A New Class of Materials: Sequence-Defined Macromolecules and Their Emerging Applications

Michael A. R. Meier* and Christopher Barner-Kowollik*

Emerging applications of a new class of materials, sequence-defined macromolecules, are explored. Such molecularly highly defined macromolecules require stringent synthesis and purification procedures, yet offer unprecedented application possibilities. The first examples of molecular data storage and related technologies are already starting to emerge today. From a more fundamental point of view, such macromolecules offer a unique opportunity to determine quantitative structure–property relationships (QSPR), which critically aids in designing materials with applications ranging from catalysis to artificial enzymes.

1. Introduction

Ever since Hermann Staudinger’s macromolecular hypothesis in the early 1920s, polymer chemists were not only driven by enhancing the accessible property profiles of synthetic macromolecules, but also by enhancing the level of structural control that can be exerted over the polymer chain during the synthetic processes (Figure 1).[1] Over the past 70 years, methodologies have emerged that allow to limit the molecular weight dispersity of polymer chains, yet none of these have been able to reduce it to unity. The situation is even more challenging if more than one monomer is incorporated, leading to statistically distributed repeating units, i.e., inducing even more dispersity (Figure 1, left). Achieving monodisperse, sequence-defined synthetic macromolecules requires nonstatistical and highly efficient synthetic processes (Figure 1, right), effectively using the entire repertoire of classic organic chemistry to construct large molecules where every building block has a defined and pre-determined place within the polymer chain—similar to naturally occurring polymers consisting of highly defined peptide sequences or DNA. While the synthetic effort thus increases substantially, the properties of the resulting soft matter materials become highly defined and lend themselves for applications that classical polymers are unable to fulfill. It is the aim of this Research News article to highlight current and emerging applications of such highly defined systems that are poised to change the way we view synthetic macromolecules.

2. Synthetic Approaches

While we will herein focus on emerging and future material applications of synthetic sequence-defined macromolecules, we briefly highlight contemporary synthetic processes to sequence-defined polymers as they critically influence the building block variety and achievable chain length, which in turn dictate the resulting material properties and thus applications. The synthetic approaches toward sequence-defined macromolecules were recently summarized and can generally be divided into liquid-phase approaches, solid-phase approaches, and fluorous-phase or polymer-tethered methodologies,[2] with all their individual advantages and disadvantages. Linear synthetic approaches can install one monomer unit per iterative cycle (Figure 2), be it in a submonomer strategy or by installing the monomer completely, for instance, via single unit monomer insertion (SUMI). By contrast, bidirectional growth (Figure 2) commences from not only one reactive site but two, thus leading to higher degrees of polymerization more rapidly, and symmetric macromolecules. Of course, also multidirectional approaches, leading to star-shaped macromolecules, are possible. The strategy leading to the fast build-up of molecular weight is the so-called iterative exponential growth (IEG, Figure 2), often termed divergent/convergent approach, which, however, offers the least control over the obtained sequence (i.e., only one type of monomer is usually used). The latter two approaches are not well-suited for solid-phase or fluorous-phase synthesis and the linear approach offers the highest possible degree of definition. Thus, several prime examples following this linear
approach were described over the last few years. For selected examples, refer to references [3–9], which are also compared in detail in a recent review [2]. Interestingly, recently researchers started to combine different synthetic approaches [10,11] in their quest for the best balance of achievable yield, degree of polymerization, structural variety, scalability, and other parameters, which all certainly depends on the target application.

3. Sequence Enables Data Storage with First Applications in Compound Tagging and Cryptography

Nature exploits sequence-defined polymers for good reason: It stores different types of information into these entities, ranging from the instruction on how to fold a complex macromolecule into an effective nanoreactor (i.e., an enzyme) to the storage of genetic information. It thus appears attractive to exploit sequence-defined macromolecules as a medium for coding and decoding of information at the molecular level [12,13]. Compared to DNA, synthetic polymer chemistry offers a unique opportunity to significantly increase the storage density. Computers use a binary system to store data. In terms of repeating units in sequence-defined macromolecules, this translates to the storage of one bit per repeat unit if a binary system (i.e., two different repeat units in a sequence-defined macromolecule) is used. DNA, on the other hand, uses four base pairs to code information in a quaternary system. Thus, if 256 permutations (i.e., 8 bit = 1 byte) are stored in DNA, a sequence of theoretically only four nucleobases (i.e., a tetramer) is necessary, whereas an octamer is necessary in the binary system. The information density thus directly correlates with the variation of possibilities per repeat unit, i.e., the $X$ in the $X^Y$ notation of possible permutations (where $X$ is the base describing the available different repeat units (often different side-chains) and $Y$ is the degree of polymerization). It is obvious that the actual amount of stored information (i.e., the number of possible permutations) even more depends on the degree of polymerization, i.e., the exponent $Y$ in the above simple equation. Thus, high degrees of polymerization, while maintaining monodispersity and absolute control over each and every repeat unit, are certainly a target of ongoing research. The current record in this sense is the storage of 24 bits per repeat unit and the formal storage of almost 100 bits in a sequence-defined tetramer that was obtained via two multi-component reactions offering six variation sites per repeat unit and thus a high number of possible permutations (Figure 3) [14]. As important as storing data in a macromolecule is the read-out of the stored information. To date, MS/MS techniques are exclusively described, but in the future, more possibilities will certainly be demonstrated. MS/MS-based read-out strategies exploiting tandem mass spectrometry can be compared to

![Figure 1. Current situation in polymer synthesis leading to disperse polymers (left) and the transition to sequence-defined, unique macromolecules (right).](https://www.advancedsciencenews.com)

![Figure 2. Schematic representation of the main synthetic approaches toward sequence-defined and monodisperse macromolecules. IEG: iterative exponential growth (often also termed divergent/convergent approach); PG: protecting group.](https://www.advancedsciencenews.com)
MS-sequencing of peptides or DNA and it was demonstrated for several synthetic sequence-defined systems as well.\cite{8,15,16} Of course, designing macromolecules with distinct and facile fragmentation,\cite{17,18} a computer assisted data interpretation\cite{19} as well as a possible erasing of the data,\cite{20} are essential to achieve practical application possibilities in the future. Recently, Lutz and co-workers also demonstrated the coding of 64 bit within one macromolecule using a binary system, where the readout was simplified by inter-byte fragmentation enabled by exactly positioned mass tags (Figure 3).\cite{21} Concerning foreseeable applications beyond data storage, where synthetic macromolecules offer the possibility to be designed inertly and thus featuring long-term stability, recently, anticounterfeit tags and applications in cryptography were discussed. As unique product identification is a highly relevant issue in our modern society, sequence-coded oligourethanes were used as barcodes to demonstrate the labeling of polymeric materials\cite{22,23} as well as of 3D-printed objects\cite{23} successfully. Very recently, a cryptography method using “molecular keys” (obtained via efficient Ugi-reactions) for data encryption and decryption was introduced.\cite{24} Since these molecular keys can also be hidden, for instance on paper or in a perfume, the combination of cryptography and steganography offers a new and nondigital channel for key distribution. If the key is considered unbreakable by current methods (at the moment 100 bits are considered safe), an unlimited amount of data can be transferred digitally in a highly secured fashion.

4. Sequence Leads to Properties and Material Applications in Catalysis and Optoelectronic Applications

Nature is a master of catalysis. Exploiting precision macromolecules that feature no dispersity (apart from the inherent isotope distribution) and a perfectly arranged order of building blocks, allows generating precisely folded structures that present catalytic pockets for highly specific reactions. Without such natural catalysts, life would not exist. The key concepts that nature employs in biocatalysis can not only serve as a blueprint to emulate enzymes, but also generates other new application possibilities, using synthetically generated macromolecules. As is the case for their natural counterparts, synthetic sequence-defined macromolecules also allow a precise design of inter- and intramolecular interactions and thus also certain functions.\cite{25} As a recent example, using synthetic sequence-defined peptides, specific solubilizers for a potential anti-Alzheimer disease drug were designed.\cite{26} Also in this context, peptoids, showing a similar structure to natural peptides but without chirality and with substituted nitrogen atoms, have advanced to being interesting candidates of peptidomimetics with application possibilities in the pharmaceutical industry.\cite{27,28} Considerable efforts have been directed at the aim of precision naturally analogue nanoreactors, yet the employed macromolecules typically rest on reversible radical deactivation polymerization (RDRP) and thus feature an inherent dispersity as well as a variation in the placement of the catalytically active ligation units or binding sites that affect the folding.\cite{29} The inherent dispersity of the chain ensemble leads to the effect that every folded nanoparticle (or enzyme mimic) is slightly different as that of this ensemble; only a fraction of molecules feature the correct folding geometry, thus creating a catalytic pocket ideally suited for the substrate. While these critical deficiencies are to a certain extent off-set by the relative ease of synthesis of the functional precursor polymers, the situation is far from ideal. Ideally, sequence-defined functional polymers will allow placing folding moieties as well as catalytically active units at exact points within the polymer chain, thus leading to folded entities with unique geometries and highly defined catalytic pockets (Figure 4). Designing such entities...
requires close collaboration of polymer scientists with bio- and theoretical chemists. Contemporary synthetic technology allows only the generation of relatively short sequences in the majority of cases, certainly not within the realm of proteins or enzymes (typically exceeding 60 kDa). However, catalysts are typically only required in small amounts and thus the limited quantities of sequence-defined polymers accessible today make their application in precision folded catalytic entities a viable prospect. Moreover, a combination of sequence-definition with methods of sequence-control might lead to more practical synthesis routes, as, for instance, the surrounding of the catalytic pocket in terms of solubility, stability, and other important factors might be achieved using more classical synthesis methods. Advantages of synthetic (metallo)enzyme mimetics include increased stability toward temperature, as well as solvents and catalysts that can be engineered for ready recovery, which sets these systems apart from less stable micellar constructs. In addition, such structures have the potential ability to be switched on and off in their catalytic activity by photonic and electric fields, thus enabling precision catalysis that can be halted and recommenced at will.\(^{[10]}\)

We submit that the realm of bio-inspired catalysis is a key area to benefit from sequence-defined polymers, yet not before these become available in a similar structural variety of contemporary synthetic polymers. Functional sequence-defined polymer systems that have, to-date, come close to providing an entry point into bio-inspired catalysis are those that provide the longest possible sequences, i.e., in excess of ten sequence-defined units.

Far less developed is sequence definition in conjugated, stiff oligomers that should allow the precise determination of structure-activity relationships for achieving photophysical properties for applications ranging from sensors to solar cells and organic light-emitting diodes (OLEDs). High degrees of polymerization in these stiff and conjugated molecules were already achieved.\(^{[2]}\) For instance, monodisperse oligofluorenes were obtained in an IEG approach (refer to Figure 2) up to the 64-mer stage and their degree of polymerization was correlated to their solid-phase morphology.\(^{[11]}\) This and other examples\(^{[2]}\) are highly impressive, but true sequence-definition in this subfield was only achieved for some architectures up to the trimer stage for linear approaches.\(^{[12–15]}\) Apart from the discussed OLEDs and solar cells, such highly defined conjugated macro-molecules should, for instance, allow applications as molecular wires.\(^{[16]}\) In all the cases, an exact positioning of donor and acceptor units, as well as the high definition of the conjugation lengths should allow for tuning of the application properties to the respective requirements.

Beyond the above discussed application fields, we further envisage that the realm of precision network formation will critically benefit from sequence-defined polymers. These macromolecules allow constructing networks that have exactly defined building blocks, with exactly placed functionalities within them. For instance, the exact placement of functionality (such as lanthanides) within a well-ordered network will allow constructing next generation NMR alignment media for the elucidation of complex chiral molecules,\(^{[17]}\) as well as tailored precision environments for cells based on well-defined hydrogels. Further, considerable innovation potential exists for applications where the mesh size of polymer networks needs to be finely controlled, e.g., in applications that target ion capture such as sea water desalination.\(^{[18]}\) Here, sequence-defined polymers will enable the design of highly regular and repetitive molecular cavities within the network structure to enable a perfect fit for specific ions. Such highly controlled networks are inaccessible based on contemporary polymer synthesis.

**Acknowledgements**

M.A.R.M. and C.B.-K. acknowledge the German Research Foundation (DFG) for continued funding in the context of the Collaborative Research Centre SFB1176 (project A3). C.B.-K. acknowledges an Australian Research Council (ARC) Laureate Fellowship enabling his photochemical research program as well as continued key support from the Queensland University of Technology (QUT) and by the Karlsruhe Institute of Technology (KIT) via the STN program of the Helmholtz association.

**Conflict of Interest**

The authors declare no conflict of interest.

**Keywords**

sequence definition, macromolecules, quantitative structure–property relationships (QSPR), molecular data storage

Received: September 16, 2018
Revised: October 17, 2018
Published online: January 2, 2019
