Preparation of Information-Containing Macromolecules by Ligation of Dyad-Encoded Oligomers

Article *in* Chemistry - A European Journal · July 2015

DOI: 10.1002/chem.201502414

CITATIONS
21

READS
66

5 authors, including:

Tam Thanh Trinh
University of Strasbourg
7 PUBLICATIONS **93** CITATIONS
SEE PROFILE

Delphine Chan-Seng
French National Centre for Scientific Research
30 PUBLICATIONS **619** CITATIONS
SEE PROFILE

Laurence Charles
Aix-Marseille Université
174 PUBLICATIONS **1,605** CITATIONS
SEE PROFILE

Jean-François Lutz
French National Centre for Scientific Research
200 PUBLICATIONS **11,323** CITATIONS
SEE PROFILE

ChemPubSoc Europe

DOI: 10.1002/chem.201502414

CHEMISTRY A European Journal Communication

Convergent Synthesis

Preparation of Information-Containing Macromolecules by Ligation of Dyad-Encoded Oligomers

Thanh Tam Trinh,[a] Laurence Oswald,[a] Delphine Chan-Seng,[a] Laurence Charles,[b] and Jean-François Lutz*[a]

Abstract: A simplified strategy for preparing non-natural information-containing polymers is reported. The concept relies on the successive ligation of oligomers that contain minimal sequence motifs. It was applied here to the synthesis of digitally-encoded poly(triazole amide)s, in which propyl and 2-methyl propyl motifs are used to code 0 and 1, respectively. A library of four oligo(triazole amide)s containing the information dyads 00, 01, 10, and 11 was prepared. These oligomers contain two reactive functions, that is, an alkyne and a carboxylic acid. Thus, they can be linked to another with the help of a reactive spacer containing azide and amine functions. Using two successive chemoselective steps, that is, azide-alkyne Huisgen cycloaddition and carboxylic acid-amine coupling, monodisperse polymers can be obtained. In particular, the library of dyads permits the implementation of any desired sequence using a small number of steps. As a proof-of-concept, the synthesis of molecular bytes 00000000 and 00000110 is described.

The development of information-containing macromolecules is an exciting new trend in synthetic polymer science.[1] Similarly to DNA, such polymers contain a message that is encoded in their chains by a controlled comonomer sequence.[2–5] The monomer-coded message can be deciphered using sequencing technologies, for example, analytical techniques developed for genomics and proteomics.[2,5–8] This property was unexplored until very recently and opens up a brand new scope of applications for synthetic polymers, for example, in the fields of data storage, molecular informatics, and molecular identification.[9] In order to design such macromolecules, a readable comonomer code should be encrypted in the chains.[2] Although many different types of codes could be potentially written in polymers, the simplest possibility is certainly the design of a monomer-based binary code. Such a code can be easily implemented in a polymer chain using two comonomers defined intentionally as 0-bit and 1-bit. It is, however, very important to note that the macromolecular string of information should be defect-free. That implies that classical polymerization processes leading to polydisperse polymers, such as chain-growth or step-growth polymerizations, are not suitable for preparing information-containing macromolecules. On the other hand, such polymers can be prepared using multistep-growth approaches, that is, iterative chemistry on a solid or soluble support.[10,11] However, such methods are usually time-consuming and limited to the synthesis of short sequences.[12,13] In order to simplify these syntheses, protecting-group-free iterative strategies have been reported in recent years.[14–20] For instance, our group has studied so-called "AB + CD" approaches, in which two successive chemoselective reactions are used to attach two building blocks containing four different reactive functions A, B, C, and D.[16] For example, copper-catalyzed Huisgen cycloaddition and amidification steps can be used to link a building block, AB, containing carboxylic acid (A) and terminal alkyne (B) functions and a building block, CD, containing primary amine (C) and azide (D) groups. It was found that this approach allows the synthesis of monodisperse sequence-defined polymers.[16] Moreover, this strategy can be used to synthesize digitally-encoded macromolecules.[21] To do so, three monomers have to be used: one CD spacer and two interchangeable AB building blocks defined as 0-bit and 1-bit. It was demonstrated that any desired binary sequences can be written in a polymer using this simple comonomer language. However, despite its easiness, this iterative approach remains time-consuming and might be difficult to apply for the synthesis of long information sequences.

Here, a macromolecular coupling strategy is proposed to simplify the synthesis of long sequence-encoded monodisperse polymers. The ligation of oligomers is a strategy that is extensively used in biopolymer synthesis for achieving long sequence-defined polymers. For example, native chemical ligation, introduced by Kent and co-workers, is a useful chemistry for protein total synthesis.[22] In such approaches, monodisperse polymers are attained using oligomers with protected end-groups that are linked to each other through convergent strategies. Oligomer ligation has also been successfully utilized for the synthesis of long oligonucleotide sequences.[23] Comparable methodologies can be used to prepare monodisperse non-natural macromolecules. For instance, Hawker and co-

[a] Dr. T. T. Trinh, L. Oswald, Dr. D. Chan-Seng, Dr. J.-F. Lutz
Precision Macromolecular Chemistry
Institut Charles Sadron, UPR-22 CNRS
23 rue du Loess, 67034 Strasbourg (France)
E-mail: jflutz@unistra.fr

[b] Prof. Dr. L. Charles
Aix-Marseille Université- CNRS
Institute of Radical Chemistry, UMR 7273
Avenue Escadrille Normandie-Niemen
13397 Marseille (France)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201502414.

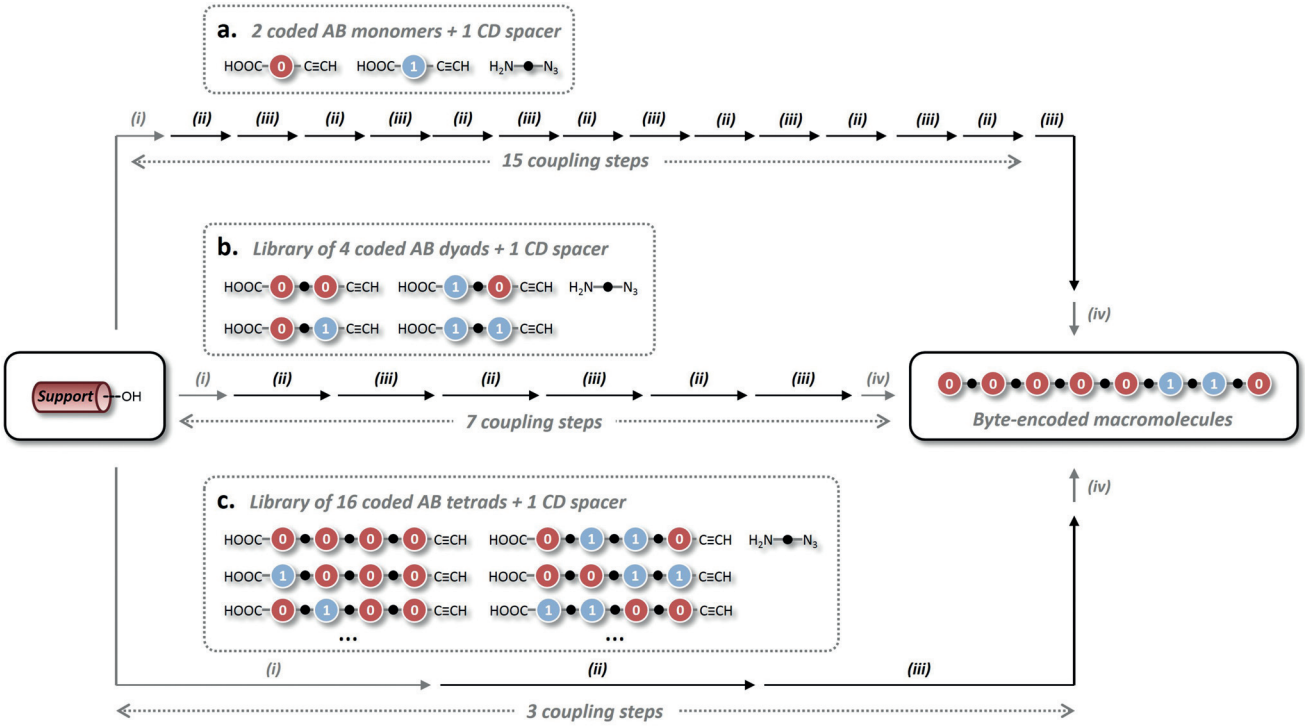

Figure 1. Possible "AB + CD" iterative strategies that could be used for the synthesis of a sequence-encoded macromolecule containing a byte of information. a) Stepwise strategy relying on three monomers, that is, two interchangeable coded AB monomers and one CD spacer. b) Stepwise ligation strategy using a library of four coded AB dyads and a CD spacer. c) Stepwise ligation strategy using a library of sixteen coded AB tetrads and a CD spacer.

workers have described an elegant convergent approach for synthesizing monodisperse aliphatic polyesters.[24] In the present work, a stepwise ligation approach, based on "AB + CD" chemistry, was studied for the synthesis of non-natural information-containing macromolecules. In particular, the aim of this study was to prepare a model macromolecule containing a byte of information (i.e., eight coded molecular bits). Figure 1 compares three possible iterative routes that could be used to prepare such a macromolecular byte. As described above, a first option is a stepwise approach employing two coded AB monomers, defined as 0 and 1, and a CD spacer (Figure 1 a). However, this approach requires 15 successive coupling steps to reach the targeted structure.

The number of steps can be reduced by using libraries of AB oligomers containing preformed information dyads or tetrads. For instance, using a library of AB dyads (i.e., four oligomers) and a CD spacer, only seven steps are necessary to synthesize a macromolecular byte (Figure 1 b). As shown in Figure 1 c, only three steps would be needed using a library of tetrads. However, this last strategy was not selected in the present work for two main reasons. First of all, it requires a broad library of building blocks (i.e., 16 oligomers). In addition, it is well-known that the yields and kinetics of macromolecular ligation may depend on the molecular weight of the used building blocks.[25] In this context, the approach shown in Figure 1 b appears as an interesting compromise for the synthesis of information-containing macromolecules. This approach was tested herein.

The investigation started with the synthesis of a library of four dyad-encoded oligomers obtained from two AB coding building blocks, that is, 5-hexynoic acid (**0** in Figure 2 a) and 3-methyl-5-hexynoic acid (**1** in Figure 2 a), and 11-azido-3,6,9-trioxaundecan-1-amine as CD spacing unit (● in Figure 2 a). It should be remarked that the AB coding building blocks differs from those used in our previous works, which were pentynoic-[21] or heptynoic-acid[16] derivatives. The reasons for that choice are related to the ease of monomer synthesis and to the efficacy of carboxylic acid-amine coupling. The oligomers were prepared on a Wang polystyrene-based soluble support ($\mathbf{P_w}$ in Figure 2 a) whose synthesis was previously described by our group.[26] Typically, a four-step procedure was used to synthesize the dyad-encoded oligomers, as shown in Figure 2 b. The hydroxy terminal group of $\mathbf{P_w}$ was first reacted with the carboxylic acid group of an AB building block (either **0** or **1**) using *N*,*N*-dicyclohexylcarbodiimide (DCC) as coupling agent in the presence of 4-(dimethylamino)pyridine (DMAP; step (i) in Figure 2 b). The resulting alkyne-functional support was afterwards reacted with a CD spacing unit by copper-assisted alkyne-azide cycloaddition (CuAAC) using a combination of CuBr and 4,4′-di-n-nonyl-2,2′-bipyridine (dNbipy) as catalyst (step (ii) in Figure 2 b). Lastly, a second AB building block that can be identical or different from the first one was linked to the support by carboxylic acid-amine coupling. The oligomers were then cleaved from the soluble support using a standard TFA/DCM protocol (step (iv) in Figure 2 b). The four possible dyads, that is, **00**, **01**, **10**, and **11**, were prepared following this

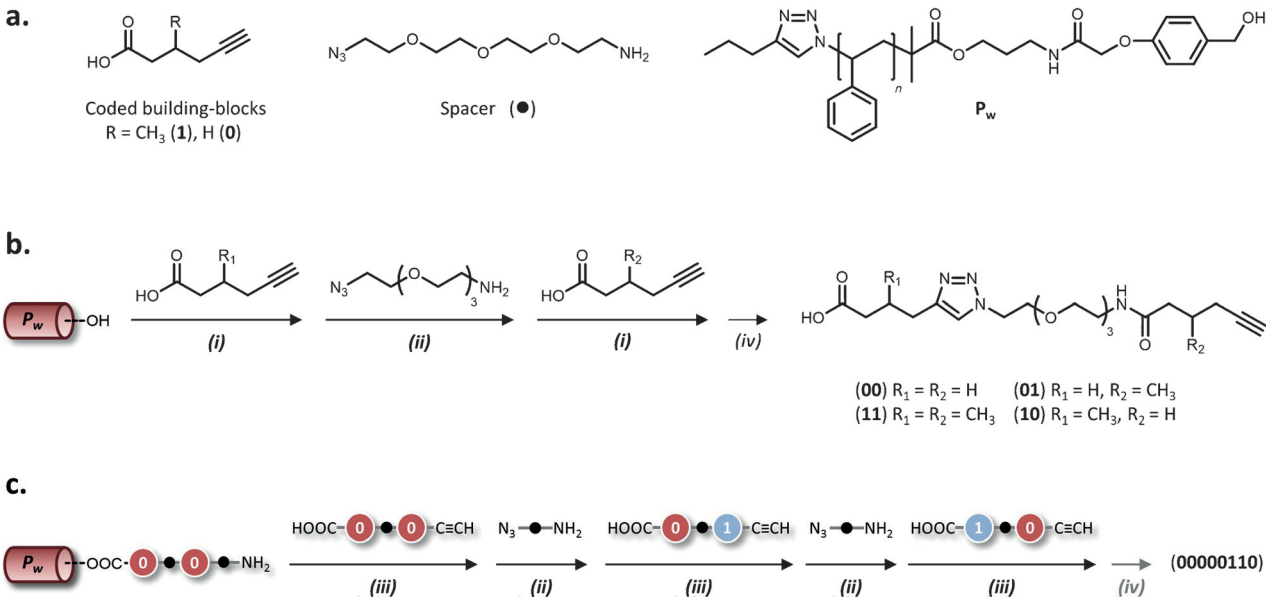

Figure 2. a) Molecular structures of the building blocks and of the soluble polymer support used in this work. b) General route followed for the synthesis of dyad-encoded oligomers. c) Specific strategy used in this work for the synthesis of a byte-encoded polymer. The displayed example corresponds to the coded polymer **00000110**. Experimental conditions: (i) DCC, DMAP, DCM, RT, (ii) CuBr, dNbipy, THF, RT, (iii) PyBOP, DIPEA, NMP, 60 °C, (iv) TFA, DCM, RT. Full details, including concentrations used, can be found in the Supporting Information.

procedure and characterized by ^{1}H and ^{13}C NMR as well as mass spectrometry to confirm their chemical integrity. The different dyad sequences can be easily distinguished by ^{1}H NMR (Figures S1–S4 in the Supporting Information) through the chemical shift and presence/absence of the methylene in the β position to the carbonyl moiety (**0**) and of the methyl protons (**1**). Practically speaking, the methylene signal is observed at 1.80 and 1.64 ppm when **0** is inserted as first and last building block respectively, whereas the methyl of **1** is identified at 0.9 ppm. Thus, the presence of characteristic peaks at 1.80 and 1.64 ppm accounts for dyad **00**, at 1.80 and 0.9 ppm for dyad **01**, at 1.64 and 0.9 ppm for dyad **10**, and solely at 0.90 ppm for dyad **11**. Furthermore, mass spectrometry measurements supported the successful synthesis of all the dyad-encoded oligomers (Figures S5–S8 in the Supporting Information).

The sequential coupling reactions of the dyad-encoded oligomers were also performed on support $\mathbf{P_w}$. As mentioned in the introduction, any desired sequence can be potentially created using the dyad library. As a proof-of-principle, two model sequences, that is, **00000000** and **00000110**, were prepared and characterized. Both start with the dyad **00**. Because the dyads were also synthesized on $\mathbf{P_w}$, as described above, it is unnecessary to cleave them from the support and to couple them back. In other words, the macromolecular bytes **00000000** and **00000110** were not prepared in seven iterative steps, as shown in Figure 1b, but using only six steps (Figure 2c). Strictly-speaking, they are obtained in ten steps, that is, three to construct the first dyad **00** on $\mathbf{P_w}$ (the first three steps in Figure 2b), one to add a spacer ●, and six additional steps to finalize the sequence. However, it should be remarked that $\mathbf{P_w}$-**00**● can be prepared in large scale and used in multiple syntheses. As detailed in Figure 2c, the macromolecular bytes were prepared using an "AB + CD" strategy employing successive carboxylic acid-amine (step (iii) in Figure 2c) and CuAAC (step (ii) in Figure 2c) coupling steps. It is noteworthy to mention that the four dyad-encoded oligomers all exhibit AB-type molecular structure. Thus, their successive attachment on a support requires intermediate coupling steps with a CD spacer, namely 11-azido-3,6,9-trioxaundecan-1-amine. The synthesis of the macromolecular bytes was monitored by MALDI-TOF-MS. After each dyad ligation, a small portion of the formed polymer was cleaved from $\mathbf{P_w}$ and analyzed by mass spectrometry. The same reaction conditions as those employed for the synthesis of the dyads, that is, CuBr/dNbipy for CuAAC and DCC/DMAP for the amidification reaction, were first tested. However, dyad-ligations on the support were found to be incomplete when DCC/DMAP was used. Various experimental conditions were screened in order to increase the yields. It was found that the use of benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBOP) in the presence of *N*,*N*-diisopropylethylamine (DIPEA) for the carboxylic acid-amine coupling steps (step (iii) in Figure 2c) led overall to better results. However, when reaching the third coupling reaction of the dyad to the support, incompleteness of the reaction was observed. Nevertheless, this problem was easily solved by repeating the amidification step. Using these optimized conditions, the macromolecular bytes were successfully synthesized. Figure 3 and Figure S11 in the Supporting Information show the MALDI-TOF-MS monitoring of the synthesis of **00000000** and **00000110**, respectively. In both cases, the desired coded monodisperse polymer was obtained. The effectiveness of the oligomer stepwise strategy was also confirmed by ESI-HR-MS analysis of the formed polymers after each dyad

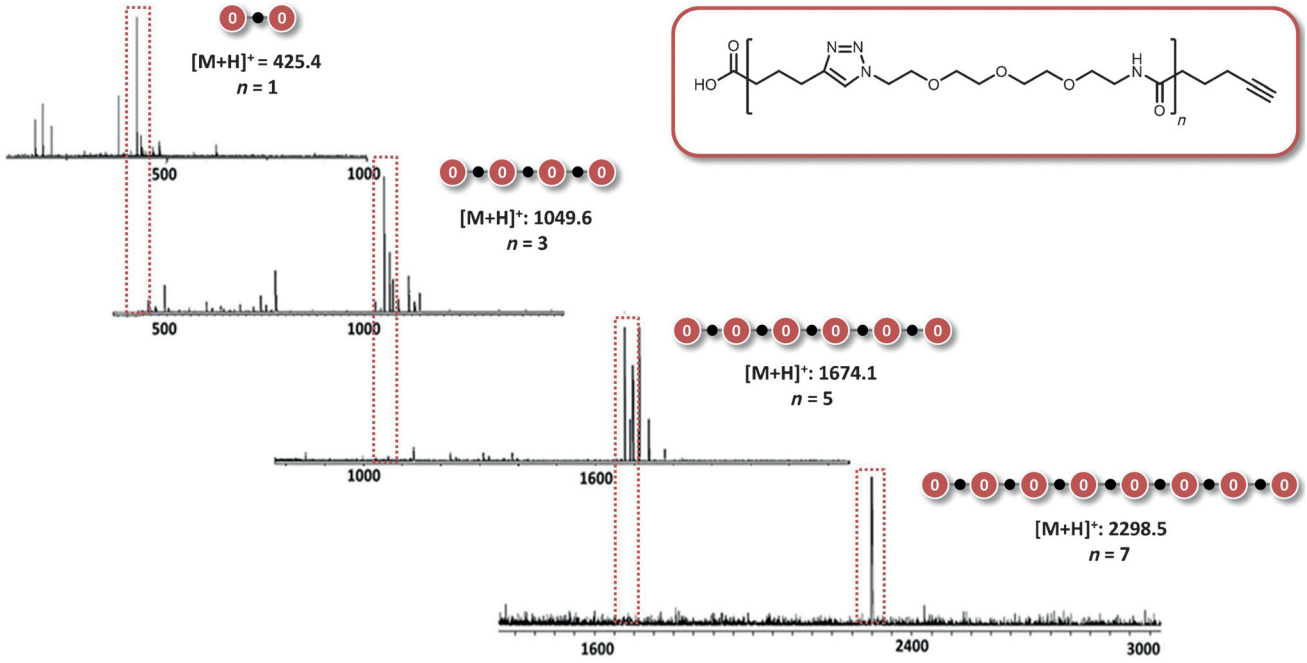

Figure 3. MALDI-TOF-MS monitoring of the stepwise synthesis of 00000000. MALDI-TOF-MS mass spectra obtained after each dyad ligation. All samples shown in this Figure were cleaved from the support $\mathbf{P_w}$ prior to MALDI-TOF-MS analysis. The different peaks observed for $n=3$ and $n=5$ species are interpreted in detail in Figures S9 and S10 in the Supporting Information.

Table 1. ESI-HR-MS monitoring of the stepwise synthesis of 00000000 and 00000110. Comparison of theoretical and experimental *m/z* values for polymers isolated at different stages of the synthesis. All polymers were cleaved from $\mathbf{P_w}$ prior to ESI-HR-MS analysis

	m/z_{th}	m/z_{exp}	Error [mDa]	Error [ppm]
00●	643.3774 $[M+H]^+$	643.3783	+0.9	+1.4
0000	525.3031 $[M+2H]^{2+}$	525.2998	−3.3	−6.3
000000	558.6577 $[M+3H]^{3+}$	558.6574	−0.3	−0.6
00000000	575.3350 $[M+4H]^{4+}$	575.3346	−0.4	−0.7
000001	563.3296 $[M+3H]^{3+}$	563.3306	+1.0	+1.7
00000110	582.3428 $[M+4H]^{4+}$	582.3322	−0.6	−1.0

ligation (Table 1). In all cases, the expected species were identified with a very high accuracy.

These results indicate that the use of a dyad-encoded oligomer library is an interesting option for preparing sequence-coded macromolecules. Indeed, using such a strategy, complex sequences can be synthesized using a small number of coupling steps as depicted in Figure 1 b. However, the experimental work and the synthesis steps that are used for creating the initial oligomer library should, of course, also be taken into account to assess the usefulness of the concept. Altogether, the strategy is undoubtedly advantageous when compared to a classical monomer-by-monomer iterative approach (Figure 1 a). Indeed, the dyad-encoded oligomers can be synthesized in large scales and used as essential building blocks in numerous syntheses.

For instance, the two model polymers **00000000** and **00000110** prepared in this work share some common synthons. Thus, the process possesses the classical advantages of a convergent synthesis.[24,27] Another important advantage of this approach is indeed the fact that any desired sequence can be built. For instance, in the case of a macromolecular byte having different α and ω termini, $2^8=256$ sequence combinations can be potentially synthesized using five building blocks, that is, four AB dyad-oligomers and one CD spacer. It should be remarked that this number of possibility is based on monomer sequence only and does not take into account the stereoconfiguration of the **1** units in the chains. The fact that many different sequences can potentially be prepared make these polymers appealing for applications in the field of data storage and product identification.[9] Indeed, such oligomers could be used, for example, as molecular barcodes. It is nevertheless important to clarify that such applications do not require that all combinations are synthesized. The crucial fact is that any chosen sequence can be synthesized.

Altogether, the utilization of convergent concepts in chemoselective "AB+CD" iterative synthesis seems very useful for preparing information-containing macromolecules. However, there are still some drawbacks. For instance, the strategy shown in this paper relies on polymer-oligomer end-group reactions in solution. Such chemistry can be influenced by the molecular weight of the polymeric reactants. This is mostly true in the case of diffusion-controlled reactions but it was also suggested that it may play a role in activation-controlled reactions. Morawetz and co-workers have extensively studied the reactivity of polymeric reagents[28] and have proposed that end-group reactivity is affected by a kinetic exclusion-volume effect,[25] which is due to the shielding behavior of polymer random coils. Although conflicting data exist on this topic,[29–31]

it is a fact that the coupling of reactive polymer chain-ends in solution is not trivial. This aspect is particularly crucial in iterative syntheses, in which quantitative yields are required at each step. It explains, for instance, the relatively long reaction times used in the present work. However, this aspect could potentially be optimized by varying the experimental conditions. For instance, we studied here the carboxylic acid-amine coupling reaction for performing macromolecular ligation. However, the "AB + CD" concepts relies on two chemical successive reactions and therefore CuAAC could also be considered for macromolecular ligation.[32] In this case, microwave-assisted chemistry could also be used to reduce coupling times.[33] It is also important to state that the type of support may play an important role. Soluble polystyrene supports were selected in this work because it was previously found that they are much more advantageous than resins for performing macromolecular ligation in solution.[34] However, low molecular weight supports such as fluorous tags may be even more practical.[19,35] Nevertheless, even though the concept can be certainly optimized, the first proof-of-principle shown in this article clearly indicates that the use of coded oligomer libraries is a very promising option for preparing monodisperse sequence-controlled polymers. These results further highlight the significance of convergent strategies in synthetic polymer chemistry.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013)/ERC grant agreement n°258593. The PhD position of T.T.T. was paid by the ERC. The CNRS and the excellence network Chemistry of Complex Systems (LabEx CSC) are also acknowledged for financial support. The authors thank Jean-Marc Strub and Hélène Nierengarten for the MALDI-TOF measurements as well as Catherine Foussat and Mélanie Legros for the SEC measurements.

Keywords: convergent synthesis · information-containing macromolecules · iterative chemistry · sequence-controlled polymers · soluble supports

[1] J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science* **2013**, *341*, 1238149.
[2] H. Colquhoun, J.-F. Lutz, *Nat. Chem.* **2014**, *6*, 455 – 456.
[3] D. Andrieux, P. Gaspard, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9516 – 9521.
[4] A. Al Ouahabi, L. Charles, J.-F. Lutz, *J. Am. Chem. Soc.* **2015**, *137*, 5629 – 5635.
[5] R. K. Roy, A. Meszynska, C. Laure, L. Charles, C. Verchin, J.-F. Lutz, *Nat. Commun.* **2015**, *6*, 7237.
[6] Z. Zhu, C. J. Cardin, Y. Gan, H. M. Colquhoun, *Nat. Chem.* **2010**, *2*, 653 – 660.
[7] H. Mutlu, J.-F. Lutz, *Angew. Chem. Int. Ed.* **2014**, *53*, 13010 – 13019; *Angew. Chem.* **2014**, *126*, 13224 – 13233.
[8] L. Charles, C. Laure, J.-F. Lutz, R. K. Roy, *Macromolecules in press,* DOI: 10.1021/acs.macromol.5b01051.
[9] J.-F. Lutz, unpublished results.
[10] R. B. Merrifield, *Angew. Chem. Int. Ed. Engl.* **1985**, *24*, 799 – 810; *Angew. Chem.* **1985**, *97*, 801 – 812.
[11] D. J. Gravert, K. D. Janda, *Chem. Rev.* **1997**, *97*, 489 – 510.
[12] N. Badi, J.-F. Lutz, *Chem. Soc. Rev.* **2009**, *38*, 3383 – 3390.
[13] J.-F. Lutz, *Polym. Chem.* **2010**, *1*, 55 – 62.
[14] T. T. Trinh, C. Laure, J.-F. Lutz, *Macromol. Chem. Phys* **2015**, in press, DOI: 10.1002/macp.201500072.
[15] R. N. Zuckermann, J. M. Kerr, S. B. H. Kent, W. H. Moos, *J. Am. Chem. Soc.* **1992**, *114*, 10646 – 10647.
[16] S. Pfeifer, Z. Zarafshani, N. Badi, J.-F. Lutz, *J. Am. Chem. Soc.* **2009**, *131*, 9195 – 9197.
[17] P. Espeel, L. L. G. Carrette, K. Bury, S. Capenberghs, J. C. Martins, F. E. Du Prez, A. Madder, *Angew. Chem. Int. Ed.* **2013**, *52*, 13261 – 13264; *Angew. Chem.* **2013**, *125*, 13503 – 13506.
[18] S. C. Solleder, M. A. R. Meier, *Angew. Chem. Int. Ed.* **2014**, *53*, 711 – 714; *Angew. Chem.* **2014**, *126*, 729 – 732.
[19] M. Porel, C. A. Alabi, *J. Am. Chem. Soc.* **2014**, *136*, 13162 – 13165.
[20] S. C. Solleder, K. S. Wetzel, M. A. R. Meier, *Polym. Chem.* **2015**, *6*, 3201 – 3204.
[21] T. T. Trinh, L. Oswald, D. Chan-Seng, J.-F. Lutz, *Macromol. Rapid Commun.* **2014**, *35*, 141 – 145.
[22] S. B. H. Kent, *Chem. Soc. Rev.* **2009**, *38*, 338 – 351.
[23] A. Wolter, J. Biernat, H. Köster, *Nucleosides Nucleotides* **1986**, *5*, 65 – 77.
[24] K. Takizawa, C. Tang, C. J. Hawker, *J. Am. Chem. Soc.* **2008**, *130*, 1718 – 1726.
[25] H. Morawetz, J.-R. Cho, P. J. Gans, *Macromolecules* **1973**, *6*, 624 – 627.
[26] L. Oswald, T. T. Trinh, D. Chan-Seng, J.-F. Lutz, *Polymer* **2015**, in press, DOI: 10.1016/j.polymer.2015.02.057.
[27] S. M. Grayson, J. M. J. Fréchet, *Chem. Rev.* **2001**, *101*, 3819 – 3868.
[28] H. Morawetz, *J. Polym. Sci. Polym. Sym.* **1985**, *72*, 9 – 15.
[29] P. E. Black, D. J. Worsfold, *J. Polym. Sci. Polym. Chem.* **1981**, *19*, 1841 – 1846.
[30] A. Okamoto, Y. Shimanuki, I. Mita, *Eur. Polym. J.* **1982**, *18*, 545 – 548.
[31] H. K. Jeon, C. W. Macosko, B. Moon, T. R. Hoye, Z. Yin, *Macromolecules* **2004**, *37*, 2563 – 2571.
[32] J. A. Opsteen, J. C. M. van Hest, *Chem. Commun.* **2005**, 57 – 59.
[33] P. Appukkuttan, W. Dehaen, V. V. Fokin, E. Van der Eycken, *Org. Lett.* **2004**, *6*, 4223 – 4225.
[34] C. Stutz, A. Meszynska, J.-F. Lutz, H. G. Börner, *ACS Macro Lett.* **2013**, *2*, 641 – 644.
[35] W. Zhang, D. P. Curran, *Tetrahedron* **2006**, *62*, 11837 – 11865.

Received: June 22, 2015
Published online on July 16, 2015