Coding in 2D: Using Intentional Dispersity to Enhance the Information Capacity of Sequence-Coded Polymer Barcodes

Chloé Laure, Denise Karamessini, Olgica Milenkovic, Laurence Charles,* and Jean-François Lutz*

Abstract: *A 2D approach was studied for the design of polymer-based molecular barcodes. Uniform oligo(alkoxyamine amide)s, containing a monomer-coded binary message, were synthesized by orthogonal solid-phase chemistry. Sets of oligomers with different chain-lengths were prepared. The physical mixture of these uniform oligomers leads to an intentional dispersity (1st dimension fingerprint), which is measured by electrospray mass spectrometry. Furthermore, the monomer sequence of each component of the mass distribution can be analyzed by tandem mass spectrometry (2nd dimension sequencing). By summing the sequence information of all components, a binary message can be read. A 4-bytes extended ASCII-coded message was written on a set of six uniform oligomers. Alternatively, a 3-bytes sequence was written on a set of five oligomers. In both cases, the coded binary information was recovered.*

Information-containing polymers have recently emerged as an interesting new class of synthetic macromolecules.[1] Similarly to DNA in biology, these polymers contain information, which is coded at the molecular level by a controlled sequence of monomers.[2] Efficient approaches have been recently reported to write, erase, and read such monomer-coded sequences.[3] In particular, binary sequences can be synthesized using two monomers defined as 0-bit and 1-bit. Our group has recently reported the synthesis of digitally-encoded poly(phosphodiesters),[4] poly(triazole amide)s,[5] and poly(alkoxyamine amide)s.[6] Sequencing by tandem (MS/MS) mass spectrometry has also been reported for the poly(alkoxyamine amide)s class of polymers.[7]

[*] C. Laure, D. Karamessini, Dr. J.-F. Lutz
    Precision Macromolecular Chemistry Group
    Institut Charles Sadron, CNRS-UPR 22
    23 rue du Loess, 67034, Strasbourg Cedex 2 (France)
    E-mail: jflutz@unistra.fr

    Prof. O. Milenkovic
    University of Illinois
    Department of Electrical and Computer Engineering
    Urbana, IL 61801 (USA)

    Prof. L. Charles
    Aix-Marseille Univ, CNRS
    ICR, Institut de Chimie
    Radicalaire
    Marseille (France)
    E-mail: laurence.charles@univ-amu.fr

 Supporting information and the ORCID identification number(s) for
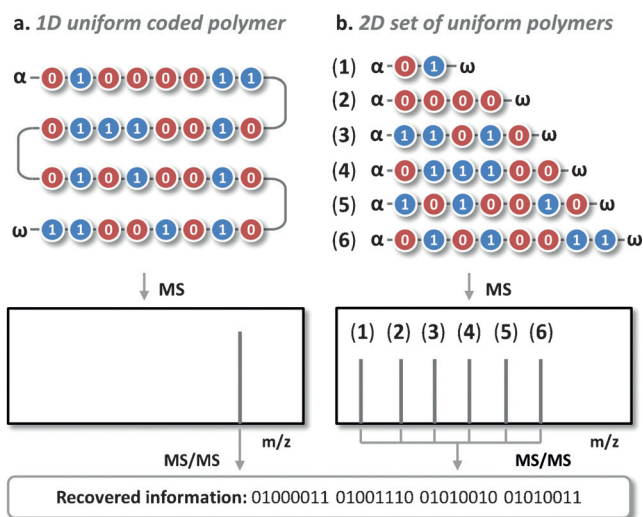 the author(s) of this article can be found under http://dx.doi.org/10.
 1002/anie.201605279.

Among the potential applications of information-containing polymers, it has been proposed that these polymers could be used as molecular barcodes in anti-counterfeit technologies.[3] Various approaches have been proposed in recent years to create coded patterns that can be used to label high-value products.[8] However, most of the reported concepts rely on nano- or micro-structured materials. Sequence-coded oligomers could be an interesting alternative because they would allow product labeling at the molecular level. Indeed, such polymers could be easily blended, adsorbed or covalently-attached to a wide range of organic or inorganic materials. Some examples of such molecular barcodes have already been reported using DNA,[9] although these developments should not be confused with the unrelated field of DNA barcoding, which deals with species detection in biology. Synthetic sequence-coded polymers could be an interesting alternative to DNA for anti-counterfeit materials.[3] Indeed, the molecular structure of non-natural polymers can be varied to ease synthesis and sequencing;[10] something which is, of course, not possible with biopolymers. For instance, we have recently reported that sequence-coded poly(alkoxyamine amide)s are relatively easy to synthesize.[6a] These polymers are prepared by a protecting-group-free solid-phase approach involving two repeating orthogonal coupling steps. Moreover, their MS/MS sequencing was found to be greatly facilitated by predictable alkoxyamine fragmentations.[7]

Sequence-coded poly(alkoxyamine amide)s are therefore interesting candidates for the development of molecular barcodes.[6a] As shown in Figure 1a, the simplest type of polymer barcode that can be conceived is a 1D linear chain that contains a monomer-coded string of binary information. However, the manual solid-phase synthesis of poly(alkoxyamine amide)s is currently limited to relatively short chains. The longest sequence reported to date contained 10 coded bits.[6b] As recently demonstrated with other types of information-containing macromolecules,[4b] it is conceivable to prepare longer sequence-defined poly(alkoxyamine amide)s by automated synthesis. However, such robotic syntheses would require chemical engineering developments that have not been investigated yet. In this context, a strategy that permits to increase the information capacity of poly(alkoxyamine amide)s but does not require the synthesis of long coded polymer chains is investigated herein. This concept is based on the fact that the MS analysis of 1D barcodes does not exploit all the possibilities offered by mass spectrometry. As shown in Figure 1a, 1D sequence-coded polymers lead typically to a single peak in MS, which is afterwards sequenced by MS/MS to access binary information. This implies that the large majority of the accessible m/z area in

a. *1D uniform coded polymer*   **b.** *2D set of uniform polymers*

Figure 1. Possible strategies for preparing digitally encoded barcodes: a) 1D strategy employing a long coded polymer chain. b) 2D barcode employing a set of monodisperse oligomers of different mass.[11] In both cases, the message is written using a binary monomer code and read by tandem mass spectrometry.

the MS spectrum is empty. This empty space could be exploited in order to increase the information capacity of polymer barcodes. For instance, polydisperse samples containing chains of different lengths should permit to store more information than a single 1D polymer chain. Of course, such materials cannot be regular polydisperse polymers synthesized by non-sequence-specific polymerization methods. However, as shown in Figure 1b, it is possible to create an intentional polydisperse material by mixing uniform sequence-coded oligomers of different lengths. Such mixtures lead to polydisperse patterns in MS (1st dimension fingerprint), in which each peak can be individually addressed and sequenced by MS/MS (2nd dimension sequencing). By summing the sequence information of all components of the mass distribution, a binary message can be read.

The preparation and sequencing of poly(alkoxyamine amide)s 2D barcodes is described herein. Scheme 1 shows the general molecular structure of the studied polymers. As described in previous publications,[6a,7a] these polymers contain a monomer-based digital code. Their repeat units consist of an alkyl amide synthon connected to a tetramethylpiperidine oxyl based alkoxyamine synthon (T). The side-chains substituents of the alkyl amide synthon define the binary code, that is, 0 and 1 bits are coded by a proton and a methyl, respectively. These macromolecules also contain defined α-

(a) Reading direction $\alpha \rightarrow \omega$ + $m/z \rightarrow$

(b) Reading direction $\omega \rightarrow \alpha$ + $\leftarrow m/z$
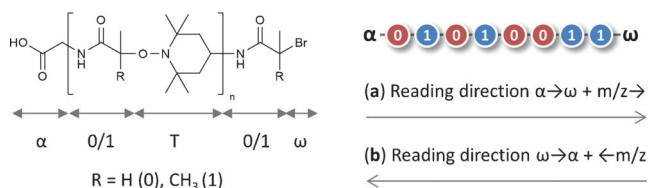
Scheme 1. Molecular structure and terminology of the sequence-coded oligo(alkoxyamine amide)s.

and ω- termini and therefore their coded monomer sequences are directional. Thus, in order to develop an efficient 2D barcode, writing and reading rules have to be defined. Two different sets of rules were studied in the present work. In a first approach (scenario (a) in Scheme 1), the oligomers were sorted out in the first dimension by increasing molar mass ($m/z \rightarrow$). In addition, each oligomer of the mass distribution was read in the second dimension from α to ω terminus ($\alpha \rightarrow \omega$). Thus, this writing/reading instruction ($m/z \rightarrow + \alpha \rightarrow \omega$) permits to form long binary sequences using a polydisperse set of monodisperse oligomers.[11] As a proof of concept, an extended ASCII coded message was implemented on a library of six different oligomers. The digital sequence 01000011 01001110 01010010 01010011, which codes for the acronym CNRS, was stored in this library. To do so, the sequence was decomposed in six blocks of different length, that is, α-01-ω, α-0000-ω, α-11010-ω, α-011100-ω, α-1010010-ω and α-01010011-ω. Alternatively, a second approach was studied (scenario (b) in Scheme 1), in which the oligomers were sorted out in the first dimension by decreasing molar mass ($\leftarrow m/z$) and read in the second dimension from ω to α terminus ($\omega \rightarrow \alpha$). To verify the validity of the language ($\leftarrow m/z + \omega \rightarrow \alpha$), a model ASCII sequence 01010000 01001101 01000011, coding the acronym PMC was decomposed into five coded blocks α-0001010-ω, α-100100-ω, α-10101-ω, α-0000-ω, and α-11-ω and stored on five different oligomers.[11]

All the oligomers of the studied libraries were synthesized individually by solid-phase chemistry using iterative orthogonal coupling steps, as previously reported.[6a] After cleavage from the resin and purification, each oligomer was characterized by high-resolution electrospray mass spectrometry (ESI-HRMS; Table 1 and Figures S1–S11 in the Supporting

Table 1: ESI-HRMS individual characterization of the sequence-coded oligo(alkoxyamine amide)s.[11]

	Sequence[a]	m/z_{th}	m/z_{exp}[b]
1	α-0T1-ω	448.1453	448.1469
2	α-0T0T0T0-ω	886.4659	886.4683
3	α-1T1T0T1T0-ω	1154.6809	1154.6814
4	α-0T1T1T1T0T0-ω	1380.8491	1380.8481
5	α-1T0T1T0T0T1T0-ω	1607.0172	1607.0178
6	α-0T1T0T1T0T0T1T1-ω	1847.2010	1847.2034
1'	α-0T0T0T1T0T1T1T0-ω	1593.0015	1593.0008
2'	α-1T0T0T1T0T0-ω	1366.8334	1366.8323
3'	α-1T0T1T0T1-ω	1154.6809	1154.6799
4'	α-0T0T0T0-ω	886.4659	886.4683
5'	α-1T1-ω	462.1609	462.1615

[a] Entries **2** and **4'** describe the same sample but are displayed twice for clarity. [b] The values correspond to $[M-H]^-$.

Information). In all cases, the desired oligomers were detected. It should be however remarked that some samples exhibited slight dispersity. As pointed out in earlier works,[6a] quantitative coupling yields are more difficult to obtain with monomer 0 than 1. Thus, 0-rich oligomers may contain small amounts of truncated chains. There are, however, several approaches to mitigate this problem. The simplest solution

would be to perform distribution shaping, a commonly used coding technique for mapping an information sequence into a coded sequence with a desired distribution.[12] Encoding messages using shaping schemes is most easily accomplished via lookup tables, as the number of entries in the table is rather small. For future studies, which will inevitably involve significantly longer encoded messages and consequently avoid the use of block-to-block mappings and lookup tables, methods such as Knuth's balancing schemes and variants thereof or decompression-based shaping may be more practical.[13] Other balancing methods were also reported in the context of coding for DNA-based storage.[14] Nevertheless, in the present work, 0-rich extended ASCII sequences were studied as such in order to assess the reliability of the method. In all cases, near monodisperse samples were observed by ESI-HRMS. In addition, the monomer sequences of each individual oligomer were characterized by both positive mode and negative mode MS/MS (Figures S12-S22).

Afterwards, 2D coding was studied using mixtures of oligomers **1–6** (CNRS-coding mixture) and oligomers **1′–5′** (PMC-coding mixture). In order to establish a readable 1st dimension fingerprint as shown in Figure 1 b, ESI-MS spectra of the two mixtures had to be recorded in the negative ion mode where all oligomers were mainly ionized as singly deprotonated species (as they contain a single acidic group), in contrast to the positive ion mode which readily allowed the protonation state of these molecules to increase with their size,[6b] hence not permitting their m/z values to increase linearly as a function of their mass. Moreover, ionization yield of the oligo(alkoxyamine amide)s was observed to rapidly decrease, probably due to their lower acidity as their size increased (Figure S23). Accordingly, concentration of each component in the mixtures had to be optimized to target the ideal barcode pattern illustrated in Figure 1 b, with peaks of similar abundance for all deprotonated molecules. However, ionization yield of each species as observed during mixture analysis was no longer the same as established during ESI-MS of individual solutions. More precisely, this suppression effect was stronger for smaller oligomers (Table S1), consistent with the model proposed by Enke relating abundance of ions measured in mass spectra to their relative concentration at the

surface of charged droplets generated during ESI.[15] As these suppression effects varied as a function of relative but also absolute concentration levels of oligomers (Table S1), mixtures had to be prepared empirically. A best compromise was achieved when targeting peaks of decreasing intensity as m/z increased, as illustrated in Figure 2 with the example of the CNRS-coding mixture. Such an optimized mixture was found to be practical for recovering ASCII-coded information by MS/MS sequencing. Indeed, the decreasing intensity of the peaks in the mixture may serve as an additional guide to read the message. It should be noted, however, that this optimization step has to be conducted by the "message writer" only, when preparing the coded mixture and that suppression effect is no longer an issue for the "message reader", who will obtain the same desired 2D MS pattern when operating the ESI source in soft conditions.

As shown in Figure 2 and Figure S24, each oligomer of the mixture can be individually addressed and comprehensively sequenced by MS/MS (i.e. 2nd dimension sequencing). Thus, the sequence-coded textual message "CNRS" was efficiently extracted from this mixture following the established reading rules $m/z \rightarrow$ and $\alpha \rightarrow \omega$, and thus validating the 2D coding concept. Still, it should be mentioned that the sequencing of these complex mixtures can be complicated by some secondary peaks. As mentioned in the introduction, it is important to use near-monodisperse oligomers in this 2D strategy. Due to incomplete coupling yields during iterative solid-phase synthesis, some oligomer samples may contain trace amounts of shorter incomplete sequences. Such impurities are not a problem at all when sequencing is performed individually (Figures S12–S22) but may interfere when sequencing is performed from a mixture. Indeed, these defects can be isomers of other sequence-coded oligomers used in the mixture. Secondary signals due to contamination by such truncated sequences are indicated by blue symbols in Figure S24. However, if the intensity of these impurities is lower than the one of the analyzed oligomer, they can be easily identified and dismissed (Figure S25). Perhaps more intriguing is the possibility to obtain "methyl scrambling" in some species activated upon collision in MS/MS experiments. It is known from proteomics that such rearrangement may occur
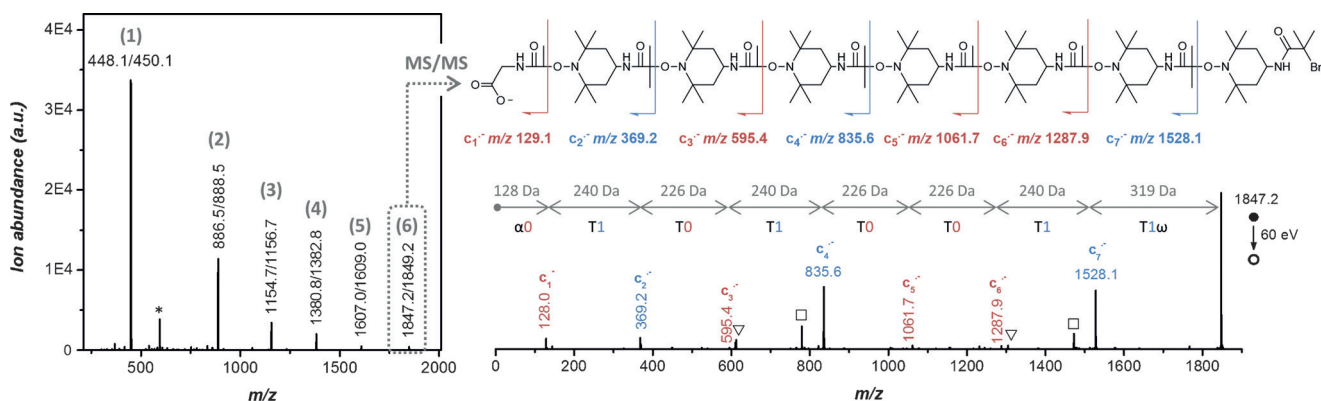

Figure 2. ESI-MS characterization of a mixture of ASCII-coded oligomers **1–6** that code for the acronym "CNRS". The peak indicated by a star is a non-brominated impurity coming from sample **6** (see Figure S6). Each brominated oligomer can be individually addressed from the mixture and sequenced by MS/MS. For clarity, only the sequencing of the longest oligomer **6** is shown, while the sequencing of oligomers **1–5** is presented in Figure S23. The symbols □ and ▽ indicate secondary fragments.

in MS/MS sequencing.[16] Although such a behavior was never observed so far with sequence-coded poly(alkoxyamine amide)s, it was found in this work that oligomers starting with the sequence α-10X-, where X can be either a 0 or a 1, may undergo methyl scrambling in negative ion mode MS/MS. This may not happen with short sequences but was observed with oligomers containing more than 5 coded units (e.g. **2′**, **3′** and **5**). Such sequencing defects are indicated by red symbols in Figure S24. The identification of this behavior is important and permits to avoid sequencing mistakes. In fact, some approaches can be foreseen to avoid or reduce the number of errors caused by scrambling. First of all, the polydisperse set of oligomers can be indeed modified. For instance, the CNRS ASCII-sequence could also be decomposed into 01, 000, 01101, 001110, 0101001, 001010011, thus avoiding the presence of long oligomers starting with the sequence α-10X. Alternatively, it is possible to reduce the number of appearances of the substring 10 in the information sequence by using a finite state machine transition diagram to generate the encoded sequence (Figure S26). The other reading language studied in this work (←*m*/*z* + ω→α) was also validated by the PMC-coding mixture of oligomers **1′–5′** (Figure S27). As for the CNRS mixture, each oligomer could be sequenced from the mixture (Figure S28) and the ASCII-encoded acronym "PMC" was retrieved.

In summary, it was found that mixtures of digitally encoded oligomers of different molar mass can be analyzed by negative-ion mode ESI-MS (1D) and sequenced by MS/MS (2D). This strategy permits the storage of several bytes of information at the molecular scale using a set of easy-to-synthesize oligomers. This study indicates further the relevance of synthetic information-containing macromolecules for the development of new polymer materials, such as anti-counterfeiting barcodes. Furthermore, the 2D analytical method reported herein is unprecedented and may also have relevance for biopolymer analysis, for instance in the important emerging field of DNA data storage[2b] and DNA *anti*-counterfeiting barcodes.[9b]

[1] a) J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science* **2013**, *341*, 1238149; b) H. Colquhoun, J.-F. Lutz, *Nat. Chem.* **2014**, *6*, 455–456.

[2] a) D. Andrieux, P. Gaspard, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9516–9521; b) V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church, W. L. Hughes, *Nat. Mater.* **2016**, *15*, 366–370.

[3] J.-F. Lutz, *Macromolecules* **2015**, *48*, 4759–4767.

[4] a) A. Al Ouahabi, L. Charles, J.-F. Lutz, *J. Am. Chem. Soc.* **2015**, *137*, 5629–5635; b) A. Al Ouahabi, M. Kotera, L. Charles, J.-F. Lutz, *ACS Macro Lett.* **2015**, *4*, 1077–1080.

[5] a) T. T. Trinh, L. Oswald, D. Chan-Seng, J.-F. Lutz, *Macromol. Rapid Commun.* **2014**, *35*, 141–145; b) T. T. Trinh, L. Oswald, D. Chan-Seng, L. Charles, J.-F. Lutz, *Chem. Eur. J.* **2015**, *21*, 11961–11965.

[6] a) R. K. Roy, A. Meszynska, C. Laure, L. Charles, C. Verchin, J.-F. Lutz, *Nat. Commun.* **2015**, *6*, 7237; b) R. K. Roy, C. Laure, D. Fischer-Krauser, L. Charles, J.-F. Lutz, *Chem. Commun.* **2015**, *51*, 15677–15680.

[7] a) L. Charles, C. Laure, J.-F. Lutz, R. K. Roy, *Macromolecules* **2015**, *48*, 4319–4328; b) L. Charles, C. Laure, J.-F. Lutz, R. K. Roy, *Rapid Commun. Mass Spectrom.* **2016**, *30*, 22–28.

[8] a) M. Ecker, T. Pretsch, *RSC Adv.* **2014**, *4*, 286–292; b) H. J. Bae, et al., *Adv. Mater.* **2015**, *27*, 2083–2089; c) X. Hou, C. Ke, C. J. Bruns, P R. McGonigal, R. B. Pettman, J. F. Stoddart, *Nat. Commun.* **2015**, *6*, 6884; d) T. Sarkar, K. Selvakumar, L. Motiei, D. Margulies, *Nat. Commun.* **2016**, *7*, 11374.

[9] a) D. Paunescu, R. Fuhrer, R. N. Grass, *Angew. Chem. Int. Ed.* **2013**, *52*, 4269–4272; *Angew. Chem.* **2013**, *125*, 4364–4368; b) D. Paunescu, W. J. Stark, R. N. Grass, *Powder Technol.* **2016**, *291*, 344–350.

[10] a) R. N. Zuckermann, J. M. Kerr, S. B. H. Kent, W. H. Moos, *J. Am. Chem. Soc.* **1992**, *114*, 10646–10647; b) H. Mutlu, J.-F. Lutz, *Angew. Chem. Int. Ed.* **2014**, *53*, 13010–13019; *Angew. Chem.* **2014**, *126*, 13224–13233; c) M. Porel, C. A. Alabi, *J. Am. Chem. Soc.* **2014**, *136*, 13162–13165; d) T. T. Trinh, C. Laure, J.-F. Lutz, *Macromol. Chem. Phys.* **2015**, *216*, 1498–1506.

[11] In general, the frontier between an oligomer and a polymer is considered to be 1000 Da but also sometimes 20 monomer units, depending on the sources. Examples and definitions can be found in: C. V. Uglea, I. I. Negulescu, *Synthesis and characterization of oligomers*, CRC Press, Boca Raton, FL, **1991**.

[12] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, **1968**.

[13] a) B. Vasić, O. Milenkovic, S. McLaughlin, *Electron. Lett.* **1996**, *32*, 1551–1552; b) J. H. Weber, K. A. S. Immink, *IEEE Trans. Inf. Theory* **2010**, *56*, 1673–1679.

[14] S. M. H. T. Yazdi, Y. Yuan, J. Ma, H. Zhao, O. Milenkovic, *Sci. Rep.* **2015**, *5*, 14138.

[15] a) C. G. Enke, *Anal. Chem.* **1997**, *69*, 4885–4893; b) N. B. Cech, C. G. Enke, *Anal. Chem.* **2000**, *72*, 2717–2723; c) N. B. Cech, C. G. Enke, *Anal. Chem.* **2001**, *73*, 4632–4639.

[16] a) S. Molesworth, S. Osburn, M. Van Stipdonk, *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2174–2181; b) L. Xiong, L. Ping, B. Yuan, Y. Wang, *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1172–1181; c) N. B. Borotto, N. Degraan-Weber, Y. Zhou, R. W. Vachet, *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1739–1746; d) R. Chawner, S. W. Holman, S. J. Gaskell, C. E. Eyers, *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1927–1938.