthioester to the free thiol only induced a moderate shift in the hydrodynamic volume, the base-catalyzed oxidation to the corresponding disulfides resulted in a distinct peak shift in SEC measurements (see **Figure S13**).^[27] These measurements indicated that near-complete removal of the thioester end-group was achieved within two hours. Nevertheless, all batches were left to react overnight in order to ensure full conversion. Subsequently, the new position tags were added to the different macromolecules to switch the order of the words and

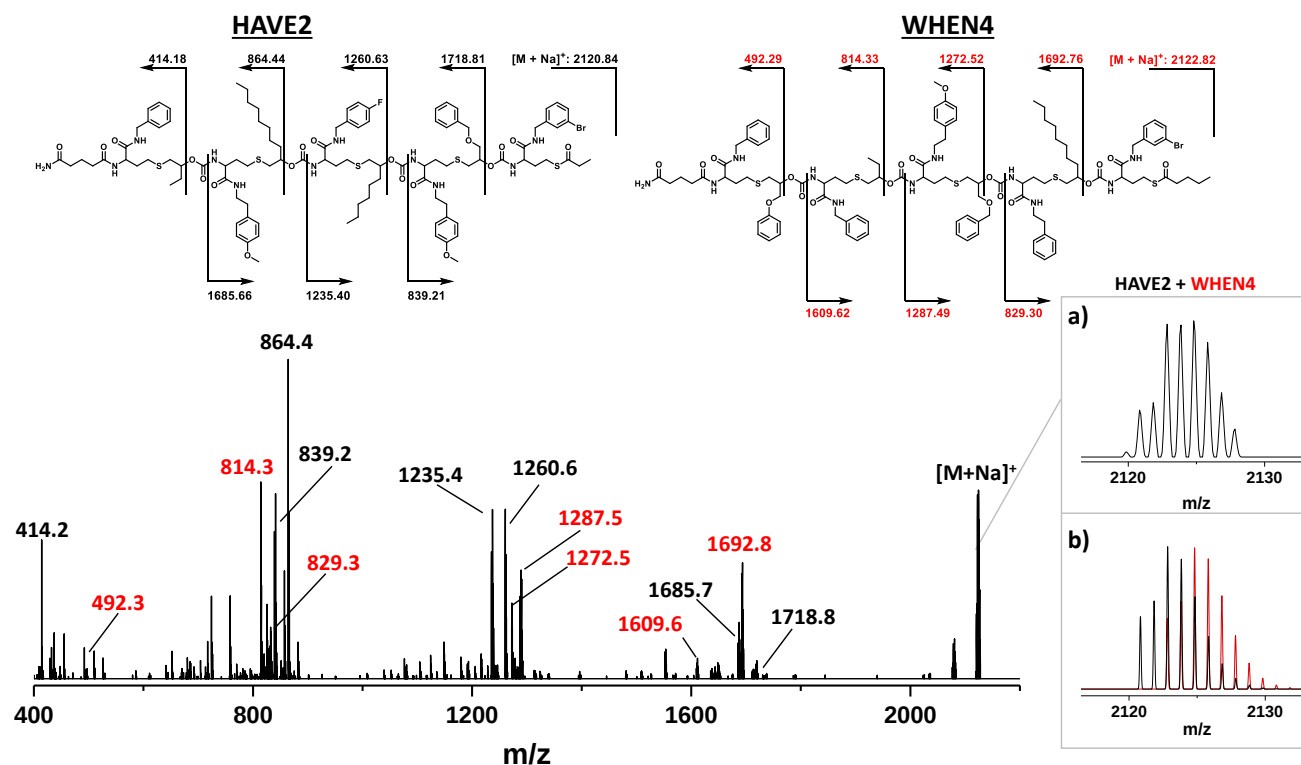


Figure 6: MS/MS spectrum of the precursor ion (2123.83 ± 5 Da). a) experimentally observed isotopic pattern for HAVE2 and WHEN4. b) theoretical isotopic envelope of HAVE2 (black curve) and WHEN4 (red curve). For simplicity reasons, only fragments from the decarboxylation pathway are indicated on the chemical structures. Note that HAVE2 and WHEN4 are almost isobaric, hence the MS/MS spectrum contains fragments of both macromolecules, yet Oligoreader is able to determine the information stored in both macromolecules based on this single MS² spectrum.

RESEARCH ARTICLE

thus to effectively rewrite the sentence into “WE HAVE FUN WHEN WE WORK HARD”. Analysis of the resulting macromolecules *via* ESI-MS showed that the reactions occurred efficiently, as no significant amounts of the original products could be observed (see **Figure S30-S37**), which was further confirmed by Oligoreader (*vide supra*). In contrast to conventional approaches, whereby the re-synthesis of four macromolecules (HAVE, FUN, WORK, HARD) would be required in order to alter the meaning of this specific sentence, *i.e.* to rewrite the data, a straightforward two-step post-modification was sufficient to obtain the same result.

During this project, we also observed that the macromolecules HAVE2 and WHEN4 are (unintentionally) almost isobaric, as their monoisotopic masses correspond to 2120.84 Da and 2122.82 Da, respectively. While this is not an issue when using a conventional read-out approach (*i.e.* ionize and fragment all macromolecules separately), it does increase the complexity of the presented read-out process. Indeed, the full MS/dd-MS² analysis is set up to isolate an *m/z* window of 10 for MS² analysis, which allows to record a fragmentation spectrum of (nearly) the complete isotopic envelope of the molecule irrespective from the specific isotope selected by the mass spectrometer. While this proved to be important for the automated sequencing algorithm, any MS² spectrum of either HAVE2 or WHEN4 will always contain fragments of both macromolecules (see **Figure 6**). In other words, Oligoreader must identify two distinct macromolecules from a single MS² spectrum in order to successfully decode the data. Despite the potential pitfalls, Oligoreader was still able to convert the macromolecular information to the data “WE HAVE FUN WHEN WE WORK HARD” (see supporting information for full algorithm output). Finally, no traces of the original macromolecules could be identified, thereby further confirming the efficiency of the herein developed rewriting approach.

Conclusion

We have introduced a simple post-modification strategy for the rewriting of data stored on synthetic macromolecules. This approach makes use of the orthogonal cleavage of a thioester moiety by an amidation reaction, while a new label can be readily added *via* a thioesterification. Apart from a series of model studies, the feasibility of this method was also confirmed on a macromolecular level by means of a demonstrator. In doing so, a simple sentence, which was encoded on seven different macromolecules, could be effectively edited by altering part of the macromolecular information. Even though this specific example serves as a proof of concept, we anticipate that the presented work can also be used in a more general rewriting context, thereby further exploiting the enormous potential of synthetic macromolecules. Simultaneously, we also successfully developed a novel algorithm, Oligoreader, that is able to decode large portions of macromolecular data. In addition, Oligoreader has the ability to scan any MS¹ spectrum for potential macromolecules and decode their corresponding MS² spectra, thereby minimizing interference from the operator and increasing the possibility for blind sequencing. Finally, Oligoreader is available free of charge and is not limited to the specific macromolecules presented in this manuscript. Based on a user-defined molecular alphabet and a set of fragmentation patterns, it can be readily adapted to other macromolecular

backbones that incorporate different encoding schemes. Consequently, the herein presented work introduces both a straightforward rewriting approach, as well as an efficient method for reading out macromolecular information.

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. Chiel Mertens, Dr. Resat Aksakal, Dr. Nezha Badi, and Filip Jr. Van Lysebetten 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.

Keywords: molecular data storage • sequence-defined macromolecules • rewriting data • Oligoreader • automated read-out

- [1] V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church, W. L. Hughes, *Nat. Mater.* **2016**, *15*, 366–370.
- [2] R. Aksakal, C. Mertens, M. Soete, N. Badi, F. Du Prez, *Adv. Sci.* **2021**, *8*, 2004038.
- [3] L. Ceze, J. Nivala, K. Strauss, *Nat. Rev. Genet.* **2019**, *20*, 456–466.
- [4] C. N. Takahashi, B. H. Nguyen, K. Strauss, L. Ceze, *Sci. Rep.* **2019**, *9*, 4998.
- [5] R. Szweda, M. Tschopp, O. Felix, G. Decher, J.-F. Lutz, *Angew. Chemie Int. Ed.* **2018**, *57*, 15817–15821.
- [6] K. S. Wetzel, M. Frölich, S. C. Solleder, R. Nickisch, P. Treu, M. A. R. Meier, *Commun. Chem.* **2020**, *3*, 63.
- [7] X. Zhang, F. Gou, X. Wang, Y. Wang, S. Ding, *ACS Macro Lett.* **2021**, *10*, 551–557.
- [8] G. Cavallo, S. Poyer, J.-A. Amalian, F. Dufour, A. Burel, C. Carapito, L. Charles, J.-F. Lutz, *Angew. Chemie Int. Ed.* **2018**, *57*, 6266–6269.
- [9] B. Liu, Q. Shi, L. Hu, Z. Huang, X. Zhu, Z. Zhang, *Polym. Chem.* **2020**, *11*, 1702–1707.
- [10] N. Zydziak, W. Konrad, F. Feist, S. Afonin, S. Weidner, C. Barner-Kowollik, *Nat. Commun.* **2016**, *7*, 13672.
- [11] 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.
- [12] L. Organick, S. D. Ang, Y.-J. Chen, R. Lopez, S. Yekhanin, K. Makarychev, M. Z. Racz, G. Kamath, P. Gopalan, B. Nguyen, C. N. Takahashi, S. Newman, H.-Y. Parker, C. Rashtchian, K. Stewart, G. Gupta, R. Carlson, J. Mulligan, D. Carmean, G. Seelig, L. Ceze, K. Strauss, *Nat. Biotechnol.* **2018**, *36*, 242–248.
- [13] S. Martens, A. Landuyt, P. Espeel, B. Devreese, P. Dawyndt, F. Du Prez, *Nat. Commun.* **2018**, *9*, 4451.
- [14] 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.
- [15] E. Laurent, J.-A. Amalian, M. Parmentier, L. Oswald, A. Al Ouahabi, F. Dufour, K. Launay, J.-L. Clément, D. Gimes, M.-A. Delsuc, L. Charles, J.-F. Lutz, *Macromolecules* **2020**, *53*, 4022–4029.
- [16] J. Steinkoenig, R. Aksakal, F. Du Prez, *Eur. Polym. J.* **2019**, *120*, 109260.
- [17] C. Mayer, G. R. McInroy, P. Murat, P. Van Delft, S. Balasubramanian, *Angew. Chemie - Int. Ed.* **2016**, *55*, 11144–11148.
- [18] J. Bonnet, P. Subsoontorn, D. Endy, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 8884–8889.
- [19] R. K. Roy, A. Meszynska, C. Laure, L. Charles, C. Verchin, J. F. Lutz, *Nat. Commun.* **2015**, *6*, 7237–7245.

RESEARCH ARTICLE

- [20] J. O. Holloway, F. Van Lijsebetten, N. Badi, H. A. Houck, F. E. Du Prez, *Adv. Sci.* **2020**, *7*, 1903698.
- [21] N. F. König, A. Al Ouahabi, L. Oswald, R. Szveda, L. Charles, J. F. Lutz, *Nat. Commun.* **2019**, *10*, 3774.
- [22] M. G. T. A. Rutten, F. W. Vaandrager, J. A. A. W. Elemans, R. J. M. Nolte, *Nat. Rev. Chem.* **2018**, *2*, 365–381.
- [23] M. Soete, C. Mertens, R. Aksakal, N. Badi, F. Du Prez, *ACS Macro Lett.* **2021**, *10*, 616–622.
- [24] S. Martens, J. Van den Begin, A. Madder, F. E. Du Prez, P. Espeel, J. *Am. Chem. Soc.* **2016**, *138*, 14182–14185.
- [25] C. Mertens, R. Aksakal, N. Badi, F. E. Du Prez, *Polym. Chem.* **2021**, *12*, 4193–4204.
- [26] V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoët, J.-C. M. Monbaliu, O. Melnyk, *Chem. Rev.* **2019**, *119*, 7328–7443.
- [27] 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.
- [28] K. De Bruycker, T. Krappitz, C. Barner-Kowollik, *ACS Macro Lett.* **2018**, *7*, 1443–1447.

Entry for the Table of Contents



The rewriting of information stored on synthetic macromolecules is achieved *via* the introduction of an editable position tag. The automated read-out is facilitated by making use of an algorithm that is able to decode large portions of macromolecular information without interference from the operator.

Institute and/or researcher Twitter usernames: Filip_Du_Prez