ACS Macro Letters

Letter

pubs.acs.org/macroletters

Synthesis of Monodisperse Sequence-Coded Polymers with Chain Lengths above DP100

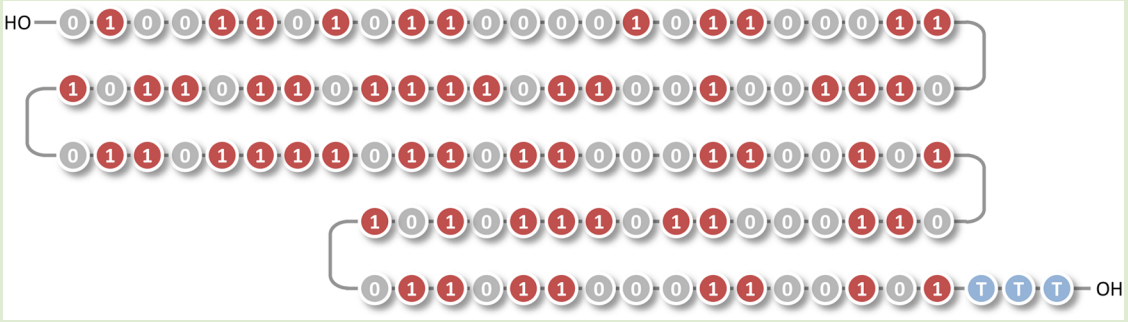
Abdelaziz Al Ouahabi,[†] Mitsuharu Kotera,[‡] Laurence Charles,[§] and Jean-François Lutz*[,†]

[†]Precision Macromolecular Chemistry Group, Institut Charles Sadron, UPR22-CNRS, 23 rue du Loess, 67034 Strasbourg, France

[‡]Laboratoire de Conception et Application de Molécules Bioactives, V-SAT Group, UMR 7199 CNRS, Université de Strasbourg, Faculté de Pharmacie, 74 route du Rhin, 67401 Illkirch Cedex, France

[§]Aix-Marseille Université − CNRS, UMR 7273, Institute of Radical Chemistry, 13397 Marseille Cedex 20, France

Supporting Information



ABSTRACT: Non-natural, sequence-encoded polyphosphates were prepared using the phosphoramidite approach on a DNA synthesizer. Two phosphoramidite monomers, namely, 2-cyanoethyl (3-dimethoxytrityloxy-propyl) diisopropylphosphoramidite (**0**) and 2-cyanoethyl (3-dimethoxytrityloxy-2,2-dimethyl-propyl) diisopropylphosphoramidite (**1**), were used in this approach to form binary-coded sequences. Using 1000 Å controlled pore glass as a support and a large excess of monomers at each step, it was possible to synthesize homopolymers and sequence-coded copolymers of high chain-length. For instance, monodisperse polymers containing 16, 24, 56, and 104 coded monomer units were synthesized and characterized in this work. These results indicate that highly efficient phosphoramidite steps are suitable for the synthesis of long non-natural information-containing macromolecules.

It was recently shown that information can be stored in polymers using monomer-based codes.[1,2] The first proofs-of-concept were obtained using coded nucleotide sequences in artificial DNA strands.[3] For instance, Church and co-worker and Goldman and co-workers have lately reported that kilobytes of data can be stored in synthetic DNA.[4,5] Interestingly, these works do not rely on high molecular weight polymers, but on libraries of macromolecules with chain lengths in the range of 100 to 150 monomer units. Very recently, our group has proposed that monodisperse synthetic polymers could be an interesting alternative to DNA for storing information.[2] For instance, it was shown that binary-coded messages can be stored in oligo(triazole amide)s,[6−8] oligo(alkoxyamine amide)s,[9] and polyphosphates.[10] In order to implement digital information in a polymer chain, two monomers have to be defined as 0 and 1 bits.[11] Coded sequences are then created by attaching these monomers one-by-one on a solid support. After synthesis, the information encrypted in the chains can be deciphered using sequencing technologies that have been developed for biopolymer analysis, such as tandem mass spectrometry.[9,12,13] Such non-natural information-containing macromolecules could be interesting for data storage, but also for other applications, such as anticounterfeit technologies.[2]

Still, the coded polymers that have been synthesized to date are relatively short. In general, the frontier between an oligomer and a polymer is considered to be 1000 Da, but also sometimes DP20, depending on the sources.[14] In previous works, sequence-coded polymers with molar mass in the range 2000−3500 Da have been synthesized.[8−10] For instance, polyphosphates containing up to 24 coded monomer units have been prepared using manual iterative phosphoramidite protocols.[10] These results are encouraging, but it would be interesting to prepare longer sequences. Here, it is important to mention that the upper limit of iterative synthesis has not yet been reached. The polymers reported so far were intentionally kept short because they have been synthesized manually. However, it is well-known from peptide and oligonucleotide chemistry that iterative synthesis can be greatly facilitated by the use of solid-phase synthesizers.[15,16] For instance, poly-

Received: August 26, 2015
Accepted: September 8, 2015
Published: September 10, 2015

DOI: 10.1021/acsmacrolett.5b00606
ACS Macro Lett. 2015, 4, 1077−1080

nucleotides containing as many as 150 monomer units can be synthesized using automated protocols.[17,18] Such robotic approaches are, of course, also relevant for the synthesis of non-natural sequence-defined polymers. In particular, commercially available DNA synthesizers can be used to prepare non-natural polyphosphates via the phosphoramidite approach. For example, Behr and co-workers have reported the automated synthesis of oligonucleotide−oligospermine biohybrids.[19] More recently, Sleiman and co-workers have used an automated synthesizer to link amphiphilic non-natural polyphosphates to DNA.[20] Here, we describe the synthesis of long binary-coded polyphosphates using a DNA synthesizer. In particular, we report for the first time the preparation of sequence-defined non-natural polymers with chains lengths containing more than 100 monomer units.

The phosphoramidite approach was introduced some decades ago by Beaucage and Caruthers for the solid-phase synthesis of oligonucleotides.[21] A three-step cycle is typically used in this strategy, as shown in Figure 1.[22] A monomer

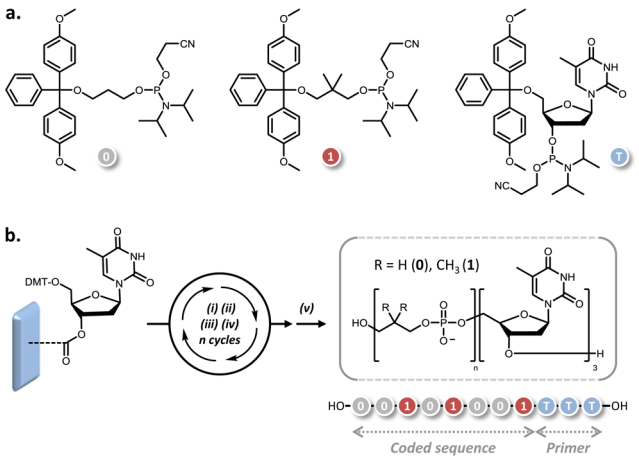


Figure 1. (a) Molecular structure of the phosphoramidite monomers used in the present work. (b) Iterative synthesis on a controlled pore glass support (blue surface). The sequences are shown with the 3′-end of the TTT sequence on the right-side following DNA conventions. Experimental conditions: (*i*) DMT deprotection: CCl_3-COOH, CH_2Cl_2; (*ii*) coupling step: RT, ACN, tetrazole; (*iii*) oxidation: RT, I_2, H_2O/pyridine/THF; (*iv*) capping step: RT, acetic anhydride, pyridine; (*v*) cleavage: RT, NH_3, CH_3NH_2.

containing a phosphoramidite group and a dimethoxytrityl (DMT)-protected alcohol is first reacted with a resin-bound OH group to form a phosphite that is afterward oxidized in a phosphate. Then, the DMT group of the bound monomer is removed to allow next coupling. As mentioned above, non-natural polymers can be constructed using this approach.[19,20,23] For instance, as shown in our previous work,[10] sequence-coded polymers can be created using the phosphoramidite approach. A digital code can be implemented using non-nucleoside phosphoramidite monomers, such as 2-cyanoethyl (3-dimethoxytrityloxy-propyl) diisopropylphosphoramidite (**0**) and 2-cyanoethyl (3-dimethoxytrityloxy-2,2-dimethyl-propyl) diisopropylphosphoramidite (**1**; Figure 1a).[10] Moreover, the resulting coded sequences can be analyzed by tandem mass spectrometry.[24] In the present work, monomers **0** and **1** were selected to build the sequence-coded polymers. As mentioned above, the synthesis of long sequences was performed on an automated DNA synthesizer.[16] As a consequence, the synthesis conditions were modified compared to previous work. First, the synthesis was performed using controlled pore glass (CPG) as a support instead of cross-linked polystyrene beads.[25] It is well-known that the pore size of porous glass is an important parameter for the synthesis of long polymer chains. For instance, pore size of 1000 Å or higher are needed for the synthesis of chains containing about 100 residues. In addition, a large excess of monomers (i.e., 10 molar equiv instead of 1.5 equiv in previous work) was used in each step in order to obtain high coupling yields. Capping steps using acetic anhydride were also used in this work. Although not strictly mandatory for the synthesis of short sequences,[10] capping steps cannot be avoided for the preparation of long chains. Besides experimental conditions, the design of the polymer chains was also slightly modified as compared to previous work. Instead of starting the sequence directly with a coding unit **0** or **1** as previously reported,[10] a primer sequence containing three thymine nucleotides was first synthesized (Figure 1b).[26] This TTT sequence facilitates the high performance liquid chromatography (HPLC) characterization of the coded polymers, but also permits to quantify the formed polymer by UV spectroscopy.

Table 1 shows the monomer sequences and the theoretical molar mass of the different polymers prepared in this work. In

Table 1. Polymers Prepared in the Present Work

	Monomer sequence [a]	Mass (Da)[b]	
		theo.	exp.[c]
H1	$(0)_{16}(T)_3$	3058.3	3058.3
H2	$(0)_{24}(T)_3$	4163.4	4163.2
H3	$(0)_{56}(T)_3$	8581.7	8581.3
H4	$(0)_{104}(T)_3$	15208.1	15207.7
C1	01010000·01001101·01000011$(T)_3$	4499.8	4499.6
C2	01001101·01100001·01110101·01110010·01101001·01100011·01100101$(T)_3$	9366.5	9366.5
C3	01001101·01100001·01100011·01110010·01101111·01101101·01101111·01101100·01100101·01100011·01110101·01101100·01100101$(T)_3$	16806.8	16806.2

[a]The sequences are shown with the 3′-end of the TTT sequence on the right side, following DNA conventions. The dots between each byte of the coded copolymer sequences are just displayed for clarity and do not represent a molecular spacer. [b]Values at isotopic maximum. [c]Data from charge state deconvoluted mass spectra.

order to assess the efficiency of the DNA synthesizer for preparing long-chain non-natural polyphosphates, a series of homopolymers of **0** (entries **H1**−**H4** in Table 1) of different chain lengths was first synthesized and characterized.

The automated syntheses were performed in trityl-on (DMT-ON) mode. This means that the terminal DMT was kept in the structure, and all other DMT groups deprotected during the iterative steps were collected in different fractions and analyzed by UV spectroscopy. The coupling efficiency was estimated by measuring the absorbance of the first two and the last fractions (Table S1), as previously reported.[27] In all cases, high average coupling efficiencies were obtained. Moreover, the

DOI: 10.1021/acsmacrolett.5b00606
ACS Macro Lett. 2015, 4, 1077−1080

automated DNA synthesizer allowed synthesis of these homopolymers in relatively short times. Polymers **H1** and **H2** were synthesized in less than 3 h, whereas longer samples **H3** and **H4** were synthesized in about 6 and 12 h, respectively. After synthesis, the homopolymers were cleaved from the CPG support using a solution of ammonia and methylamine and were subsequently purified on a reverse-phase column. This purification step was not performed in our previous works but is important when long monodisperse chains are targeted. Indeed, it permits to separate the targeted final sequence from the capped chains obtained earlier in the iterative process (Figure S1). Afterward, the formed polymers were characterized by ion exchange HPLC, UV spectroscopy (Table S2), and electrospray ionization mass spectrometry (ESI-MS). Figure 2 shows the HPLC traces recorded for the different

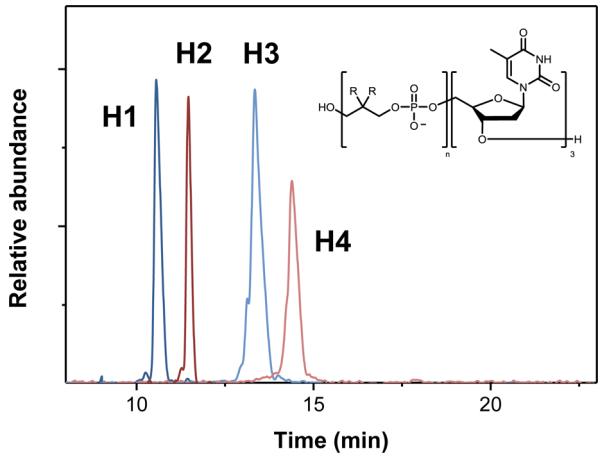


Figure 2. Ion-exchange HPLC traces recorded for the different homopolymers **H1**−**H4**. The analysis was performed with 10% ACN and 20% of 2 M NH_3 in water using a gradient of 2.5 M NaCl.

homopolymers. In each case, a single peak was observed, thus, indicating a well-controlled synthetic process. This was confirmed by ESI-MS measurements. Table S3 shows mass data obtained after negative-mode ESI-MS characterization of homopolymers **H1**−**H4**. While small polyphosphate polymers (ca. below 5000 Da) could also be observed in the positive ion mode,[10] negative mode electrospray was required for mass detection of larger macromolecules, as typically performed for DNA.[28] In all cases, main signals could be safely assigned to multideprotonated species corresponding to the expected macromolecules (Table S3). Ammonium salt was supplemented to the electrosprayed sample in order to prevent proton/cation exchange to occur in solution.[29] As a result, only minor signals corresponding to polyphosphates with only one cation counterion ($[M - (z + 1)H + C]^{z-}$, with C = Na, K) were detected (Figure S2). While ESI allows large macromolecules to be detected with low m/z range mass analyzer thanks to multiple charging, interpretation of such complex mass spectra can be tedious unless analyzed using a deconvolution technique,[30] which combines peaks corresponding to the same molecule detected under different charge state. Deconvolution of all ESI mass spectra permitted to confirm univocally the existence of monodisperse homopolymers (Table 1 and Figure S3).

Sequence-coded copolymers were then prepared on the DNA synthesizer (entries **C1**−**C3** in Table 1). These copolymers contain a message encrypted by an extended ASCII code, in which a byte, that is, 8 binary bits, codes for a specific character. For example, 3-bytes copolymer **C1** codes the initials of our research group "PMC", that is, 01010000 (P) 01001101 (M) 01000011 (C). The longer polymers **C2** and **C3** contain 7 and 13 bytes, respectively, and code longer words, namely, "Maurice" and "Macromolecule". The coded sequences were synthesized following DNA conventions. Therefore, the extended ASCII message is written inversely as compared to the sense of synthesis, that is, the message starts at the ω-hydroxy end group and is terminated by the TTT sequence (Table 1). Figure 3 shows the HPLC traces recorded for the

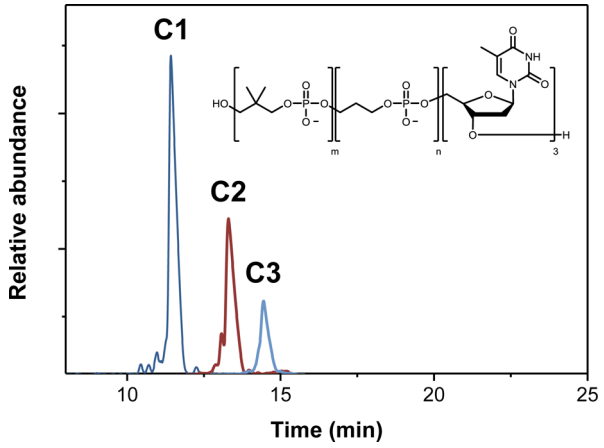


Figure 3. Ion-exchange HPLC traces recorded for the sequence-coded copolymers **C1**−**C3**. The analysis was performed with 10% ACN and 20% of 2 M NH_3 in water using a gradient of 2.5 M NaCl.

three different copolymers **C1**−**C3**. Defined peaks were observed for all expected species. In the case of **C1**, a slight multimodality was observed. This could reflect some polydispersity but could also be an artifact due to the type of ion-exchange columns used in this work. Monodispersity was confirmed by mass spectrometry analysis. Negative mode ESI-MS indicated in all cases the formation of monodisperse species with the expected molar mass (Tables 1 and S3 and Figure S4). As an example, Figure 4 shows the ESI-MS spectrum and the corresponding deconvolution observed for the longest copolymer **C3**. MS/MS sequencing was also performed for these polymers. Preliminary data indicate that correct coded sequences were formed in the studied samples. These data will be reported in detail in a forthcoming publication.[24]

In summary, sequence-coded polyphosphates were synthesized using automated phosphoramidite chemistry. The use of a DNA synthesizer in conjunction with high monomer excess and capping steps allowed preparation of long homopolymer and copolymer chains in a relatively short time. In particular, sequence-defined polymers containing more than 100 coding monomers were obtained in this work. This constitutes an important milestone in the field of non-natural information-containing macromolecules since chain lengths around DP100 are useful for information-related technologies.[11] Indeed, since about a decabyte of data can now be stored in a single chain, it seems plausible to write hectobytes or kilobytes of information in synthetic polymers using addressable chain libraries. Altogether, these results indicate that the development of synthetic sequence-coded polymers is a fast-moving area of research.

DOI: 10.1021/acsmacrolett.5b00606
ACS Macro Lett. 2015, 4, 1077−1080

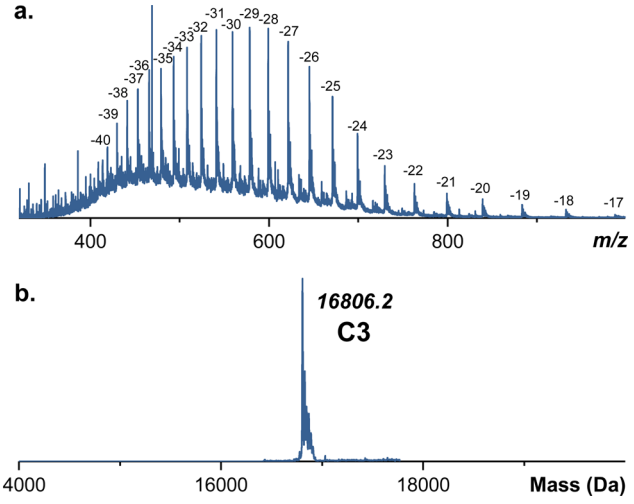


Figure 4. Mass spectrometry analysis of the sequence-coded copolymer **C3**: (a) Negative-mode ESI-MS spectrum, with peaks annotated with the corresponding ion charge state; (b) Deconvolution of the spectrum.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsmacrolett.5b00606.

Experimental procedures and additional data, tables and figures (PDF).

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jflutz@unistra.fr.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the IDEX program of the University of Strasbourg and the CNRS. The authors thank Jean-Marc Strub for some preliminary mass spectrometry measurements.

■ REFERENCES

(1) Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. *Science* **2013**, *341*, 1238149.
(2) Lutz, J.-F. *Macromolecules* **2015**, *48*, 4759−4767.
(3) Bancroft, C.; Bowler, T.; Bloom, B.; Clelland, C. T. *Science* **2001**, *293*, 1763−1765.
(4) Church, G. M.; Gao, Y.; Kosuri, S. *Science* **2012**, *337*, 1628−1628.
(5) Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; LeProust, E. M.; Sipos, B.; Birney, E. *Nature* **2013**, *494*, 77−80.
(6) Pfeifer, S.; Zarafshani, Z.; Badi, N.; Lutz, J.-F. *J. Am. Chem. Soc.* **2009**, *131*, 9195−9196.
(7) Trinh, T. T.; Oswald, L.; Chan-Seng, D.; Lutz, J.-F. *Macromol. Rapid Commun.* **2014**, *35*, 141−145.
(8) Trinh, T. T.; Oswald, L.; Chan-Seng, D.; Charles, L.; Lutz, J.-F. *Chem. - Eur. J.* **2015**, *21*, 11961−11965.
(9) Roy, R. K.; Meszynska, A.; Laure, C.; Charles, L.; Verchin, C.; Lutz, J.-F. *Nat. Commun.* **2015**, *6*, 7237.
(10) Al Ouahabi, A.; Charles, L.; Lutz, J.-F. *J. Am. Chem. Soc.* **2015**, *137*, 5629−5635.
(11) Colquhoun, H. M.; Lutz, J.-F. *Nat. Chem.* **2014**, *6*, 455−456.
(12) Mutlu, H.; Lutz, J.-F. *Angew. Chem., Int. Ed.* **2014**, *53*, 13010−13019.
(13) Charles, L.; Laure, C.; Lutz, J.-F.; Roy, R. K. *Macromolecules* **2015**, *48*, 4319−4328.
(14) Uglea, C. V.; Negulescu, I. I. *Synthesis and Characterization of Oligomers*; CRC Press: Boca Raton, FL, 1991.
(15) Merrifield, R. B. *Angew. Chem., Int. Ed. Engl.* **1985**, *24*, 799−810.
(16) Caruthers, M. H. *Science* **1985**, *230*, 281−285.
(17) Wolter, A.; Biernat, J.; Köster, H. *Nucleosides Nucleotides* **1986**, *5*, 65−77.
(18) LeProust, E. M.; Peck, B. J.; Spirin, K.; McCuen, H. B.; Moore, B.; Namsaraev, E.; Caruthers, M. H. *Nucleic Acids Res.* **2010**, *38*, 2522−2540.
(19) Noir, R.; Kotera, M.; Pons, B.; Remy, J.-S.; Behr, J.-P. *J. Am. Chem. Soc.* **2008**, *130*, 13500−13505.
(20) Edwardson, T. G. W.; Carneiro, K. M. M.; Serpell, C. J.; Sleiman, H. F. *Angew. Chem., Int. Ed.* **2014**, *53*, 4567−4571.
(21) Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981**, *22*, 1859−1862.
(22) Beaucage, S. L.; Iyer, R. P. *Tetrahedron* **1992**, *48*, 2223−2311.
(23) Häner, R.; Garo, F.; Wenger, D.; Malinovskii, V. L. *J. Am. Chem. Soc.* **2010**, *132*, 7466−7471.
(24) Charles, L.; Al Ouahabi, A.; Lutz, J.-F. Manuscript in preparation.
(25) Köster, H. *Tetrahedron Lett.* **1972**, *13*, 1527−1530.
(26) The word primer is not used here with its biological meaning. It simply refers to a short starting sequence of thymine nucleotides.
(27) Voirin, E.; Behr, J.-P.; Kotera, M. *Nat. Protoc.* **2007**, *2*, 1360−1367.
(28) Murray, K. K. *J. Mass Spectrom.* **1996**, *31*, 1203−1215.
(29) Stults, J. T.; Marsters, J. C. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 359−363.
(30) Mann, M.; Meng, C. K.; Fenn, J. B. *Anal. Chem.* **1989**, *61*, 1702−1708.

DOI: 10.1021/acsmacrolett.5b00606
ACS Macro Lett. 2015, 4, 1077−1080