Macromolecules

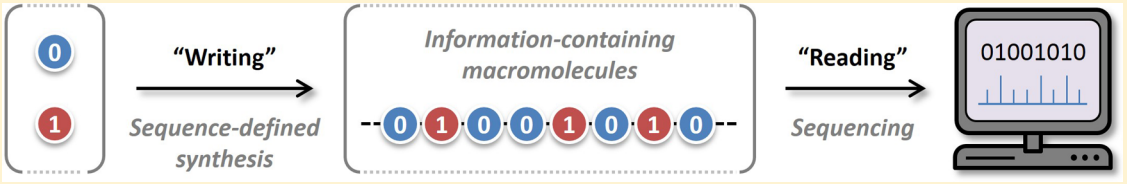
Perspective

pubs.acs.org/Macromolecules

Coding Macromolecules: Inputting Information in Polymers Using Monomer-Based Alphabets

Jean-François Lutz*

Precision Macromolecular Chemistry, Institut Charles Sadron, UPR22-CNRS, 23 rue du Loess, BP84047, 67034 Strasbourg, Cedex 2, France



ABSTRACT: The monomer units of a polymer can be used to encode a message. This property is used, for instance, by nature to store genetic information in DNA macromolecules. Therefore, during the past decades, many researchers have aimed to recreate *in vitro* or *in vivo* the properties of nucleic acids. Peptide nucleic acids, or more generally speaking xeno-nucleic acids, are interesting examples of man-made genetic polymers. However, the genetic code is surely not the only type of code that can be "written" in a polymer. In principle, many other monomer-based codes could be developed. For example, a binary code can be potentially implemented in a synthetic macromolecule using two comonomers defined as 0 and 1 bit. This possibility is exciting because it would permit to develop a full new class of synthetic polymers, which contain sequence-coded information. Such polymers could be interesting for a variety of new applications, for example in the field of data storage and product identification. However, these tempting options are currently underexplored. It should be clarified that the development of information-containing macromolecules is not trivial. First of all, in order to contain "readable" information, such polymers should possess perfectly controlled comonomer sequences. Moreover, chemical and analytical methods that allow deciphering of sequence-coded information have to be developed. The aim of the present Perspective is to show that significant progress has been done in that direction during the past two years. For instance, convenient strategies have been reported for the preparation of monodisperse sequence-defined macromolecules. In addition, encouraging advances have been made for the sequencing of non-natural polymers. These recent results are discussed and critically analyzed herein. Altogether, monomer-based information storage should be regarded as a new property of synthetic matter.

INTRODUCTION

Most of the important properties of synthetic polymers have been identified, and many of them were discovered long ago. For instance, the thermal, optical, electrical, mechanical, and rheological properties of polymer materials have been extensively studied.[1] In addition, the self-assembly,[2] supramolecular,[3] stimuli-responsive,[4] shape-memory,[5] and self-healing[6,7] behaviors of polymers are widely investigated. However, one important property of synthetic polymers has been overlooked, namely their information-storage capacity.[8] This property is essential in biology. Indeed, in all known life forms, genetic information is stored in DNA chains in the form of nucleotide sequences.[9] This is an efficient storage strategy, in which only four monomers—the well-known ATGC code—are used to encrypt information. Although central in natural science for eons, this property was never mimicked in synthetic polymers. Of course, many important features of DNA have been recreated by chemists. For instance, examples of artificial self-replicators[10,11] and non-natural double-helices[12] have been reported. However, the possibility to form monomer-coded strings of information on synthetic polymers was not considered until very recently.[8,13] Yet, such a property could open up a full new range of applications for synthetic polymers, e.g., in the fields of data storage, information processing, and product identification. In this context, the aim of the present Perspective is to discuss recent progress in polymer science that makes this new class of polymers attainable.

The field of sequence-controlled polymers has been neglected for a long time.[14−16] During the early years of polymer science, different types of copolymers such as statistical, alternating, and block copolymers were rapidly discovered.[17−20] Although significant progress has been made to simplify their synthesis, these categories of copolymers have more or less remained the state-of-the-art for almost five decades.[21] In fact, during that period of time, most progress in sequence definition has been made in biochemistry, in particular with the development of methods allowing synthesis of sequence-defined peptides[22] and oligonucleotides,[23] as well as the sequencing of biopolymers.[24−27] Only recently have examples of non-natural sequence-defined polymers been reported.[28] This discipline has a part of its roots in the fields of foldamers and peptidomimetics. Early examples of sequence-

Received: April 29, 2015
Revised: June 11, 2015
Published: July 15, 2015

DOI: 10.1021/acs.macromol.5b00890
Macromolecules 2015, 48, 4759−4767

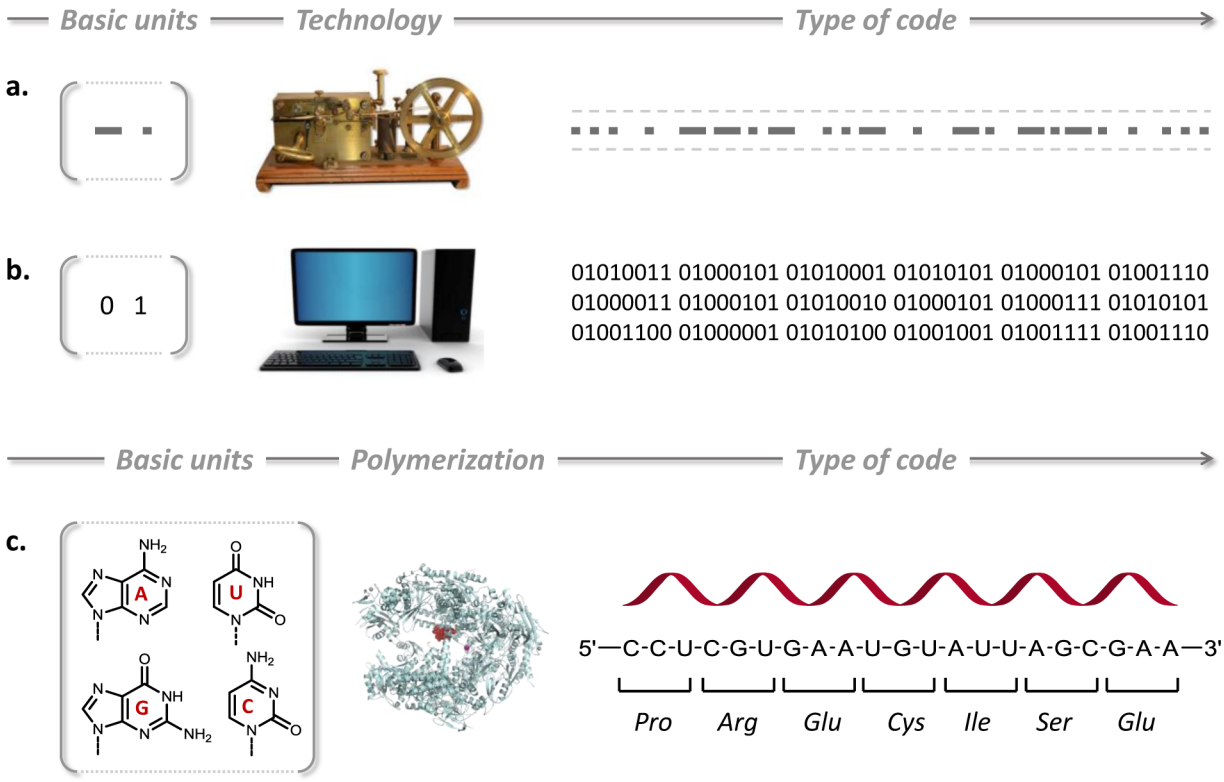


Figure 1. Examples of manmade and biological codes: (a) Morse code created in 1832 and used in telegraphic communications in the 19th and 20th century. (b) Binary code introduced in 1679 by Leibniz[68] and used in digital communication and computer technologies in the 20th and 21st century. The displayed digital sequence is an 8-bit extended ASCII-encoded text. (c) Genetic code used in biology for protein synthesis. This cartoon is not comprehensive and represents only some of the 64 codons that constitute the genetic code.

defined oligomers were reported by Moore,[29] Zuckerman,[30] Schultz,[31] Fréchet,[32] and others.[14] In addition, the synthesis of sequence-defined polypeptides containing noncanonical monomers by protein engineering has been reported by Tirrell[33] and Schultz.[34] More recently, strategies for controlling comonomer sequences in conventional polymerization processes such as chain-growth and step-growth polymerizations have been increasingly studied in the literature. For example, interesting concepts for regulating sequences in controlled radical polymerizations,[35−41] anionic polymerization,[42,43] ring-opening polymerization of lactones,[44] ring-opening metathesis polymerization,[45] multicomponent polymerizations,[46] step-growth radical polymerization,[47] and acyclic diene metathesis polymerization[48] have been described. Furthermore, advanced concepts for sequence regulation based on templates[49,50] or molecular machines[51] have been reported. Overall, the field has drastically evolved during the past few years.[28] Numerous options are now available for the synthesis of sequence-controlled polymers.

In terms of properties and applications, the motivations for preparing sequence-controlled polymers are diverse. For example, one important aim is the formation of folded macromolecules[52,53] and, more broadly speaking, the preparation of complex single-chain objects that can mimic the properties of globular proteins.[54−60] It was also observed that monomer sequence regulation allows fine-tuning of macroscopic properties such as crystallinity,[61] biodegradability,[62,63] and optoelectronic properties.[64] Apart from that, and as mentioned above, the molecular storage of information will be a new application for sequence-controlled polymers. In such information-containing polymers, the monomer units are used as molecular bits to encode a message.[8] For example, a digital code can be implemented in a chain using two monomers defined intentionally as 0 and 1 bit. It has already been demonstrated that coded messages can be encrypted in DNA chains.[65,66] These messages can be easily analyzed using sequencing tools developed for genomics. However, for such an application which is more related to materials science than biology, DNA is most probably not the only polymer that could be used. As pointed out in a recent essay,[8] synthetic polymers could be an interesting alternative to DNA for molecular information storage. However, this implies that technical tools for coding and reading information on synthetic polymers should be developed. This is definitely not a trivial task because, similarly to biopolymers, synthetic information-containing polymers should have perfectly controlled molecular structures, i.e., being monodisperse and sequence-defined. In addition, the sequencing of synthetic macromolecules is still an underdeveloped field of research.[67] However, all the necessary elements for achieving such macromolecules are probably already available. These recent progresses are summarized and critically analyzed in the present Perspective article.

MONOMER-BASED CODES

As described above, monomers could be used as coding units to implement a message in a polymer. This concept is very general, and in principle many different types of monomer-based codes could be created. Figure 1 shows examples of common codes found in nature or used in human communications. All these codes rely on two or more basic units (e.g., bits or trits) or eventually (e.g., bits, trits ...) that are arranged in a string of information. Binary codes are used in most human information technologies.[68] For example, the

DOI: 10.1021/acs.macromol.5b00890
Macromolecules 2015, 48, 4759−4767

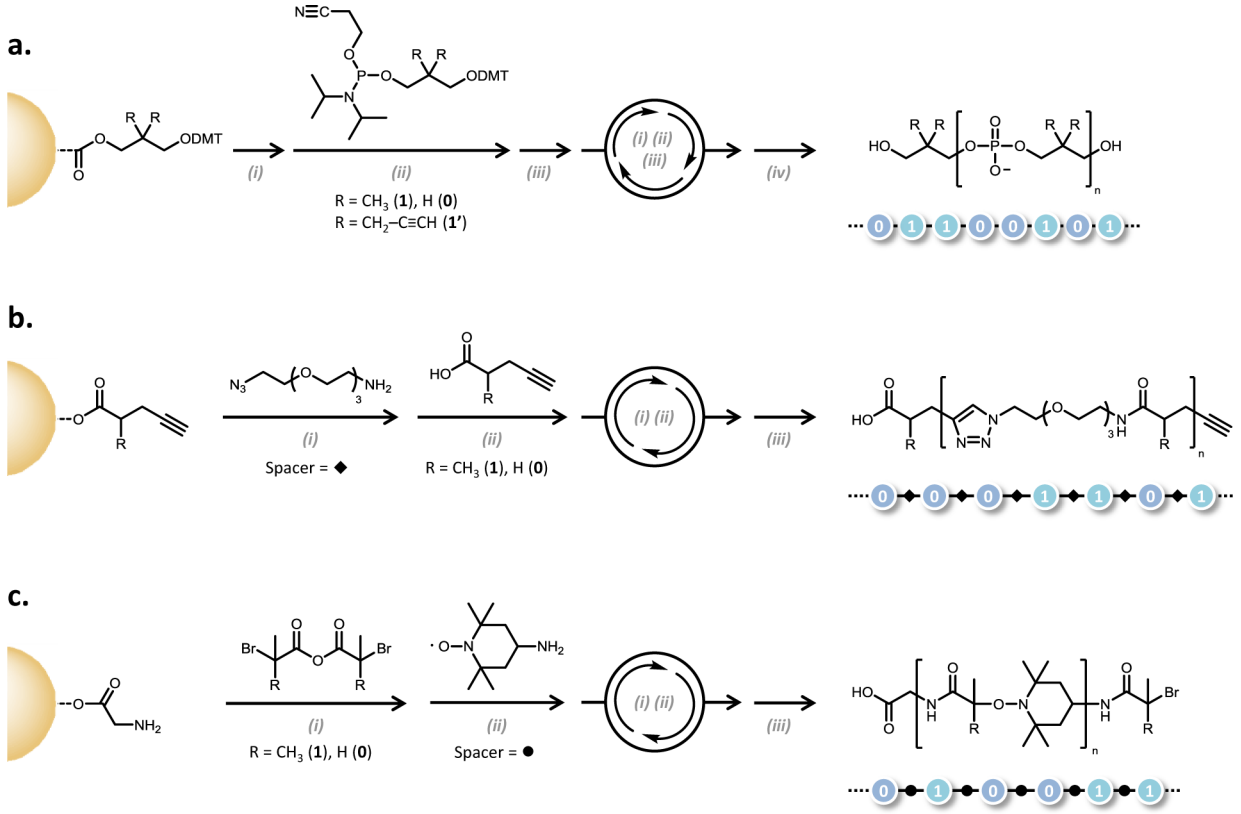


Figure 2. Strategies reported by our group for the synthesis of information-containing macromolecules: (a) Synthesis of sequence-coded non-natural polyphosphates by phosphoramidite chemistry. Experimental conditions: (i) DMT deprotection: CCl_3–COOH, CH_2Cl_2; (ii) coupling step: RT, ACN, tetrazole; (iii) oxidation: RT, I_2, H_2O/pyridine/THF; (iv) cyanoethyl deprotection and cleavage: piperidine, ACN then NH_3, H_2O, dioxane. (b) Synthesis of sequence-coded oligo(triazole amide)s via an "AB + CD" chemoselective approach involving successive copper-catalyzed azide–alkyne Huisgen cycloaddition and carboxylic acid-amine coupling steps.[76,77] Experimental conditions: (i) Huisgen cycloaddition: CuBr; dNbipy, THF; (ii) amidification: PyBOP, DIPEA, CH_2Cl_2; (iii) cleavage: TFA/CH_2Cl_2. (c) Synthesis of sequence-coded oligo(alkoxyamine amide)s via an accelerated "AB + CD" chemoselective approach involving successive anhydride–amine and nitroxide radical coupling steps.[79] Experimental conditions: (i) anhydride–amine coupling: THF, DIPEA or K_2CO_3; (ii) radical–radical coupling: CuBr, Me_6TREN, DMSO; (iii) cleavage: TFA, CH_2Cl_2.

popular Morse code (Figure 1a) and digital codes (Figure 1b) are both based on two basic units (i.e., dot/dash and 0/1). The widespread use of such binary codes is not only due to their simplicity but also in some cases to technological reasons. For instance, computers utilize binary languages because they interpret electrical or optical signals with only two different on/off states. Biology, on the other hand, uses more complex codes. As mentioned in the Introduction, DNA utilizes a quaternary code based on four basic units A, T (or U), G, and C (Figure 1c), although in principle a binary code could be used as well for storing genetic information. However, a binary genetic code would be disadvantageous for multiple reasons. For instance, longer DNA strands would be needed to store all genetic information, and protein synthesis would be greatly complicated, if not even impossible. Indeed, in codon-based translation, 64 possibilities—coding the 21 natural amino acids—are created with trimers of the four letters AUGC (i.e., $4^3 = 64$). If only two letters were used, hexamers would be needed to create the same number of possibilities (i.e., $2^6 = 64$). In other words, the use of a four-letter code permits Nature to save monomers and atoms.

The design of non-natural information-containing macromolecules is somehow inspired by both man-made and biological codes. However, such polymers are primarily conceived for nonbiological applications, e.g., in the field of nanotechnology and materials science. Thus, their molecular coding should be to some degree related to human technologies. That is the reason why binary-encoded polymers have been mainly studied to date and will be chiefly discussed in the following sections of this Perspective. Nevertheless, codes based on three (ternary), four (quaternary), or more monomers could be also envisioned. As learned from biology, such extended codes could be monomer-economical and simplify polymer design. In other words, they would permit to encode a large number of possibilities using polymers with short chain lengths. For example, considering a model heterotelechelic decamer (DP = 10) with distinct α- and ω-termini, $2^{10} = 1024$ different sequences can be created with a binary monomer code, whereas 59049 (3^{10}) and 1048576 (4^{10}) situations can be attained with a ternary or quaternary code, respectively. More than 6 billion possibilities could be obtained using six different monomers. Such information-storage capacities are immense and underexplored. It should be moreover remarked that these calculations are made assuming that one monomer is a basic unit. The overall number of possibilities could be even larger if regiochemistry and stereochemistry are taken into account. These simple calculations suggest that extended monomer sets are probably advantageous for polymer coding. It would be possible to write sentences of text on a polymer using 26 different monomers

DOI: 10.1021/acs.macromol.5b00890
Macromolecules 2015, 48, 4759–4767

that code for the 26 letters of the Latin alphabet. Such a goal would be achievable using, for example, 21 natural and 5 noncanonical amino acids. However, the use of a large number of different monomers may also greatly complicate polymer synthesis and characterization. The synthesis of sequence-defined polymers using complex monomer alphabets would require the use of side-chain protecting groups and orthogonal chemistries. In addition, polymer sequencing can be difficult if complicated macromolecular structures are used. Thus, the design of non-natural information-containing macromolecules requires somehow a good balance between chemical simplicity and monomer economy.

■ SYNTHESIS OF CODED MACROMOLECULES

Sequence-coded copolymers are easy to conceptualize but not trivial to synthesize. As defined in the Introduction, a string of information can be possibly implemented in a polymer chain using a monomer code. However, if one aims to implement a readable message in a polymer, the encoded monomer sequence has to be perfectly controlled. In other words, the primary structure of the polymer should be defect-free and exactly the same in all chains.[69] Thus, such macromolecules cannot be synthesized using classical methods such as chain-growth or step-growth polymerizations. Even though these techniques enable some degree of sequence regulation, they do not allow synthesis of perfectly monodisperse sequence-defined polymers.[70] On the other hand, solid-phase iterative chemistry is an adequate tool for preparing information-containing macromolecules. Although initially developed for the synthesis of sequence-defined oligopeptides,[22] this technique has been extended to the preparation of a wide variety of non-natural oligomers.[14] In this approach, monomers are covalently bound one-by-one on a solid support. This strategy is efficient but time-consuming since it often requires deprotection and capping steps. Hence, solid-phase chemistry is generally used for the synthesis of short oligomers with fewer than 10 monomer units.[28] Information-containing polymers should typically exhibit longer chain lengths in order to store substantial information (*vide infra*).[8] Therefore, fast and efficient iterative approaches, which resemble real polymerization processes, have to be selected for preparing them. Figure 2 shows three straightforward strategies that have been recently studied by our group for the synthesis of information-containing macromolecules. The first approach utilizes phosphoramidite coupling steps (Figure 2a), which have been initially introduced for oligonucleotide synthesis.[71] This strategy relies on a three-step cycle: (i) deprotection of a resin-bound hydroxy group, (ii) coupling of the hydroxy function with a phosphoramidite monomer, and (iii) oxidation of the resulting phosphite into a phosphate. Since this chemistry has been highly optimized during the past decades, each cycle can be quantitatively completed within a few minutes.[72] As a consequence, sequence-defined polynucleotides containing more than 100 residues can be synthesized using this approach.[73] We have shown in a recent publication that phosphoramidite chemistry also allows convenient synthesis of non-natural sequence-coded polyphosphates.[74] Non-nucleoside phosphoramidite monomers were studied in this work (Figure 2a). In order to implement a code in the polymer chains, two phosphoramidite monomers containing a propyl and a 2,2-dimethylpropyl synthon were used as 0 and 1 bit, respectively. Long encoded sequences were easily synthesized using these monomers. In addition, a monomer containing a 2,2-dipropargylpropyl synthon was studied (structure **1′** in Figure 2a). This monomer allows postpolymerization modification of the polymers by copper-catalyzed azide−alkyne cycloaddition (CuAAC). As a consequence, the molecular structure of the coding monomer units can be varied after polymer synthesis. Altogether, phosphoramidite chemistry seems to be a very interesting platform for preparing information-containing polymers. Still, it involves deprotection and oxidation steps that somehow complicate polymer synthesis. Over the past years, interesting protecting-group-free iterative strategies have been reported in the literature.[75] For instance, our group as introduced an efficient "AB + CD" concept that allow synthesis of sequence-coded polymers (Figure 2b).[76] This approach utilizes two building blocks AB and CD. The first one possesses reactive carboxylic acid (A) and alkyne (B) moieties, whereas the other contains an azide (C) and a primary amine (D). The building blocks are attached together using two consecutive chemoselective steps: (i) CuAAC and (ii) carboxylic acid−amine coupling. Protecting groups are not required in this approach because function A reacts only with function D, whereas function B reacts exclusively with function C. This concept allows preparation of monodisperse oligo(triazole amide)s[76] and was recently extended to the synthesis of coded macromolecules.[77] In this latter study, two different AB building blocks were used as 0 and 1 coding units. Their interchangeable use in step ii allowed synthesis of sequence-encoded oligo(triazole amide)s. Very recently, a convergent strategy involving the successive ligation of dyad-encoded oligomers was also introduced to simplify this approach.[78] However, this strategy remains, overall, relatively time-consuming. Our group has recently identified an even more efficient AB + CD method for preparing information-containing macromolecules (Figure 2c).[79] This approach utilizes an AB building block, which is an acyclic symmetric acid anhydride (A) containing alkyl bromides (B), and a CD building block, which is a nitroxide (C) bearing a primary amine (D). Two successive chemoselective steps can be performed with these monomers: (i) the reaction of the anhydride with a primary amine to form an amide bond and (ii) the coupling of a carbon-centered radical, obtained by copper activation of an alkyl bromide, with a nitroxide to afford an alkoxyamine. It was found that the repetition of these two steps allow synthesis of monodisperse sequence-coded oligo(alkoxyamine amide)s. Furthermore, the formed coded polymers can be easily sequenced by tandem mass spectrometry (see next section for details).

■ SEQUENCING

The word "sequencing" refers to the complete decoding of a sequence-defined macromolecule. It should not be confused with sequence analysis using NMR and other methods, which has been classically used in polymer science for quantifying the fraction of dyads or triads in alternating or statistical copolymers.[80−82] The differences between sequencing and sequence analysis were discussed in details in a recent review.[67] The readers may refer to that text for an overview on these topics. Sequencing methodologies have been primarily developed for the analysis of biopolymers such as proteins and DNA. Pioneer methods for protein sequencing have been reported independently by Edman[24] and Sanger[25] in the early 1950s, while efficient DNA sequencing methods have been reported about 25 years later by Gilbert[26] and Sanger.[27] After that, a broad spectrum of optimized methodologies has been

DOI: 10.1021/acs.macromol.5b00890
Macromolecules 2015, 48, 4759−4767

reported.[67,83−87] As a consequence, biopolymers can be sequenced today in a fast and efficient way using automated sequencers. However, these available sequencing methodologies have been rarely used for characterizing non-natural polymers.[67] Some of these methodologies rely on biological concepts (e.g., DNA amplification, polymerase-based sequencing, enzymatic cleavage) and cannot be applied to all polymers. However, some other are more universal. For instance, tandem mass spectrometry is a sequencing method that can be applied to both natural and non-natural polymers.[67,88−91] In MS/MS measurements, sequence-defined oligomers are fragmented in a mass spectrometer, and the resulting fragments are characterized. This technique has been extensively used for peptide and oligonucleotide sequencing.[88,89] In comparison, it has been less frequently applied to non-natural polymers.[92−94] However, some synthetic polymers could be much easier to sequence than biopolymers. Indeed, the molecular structure of synthetic macromolecule can be tailored for a particular sequencing technology, whereas in biopolymer sequencing, the read-out device has to be adapted to a molecular structure that is imposed by biology. For instance, in collaboration with the group of Laurence Charles in Marseille, we have recently discovered that poly(alkoxyamine amide)s are remarkably convenient to sequence by ESI MS/MS.[79] Figure 3 shows

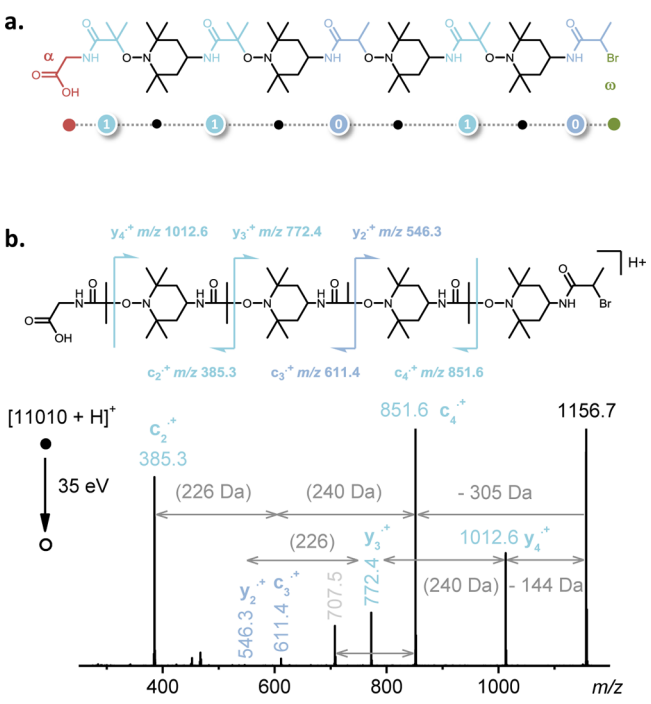


Figure 3. Sequencing of an oligo(alkoxyamine amide) containing the monomer-coded sequence 11010. (a) Molecular structure of the monodisperse oligomer containing defined α- and ω-termini. (b) ESI-MS/MS spectrum obtained after collisional activation of the m/z 1156.7 precursor ion containing the ^{79}Br isotope. This spectrum was recorded by Laurence Charles at the University of Marseille. The figure is adapted from ref 79.

the tandem sequencing of a coded oligo(alkoxyamine amide)s containing the binary sequence 11010. Because of the presence of alkoxyamine "weak links" in the polymer backbone, the fragmentation of the oligomers in the mass spectrometer is straightforward because no alternative dissociation reaction competes with the low-energy homolytic cleavage of the C−ON bond. As a consequence, the sequencing of these polymers can be can be performed in a clean fashion within a few minutes. Of course, not all synthetic polymers are easy to sequence by tandem mass spectrometry. Preliminary measurements made with sequence-coded polyphosphates (Figure 2a) and oligo(triazole amide)s (Figure 2b) indicated that the MS/MS fragmentation pathways of these polymers is more complex than the one found for oligo(alkoxyamine amide)s. However, other sequencing techniques can be considered for these polymers.[67] For example, water-soluble sequence-coded polyphosphates are interesting analytes for nanopore sequencing. In this technique, the polymer is analyzed in a pore, which can be biological or synthetic. For instance, membrane protein channels such as α-hemolysin have been extensively used for DNA sequencing.[95] When passing through the protein pore, the DNA analyte induces a blockade in the channel current, which can be recorded and correlated to a primary structure.[96,97] Although numerous studies have been done with nucleic acids, only a few other polymers have been studied in protein nanopores.[98−102] In particular, there are almost no studies dedicated to synthetic copolymers.[103] Nanopore sequencing is certainly an underexplored option for the sequencing of non-natural sequence-coded polymers. Indeed, the molecular structure of synthetic polymers could be adapted to simplify their readability in pores. Still, it should be remarked that pore/chain interactions depend on the charge, the stiffness, and the solution conformation of the polymeric analytes, and therefore the nanopore technique is probably not suitable for all types of water-soluble polymers. Besides mass spectrometry and nanopores, many other analytical techniques could be considered for the sequencing of sequence-coded polymers.[67] For instance, nonconventional NMR approaches can be used. In that regard, the tweezer technique recently developed by Colquhoun and co-workers has a lot of potential.[104−106] In this approach, supramolecular reporters, i.e., the so-called tweezers, are used. These molecules can bind specific sequence motifs and thus increase the readability of NMR signals. This technique has been so far mostly applied to random copolymers but could be extended to intentionally coded synthetic macromolecules.

SEQUENCE MANIPULATION

Synthetic information-containing macromolecules may also exhibit properties that are not available in nature. For instance, the information stored in their chains can be degraded on demand using a specific trigger. In that regard, synthetic polymers are probably more versatile than biopolymers because they can be designed on purpose for exhibiting such a behavior. Our group has shown in a recent publication that sequence-encoded oligo(alkoxyamine amide)s can be thermally degraded.[79] These materials are stable at room temperature, and digitally encoded information can be stored for long period of times in standard laboratory conditions. However, because of the homolysis of the C−ON bonds,[107] polyalkoxyamines exhibit a dynamic behavior at elevated temperatures.[108] For example, TEMPO-based sequence-encoded polymers shown in Figure 1c decompose above 60−70 °C. Thermal degradation was observed in the solid state and in solution. In the latter case, the thermal degradation can be controlled by adding a spin-trap that reduces the probability of some degradation reactions and favor some others.[79] As a result, the polymers can be degraded into a few low-molecular-weight residues. Thus, temperature can be used as a simple trigger to erase sequence-

DOI: 10.1021/acs.macromol.5b00890
Macromolecules 2015, 48, 4759−4767

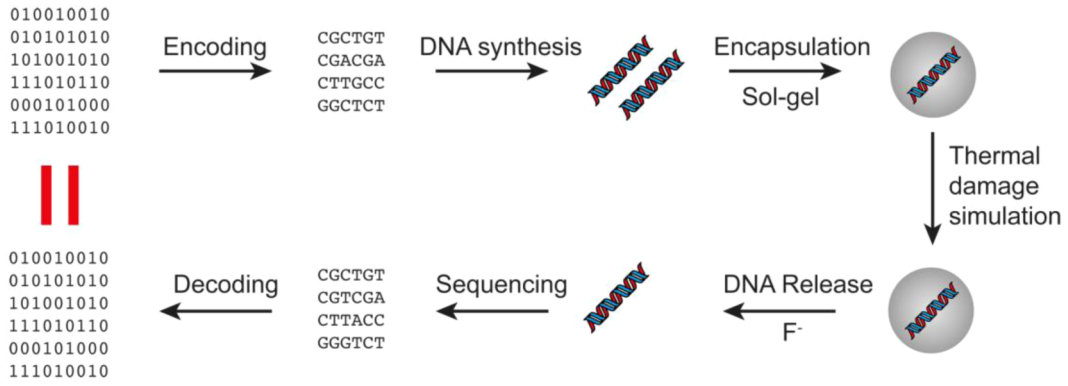


Figure 4. Concept for preservation of a DNA-encoded text in silica spheres. After DNA synthesis, the DNA strands are encapsulated by sol−gel chemistry. The silica shell can be selectively degraded, and the intact DNA strands are sequenced. Reproduced with permission from ref 115. Copyright 2015 Wiley-VCH.

encoded information. Comparable behaviors may also be obtained with other types of stimuli, for example using light-sensitive polymers. Beyond permanent degradation, more complex sequence manipulations can be foreseen. For instance, dynamic polymers introduced by Lehn and co-workers could be interesting macromolecules for scrambling or rewriting encoded sequences.[109] During the past decades, many examples of dynamic covalent linkages have been reported in the literature.[110,111] These chemistries could be used to build the backbone of information-containing macromolecules. However, based on the current state-of-the-art, the idea to rewrite efficiently sequence information seems very difficult to achieve without using a template.

RELEVANCE FOR APPLICATIONS

Synthetic sequence-encoded polymers may open a brand new range of applications in materials science and nanotechnology. As mentioned in the Introduction, data storage could be an obvious application of such polymers. However, as discussed in a recent publication,[8] a clear distinction should be made between data storage and data processing. It is for the moment very unlikely to imagine molecular memory devices based on information-containing macromolecules. However, such polymers could be certainly used for storing information for long period of times in limited volumes. In that regard, synthetic polymers could be even more interesting than DNA. Indeed, the half-life time of DNA in bone was measured to be 521 years.[112] That explains why intact archeological DNA strands older than a few million years cannot be found on earth. The molecular structure of synthetic information-containing polymers could be certainly optimized for long-term storage, for instance using poorly cleavable polymer backbones. In addition, similarly to DNA in bone, these polymers could be preserved in inorganic matrices that protect them from light as well as thermal and oxidative degradation. In that regard, the recent concept reported by Grass and co-workers for DNA preservation in silica (Figure 4) is very interesting and could be transposed to non-natural sequence-coded macromolecules.[113−115]

The development of molecular barcodes is another potential area of application for sequence-coded polymers. Anticounterfeiting technologies have become increasingly important during the past 15 years in many industrial areas such as pharmaceutical, cosmetics, food, chemical, and automotive industry. Labeling technologies are also particularly important for the manufacturing and identification of high-value goods such as luxury products or banknotes. In this context, a broad spectrum of coded nanomaterials allowing product identification have been proposed during the past years.[113,116−118] Cheap synthetic sequence-encoded polymers could be an interesting option for tagging industrial products. Indeed, oligomer barcodes could be easily blended, encapsulated, or covalently linked to organic or inorganic materials.

SUMMARY AND OUTLOOK

Information-containing macromolecules constitute undoubtedly a promising new class of synthetic polymers. Indeed, these polymers permit to store a message that can be decoded. As described above, this new property has a real relevance for applications in nanotechnology and materials science. Moreover, the molecular structure of information-containing polymers can be varied and adapted to technological developments. In other words, the thermal, mechanical, and degradation properties of these polymers can be tuned by synthesis. In that regard, synthetic information-containing macromolecules are probably much more versatile than biopolymers. Still, this field of research is very recent, and synthetic sequence-coded polymers are far from being as convenient as artificial DNA. First of all, synthetic polymers cannot be replicated and amplified as DNA can be. Although amplification is not strictly required for most of the applications discussed in this Perspective, it is undoubtedly a very practical option. Furthermore, the iterative synthesis and sequencing of oligonucleotides is a highly established area of research. In comparison, the development of synthetic coded polymers is still in its infancy. However, the preparation of long encodable/decodable synthetic polymers is probably only a matter of time and optimization. As shown in this Perspective, practical routes for the synthesis of defect-free information-containing polymers exist. With proper automation and engineering, it is almost certain that long synthetic sequence-coded polymers can be attained. Similarly, the sequencing of non-natural polymers is possible and will be probably become easier and faster in the near future. In summary, the monomer units of a synthetic copolymer can be much more than randomly distributed building blocks leading to globally adjusted properties. Indeed, they are also the letters of an almost infinite alphabet. This opens up vertiginous possibilities for the field of polymer science.

DOI: 10.1021/acs.macromol.5b00890
Macromolecules 2015, 48, 4759−4767

AUTHOR INFORMATION

Corresponding Author

*E-mail: jflutz@unistra.fr.

Notes

The authors declare the following competing financial interest(s): J.-F.L. is named inventor on a patent application related to the synthesis of sequence-coded poly(alkoxyamine amide)s.

Biography

Dr. Jean-François Lutz is CNRS research director, deputy director of the Institut Charles Sadron in Strasbourg, and head of the Precision Macromolecular Chemistry group. He received his doctoral degree from the University of Montpellier II in 2000 and his habilitation degree from the University of Potsdam in 2009. Before joining the CNRS, he was postdoctoral fellow in the group of Krzysztof Matyjaszewski at Carnegie Mellon University (2001−2003) and afterwards leader of the research group Nanotechnology for Life Science at the Fraunhofer Institute for Applied Polymer Research (2003−2010). He is author of over 150 publications, is listed as coinventor in about 10 patents, and serves as an executive advisory board member for several journals including *Progress in Polymer Science*, *Polymer Chemistry*, *Macromolecular Chemistry, and Physics*, *Macromolecular Rapid Communications*, *European Polymer Journal*, and *Designed Monomers & Polymers*. He received in 2008 the prize of the polymer division of the French Chemical Society. He is also an ERC laureate since 2010 through successive starting (StG 2010) and proofs of concept (PoC 2015) grants. His current research interests include the synthesis of sequence-controlled polymers, single-chain technologies, and the preparation of information-containing macromolecules.

ACKNOWLEDGMENTS

The research of J.F.L. on information-containing macromolecules is supported by the H2020 program of the European Union (project Euro-Sequences, H2020-MSCA-ITN-2014, grant agreement no. 642083), the Cluster of Excellence Chemistry of Complex Systems (LabEx CSC), and the CNRS. J.F.L. thanks Abdelaziz Al Ouahabi, Nezha Badi, Delphine Chan-Seng, Dalila Chouikhi, Chloé Laure, Anna Meszynska, Laurence Oswald, Sebastian Pfeifer, Raj Kumar Roy, Thanh Tam Trinh, Claire Verchin, and Zoya Zarafshani for their contributions to the synthesis of sequence-coded macromolecules. Laurence Charles (Université Aix Marseille) and Jan C. Behrends (Freiburg Universität) are greatly acknowledged for fruitful collaborations on polymer sequencing. Robert Grass (ETH Zürich) is also thanked for kindly providing the artwork of Figure 4.

REFERENCES

(1) Van Krevelen, D. W.; Te Nijenhuis, K. *Properties of Polymers*, 4th ed.; Elsevier: Amsterdam, The Netherlands, 2009; p 1032.
(2) Ikkala, O.; ten Brinke, G. *Science* **2002**, *295*, 2407−2409.
(3) Aida, T.; Meijer, E. W.; Stupp, S. I. *Science* **2012**, *335*, 813−817.
(4) Stuart, M. A. C.; et al. *Nat. Mater.* **2010**, *9*, 101−113.
(5) Lendlein, A.; Kelch, S. *Angew. Chem., Int. Ed.* **2002**, *41*, 2034−2057.
(6) Cordier, P.; Tournilhac, F.; Soulie-Ziakovic, C.; Leibler, L. *Nature* **2008**, *451*, 977−980.
(7) Montarnal, D.; Capelot, M.; Tournilhac, F.; Leibler, L. *Science* **2011**, *334*, 965−968.
(8) Colquhoun, H. M.; Lutz, J.-F. *Nat. Chem.* **2014**, *5*, 455−456.
(9) Hood, L.; Galas, D. *Nature* **2003**, *421*, 444−448.
(10) Orgel, L. E. *Nature* **1992**, *358*, 203−209.
(11) Lee, D. H.; Granja, J. R.; Martinez, J. A.; Severin, K.; Ghadiri, M. R. *Nature* **1996**, *382*, 525−528.
(12) Yashima, E.; Maeda, K.; Iida, H.; Furusho, Y.; Nagai, K. *Chem. Rev.* **2009**, *109*, 6102−6211.
(13) Andrieux, D.; Gaspard, P. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 9516−9521.
(14) Badi, N.; Lutz, J.-F. *Chem. Soc. Rev.* **2009**, *38*, 3383−3390.
(15) Lutz, J.-F. *Polym. Chem.* **2010**, *1*, 55−62.
(16) Lutz, J.-F. *Nat. Chem.* **2010**, *2*, 84−85.
(17) Carothers, W. H. Linear polyamides and their production. US 2130523 A, 1938.
(18) Mayo, F. R.; Lewis, F. M. *J. Am. Chem. Soc.* **1944**, *66*, 1594−1601.
(19) Mayo, F. R.; Lewis, F. M.; Walling, C. *J. Am. Chem. Soc.* **1948**, *70*, 1529−1533.
(20) Dunn, A. S.; Melville, H. W. *Nature* **1952**, *169*, 699−700.
(21) Lutz, J.-F.; Meyer, T.; Ouchi, M.; Sawamoto, M. *Sequence-Controlled Polymers: Synthesis, Self-Assembly, and Properties*; American Chemical Society: Washington, DC, 2014; Vol. *1170*.
(22) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149−2154.
(23) Letsinger, R. L.; Ogilvie, K. K. *J. Am. Chem. Soc.* **1969**, *91*, 3350−3355.
(24) Edman, P. *Acta Chem. Scand.* **1950**, *4*, 283−293.
(25) Sanger, F.; Tuppy, H. *Biochem. J.* **1951**, *49*, 481−490.
(26) Maxam, A. M.; Gilbert, W. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 560−564.
(27) Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 5463−5467.
(28) Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. *Science* **2013**, *341*, 1238149.
(29) Young, J. K.; Nelson, J. C.; Moore, J. S. *J. Am. Chem. Soc.* **1994**, *116*, 10841−10842.
(30) Zuckermann, R. N.; Kerr, J. M.; Kent, S. B. H.; Moos, W. H. *J. Am. Chem. Soc.* **1992**, *114*, 10646−10647.
(31) Cho, C. Y.; Moran, E. J.; Cherry; Stephans, J. C.; Fodor, S. P.; Adams, C. L.; Sundaram, A.; Jacobs, J. W.; Schultz, P. G. *Science* **1993**, *261*, 1303−1305.
(32) Malenfant, P. R. L.; Fréchet, J. M. J. *Chem. Commun.* **1998**, 2657−2658.
(33) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727−733.
(34) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498−500.
(35) Pfeifer, S.; Lutz, J.-F. *J. Am. Chem. Soc.* **2007**, *129*, 9542−9543.
(36) Ida, S.; Terashima, T.; Ouchi, M.; Sawamoto, M. *J. Am. Chem. Soc.* **2009**, *131*, 10808−10809.
(37) Satoh, K.; Matsuda, M.; Nagai, K.; Kamigaito, M. *J. Am. Chem. Soc.* **2010**, *132*, 10003−10005.
(38) Hibi, Y.; Ouchi, M.; Sawamoto, M. *Angew. Chem., Int. Ed.* **2011**, *50*, 7434−7437.
(39) Lutz, J.-F. *Acc. Chem. Res.* **2013**, *46*, 2696−2705.
(40) Gody, G.; Maschmeyer, T.; Zetterlund, P. B.; Perrier, S. *Nat. Commun.* **2013**, *4*, 2505.

DOI: 10.1021/acs.macromol.5b00890
Macromolecules 2015, 48, 4759−4767

(41) Vandenbergh, J.; Reekmans, G.; Adriaensens, P.; Junkers, T. *Chem. Commun.* **2013**, *49*, 10358−10360.
(42) Hutchings, L. R.; Brooks, P. P.; Parker, D.; Mosely, J. A.; Sevinc, S. *Macromolecules* **2015**, *48*, 610−628.
(43) Li, J.; He, J. *ACS Macro Lett.* **2015**, 372−376.
(44) Kramer, J. W.; Treitler, D. S.; Dunn, E. W.; Castro, P. M.; Roisnel, T.; Thomas, C. M.; Coates, G. W. *J. Am. Chem. Soc.* **2009**, *131*, 16042−16044.
(45) Zhang, J.; Matta, M. E.; Hillmyer, M. A. *ACS Macro Lett.* **2012**, *1*, 1383−1387.
(46) Deng, X.-X.; Li, L.; Li, Z.-L.; Lv, A.; Du, F.-S.; Li, Z.-C. *ACS Macro Lett.* **2012**, *1*, 1300−1303.
(47) Satoh, K.; Ozawa, S.; Mizutani, M.; Nagai, K.; Kamigaito, M. *Nat. Commun.* **2010**, *1*, 6.
(48) Atallah, P.; Wagener, K. B.; Schulz, M. D. *Macromolecules* **2013**, *46*, 4735−4741.
(49) McKee, M. L.; Milnes, P. J.; Bath, J.; Stulz, E.; Turberfield, A. J.; O'Reilly, R. K. *Angew. Chem., Int. Ed.* **2010**, *49*, 7948−7951.
(50) Niu, J.; Hili, R.; Liu, D. R. *Nat. Chem.* **2013**, *5*, 282−292.
(51) Lewandowski, B.; et al. *Science* **2013**, *339*, 189−193.
(52) Hill, D. J.; Mio, M. J.; Prince, R. B.; Hughes, T. S.; Moore, J. S. *Chem. Rev.* **2001**, *101*, 3893−4012.
(53) Guichard, G.; Huc, I. *Chem. Commun.* **2011**, *47*, 5933−5941.
(54) Ouchi, M.; Badi, N.; Lutz, J.-F.; Sawamoto, M. *Nat. Chem.* **2011**, *3*, 917−924.
(55) Schmidt, B. V. K. J.; Fechler, N.; Falkenhagen, J.; Lutz, J.-F. *Nat. Chem.* **2011**, *3*, 234−238.
(56) Terashima, T.; Mes, T.; De Greef, T. F. A.; Gillissen, M. A. J.; Besenius, P.; Palmans, A. R. A.; Meijer, E. W. *J. Am. Chem. Soc.* **2011**, *133*, 4742−4745.
(57) Giuseppone, N.; Lutz, J.-F. *Nature* **2011**, *473*, 40−41.
(58) Altintas, O.; Barner-Kowollik, C. *Macromol. Rapid Commun.* **2012**, *33*, 958−971.
(59) Hosono, N.; Gillissen, M. A. J.; Li, Y.; Sheiko, S. S.; Palmans, A. R. A.; Meijer, E. W. *J. Am. Chem. Soc.* **2012**, *135*, 501−510.
(60) Roy, R. K.; Lutz, J.-F. *J. Am. Chem. Soc.* **2014**, *136*, 12888−12891.
(61) Srichan, S.; Kayunkid, N.; Oswald, L.; Lotz, B.; Lutz, J.-F. *Macromolecules* **2014**, *47*, 1570−1577.
(62) Stayshich, R. M.; Meyer, T. Y. *J. Am. Chem. Soc.* **2010**, *132*, 10920−10934.
(63) Thomas, C. M.; Lutz, J.-F. *Angew. Chem., Int. Ed.* **2011**, *50*, 9244−9246.
(64) Norris, B. N.; Zhang, S.; Campbell, C. M.; Auletta, J. T.; Calvo-Marzal, P.; Hutchison, G. R.; Meyer, T. Y. *Macromolecules* **2013**, *46*, 1384−1392.
(65) Church, G. M.; Gao, Y.; Kosuri, S. *Science* **2012**, *337*, 1628−1628.
(66) Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; LeProust, E. M.; Sipos, B.; Birney, E. *Nature* **2013**, *494*, 77−80.
(67) Mutlu, H.; Lutz, J.-F. *Angew. Chem., Int. Ed.* **2014**, *53*, 13010−13019.
(68) Leibniz, G. W. *Explication de l'arithmétique binaire*; Jean Boudot: Marseille, France, 1703.
(69) It should be mentioned that this aspect could be improved by the use of error-correcting codes. An example of that can be found in ref 114.
(70) Lutz, J.-F. *ACS Macro Lett.* **2014**, *3*, 1020−1023.
(71) Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981**, *22*, 1859−1862.
(72) Beaucage, S. L.; Iyer, R. P. *Tetrahedron* **1992**, *48*, 2223−2311.
(73) LeProust, E. M.; Peck, B. J.; Spirin, K.; McCuen, H. B.; Moore, B.; Namsaraev, E.; Caruthers, M. H. *Nucleic Acids Res.* **2010**, *38*, 2522−2540.
(74) Al Ouahabi, A.; Charles, L.; Lutz, J.-F. *J. Am. Chem. Soc.* **2015**, *137*, 5629−5635.
(75) Trinh, T. T.; Laure, C.; Lutz, J.-F. *Macromol. Chem. Phys.* **2015**, DOI: 10.1002/macp.201500072.
(76) Pfeifer, S.; Zarafshani, Z.; Badi, N.; Lutz, J.-F. *J. Am. Chem. Soc.* **2009**, *131*, 9195−9196.
(77) Trinh, T. T.; Oswald, L.; Chan-Seng, D.; Lutz, J.-F. *Macromol. Rapid Commun.* **2014**, *35*, 141−145.
(78) Trinh, T. T.; Oswald, L.; Chan-Seng, D.; Charles, L.; Lutz, J.-F. *Chem. Eur. J.* 2015, DOI: 10.1002/chem.201502414.
(79) Roy, R. K.; Meszynska, A.; Laure, C.; Charles, L.; Verchin, C.; Lutz, J.-F. *Nat. Commun.* **2015**, *6*, 7237.
(80) Harwood, H. J. *Angew. Chem., Int. Ed.* **1965**, *4*, 1051−1060.
(81) Randall, J. C. *Polymer Sequence Determination: Carbon-13 NMR Method*; Academic Press: New York, 1977; p 155.
(82) Bovey, F. A.; Mirau, P. A. *NMR of Polymers*; Academic Press: San Diego, 1996; p 459.
(83) Church, G. M.; Gilbert, W. *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81*, 1991−1995.
(84) Chait, B.; Wang, R.; Beavis, R.; Kent, S. *Science* **1993**, *262*, 89−92.
(85) Venter, J. C.; et al. *Science* **2001**, *291*, 1304−1351.
(86) Steen, H.; Mann, M. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 699−711.
(87) Shendure, J.; Ji, H. *Nat. Biotechnol.* **2008**, *26*, 1135−1145.
(88) Murray, K. K. *J. Mass Spectrom.* **1996**, *31*, 1203−1215.
(89) Yates, J. R. *J. Mass Spectrom.* **1998**, *33*, 1−19.
(90) Gruendling, T.; Weidner, S.; Falkenhagen, J.; Barner-Kowollik, C. *Polym. Chem.* **2010**, *1*, 599−617.
(91) Altuntaş, E.; Schubert, U. S. *Anal. Chim. Acta* **2014**, *808*, 56−69.
(92) Paulick, M. G.; Hart, K. M.; Brinner, K. M.; Tjandra, M.; Charych, D. H.; Zuckermann, R. N. *J. Comb. Chem.* **2006**, *8*, 417−426.
(93) Thakkar, A.; Cohen, A. S.; Connolly, M. D.; Zuckermann, R. N.; Pei, D. *J. Comb. Chem.* **2009**, *11*, 294−302.
(94) Porel, M.; Alabi, C. A. *J. Am. Chem. Soc.* **2014**, *136*, 13162−13165.
(95) Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 13770−13773.
(96) Kumar, H.; Lansac, Y.; Glaser, M. A.; Maiti, P. K. *Soft Matter* **2011**, *7*, 5898−5907.
(97) Ying, Y.-L.; Zhang, J.; Gao, R.; Long, Y.-T. *Angew. Chem., Int. Ed.* **2013**, *52*, 13154−13161.
(98) Murphy, R. J.; Muthukumar, M. *J. Chem. Phys.* **2007**, *126*, 051101.
(99) Brun, L.; Pastoriza-Gallego, M.; Oukhaled, G.; Mathe, J.; Bacri, L.; Auvray, L.; Pelta, J. *Phys. Rev. Lett.* **2008**, *100*, 158302.
(100) Reiner, J. E.; Kasianowicz, J. J.; Nablo, B. J.; Robertson, J. W. F. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 12080−12085.
(101) Baaken, G.; Ankri, N.; Schuler, A. K.; Ruhe, J.; Behrends, J. C. *ACS Nano* **2011**, *5*, 8080−8088.
(102) Sakaue, T.; Brochard-Wyart, F. *ACS Macro Lett.* **2014**, *3*, 194−197.
(103) Mirigian, S.; Wang, Y. B.; Muthukumar, M. *J. Chem. Phys.* **2012**, *137*, 064904.
(104) Colquhoun, H. M.; Zhu, Z. *Angew. Chem. Int., Ed.* **2004**, *43*, 5040−5045.
(105) Colquhoun, H. M.; Zhu, Z.; Cardin, C. J.; Gan, Y.; Drew, M. G. B. *J. Am. Chem. Soc.* **2007**, *129*, 16163−16174.
(106) Zhu, Z.; Cardin, C. J.; Gan, Y.; Colquhoun, H. M. *Nat. Chem.* **2010**, *2*, 653−660.
(107) Marque, S.; Le Mercier, C.; Tordo, P.; Fischer, H. *Macromolecules* **2000**, *33*, 4403−4410.
(108) Otsuka, H.; Aotani, K.; Higaki, Y.; Takahara, A. *J. Am. Chem. Soc.* **2003**, *125*, 4064−4065.
(109) Skene, W. G.; Lehn, J.-M. P. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8270−8275.
(110) Lehn, J.-M. *Prog. Polym. Sci.* **2005**, *30*, 814−831.
(111) Maeda, T.; Otsuka, H.; Takahara, A. *Prog. Polym. Sci.* **2009**, *34*, 581−604.
(112) Allentoft, M. E.; et al. *Proc. R. Soc. B* **2012**, *279*, 4724−4733.
(113) Paunescu, D.; Fuhrer, R.; Grass, R. N. *Angew. Chem., Int. Ed.* **2013**, *52*, 4269−4272.
(114) Paunescu, D.; Puddu, M.; Soellner, J. O. B.; Stoessel, P. R.; Grass, R. N. *Nat. Protoc.* **2013**, *8*, 2440−2448.

DOI: 10.1021/acs.macromol.5b00890
Macromolecules 2015, 48, 4759−4767

(115) Grass, R. N.; Heckel, R.; Puddu, M.; Paunescu, D.; Stark, W. J. *Angew. Chem., Int. Ed.* **2015**, *54*, 2552−2555.
(116) Ecker, M.; Pretsch, T. *RSC Adv.* **2014**, *4*, 286−292.
(117) Bae, H. J.; et al. *Adv. Mater.* **2015**, *27*, 2083−2089.
(118) Hou, X.; Ke, C.; Bruns, C. J.; McGonigal, P. R.; Pettman, R. B.; Stoddart, J. F. *Nat. Commun.* **2015**, *6*, 6884.

DOI: 10.1021/acs.macromol.5b00890
Macromolecules 2015, 48, 4759−4767