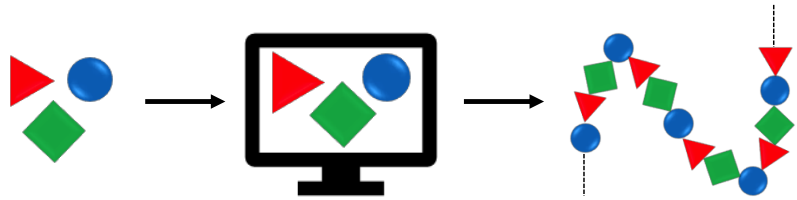
Exploring strategies to bias sequence in natural and synthetic oligomers and polymers

Jan K. Szymański,[†] Yousef M. Abul-Haija,[†] and Leroy Cronin*

WestCHEM, School of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK

CONSPECTUS



Millions of years biological evolution has driven the development of the highly sophisticated molecular machinery found within living systems. These systems produce polymers such as proteins and nucleic acids with incredible fidelity and function. In nature, the precise molecular sequence is the factor that determines the function of these macromolecules. Given that the ability to precisely define sequence emerges naturally, the fact that biology achieves unprecedented control over the unit sequence of the monomers through evolved enzymatic catalysis is incredible. Indeed, the ability to engineer systems that allow polymer synthesis with precise polymer sequence control is a feat that technology is yet to replicate in artificial synthetic systems. This is because, without access to evolutionary control for finely-tuned biological catalysts, the ability to correct errors, harness multiple competing processes, means that the prospects for digital control of polymerization have been firmly bootstrapped to biological systems or limited to stepwise synthetic protocols.

Here we give an overview of strategies that have been used over the last five years in efforts to program polymer synthesis with sequence control in the laboratory. We also briefly explore how the use of robotics, algorithms, and employing stochastic chemical processes

might lead to new understanding, mechanism, and strategies to achieve full digital control. The aim is to see if it is possible to go beyond bootstrapping to biological polymers, or step-wise chemical synthesis. We start by describing non-enzymatic techniques used to obtain sequence-controlled natural polymers, a field which lends itself to direct application of insights gleaned from biology. We discuss major advances, such as the use of rotaxane-based molecular machines and templated approaches, including using biological polymers as templates for purely synthetic chains. We then discuss synthetic polymer chemistry, whose array of techniques allow the production of polymers with enormous structural and their functional diversity is well established, but so far with only limited control over the unit sequence itself.

Synthetic polymers can be subdivided into multiple classes depending on the nature of processes used to synthesize them, such as by addition or condensation. Consequently, varied approaches for sequence control have been demonstrated in the area, including but not limited to click reactions, iterative solid-phase-chemistry and exploiting chemical affinity of the monomers themselves. In addition to those, we highlight the importance of environmental bias on possible control of polymerization on single-unit level, such as using catalyst switching or external stimuli.

Even the most successful experimental sequence control approach needs appropriate tools to verify its scope and validity; therefore, we devote part of the present account to possible analytical approaches to sequence readout, starting with a well-established tandem mass spectrometry techniques and touching on those more applicable in specific classes of processes, such as diffusion-ordered NMR. Finally, we discuss progress in modeling and automation of sequence-controlled polymers.

We postulate that developments in analytical chemistry, bio-informatics, and computer modelling will lead to new ways of exploring the development of new strategies for the realization of sequence control by means of sequence bias. This is because, by treating the assembly of polymers as a network of chemical reactions, it will become possible to develop control strategies which can bias the outcome of the polymer assembly. The grand aim would be the synthesis of complex polymers in one step, yet with a precisely defined digital sequence.

1. INTRODUCTION

Modern biology owes its extraordinary chemical complexity to functional oligomers and polymers including sugars, proteins and nucleic acids, which have been designed, refined, and adapted by the process of evolution. Their utility is set by the primary linear sequence of units in the chains, for instance amino acids in proteins, and nucleobases in nucleic acids. From this primary sequence, additional structural features from secondary, tertiary and even quaternary structures emerge. The fact that the pathophysiology of genetic disorders is manifested by defects in sequence attests to the immense importance carried by the sequence, down to single units whose replacement with seemingly related ones leads to widespread disruption of metabolic activities. Over the years, synthetic polymer chemists have been trying to establish an analogous relationship between structure and function of human-made macromolecules. To do this they have been trying to devise efficient ways of gaining control of the polymer sequence, but full control of every monomer identity remains a challenging, if not impossible prospect. Recent reviews describe progress from the last decade[1,2] but here we focus on progress since 2013 and show how this ambition and grand vision of sequence control at the molecular level might be achieved.

2. SEQUENCE IN NATURAL POLYMERS

Biological systems have almost complete sequence control of polymerization converting genetic code into defined protein sequences involving three steps: DNA replication, DNA to RNA transcription, and RNA to protein translation. Nature has always performed these

complex processes in a precise way, including ways to both limit and correct errors. The ability to implement sequence controlled polymerization with the precision and efficiency exhibited by biological systems would pave the way towards the development of new types of materials, that of sequence controlled 'inorganic' or non-biological matter. To achieve this aim, many different biological approaches to sequence control have been studied and explored. These include non-enzymatic templating to steer the coupling of simple monomers mainly through Watson-Crick base pairing.[3] In a more complex approach, enzymes have been used to catalyze oligomerization *in vitro*. A famous example is the polymerase chain reaction (PCR), a process that involves copying and amplification of a certain DNA sequence. Similar methods were also studied for non-natural nucleic acid polymerization. The most complex yet effective approach to develop sequence controlled polymerization is by using proxies in the form of living organisms, usually bacteria, through the introduction of artificial genes. This method is advantageous as it tolerates both natural and non-natural monomers. Chemistry-based systems have been also attempted as an alternative to biological approaches. The concept of the artificial ribosome based upon supramolecular molecular machines was recently developed.[4,5] This approach is based on rotaxane-based machine which travels along a track of amino acids coupling them in a sequence-controlled manner according to the movement direction of a thiolated ring (as illustrated in Figure 1).[5]

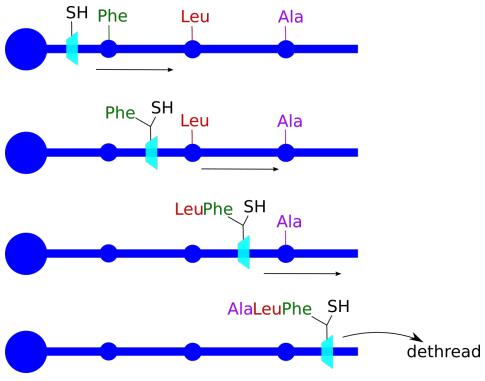


Figure 1. The concept of rotaxane-based molecular machine for controlled sequence peptide synthesis. Adapted with permission from ref. 5. Copyright 2013 AAAS.

Natural polymers as templates for sequence specific polymerization

Sequence controlled polymerization in natural polymers such as proteins is perfectly controlled by template-(macro)molecules bearing sequence information through which monomers can be selectively recognized and coupled. Inspired by that, natural template-assisted constrained peptide sequence synthesis and selection was recently achieved for an enzymatically catalyzed mixture utilizing electrostatic interactions between charged amino acids and oppositely charged polysaccharide templates[6]. In the absence of any template, peptide sequence selection could be also achieved under programmable reaction conditions. These are enzyme-assisted dipeptide polymerizations where the sequence of the most thermodynamically stable peptide was selected by the system.[7]

In a recent study, Liu and co-workers have developed an enzyme-free, DNA-templated translation system[8] which enabled translation of DNA into sequence-defined synthetic polymers, Figure 2. In this approach, macrocyclic substrates hybridize with codons on a DNA analogue (peptide nucleic acid (PNA)), allowing for polymer building block organization along the template, coupling and oligomerization. This is followed by linker cleavage, releasing the PNA adapters and liberating the product. Using this approach, 16 monomers were successfully coupled in a defined manner to form synthetic polymers with molecular weights of 26 kDa and 90 residues of densely functionalized β-amino acid.

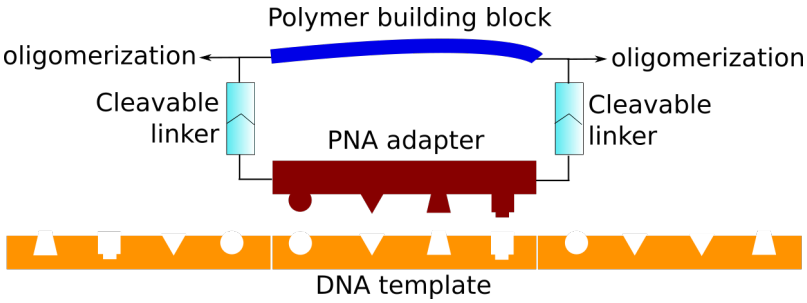


Figure 2. Schematic representation of sequence defined polymerization using enzyme-free, DNA-templated, synthesis of non-nucleic acid polymers. Adapted by permission from Macmillan Publishers Ltd: ref. 8, copyright 2013.

3. SEQUENCE CONTROL IN SYNTHETIC POLYMERS

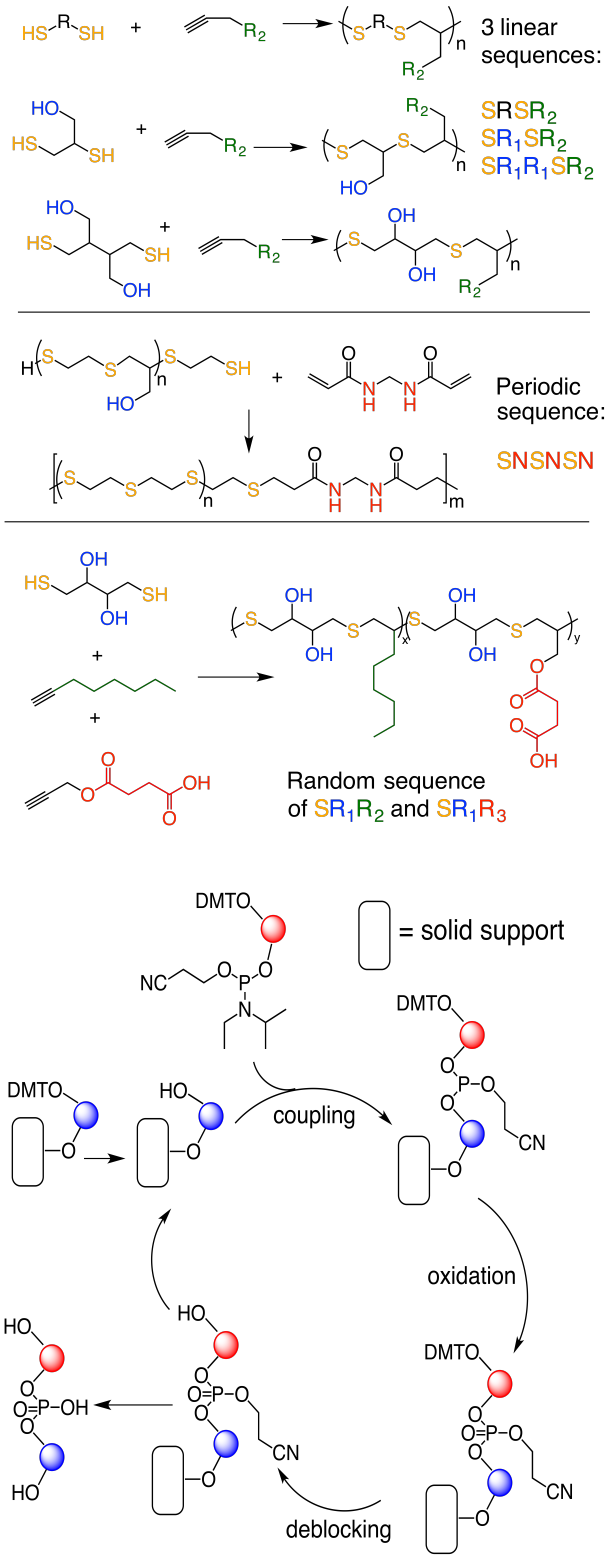
Sequence controlled polymerization controls all of biology yet today, but despite many years of developments in polymer science, molecular-level sequence control has not be achieved in any large scale technology. A key development in the future of polymer science could be the precise sequence control of polymeric materials. In this section, we discuss some promising research directions leading to sequence-controlled synthetic polymers.

3.1. Step-growth and multi-step-growth polymers

The common feature of polymerization processes described in this section is the sequential addition of monomers to a growing chain, typically involving functional group coupling and formation of byproducts. There are significant mechanistic differences among members of this group and they can be accordingly divided into step-growth and multi-step-growth processes.

Step-growth polymers. Polyesters and polyamides, used in many practical applications in daily-life, are leading examples of this type of polymers. Since the unit of addition takes place by reactions between functional groups on chain ends, sequence control has recently been demonstrated using click chemistry and by using multicomponent reactions. The click chemistry processes were envisioned by Sharpless[9] as reactions that are fast, highly specific and high-yielding, furnishing well-defined products in mild conditions. Their utility in polymer science is well known[10] and they have also found uses for control of chain sequences. In this regard the thiol-yne coupling was used by Han and coworkers[11] to produce a sequence-controlled polymer by exploiting successive additions of a thiol group to a triple carbon-

carbon bond (Figure 3, top panel). This allowed for obtaining random, periodic and linear



copolymers in controlled fashion.

Figure 3. Click approaches used in producing sequence-controlled polymers: thiol-yne coupling (top, adapted by permission from Macmillan Publishers Ltd: ref. 11, copyright 2014) in step-growth polymerization and phosphoramidite chemistry in a multi-step-growth process (bottom, adapted with permission from ref. 22. Copyright 2016 ACS Publications).

The development of multicomponent reactions for polymer science allows the inherent selectivity and atom economy, in some cases involving multiple substrates. They are useful for polymer chemistry as well-defined monomers can undergo reactions to furnish a polymer with a specific sequence. In one example, a fast Biginelli (Figure 4, left) reaction between a ketoester, an aldehyde and urea was used to produce sequence-controlled polymers under mild conditions.[12] Another multicomponent process, the Passerini (Figure 4, right) reaction between isocyanides, aldehydes and carboxylic acids, has also been developed.[13] In another example, amine-thiol-ene conjugation, followed by alkyne-azide-amine-coupling has been employed in an elegant sequence, which resulted in a well-defined polymer.[14]

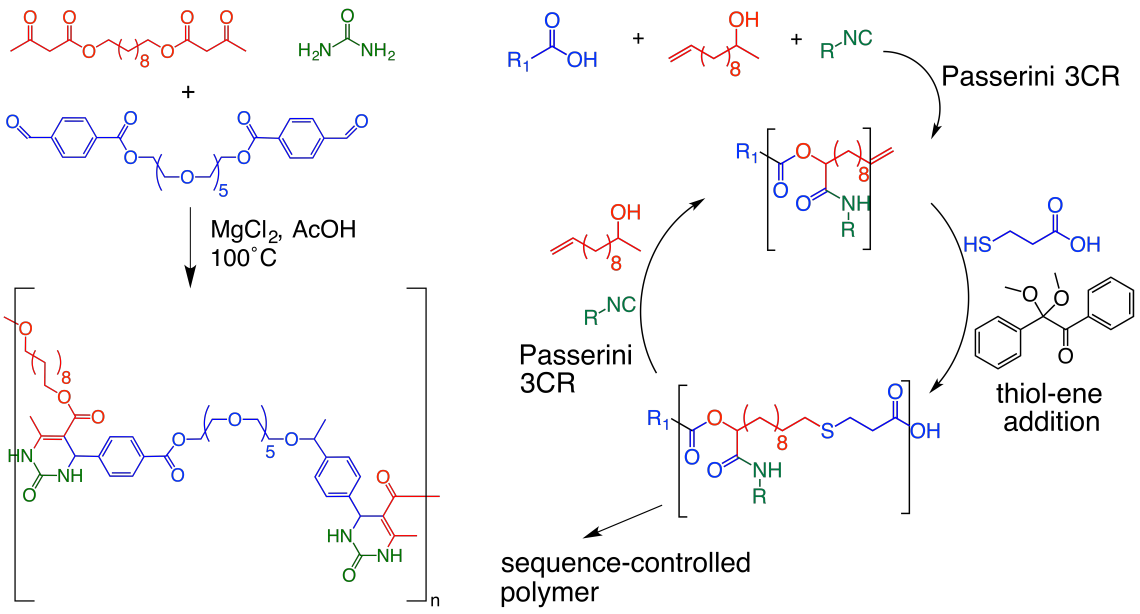


Figure 4. The three-component Biginelli (left, adapted with permission from ref. 12. Copyright 2016 ACS Publications) and Passerini (right, adapted with permission from ref 13.

Copyright 2014 Wiley) reactions and their utility for controlling sequence in step-growth polymerization.

Multi-step-growth polymers. This term refers to polymers produced using solid-phase iterative chemistry.[15] In contrast to step-growth polymers, chains grow only on one end, with the other tethered to a support. Unit sequence in such polymers has chiefly been controlled similar to solid-phase peptide synthesis, using the same cycles of binding and release.[16,17] One attractive aspect of the macromolecules used being synthetic is that they can be functionalized with groups allowing for orthogonal chemistry, and thus for dispensing with the costly binding and cleavage steps. Peptoids (N-substituted glycine polymers) are setting on the boundary of natural and synthetic polymers. Since their first synthesis using a step-wise method,[18] their use was substantially increased leading to various applications. There is an excellent review on this topic by Zuckermann and his co-workers.[19] This method can be also modified to use liquid phase-based supports, such as native polystyrene chains[20] or fluorinated hydrocarbon chains (Figure 5).[21]

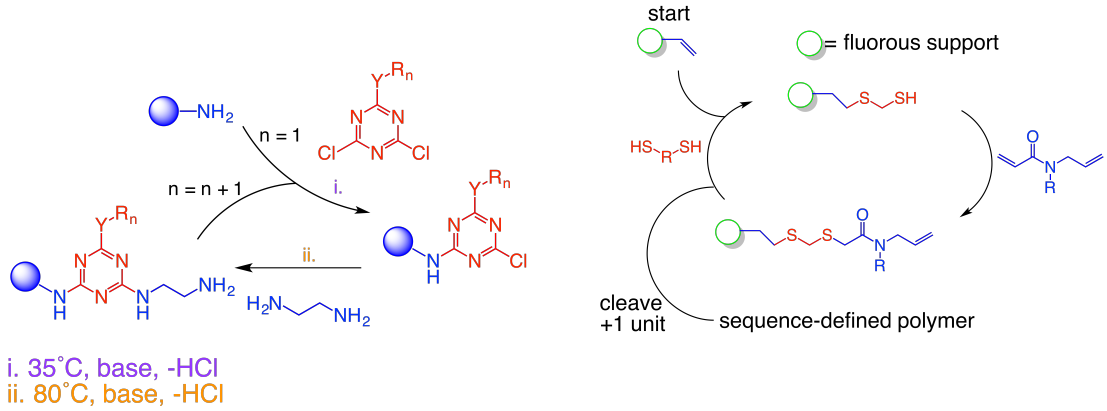


Figure 5. Selected iterative approaches to sequence control in multi-step-growth polymers based on solid (left, adapted with permission. Published by Wiley-VCH Verlag GmbH & Co. KGaA, copyright 2016 The Authors) and liquid (right, adapted with permission from ref. 21. Copyright 2014 ACS Publications) supports.

Phosphoramidite coupling,[16, 22] historically used in oligonucleotide synthesis, has also been applied in synthesis of sequence-controlled polymers (Figure 3, bottom panel), with the nature of the phosphoramidite bond allowing for easy sequence readout using mass spectrometry.

Outlook. Sequence control in iteratively synthesized polymers is an extremely dynamic field, with many more examples to be found in recent comprehensive reviews.[23] Many of the step-growth processes described above lead to polymers that cannot adequately be termed "sequence-controlled", but rather "sequence-defined". The subtle but important difference lies in the fact that the experimentalist defines the sequence by setting up the materials to take part in a multicomponent reaction, or, in other words, the sequence is defined by the functional groups and the characteristics of the reaction. On the other hand, in a true "sequence-controlled" process, the sequence emerges *e.g.* due to external factors acting in a sequential manner, or to monomers being added sequentially. Perhaps the closest to that aim are results presented by Du Prez, Madder and co-workers[24] whereby successive thiol-ene couplings were utilized to build up a polymer chain in a sequence determined by the order in which the monomers were added.

The sequential addition is at the core of multi-step-growth processes, with recent reports indicating that polymers up to 100 units long can be obtained by harnessing DNA synthesizers in a solid-phase protocol.[25] Further advances are undoubtedly to be made in liquid-only systems. For instance, the use of liquid handling robots that can produce well defined volumes of reactant could be used to control reactions at a liquid-liquid interfaces between droplets, see Figure 6.[26] A promising extension to the realm of sequence control could feature functionalized monomers dissolved in individual droplets, and chains being built up upon individual droplet collisions. This would necessarily have to take into account the timescale disparity between the collisions and the relevant reaction rates. As a result, the

coupling would have to be relatively fast – one promising candidate could be the triazolinedione chemistry.[27]

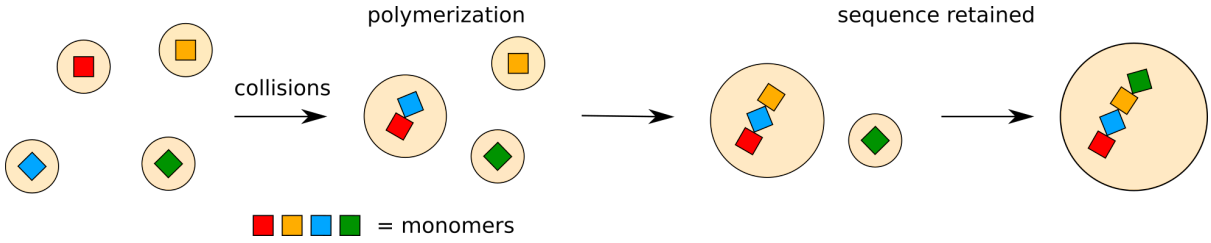


Figure 6. Collisions of monomer-containing droplets in a specific sequence leading to that sequence being retained in the resulting polymer.

3.2 Chain-growth polymers

In contrast to step-growth polymers, growth of the chain-growth polymers is relatively fast due to the presence of reactive intermediates such as carbocations, carbanions or free radicals. Ionic and radical variants both lend themselves to sequence control, typically by chemical stimuli.

Free radical polymerization. There has been an immense progress regarding molecular weight control in the free radical polymerization processes over the last 20 years and as a result most of these reactions can now be run in a controlled fashion. This is because of the emergence of "living" free radical polymerization approaches, such as atom transfer radical polymerization (ATRP), reversible activation-fragmentation-chain transfer polymerization (RAFT) and single electron transfer living radical polymerization (SET-LRP). Further control of sequence in these reactions can be imposed in multiple ways. As an example, sequence-controlled multi-block polyacrylates have been obtained by Haddleton and co-workers by simply adding the monomers, in a desired sequence, to a UV light-controlled SET-LRP medium using a copper complex as a catalyst.[28] On the other hand, using the RAFT process, Perrier and coworkers were able to demonstrate efficient sequence-controlled synthesis of multi-block copolymers, with each block up to 100 units long.[29]

Ionic polymerization. In ionic polymerization, carbocations or carbanions (as opposed to free radicals) are the active species. The chain length distribution of resulting polymers is much more tightly controlled, but the active species are much more sensitive to impurities such as oxygen. Nevertheless, these reactions are used in sequence-controlled polymerizations. In one example, Kanazawa and Aoshima[30] have demonstrated cationic terpolymerization of vinyl ethers, oxiranes and ketones, with strict selectivity of the units, resulting in repeating vinyl ether-oxirane-ketone sequences. Wurm and coworkers[31] have reported on the polymerization of a mixture of up to five different aziridine-based monomers with substituents characterized by varied electron-withdrawing strengths. The result was sequence control stemming from reactivity differences: the most reactive monomers were completely consumed in the amount of time in which the least reactive have only reached 20% conversion. In another recent example, diphenylethylene derivatives were copolymerized with either styrene or butadiene to provide perfectly alternating or telechelic copolymers.[32] This strategy was based on steric hindrance provided by diphenylethylene, which is unable to polymerize on its own.

Ring-opening polymerization. The characteristic feature of ring-opening polymerizations is the fact that monomers are cyclic and chain growth takes place through successive ring-opening and addition of the resulting segments to the active center. Lactones, lactames and cyclic carbonates are examples of typical monomers; sequence control is typically achieved by modifying the properties of polymerization catalysts in order to influence their affinity for a particular monomer class. Li, Guo and coworkers have implemented this approach by switching the catalyst between Brønsted acidic (optimized for cyclic lactones) and basic/conjugate acidic (efficient in L-lactide polymerization) by adding a Brønsted base.[33] Another approach involved encoding a sequence in a macrocycle, which was then polymerized in an entropy-driven fashion.[34]

Outlook. Other than the approaches described above, the sequence in polymer chains has been imposed by exploiting differences in reactivity between specific functional groups in

polymer chains, as demonstrated by Kamigaito in polymerization of maleimide and styrene (or limonene) units, whereby the maleimide units were further functionalized with sequence-defined side chains,[35] or by Sawamoto, whose group demonstrated individual addition of bulky methacrylate units followed by their transesterification with different alcohols to obtain sequence-defined polymers.[36]

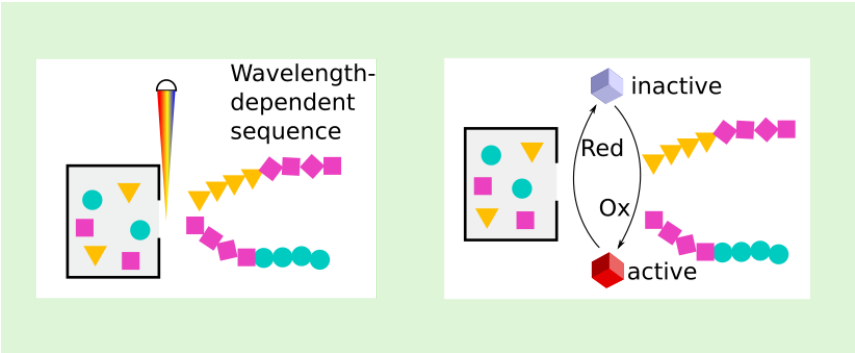


Figure 7. Different aspects of external environments as factors controlling sequence distribution: light (left) and oxidation-reduction potential (right).

External stimuli that could conceivably be used to control processes of this type are for example illumination and oxidation-reduction potential (Figure 7). RedOx-responsive catalysts for ring-opening polymerizations have been first described by Diaconescu and coworkers. Briefly, by changing the oxidation state of the catalytic metal center, not only can the catalyst be switched between active and inactive state, but it can also be made active towards different classes of monomers.[37] One can thus envision a mixture of monomers subjected to changing redox state of a catalyst that would give rise to chemoselective emergence of sequence-controlled polymers. This is indeed what was recently achieved by two groups: Williams and coworkers[38] were able to show that mixtures of monomers selected from four different classes can be selectively polymerized by switching the dizinc catalyst between different oxidation states (Figure 8), whereas Byers *et al.*[39] demonstrated a similar behavior for mixtures of lactides and epoxides in the case of an iron-based catalyst.

A major step towards illumination-controlled oligomer sequence emergence has been hinted at by the Boyer group, who have shown that illumination can impose sequence control on the level of individual monomer units.[40] A related study by the authors[41] has further shown that the light wavelength used is able to effectively select a species from a mixture of monomers. This suggests a possible generalization where the wavelength of light acts as an environmental factor controlling the sequence emerging from a mixture of monomers.

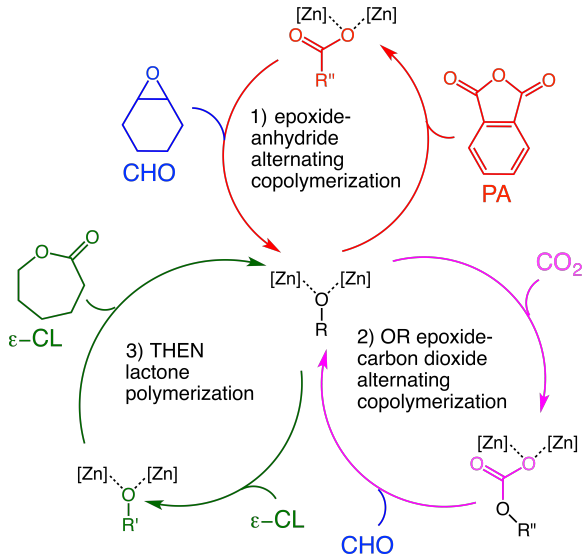


Figure 8. Example of chemoselective emergence of a sequence-controlled copolymer from a mixture of competing monomers determined by electron-withdrawing strengths of substituents in ring-opening polymerization. Adapted with permission from ref. 38. Copyright 2016 ACS Publications.

3.3. Multistep flow synthesis and iterative exponential growth (Flow-IEG)

Iterative exponential growth (Flow-IEG) combines multistep continuous flow chemistry and polymer synthesis for semi-automated synthesis of polymers.[42, 43] Jamison *et al* have chosen a copper catalyzed azide–alkyne cycloaddition reaction to polymerize an ester monomer functionalized with a triisopropylsilyl (TIPS) protected alkyne and an alkyl bromide as illustrated in Figure 9. This approach was successfully validated for achieving sequence controlled polymerization by targeting an alternating sequence $(ABAB)_n$ and a sequence with $(AABB)_n$ repeating unit. As a result, pure high molecular weight polymers were obtained.

The user-friendly nature, scalability, and modularity of Flow-IEG provides a general strategy for the automated synthesis of sequence and architecturally defined, uniform macromolecules.

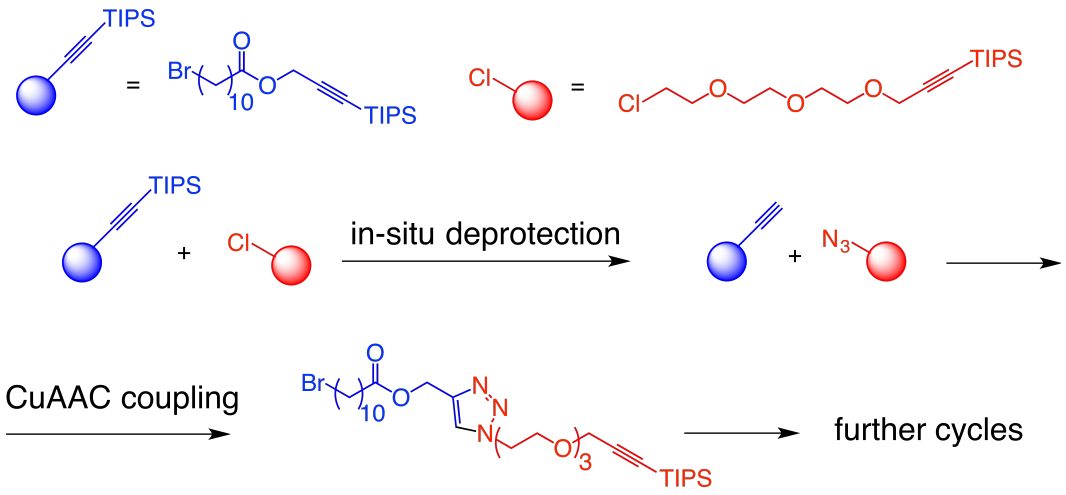


Figure 9. Example implementation of sequence-controlled polymerization based on Huisgen copper-azide-alkyne cycloaddition in a flow-IEG process. Adapted with permission from ref. 43. Copyright 2015 PNAS.

4. HYBRIDS

Polymer hybridization is of great interest especially when it comes to develop materials with new properties. However, achieving hybrid polymerization by combining natural and non-natural building blocks in a sequence defined manner is still a challenge. To achieve this, Sleiman and her team have recently coupled perfluorocarbons with nucleic acids and other non-natural polymers. This was accomplished by using an automated approach using phosphoramidite chemistry as illustrated in Figure 10.[44] As a result of introducing the perfluorocarbon chains, the thermal stability and nuclease resistance of DNA strands has

significantly improved (by up to 20 °C), which resulted in self-assembly of monodisperse micellar nanoparticles.

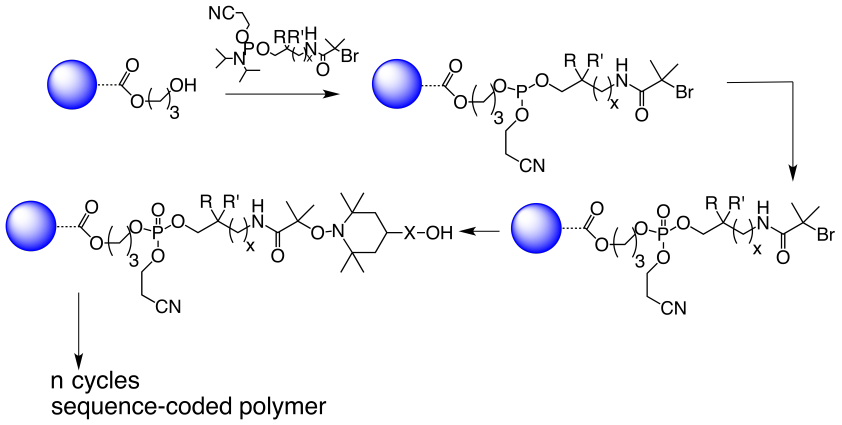


Figure 10. Sequence controlled oligomer synthesis with phosphoramidites grafted sequentially onto a growing chain attached to a solid support. Adapted with permission from ref. 44. Copyright 2016 Royal Society of Chemistry.

5. ANALYSIS, AUTOMATION AND MODELING

5.1. Analytical and sequencing methods

Natural, enzymatically controlled poly- and oligo-merization processes have been invented by evolution with numerous safety checks and correction steps which ensure that sequence fidelity is preserved. However, in the case of synthetic polymers, every novel sequence control protocol necessarily needs reliable analytical tools to confirm that the attempted sequential polymerization was indeed successful. Tandem mass spectrometry (MS/MS) has historically been the technique of choice for investigating sequences of easily-fragmented polymers such as poly(alkoxyamine amide)s or poly(triazole amide)s. In recent reports,[45, 46] Lutz, Charles and coworkers described sequence readout using MS/MS with electrospray ionization (ESI) as a means of retrieving binary information that was earlier encoded in the chains. These techniques are best suited to polymers containing easily cleavable ether or amide functions. In contrast, the main chains of common vinyl polymers are built up entirely of carbon atoms and are thus nowhere near as amenable to fragmentation. For such polymers,

information from several different techniques must be collected to obtain reliable unit sequence details. As a starting point, molecular weight, and thus the degree of polymerization, must be obtained. This is typically provided by gel permeation chromatography/size exclusion chromatography, but requires polymer purification and extensive calibration. *In situ* molecular weight measurements on complex monomer/polymer mixtures can be more conveniently conducted using diffusion-ordered NMR spectroscopy, or DOSY.[47]

Matrix-assisted laser-desorption-ionization (MALDI) is a technique that has long been used to analyze molecular weight distributions of synthetic polymers. These include copolymers produced from substrates with different reactivities and sequence analysis in alternating copolymers.[32, 48] This technique has limitations which preclude its usefulness for detailed readout of sequence in high-molecular-weight polymers, but for other cases, a combination of unit ratios obtained from NMR measurements with molecular weights measured by MALDI-MS is potentially the most powerful. If the respective monomer propagation rates, as well as addition and sequence times, are known, a descriptive model could give unambiguous averaged sequences for the produced polymers and therefore confirmation of the sequence-controlled nature of the reaction.

5.2. Automated synthesis of sequence-controlled polymers

All the methodologies described above could be easily extended by introducing a degree of automation. This would result in shifting the burden of overseeing the sequence of events leading to polymers from chemists to computer-controlled reaction setups, thus avoiding potential reproducibility issues and ensuring that variation between experiments is minimal.

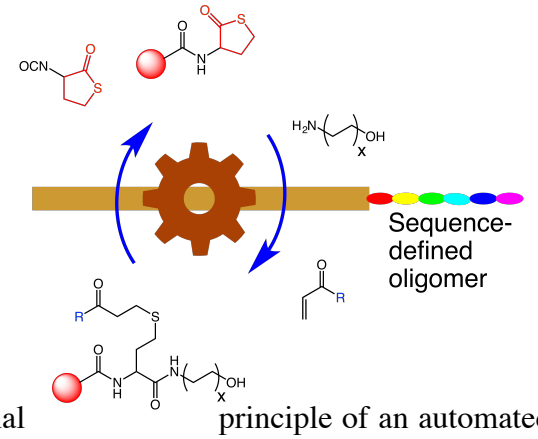


Figure 11. Operational principle of an automated system inspired by peptide synthesizers and used to produce sequence-defined oligomers. Adapted with permission from ref. 50. Copyright 2016 ACS Publications.

Reports concerning automation of polymer synthesis have been relatively scarce, but there exist several examples of automation, with both commercial and in-house setups. The former approach was taken by Matyjaszewski and coworkers, who used a commercially available DNA synthesizer to conduct photocontrolled ATRP by programming a specific sequence of monomer additions to be performed by the machine.[49] This allowed for producing well-defined homopolymers, block copolymers and DNA-polymer hybrids, but the nature of the equipment necessarily limited broader applications. A more flexible implementation, presented by Du Prez, Espeel and others,[50] involved an automated peptide synthesizer adapted to conduct sequential thiolactone ring openings and acrylate couplings (Figure 11), which led to strictly sequence-defined oligomers with diverse functional groups.

5.3. MODELING APPROACHES

Theoretical modelling is well-established in sequence studies of natural polymers such as proteins, primarily due to the existence of the PDB database, which can be used to train computational procedures such as neural networks.[51] The synthetic polymer space is, however, nowhere near as deeply explored and the analogous approach would be prohibitively expensive. As a result, alternative theoretical frameworks are being developed, sometimes with very specific optimization targets in mind.[52, 53] An important strand of

stochastic approaches to polymerization modelling relates to processes conducted in a continuous fashion, for example in flow reactors. Numerous additional parameters, such as monomer residence time and mixing rate come into play here, but models have nevertheless been developed to simulate these conditions.[54, 55]

6. PROGRAMMING SEQUENCE AND FUTURE CHALLENGES

Here we have attempted to give an overview of recent advances in polymer sequence control, along with outlook to future challenges and directions. These can be more broadly put in a context of two main strands: selection of sequence from a random mixtureand improving our understanding of the relationships between sequence and polymer properties. By analogy to natural polymers, synthetic macromolecules of the same chemical nature, but possessing different sequence, can be expected to exhibit different properties. Recent examples of this tendency include optoelectronic property differences between polyvinylenes[56] and divergent properties of random versus alternating polyesters, including hydrolytic susceptibility, fluoride ion affinity, ductility and Young's modulus.[57]

The solid-phase-based processes are being developed with an outlook to increase coupling yields and rates, since the sequential addition of units means that there is typically only one type of monomer present in the reaction medium at any given moment. An example recently reported by us[58] is based on rehydration/dehydration cycles, which allowed for efficient uncatalyzed formation of oligopeptides in unprecedented yields. Importantly, the system was fully controllable digitally: parameters such as cycle number and duration, monomer concentration, temperature and pH could be set and controlled allowing for the straightforward exploration of all the different environmental variables.

In the case of chain-growth polymers, we have discussed major advances made recently in the synthesis of multiblock polymers. We also recognize that there are additional ambitious prospects related to importance of bias in polymerization on the level of individual units. Here the understanding of how a molecular constructor might be designed from scratch to build complex self-replicating architectures might seem outlandish, but this is perhaps part of the key problem, and not just limited to chemistry, but also of relevance to computer science and technology.[59] For biology to emerge such a problem had to be solved without an explicit constructor.[60] Indeed, oligomers that can self-replicate must be able to emerge naturally (for biology even to exist),[61] and these will produce molecules and systems that are more complex than to be expected if the process forming them was random.[62] To that end, in earlier sections, we have given numerous examples of the desired sequence arising from a mixture of monomers being controlled by factors as diverse as steric hindrance, chemical reactivity, template presence or catalyst affinity. These are important because any complex mixture of many monomers has to be intrinsically biased to produce well-defined polymer sequences. Otherwise, when all monomers are equally likely to polymerize, the product will consist of perfectly random chains only. Therefore, the stochastic modeling techniques and analytical approaches summarized above can be utilized to better understand the influence of different environmental factors on complex monomer mixtures and the possibilities of "pushing" the mixtures in the direction of increasing function. An ultimate breakthrough, bringing the field closer to the molecular machinery of biology,[63] might come from incorporating the one-pot, one-instruction nature of chain-growth polymerization and sequential characteristics of solid-phase processes into one system (Figure 12).

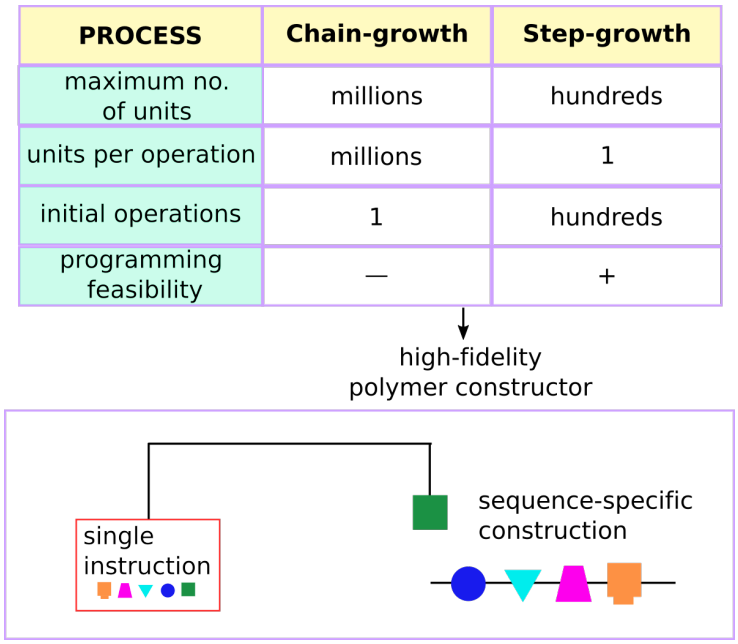


Figure 12. To truly mimic nature, high-fidelity synthesis of long polymer chains will need to be implemented in a manner that only requires a single instruction (*i.e.* what might be referred to as a "one-pot" setup). This approach stands in contrast to chain-growth and step-growth classes of processes, but incorporates elements of both.

In the former case, radical-based kinetics allows for rapid chain growth, but the incorporation of individual monomer units is harder to control. In the latter, the chain elongation kinetics is slower, but sequential addition is easier to implement. High-fidelity synthesis of long, sequence-controlled chains, determined by a unique set of initial conditions, would then pave the way for efficient exploration of the sequence space, with potential goals as diverse as (auto)catalytic activity, material properties, control of microstructure and self-assembly. It is our wish that this account will be used to define a new goal by which polymer science and molecular synthesis come together to aim for high precision assembly of millions of bonds in just one programmable operation. Even small steps toward such a feat, whilst currently far from reach, would show that the dream of molecular 'hard' nanotechnology might best be

solved using sequence controlled polymer systems. These, like those found in biology, would be 'engines of creation'.[64]

7. AUTHOR INFORMATION

Corresponding author:

*Email: Lee.Cronin@glasgow.ac.uk

ORCID:

Jan K. Szymański: orcid.org/0000-0002-7149-8315

Yousef M. Abul-Haija: orcid.org/0000-0002-0357-0653

Leroy Cronin: orcid.org/0000-0001-8035-5757

Author Contribution

†These authors contributed equally.

Notes

The authors declare no competing financial interest.

Biographical Information

Dr. Jan K. Szymański: Jan obtained a M.Sc. degree in chemistry at Adam Mickiewicz University in Poznan, Poland (2007), followed by a doctorate at the Institute of Physical Chemistry of the Polish Academy of Sciences, Warsaw (2012), working on complex reactions in small volumes and their applications for information processing. He then moved to Harvard University to study nonequilibrium formation of polymer self-asssemblies in an origins-of-life context as a member of the Harvard Origins of Life Initiative. Jan has been a member of the Cronin group since February 2016 and his work in the group involves designing complex reactions in oil droplets, with an outlook to sequence control in macromolecules and evolving the physical behavior of the droplets. His other research interests include controlling chemical reactions with external stimuli such as light and oxidation-reduction potential.

Dr. Yousef M. Abul-Haija: Yousef is a postdoctoral research associate in the Cronin research group at the University of Glasgow, currently working on exploring the emergence of peptide assemblies without biological constraints to investigate how alternative biologies might be created. He earned his PhD (2015) in soft supramolecular materials in Prof. Rein Ulijn research group at the University of Strathclyde which was followed by a one year postdoctoral position in the same group. He also worked in the industry (Hikma pharmaceuticals-Jordan) for two years. Yousef completed his BSc (2006) and MSc (2009) in Applied Chemistry at Jordan University of Science and Technology where he worked on developing polymer-based materials through means of copolymerization and crosslinking. He is interested in the design of supramolecular materials, complex chemical systems, peptide nanotechnology, and structural and functional control of chemical networks.

Prof. Leroy Cronin: Leroy (Lee) Cronin is the Regius Professor of Chemistry at the School of Chemistry, University of Glasgow. Lee was an undergraduate and DPhil student at the University of York, and a research fellow at the University of Edinburgh and the University of Bielefeld. From 2000 he was a lecturer at the University of Birmingham, before moving to the University of Glasgow in 2002. There, he was promoted to Professor (2006), Gardiner Professor (2009), and most recently to the Regius Chair (2013). He has received several awards, including the RSC Bob Hay Lectureship, the RSC Corday Morgan Medal and Prize, and the RSE/BP Hutton Prize in Energy Innovation. His research spans a range of fields under the umbrella of 'Complex Chemical Systems', focusing on understanding and controlling self-assembly and self-organisation in chemistry to develop functional molecular and nano-molecular chemical systems, along with linking architectural design with function, and engineering system-level functions.

8. ACKNOWLEDGEMENTS

We gratefully acknowledge financial support from the EPSRC (Grant Nos EP/P00153X/1, EP/J015156/1, EP/K021966/1, EP/K038885/1, EP/L015668/1, EP/L023652/1), BBSRC

(Grant No. BB/M011267/1), the EC (projects 610730 EVOPROG, 611640 EVOBLISS, 318671 MICREAGENTS), ERC (project 670467 SMART-POM).

9. REFERENCES

(1) Badi, N.; Lutz, J.-F. Sequence Control in Polymer Synthesis. *Chem. Soc. Rev.* **2009**, *38*, 3383–3390.

(2) Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-Controlled Polymers. *Science* **2013**, *341*, 1238149.

(3) Orgel, L. E. Molecular Replication. *Nature* **1992**, *358*, 203–209.

(4) De Bo, G.; Kuschel, S.; Leigh, D. A.; Lewandowski, B.; Papmeyer, M.; Ward, J. W. Efficient Assembly of Threaded Molecular Machines for Sequence-Specific Synthesis. *J. Am. Chem. Soc.* **2014**, *136*, 5811–5814.

(5) 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*, 189-193.

(6) Abul-Haija Y. M.; Ulijn R. V. Sequence Adaptive Peptide–Polysaccharide Nanostructures by Biocatalytic Self-Assembly. *Biomacromolecules* **2015**, *16*, 3473–3479.

(7) Pappas C. G.; Shafi R.; Sasselli I. R.; Siccardi H.; Wang T; Narang V.; Abzalimov R.; Wijerathne N.; Ulijn R. V. Dynamic Peptide Libraries for the Discovery of Supramolecular Nanomaterials. *Nat. Nanotechnol.* **2016**, *11*, 960-967.

(8) 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*, 282–292.

(9) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.

(10) Xi, W.; Scott, T. F.; Kloxin, C. J.; Bowman, C. N. Click Chemistry in Materials

Science. *Adv. Funct. Mater.* **2014**, *24*, 2572–2590.

(11) Han, J.; Zheng, Y.; Zhao, B.; Li, S.; Zhang, Y.; Gao, C. Sequentially Hetero-Functional, Topological Polymers by Step-Growth Thiol-Yne Approach. *Sci. Rep.* **2014**, *4*, 4387.

(12) Xue, H.; Zhao, Y.; Wu, H.; Wang, Z.; Yang, B.; Wei, Y.; Wang, Z.; Tao, L. Multicomponent Combinatorial Polymerization via the Biginelli Reaction. *J. Am. Chem. Soc.* **2016**, 138, 8690–8693.

(13) Solleder, S. C.; Meier, M. A. R. Sequence Control in Polymer Chemistry through the Passerini Three-Component Reaction. *Angew. Chem. Int. Ed.* **2014**, *53*, 711–714.

(14) Zhang, Z.; You, Y.-Z.; Wu, D.-C.; Hong, C.-Y. Syntheses of Sequence-Controlled Polymers via Consecutive Multicomponent Reactions. *Macromolecules* **2015**, *48*, 3414–3421.

(15) Lutz, J.-F.; Lehn, J.-M.; Meijer, E. W.; Matyjaszewski K. From Precision Polymers to Complex Materials and Systems. *Nat. Rev. Mat.* **2016**, *1*, 16024.

(16) Al Ouahabi, A.; Charles, L.; Lutz, J. F. Synthesis of Non-Natural Sequence-Encoded Polymers Using Phosphoramidite Chemistry. *J. Am. Chem. Soc.* **2015**, *137*, 5629–5635.

(17) Grate, J. W.; Mo, K. F.; Daily, M. D. Triazine-Based Sequence-Defined Polymers with Side-Chain Diversity and Backbone-Backbone Interaction Motifs. *Angew. Chem. Int. Ed.* **2016**, *55*, 3925–3930.

(18) Zuckermann, R. N.; Kerr, J. M.; Kent, S. B. H.; Moos, W. H. Efficient Method for the Preparation of Peptoids [Oligo(N-Substituted Glycines)] by Submonomer Solid-Phase Synthesis. *J. Am. Chem. Soc.* **1992**, *114*, 10646-10647.

(19) Knight, A. S.; Zhou, E. Y.; Francis, M. B.; Zuckermann, R. N. Sequence Programmable Peptoid Polymers for Diverse Materials Applications. *Adv. Mater.* **2015**, *27*, 5665–5691.

(20) Pfeifer, S.; Zarafshani, Z.; Badi, N.; Lutz, J.-F. Liquid-Phase Synthesis of Block

Copolymers Containing Sequence-Ordered Segments. *J. Am. Chem. Soc.* **2009**, *131*, 9195–9197.

(21) Porel, M.; Alabi, C. A. Sequence-Defined Polymers via Orthogonal Allyl Acrylamide Building Blocks. *J. Am. Chem. Soc.* **2014**, *136*, 13162–13165.

(22) 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, 9417–9420.

(23) Solleder S. C.; Schneider, R. V.; Wetzel, K. S.; Boukis, A. C.; Meier, M. A. R. Recent Progress in the Design of Monodisperse, Sequence-Defined Macromolecules. *Macromol. Rapid Commun.* **2017**, *38*, 1600711.

(24) Espeel, P.; Carrette, L. L. G.; Bury, K.; Capenberghs, S.; Martins, J. C.; Du Prez, F. E.; Madder, A. Multifunctionalized Sequence-Defined Oligomers from a Single Building Block. *Angew. Chem. Int. Ed.* **2013**, *52*, 13261–13264.

(25) Al Ouahabi, A.; Kotera, M.; Charles, L.; Lutz, J.-F. Synthesis of Monodisperse Sequence-Coded Polymers with Chain Lengths above DP100. *ACS Macro Lett.* **2015**, *4*, 1077-1080.

(26) Gutierrez, J. M. P.; Hinkley, T.; Taylor, J. W.; Yanev, K.; Cronin, L. Evolution of Oil Droplets in a Chemorobotic Platform. *Nat. Commun.* **2014**, *5*, 5571.

(27) Billiet, S.; De Bruycker, K.; Driessen, F.; Goossens, H.; Van Speybroeck, V.; Winne, J. M.; Du Prez, F. E. Triazolinediones Enable Ultrafast and Reversible Click Chemistry for the Design of Dynamic Polymer Systems. *Nat. Chem.* **2014**, *6*, 815–821.

(28) Anastasaki, A.; Nikolaou, V.; Pappas, G. S.; Zhang, Q.; Wan, C.; Wilson, P.; Davis, T. P.; Whittaker, M. R.; Haddleton, D. M. Photoinduced Sequence-Control via One Pot Living Radical Polymerization of Acrylates. *Chem. Sci.* **2014**, *5*, 3536–3542.

(29) Cody, G.; Maschmeyer, T.; Zetterlund, P. B.; Perrier, S. Rapid and Quantitative One-Pot Synthesis of Sequence-Controlled Polymers by Radical Polymerization. *Nat.*

Commun. **2013**, *4*, 2505.

(30) Kanazawa, A.; Aoshima, S. Exclusive One-Way Cycle Sequence Control in Cationic Terpolymerization of General-Purpose Monomers via Concurrent Vinyl-Addition, Ring-Opening, and Carbonyl-Addition Mechanisms. *ACS Macro Lett.* **2015**, *4*, 783–787.

(31) Rieger, E.; Alkan, A.; Manhart, A.; Wagner, M.; Wurm, F. R. Sequence-Controlled Polymers via Simultaneous Living Anionic Copolymerization of Competing Monomers. *Macromol. Rapid Commun.* **2016**, *37*, 833–839.

(32) Hutchings, L. R.; Brooks, P. P.; Parker, D.; Mosely, J. A.; Sevinc, S. Monomer Sequence Control via Living Anionic Copolymerization: Synthesis of Alternating, Statistical, and Telechelic Copolymers and Sequence Analysis by MALDI ToF Mass Spectrometry. *Macromolecules* **2015**, *48*, 610–628.

(33) Wang, X.; Liu, J.; Xu, S.; Xu, J.; Pan, X.; Liu, J.; Cui, S.; Li, Z.; Guo, K. Traceless Switch Organocatalysis Enables Multiblock Ring-Opening Copolymerizations of Lactones, Carbonates, and Lactides: By a One plus One Approach in One Pot. *Polym. Chem.* **2016**, *7*, 6297–6308.

(34) Weiss, R. M.; Short, A. L.; Meyer, T. Y. Sequence-Controlled Copolymers Prepared via Entropy-Driven Ring-Opening Metathesis Polymerization. *ACS Macro Lett.* **2015**, *4*, 1039–1043.

(35) Soejima, T.; Satoh, K.; Kamigaito, M. Sequence-Regulated Vinyl Copolymers with Acid and Base Monomer Units via Atom Transfer Radical Addition and Alternating Radical Copolymerization. *Polym. Chem.* **2016**, *7*, 4833–4841.

(36) Oh, D.; Ouchi, M.; Nakanishi, T.; Ono, H.; Sawamoto, M. Iterative Radical Addition with a Special Monomer Carrying Bulky and Convertible Pendant: A New Concept toward Controlling the Sequence for Vinyl Polymers. *ACS Macro Lett.* **2016**, *5*, 745–749.

(37) Wang, X.; Thevenon, A.; Brosmer, J. L.; Yu, I.; Khan, S. I.; Mehrkhodavandi, P.; Diaconescu, P. L. Redox Control of Group 4 Metal Ring-Opening Polymerization

Activity toward L-Lactide and ε-Caprolactone. *J. Am. Chem. Soc.* **2014**, *136*, 11264–11267.

(38) Romain, C.; Zhu, Y.; Dingwall, P.; Paul, S.; Rzepa, H. S.; Buchard, A.; Williams, C. K. Chemoselective Polymerizations from Mixtures of Epoxide, Lactone, Anhydride, and Carbon Dioxide. *J. Am. Chem. Soc.* **2016**, *138*, 4120–4131.

(39) Biernesser, A. B.; Delle Chiaie, K. R.; Curley, J. B.; Byers, J. A. Block Copolymerization of Lactide and an Epoxide Facilitated by a Redox Switchable Iron-Based Catalyst. *Angew. Chem. Int. Ed.* **2016**, *55*, 5251–5254.

(40) Xu, J.; Fu, C.; Shanmugam, S.; Hawker, C. J.; Moad, G.; Boyer, C. Synthesis of Discrete Oligomers by Sequential PET-RAFT Single-Unit Monomer Insertion. *Angew. Chem. Int. Ed.* **2017**, *129*, 8496-8503.

(41) Xu, J.; Shanmugam, S.; Fu, C.; Aguey-Zinsou, K. F.; Boyer, C. Selective Photoactivation: From a Single Unit Monomer Insertion Reaction to Controlled Polymer Architectures. *J. Am. Chem. Soc.* **2016**, *138*, 3094–3106.

(42) Barnes, J. C.; Ehrlich, D. J. C.; Gao, A. X.; Leibfarth, F. A.; Jiang, Y.; Zhou, E.; Jamison, T. F.; Johnson, J. A. Iterative Exponential Growth of Stereo- and Sequence-Controlled Polymers. *Nat. Chem.* **2015**, *7*, 810–815.

(43) Leibfarth, F. A.; Johnson, J. A.; Jamison, T. F. Scalable Synthesis of Sequence-Defined, Unimolecular Macromolecules by Flow-IEG. *Proc. Natl. Acad. Sci.* **2015**, *112*, 10617–10622.

(44) de Rochambeau, D.; Barlog, M.; Edwardson, T. G. W.; Fakhoury, J. J.; Stein, R. S.; Bazzi, H. S.; Sleiman, H. F. "DNA-Teflon" Sequence-Controlled Polymers. *Polym. Chem.* **2016**, *7*, 4998–5003.

(45) Amalian, J.-A.; Trinh, T. T.; Lutz, J.-F.; Charles, L. MS/MS Digital Readout: Analysis of Binary Information Encoded in the Monomer Sequences of Poly(triazole Amide)s. *Anal. Chem.* **2016**, *88*, 3715–3722.

(46) Al Ouahabi, A.; Amalian, J.-A.; Charles, L.; Lutz, J.-F. Mass Spectrometry Sequencing of Long Digital Polymers Facilitated by Programmed Inter-Byte

Fragmentation. *Nat. Commun.* **2017**, *8*, 967.

(47) Li, W.; Chung, H.; Daeffler, C.; Johnson, J. A.; Grubbs, R. H. Application of ^{1}H DOSY for Facile Measurement of Polymer Molecular Weights. *Macromolecules* **2012**, *45*, 9595–9603.

(48) Huijser, S.; Mooiweer, G. D.; Van Der Hofstad, R.; Staal, B. B. P.; Feenstra, J.; Van Herk, A. M.; Koning, C. E.; Duchateau, R. Reactivity Ratios of Comonomers from a Single MALDI-ToF-MS Measurement at One Feed Composition. *Macromolecules* **2012**, *45*, 4500–4510.

(49) Pan, X.; Lathwal, S.; Mack, S.; Yan, J.; Das, S. R.; Matyjaszewski, K. Automated Synthesis of Well-Defined Polymers and Biohybrids by Atom Transfer Radical Polymerization Using a DNA Synthesizer. *Angew. Chem. Int. Ed.* **2017**, *56*, 2740–743.

(50) Martens, S.; Van den Begin, J.; Madder, A.; Du Prez, F. E.; Espeel, P. Automated Synthesis of Monodisperse Oligomers, Featuring Sequence Control and Tailored Functionalization. *J. Am. Chem. Soc.* **2016**, *138*, 14182–14185.

(51) Qian, N.; Sejnowski, T. J. Predicting the Secondary Structure of Globular Proteins Using Neural Network Models. *J. Mol. Biol.* **1988**, *202*, 865–884.

(52) Meenakshisundaram, V.; Hung, J.-H.; Patra, T. K.; Simmons, D. S. Designing Sequence-Specific Copolymer Compatibilizers Using a Molecular-Dynamics-Simulation-Based Genetic Algorithm. *Macromolecules* **2017**, *50*, 1155–1166.

(53) Fierens, S. K.; Telitel, S.; Van Steenberge, P. H. M.; Reyniers, M.-F.; Marin, G. B.; Lutz, J.-F.; D'hooge, D. R. Model-Based Design to Push the Boundaries of Sequence Control. *Macromolecules* **2016**, *49*, 9336–9344.

(54) Nassar, R.; Too, J. R.; Fan, L. T. Stochastic Modeling of Polymerization in a Continuous Flow Reactor. *J. Appl. Polym. Sci.* **1981**, *26*, 3745-3759.

(55) Lemos, T.; Melo, P. A.; Pinto, J. C. Stochastic Modeling of Polymer Microstructure from Residence Time Distribution. *Macromol. React. Eng.* **2015**, *9*, 259-270.

(56) Zhang, S.; Hutchison, G. R.; Meyer, T. Y. Sequence Effects in Conjugated Donor-Acceptor Trimers and Polymers. *Macromol. Rapid Commun.* **2016**, *37*, 882–887.

(57) Peng, C.; Joy, A. Alternating and Random-Sequence Polyesters with Distinct Physical Properties. *Polym. Chem.* **2017**, *8*, 2397–2404.

(58) Rodriguez-Garcia, M.; Surman, A. J.; Cooper, G. J. T.; Suárez-Marina, I.; Hosni, Z.; Lee, M. P.; Cronin, L. Formation of Oligopeptides in High Yield under Simple Programmable Conditions. *Nat. Commun*. **2015**, *6*, 8385.

(59) Von Neumann J. Theory of Self-Reproducing Automata. University of Illinois Press, Illinois, **1966**.

(60) Kauffman, S. A.; Smith, R. G. Adaptive Automata Based on Darwinian Selection. *Physica D*. **1986**, *22*, 68-82.

(61) Patzke, V.; von Kiedrowski, G. Self-Replicating Systems, *ARKIVOC*. **2007**, 5, 293–310.

(62) Marshall, S. M.; Murray, A. R. G.; Cronin, L. A Probabilistic Framework for Identifying Biosignatures using Pathway Complexity. *Phil. Trans. R. Soc. A*. **2017,** *375*, 20160342.

(63) Greig, L. M.; Philp, D. Applying Biological Principles to the Assembly and Selection of Synthetic Superstructures. *Chem. Soc. Rev.* **2001**, *30*, 287–302.

(64) Drexler, K. E. Engines of Creation, Fourth Estate, London, **1990.**