Sequence-controlled Polymer: Design, Development and Prospects

IOP Conf. Series: Materials Science and Engineering **493** (2019) 012058 doi:10.1088/1757-899X/493/1/012058

Sequence-controlled Polymer: Design, Development and Prospects

Yuwei Zhang *

College of Materials Science and Technology, University of Science and Technology Beijing, Beijing, China

*Corresponding author e-mail: 1120763042@qq.com

Abstract. Sequence control is of crucial significance in biology. There are great prospects in the application of information storage and transmission by controlling the sequence structure so as to control sequenced compounds. At present, although there are effective synthetic methods for both biosynthesis and chemical synthesis, the severest challenge may be that there is currently no general strategy for synthesizing sequence-controlled polymers.

1. Introduction

During its nearly 100-year history of development, a core idea of polymer science is to regulate the properties of polymer materials by regulating polymer structure. However, the corresponding relation between material structure and material properties has always been vague, especially for polymer materials. This is because polymer materials are mixtures rather than compounds, polymer materials have no clear molecular weight, and the chirality is also uncontrollable.

Polymerization developed from the initial uncontrollable stepwise polymerization and chain polymerization to the later living anionic polymerization, living cationic polymerization and controlled radical polymerization. On the one hand, the development of synthetic methods has shifted the molecular weight of the polymer from a broad distribution to a narrow distribution, aiming to obtain a monodisperse polymer. On the other hand, Ziegler-Natta catalyst was used to obtain stereoregular polypropylene and then a series of chiral catalysts were developed to control the chirality during polymerization so as to obtain stereoregular polymers [1]. The two efforts share the same direction, namely, synthesizing polymer compounds and obtaining a single chiral, single molecular weight polymer rather than a mixture. Of course, the further aim is to precisely control the sequence structure of polymers. Sequence control is ubiquitous in biopolymers (such as proteins, DNA, and RNA), and specific biological functions of biomolecules are inextricably linked to their precisely controlled sequence structure. However, when the sequence structure is mismatched in the course of life, the living body will be disordered and thus invite diseases.

Therefore, sequence control is of vital significance in biology [2-6]. The genetic information of human beings and other plants and animals is stored and transmitted by specific DNA sequences A, T, G, and C on DNA. In addition, the amino acid sequence on the polypeptide chain determines the secondary structure of the polypeptide chain and even the higher-order topology, which in turn affects the biological properties and functions of proteins [7-8].

2. Synthesis Strategies

DNA and RNA, two major nucleic acids, have been discovered in biology. DNA exists in the cell nuclear and stores the genetic information of living organisms. While the types and functions of RNA are relatively abundant, such as messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA). There are two main forms of polymerization in living organisms to obtain nucleic acids of a particular sequence, namely DNA replication (DNA→DNA) and transcription (DNA → mRNA). DNA replication is equivalent to a template polymerization in which double-stranded DNA with a specific sequence structure is replicated into two DNA strands, namely, a leading strand and a lagging strand. For researchers dedicated to sequence-controlled polymerization, the synthesis of the dominant chain is an inspiring life process. During the replication process, the newly generated DNA strand forms a specific sequence with the help of DNA polymerase. This synthesis is directional, and the polymerase travels along the template strand after each monomer linkage, thus allowing its sequence to be accurately replicated. The transcription mechanism of DNA to mRNA is very similar to the synthesis of the leader strand in replication, which dissociates double-stranded DNA and helps form phosphodiester bonds by specific RNA polymerase. In addition, deoxyribose nucleoside triphosphate (dNTP) in the replication of the monomeric DNA is the equivalent based on ribose.

2.1. Biological Method

Proteins are biomacromolecules with specific sequence structure constructed by 22 different α-amino acid monomers (21 in eukaryotes). The main chain structure of proteins (namely the amino acid sequence in the peptide chain) affects the secondary, tertiary, and quaternary structure, thereby influencing the biological properties and functions of proteins. In an organism, proteins are synthesized by a translation process (mRNA → protein) in which the nucleotide sequence of the mRNA is transcribed into an amino acid sequence. From the perspective of polymer chemistry, translation may be the most reasonable and successful polymerization mechanism that has been identified so far, as it allows for the transfer of information between two completely different types of polymeric backbones, namely, from poly (phosphodiester) to polyamide.

At present, researchers pay much attention to and have great interest in how to introduce specific monomer sequences into polymer synthesis. After years of experimentation, two research trends have been developed to control the structure of monomer sequences in polymers. The first method is to use the biological concept of natural optimization for sequence regulation. For example, DNA templates, enzymes or even living organisms can be used to obtain sequence-controlled polymers. These natural mechanisms are adapted to tolerate non-natural monomers. Another trend is to prepare sequence-controlled polymers by use of synthetic chemistry such as solid phase synthesis and liquid phase synthesis. Sequence polymers of certain specific monomers can be even synthesized quickly and easily by use of an automatic synthesizer, which has many constraints on substrates, though.

Similar to the role of DNA in genes, the specific sequence structure in polymers can serve as information storage devices at the molecular level where each functional side group can be considered as an information “bit”. Therefore, oligomers with specific sequence structure play an important role in the field of polymer basic research and many technical applications. Obviously, the applicability of DNA as an operable sequence-controlled polymer is not unique, and chemistry and biology offer many interesting alternatives for the preparation of sequence-determined macromolecules. However, the main challenge lies in that no general strategy for synthesizing sequence-controlled polymers has been found.

DNA replication, DNA→RNA transcription and RNA→protein translation are three dominant sequence-controlled polymerization processes in biological organisms, and the most complex part is the translation mechanism which relies on ribosomes, large catalytic particles composed of RNA and proteins. Biosynthesis, and the three processes mentioned above demonstrate significant advantages because they are more demanding than any other known synthetic polymerizations. Orgel and his colleagues optimized the non-nucleic acid replication of nucleic acids and initiated the development of this protein-free replication system in which the activated nucleotide was linked to the oligonucleotide template by Watson-Crick base pairing [9] and polymerized through chemical means (namely in the

absence of enzyme). This protein-free replication system lacked efficiency and replication fidelity and the newly formed oligomers wound bind strongly to the template, hindering the repetition of the process [10]. The complex process of another polymerization strategy based on biology is the use of enzymes for in vitro replication, the most important example being PCR. The high temperature required for chain fragmentation in PCR largely limits the use of DNA polymerase I, and the subsequent use of thermostable Taq polymerase has successfully made PCR commercialized. However, PCR is limited to natural nucleic acids. Though it was later applied to non-natural monomers, but the fidelity was limited. The ribosome mechanism used to produce sequence-controlled polymers plays an important role in genetic engineering. Traditional genetic engineering is the one in which artificial genes of encoded proteins are integrated into plasmid DNA and the latter is introduced into a bacterial host. Later, unnatural amino acids could also be incorporated into proteins, namely, replacing natural amino acids with irregular residues. In this process, however, the yield, species and functions of polymers are restricted [11].

David R. Liu and his colleagues constructed cyclic monomer modules by linking polypeptide that could be polymerized, polyethylene glycol and other skeletons through disulfide bonds and pent peptide nucleic acid: the peptide nucleic acid and the parental DNA were coupled by complementary base pairing and then were linked with monomer modules through the Alkynyl-Azide click reaction; polymer chains were detached from the peptide nucleic acid after the disulfide bonds were broken; different sequence structure could be introduced due to the relative independence of the skeleton and the template. The construction idea was based on a stepwise process instead of a chain process [12-13] (Figure 1). In addition, Leigh and his colleagues designed the rotaxane molecular machine in which amino acid units were sequentially immobilized on a linear template, and a cyclic monomer with mercapto walked on the linear template and sequentially connected those amino acid units [14] (Figure 2).

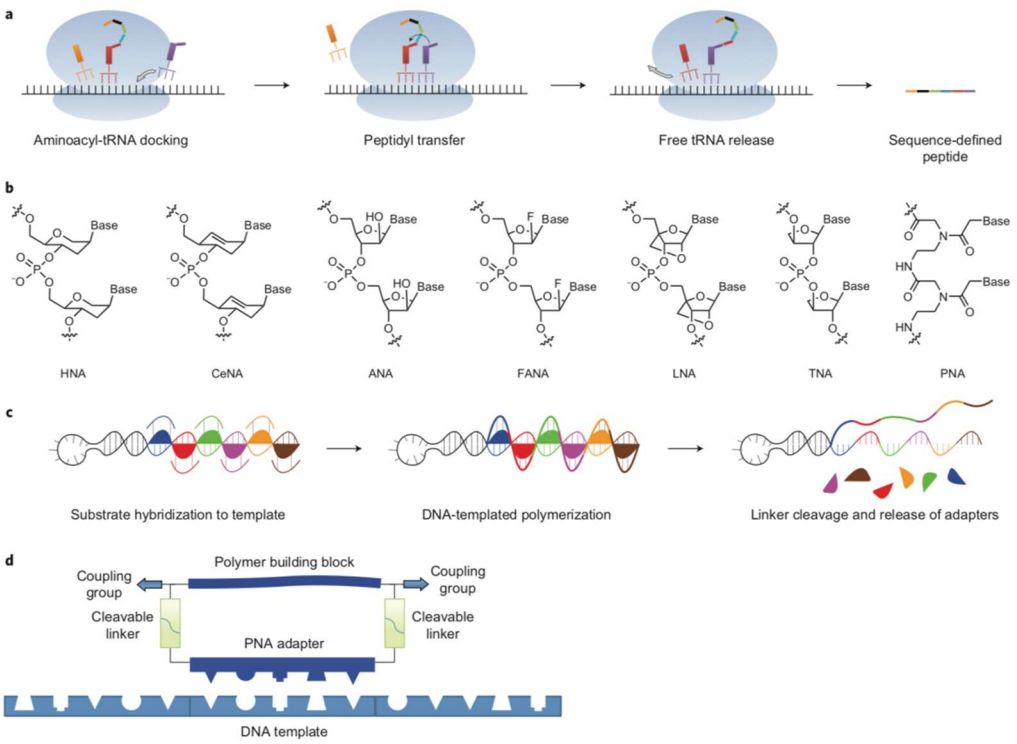

Figure 1. Synthesis of precisely sequence-controlled polymer by using DNA as a template

2.2. Chemical Method

Chemical synthesis can be used to prepare sequence-controlled macromolecules with different chemical structure. In addition, compared with DNA technology, chemical programs can produce simpler and cheaper sequence-defined materials on a larger scale. However, fully synthetic sequence-controlled schemes are more challenging than biotechnological approaches and are generally limited to the synthesis of short oligomers. The most obvious way to prepare a fragment with particular sequence is to link the monomer units one by one by following iterative chemical procedures. However, it is still a cumbersome process requiring high reaction yields and repeated purification procedures.

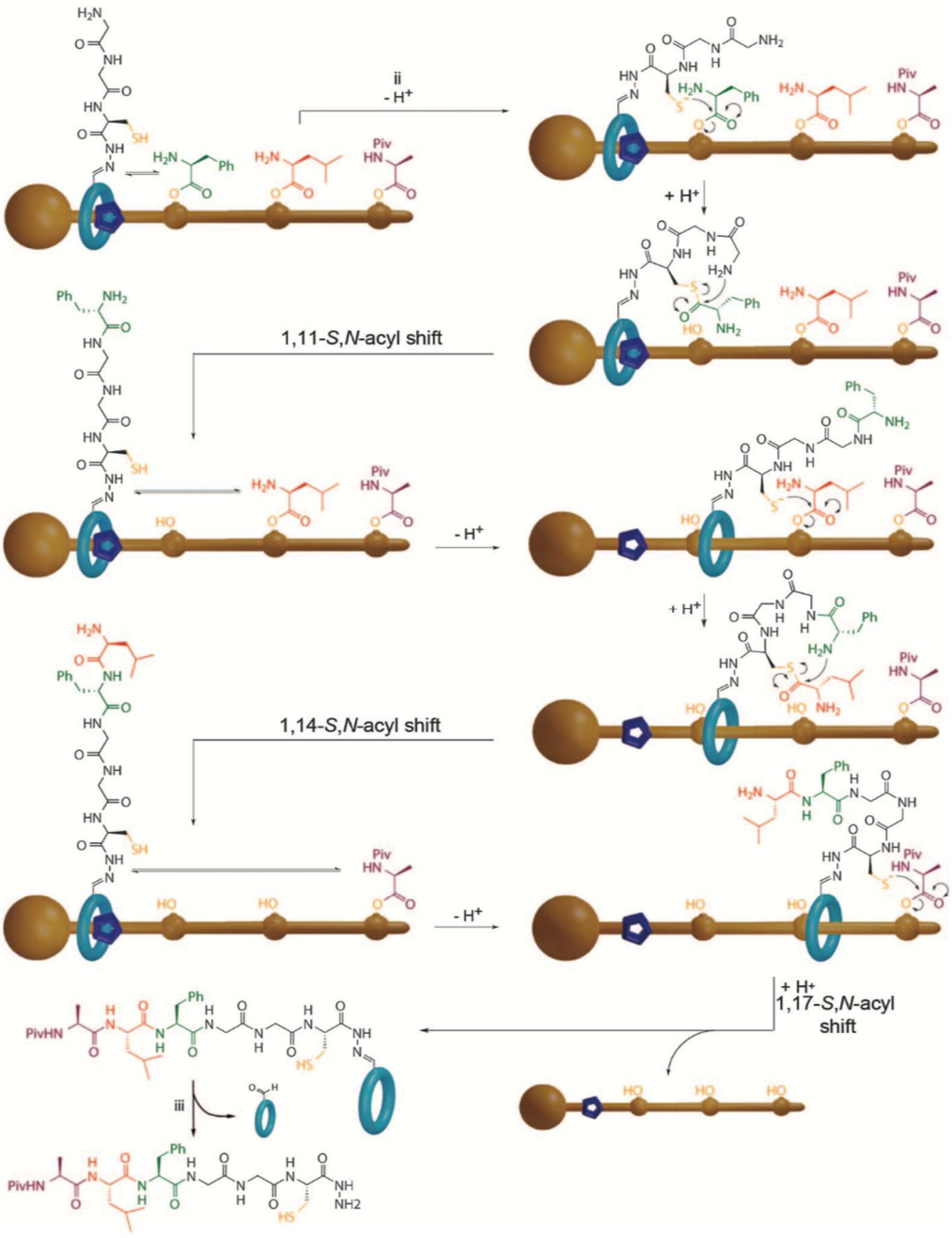

Figure 2. Synthesis of precisely sequence-controlled peptides by molecular machinery

Chemical-based polymerization processes can result in more abundant structure than biological polymerization and one of the best approaches are the iterative synthesis of insoluble supports. However, due to the limited access to the reaction sites, the coupling efficiency is affected. An effective alternative to this problem is to use a single soluble polymer chain as the support for the preparation of new non-natural sequence-controlled oligomers, which is just an innovation rather than an individual technology, though. A more important challenge confronting this area is the control of monomer sequences, such as chain propagation and gradient growth, in a "traditional" polymerization process. For example, if adding monomers in active chain growth polymerization can be controlled, the stepwise addition of equimolar amounts of monomers and initiators will allow for sequence adjustment. However, even when an active polymerization mechanism that is highly controlled and involving dormant substance is adopted, this "monomer addition" strategy is difficult to realize due to the inherent tendency of the monomer to react with itself (namely, in the communication process). One way to overcome this problem is to gradually reduce the reactivity of the monomers added.

Sequence regulation can also be realized in chain-growth polymerization by using specific comonomer pairs. By controlling the time of addition, the position of the receptor monomer units can be precisely controlled in the backbone of polymers. However, this chain-growth method results in deviations of the length, composition and order between chain segments. Although these defects cannot be completely suppressed, they can be greatly reduced by adopting an optimized polymerization scheme.

Stepwise growth polymerization also allows for sequential adjustment of macromolecular synthesis and is suitable for the preparation of periodic microstructure. In fact, sequence-defined oligomers containing a reactive chain ends can be polymerized through stepwise growth. Another strategy for sequence regulation in synthesis is catalysis, which is still in the exploratory phase.

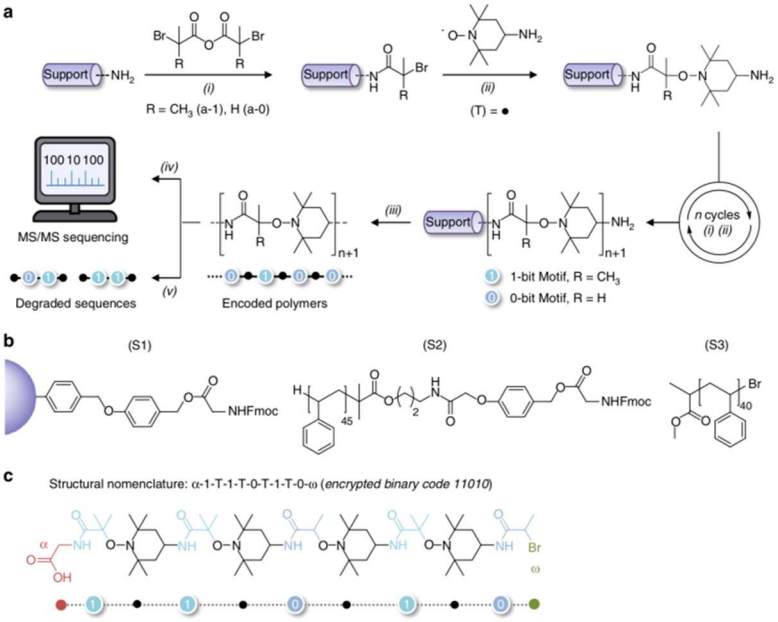

Figure 3. Sequence-readable synthetic polymers

Based on bromoacetic anhydride and amino TEMPO as the reaction structure units, Lutz and his colleagues synthesized monodisperse sequence-encoded poly (alkoxyamine amide)s through an iterative strategy involving two chemoselective steps: the reaction of a primary amine with an acid anhydride and the radical coupling of a carbon-centered radical with a nitroxide. The use of brominated anhydrides to write sequence information on the polymer chain and the weaker and easily cleavable nature of alkoxy bond made it possible to read information on the sequence polymers by use of a two-stage tandem mass

spectrometer. On this basis, by combining the synthesis of poly phosphodiester and the easily cleavable nature of poly(alkoxyamine amide)s which could read information, Lutz and his colleagues made phosphoramidite with bromoacetamide and TEMPO with hydroxyl as structural units, realized a three-step cycle strategy to achieve chain growth through the reaction of hydroxyl and phosphoramidite and its oxidation and the radical coupling reaction of bromine and TEMPO, and employed various phosphoramidite to introduced into polymer chains sequence information which could be read by secondary tandem mass spectrometry[16].

3. Conclusions and Outlook

Compared with material used on a daily basis, the synthesis of sequence polymers is characterized by clear sequence structure, and the secondary structure of the sequence polymer or even the properties of the material may be precisely predicted. Moreover, the most important property of sequence polymers is reflected in the sequence which corresponds to information stored. Currently, research has shown that the sequence information on polymers is readable. Since chemically synthesized polymers have a wide range of raw materials and structure units, it can store more information than the current binary system. Taking DNA as an example, it stores the genetic information of most living organisms in the world with only four sequence combinations of deoxyribonucleic acids. Therefore, the use of synthetic sequence polymers to store and transfer information is a very promising and challenging research direction [2-3].

Over the past century witnessing the development of polymer science, the pursuit of control over polymers' precise structure has been largely inspired by biopolymers widely found in nature. Biopolymers, such as proteins, DNA, and RNA, can form complex and defined structure due to their precise sequence structure and stereo configuration, thus playing a vital role in life activities. Compared with the life system, the regulation of precise structure in current research on polymers is still limited. With the hundred years of the development of polymer chemistry and gradually clear understanding of the life system combined, it is believed that there will be significant progress made in both synthesis methods and applications of synthetic sequence-controlled polymers.

References

[1] Corradini, P.; Guerra, G.; Cavallo, L., Do New Century Catalysts Unravel the Mechanism of Stereocontrol of Old Ziegler−Natta Catalysts? Accounts Chem Res 2004, 37 (4), 231-241.
[2] Lutz, J. F., Defining the Field of Sequence-Controlled Polymers. Macromol Rapid Commun 2017, 38 (24).
[3] Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M., Sequence-Controlled Polymers. Science 2013, 341 (6146).
[4] Matyjaszewski, K., Architecturally Complex Polymers with Controlled Heterogeneity. Science 2011, 333 (6046), 1104-1105.
[5] Matyjaszewski, K., Macromolecular engineering: From rational design through precise macromolecular synthesis and processing to targeted macroscopic material properties. Prog Polym Sci 2005, 30 (8), 858-875.
[6] Badi, N.; Lutz, J.-F., Sequence control in polymer synthesis. Chem Soc Rev 2009, 38 (12), 3383-3390.
[7] Sanger, F.; Tuppy, H., The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. Biochem J 1951, 49 (4), 481-490.
[8] Perutz, M. F., Relation between Structure and Sequence of Hæmoglobin. Nature 1962, 194, 914.
[9] Watson, J. D.; Crick, F. H. C., Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature 1953, 171, 737.
[10] Taylor, A. I.; Pinheiro, V. B.; Smola, M. J.; Morgunov, A. S.; Peak-Chew, S.; Cozens, C.; Weeks, K. M.; Herdewijn, P.; Holliger, P., Catalysts from synthetic genetic polymers. Nature 2014, 518, 427.
[11] Meng, W.; Muscat, R. A.; McKee, M. L.; Milnes, P. J.; El-Sagheer, A. H.; Bath, J.; Davis, B. G.; Brown, T.; O'Reilly, R. K.; Turberfield, A. J., An autonomous molecular assembler for

programmable chemical synthesis. Nat Chem 2016, 8, 542.
[12] Kleiner, R. E.; Brudno, Y.; Birnbaum, M. E.; Liu, D. R., DNA-templated polymerization of side-chain-functionalized peptide nucleic acid aldehydes. J Am Chem Soc 2008, 130 (14), 4646-59.
[13] Niu, J.; Hili, R.; Liu, D. R., Enzyme-free translation of DNA into sequence-defined synthetic polymers structurally unrelated to nucleic acids. Nat Chem 2013, 5 (4), 282-92.
[14] Lewandowski, B.; De Bo, G.; Ward, J. W.; Papmeyer, M.; Kuschel, S.; Aldegunde, M. J.; Gramlich, P. M. E.; Heckmann, D.; Goldup, S. M.; D'Souza, D. M.; Fernandes, A. E.; Leigh, D. A., Sequence-Specific Peptide Synthesis by an Artificial Small-Molecule Machine. Science 2013, 339 (6116), 189-193.
[15] Roy, R. K.; Meszynska, A.; Laure, C.; Charles, L.; Verchin, C.; Lutz, J.-F., Design and synthesis of digitally encoded polymers that can be decoded and erased. Nat Commun 2015, 6, 7237.
[16] Cavallo, G.; Al Ouahabi, A.; Oswald, L.; Charles, L.; Lutz, J.-F., Orthogonal Synthesis of "Easy-to-Read" Information-Containing Polymers Using Phosphoramidite and Radical Coupling Steps. J Am Chem Soc 2016, 138 (30), 9417-9420.