HAL
open science

Orthogonal synthesis of poly(alkoxyamine phosphodiester)s for data storage applications

Gianni Cavallo

To cite this version:

Gianni Cavallo. Orthogonal synthesis of poly(alkoxyamine phosphodiester)s for data storage applications. Polymers. Université de Strasbourg, 2018. English. NNT : 2018STRAF016 . tel-02946941

HAL Id: tel-02946941

https://theses.hal.science/tel-02946941v1

Submitted on 23 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE STRASBOURG

École Doctorale des Sciences Chimiques

Institut Charles Sadron, CNRS

Thèse présentée par :

Gianni Cavallo

soutenue le : **20 septembre 2018**

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline/Spécialité : Chimie des Polymères

Orthogonal synthesis of poly(alkoxyamine phosphodiester)s for data storage applications

THÈSE dirigée par :

M. Jean-François LUTZ	Directeur de recherche au CNRS, Institut Charles Sadron, Strasbourg

RAPPORTEURS :

M. Robert Häner	Professeur, Université de Berne, Berne
M. Simon Harrisson	Chargé de recherche, Université Paul Sabatier, Toulouse

AUTRES MEMBRES DU JURY :

Mme Laurence Charles	Professeur, Aix-Marseille Université, Marseille
Mme Amparo Ruiz-Carretero	Chargée de recherche, Institut Charles Sadron, Strasbourg
M. Nicolas Giuseppone	Professeur, Institut Charles Sadron, Strasbourg

Table of contents

List of abbreviations **1**

Introduction Générale **5**

General introduction **11**

Chapter I **19**

1 Introduction 21

2 Sequence-Defined Polymers 21

2.1 Biopolymers 21

2.1.1 DNA 21

2.1.2 Proteins 23

2.2 Synthetic polymers 25

2.2.1 Step-growth polymerization 25

2.2.2 Chain-growth polymerization 28

2.2.3 Multi-step growth polymerization 31

3 Synthetic strategies for the synthesis of sequence defined polymers 34

3.1 Introduction 34

3.2 Protecting group approach (AB+AB) 34

3.3 Iterative synthesis using latent reactive groups 37

3.4 (AA+BB) approach 38

3.5 Chemoselective (AB+CD) approach 39

3.6 (AA+BC) approach 41

3.7 Multicomponent strategies 42

4 Chemical tools for the synthesis of poly(alkoxyamine phosphodiester)s 44

4.1 Introduction 44

4.2 Phosphoramidite chemistry historical background and development 44

4.3 Phosphoramidite chemistry protocol 47

4.4 Nitroxide radical coupling 49

5 Digital polymers 52

5.1 Introduction 52

5.2 Oligo(triazole amide)s 53
5.3 Poly(alkoxyamine amide)s 55
5.4 Sequence-coded polyurethanes 57
5.5 Poly(phosphodiester)s 58
6. Conclusions 61

Chapter II 63
1 Introduction 65
2 Result and discussion 66
2.1 Monomers design and synthesis 66
2.2 Solid-state iterative protocol 71
2.3 Synthesis of sequence-defined poly(alkoxyamine phosphodiester)s using 4-Hydroxy-TEMPO 75
2.4 Design of optimized hydroxy-functionalized nitroxides 76
3 Conventional sequencing 81
4 Fragmentation-free sequencing approach 86
4.1 Introduction 86
4.2 General strategy for F^2S using poly(alkoxyamine phosphodiester)s 87
4.3 Results and discussion 88
4.3.1 Synthesis of the capping agent 88
4.3.2 Iterative protocol optimization and F^2S sequencing 89
5 Conclusions 94

Chapter III 95
1 Introduction 97
Results and discussion 99
2.1 First generation of poly(akoxyamine phosphodiester)s encoding dyads using alphabet 1 99
2.2 Second generation of poly(alkoxyamine phosphodiester)s encoding dyads using alphabet 2 101
2.3 Synthesis of uniform poly(alkoxyamine phosphodiester)s encoding dyads using alphabet 3 105

3 Sequencing *via* dyads extraction 110
3 Conclusions and perspectives 112

Chapter IV 113
1 Introduction 115
2 Result and discussion 117
2.1 Monomers design and synthesis 117
2.1.1 Monomers design 117
2.1.2 Synthesis of monomer Pr_2 117
2.2 Synthesis of uniform sequence-defined poly(alkoxyamine phosphodiester)s using SG1-OH/T_3 120
2.3 Synthesis of uniform sequence-defined poly(alkoxyamine phosphodiester)s using SG1-OH/PROXYL 122
3 Conclusions and perspectives 129

General conclusion 131

Experimental part 137
1 Characterization techniques and instrumentation 139
1.1 NMR 139
1.2 SEC 139
1.3 MS spectrometry 139
1.4 UV spectroscopy 139
1.5 Iterative synthesis 139
2 Materials 140
3 Experimental section 141
3.1 Syntheses Chapter 2 141
3.1.1 Solid support modification 141
3.1.1.1 Synthesis of intermediate 1 141
3.1.1.2 Synthesis of intermediate 2 141
3.1.1.3 Synthesis of intermediate 3 142
3.1.2 Monomers synthesis 142

3.1.2.1 Synthesis of intermediate HO-0_a 143
3.1.2.2 Synthesis of intermediate HO-0b 144
3.1.2.3 Synthesis of intermediate HO-0_c 144
3.1.2.4 Synthesis of intermediate HO-1_a 144
3.1.2.5 Synthesis of intermediate HO-1_b 144
3.1.2.6 Synthesis of coding monomer 0_a 144
3.1.2.7 Synthesis of coding monomer 0_b 145
3.1.2.8 Synthesis of coding monomer 0_c 145
3.1.2.9 Synthesis of coding monomer 1_a 145
3.1.2.10 Synthesis of coding monomer 1_b 145
3.1.2.11 Synthesis of intermediate 4 146
3.1.2.12 Synthesis of linker T_2 146
3.1.2.13 Synthesis of linker T_3 146
3.1.3 Solid-state iterative protocol 146
3.1.3.1 Phosphoramidite coupling (step i_1 and i_2) 146
3.1.3.2 Nitroxide-radical coupling (step ii) 147
3.1.3.3 Cleavage and deprotection (step iii) 147
3.1.4 Monomer synthesis for F^2S sequencing 147
3.1.4.1 Synthesis of intermediate 5 148
3.1.4.2 Synthesis of intermediate 6 148
3.1.4.3 Synthesis of intermediate 7 148
3.1.4.4 Synthesis of linker L_1 148
3.1.4.5 Capping agent synthesis. 149
3.1.5 Solid-state protocol for F^2S 150
3.1.5.1 FMOC deprotection (8) 150
3.1.5.2 Linker L_1 coupling (9) 150
3.1.5.3 Poly(alkoxyamine phosphodiester)s synthesis on solid support. 151
3.2 Additional figures chapter2 151
3.2.1 MS analysis and MS/MS sequencing. 151
3.2.2 SEC. 153
3.2.3 NMR. 154
3.2.3.1 Coding monomer NMR 154

3.2.3.2 Poly(alkoxyamine phosphodiester)s NMR 155
3.2.3.2 Capping agent NMR 156
3.3 Syntheses Chapter 3 156
3.3.1 Synthesis of alphabet 1 156
3.3.1.1 Synthesis of coding nitroxide 0_d 156
3.3.1.2 Synthesis of coding nitroxide 1_d 157
3.3.2 Synthesis of alphabet 2 157
3.3.2.1 Synthesis of coding nitroxide $0'_d$ 157
3.3.2.2 Synthesis of intermediate 11 157
3.3.2.3 Synthesis of intermediate 12 158
3.3.2.4 Synthesis of intermediate 13 158
3.3.2.5 Synthesis of intermediate 14 158
3.3.2.6 Synthesis of nitroxide $1'_d$ 159
3.3.3 Synthesis of alphabet 3 159
3.3.3.1 Synthesis of intermediate 15 159
3.3.3.2 Synthesis of intermediate 17 160
3.3.3.3Synthesis of intermediate 18 160
3.3.3.4 Synthesis of intermediate 19 161
3.3.3.5 Synthesis of coding nitroxide $0''_d$ 161
3.3.3.6 Synthesis of intermediate 20 161
3.3.3.7 Synthesis of intermediate 21 161
3.3.3.8 Synthesis of intermediate 22 162
3.3.3.9 Synthesis of coding nitroxide $1''_d$ 162
3.3.4 Iterative solid-state protocol 162
3.4 Additional figures chapter 3 162
3.4.1 MS analysis and sequencing 162
3.4.2 NMR 164
3.4.3 SEC 165
3.5 Syntheses Chapter 4 165
3.5.1 Synthesis of monomer Pr_1 165
3.5.2 Synthesis of monomer Pr_2 166
3.5.2.1 Synthesis of intermediate 17 166

3.5.2.2 Synthesis of intermediate 18 166
3.5.2.3 Synthesis of intermediate 23 166
3.5.2.4 Synthesis of Monomer Pr_2 166
3.5.3 Solid state Iterative protocol 167
3.5.3.1 Phosphoramidite coupling (step i_1 and i_2) 167
3.5.3.2 Nitroxide-radical coupling (step ii) 167
3.5.3.3 Cleavage and deprotection (step iii) 167
3.6 Additional figures chapter 4 167
3.6.1 MS analysis and selective cleavage 167
3.6.2 SEC 168
3.6.3 NMR 169

References 171
List of publications 177

List of figures

Figure 0.1: General concept of molecular encryption and decryption in synthetic sequence-defined polymers. 13
Figure 0.2: Summary of the projects presented in the thesis. 15
Figure I.1: DNA double helix, chemical structure and base pairing rule. 22
Figure I.2: Schematic representation of the central dogma of molecular biology..... 22
Figure I.3: Translation table from codons to amino acids. 23
Figure I.4: Representation of primary, secondary, tertiary and quaternary structures of a protein. 24
Figure I.5: General classification of the polymerization techniques. 25
Figure I.6: CuAAC step-growth polymerization. 26
Figure I.7: Synthesis of ABCC sequence regulated vinyl copolymer *via* metal-catalyzed step-growth radical polymerization. 27
Figure I.8: RACP general strategy for the polymer synthesis. 27
Figure I.9: Probability of finding monomers from each block of of an $A_5B_5A_5C_5A_5D_5A_5C_5A_5B_5$ and $A_{20}B_{20}A_{20}C_{20}A_{20}D_{20}A_{20}C_{20}A_{20}B_{20}$. 29
Figure I.10: Template-assisted single insertion. 30
Figure I.11: General concept of ultra-precise maleimide insertion..... 30
Figure I.12: Solid phase-assisted synthesis concept. 31
Figure I.13: Standard protocol for the polypeptide synthesis on solid support..... 32
Figure I.14: General strategy for (AB+AB) approach. 35
Figure I.15: General strategy for the formation of responsive 2D polymers. 35
Figure I.16: General strategy or the synthesis of sequence-defined polymers oligonucleotide hybrids. 36
Figure I.17: Strategy for the synthesis of sequence-defined oligocarbamates. 36
Figure I.18: General strategy for sequence-defined oligothiophenes**.**..... 38
Figure I.19: (AB+AB) solid support-assisted synthesis of triazine-based polymers. **(i)** triazine coupling**:** DIPEA, THF 35ºC; **(ii)** diamine coupling: DIPEA, NMP 80ºC; **(iii)** cleavage: TFA. 39
Figure I.20: Chemoselectivity concept..... 39
Figure I.21: Peptoids synthesis.. 40
Figure I.22: Development of sequence-defined polymers based on thiolactone chemistry. . 41
Figure I.23: General scheme for the synthesis of sequence-defined polymers *via p*hosphine-catalyzed Michael and thiol-ene "click" addition using fluorous tags. 42
Figure I.24: Strategy for sequence defined polymer synthesis *via* multicomponent approach. 43
Figure I.25: First dinucleotide synthesis by Todd**.** 44
Figure I.26: Phosphodiester method. 45
Figure I.27: Phosphotriester method..... 45
Figure I.28: Phosphite triester method on silica matrix. 46

Figure I.29: Mechanismsa involved in NRC and radical formation. **a)** General mechanisms of NRC; **b)** Mechanisms of radical formation. 49
Figure I.30: Chain end functionalization strategy. 51
Figure I.31: General strategy for the synthesis of triblock copolymers *via* one-pot CuAAC and NRC reactions. 52
Figure I.32: Oligo(triazole amide) synthesis. 53
Figure I.33: Fragmentation pattern for oligo(triazole amide)s and sequencing *via* tandem MS of a sequence coding for 0-0-0-0. 54
Figure I.34: General strategy for the synthesis of sequence-defined poly(alkoxyamine amide)s. 55
Figure I.35: MS/MS sequencing of a poly(alkoxyamine amide) sequence coding for 1-1-0-1-0. 56
Figure I.36: Monomer structures and iterative protocol for the synthesis of sequence-defined oligocarbamates. 57
Figure I.37: Tandem MS sequencing of digitally encoded oligocarbamates. 58
Figure I.38: Monomer structures and general strategy for the synthesis of poly(phosphodiester)s. 59
Figure I.39: General concept of molecular byte tags applied for poly(phosphodiester)s. 61
Figure II.1: a) General structure of poly(alkoxyamine phosphodiester)s. **b)** General structures of the monomers employed in the iterative protocol. 67
Figure II.2: Structures of the different coding monomers employed for the synthesis of poly(alkoxyamine phosphodiester)s. 68
Figure II.3: Example of ^{1}H-NMR (a) and ^{13}C-NMR (b) for monomer 0_a registered in $CDCl_3$. ... 70
Figure II.4: Hydroxy-functionalized nitroxide monomers tested in this work. 71
Figure II.5: UV-Spectra obtained for the deprotection solution of the modified resin using a solution of TCA 3 % in DCM. 73
Figure II.6: ESI-MS recorded for a decamer synthesized using monomer 0_a in step (*i*) and 4-Hydoxy-TEMPO in step (*ii*). 76
Figure II.7: ESI-HRMS registered for a hexamer synthesized using linker T_2 and coding monomer 0_a. 78
Figure II.8: Positive ESI-HRMS recorded for the linker T_3. 79
Figure II.9: Negative ESI-HRMS recorded for entry 3 Table II.1. 80
Figure II.10: Products originated from a hypothetical disproportionation reaction in step (*ii*). 81
Figure II.11: Example of Sequencing of poly(alkoxyamine phosphodiester)s. 85
Figure II.12: Example of sequencing for a polymer coding for one byte. 86
Figure II.13: General concept of fragmentation free sequencing. 87
Figure II.14: General strategy employed in this work. 88
Figure II.15: ESI-MS registered for a polymer synthesized using 4 glycines, L_1 as linker and performing one iterative cycle. 90
Figure II.16: Sequencing of the polymer synthesized using F^2s approach. 93

Figure III.1: General strategy employed for the synthesis of poly(alkoxyamine phosphodiester)s encoding dyads. 98
Figure III.2: Structures of the building blocks synthesized and tested for the development of Poly(alkoxyamine phosphodiester)s encoding dyads. 99
Figure III.3: Positive ESI-MS recorder for the monomers couple 0_d and 1_d (alphabet 1). 100
Figure III.4: Negative ESI-HRMS analysis of polymer synthesized using alphabet 1 coding for 0_a-1_d-0_a-0_d. 100
Figure III.5: Positive ESI-MS recorder of monomer $0'_d$ and monomer $1'_d$ (alphabet2). 103
Figure III.6: Negative ESI-HRMS analysis of oligomer synthesized using alphabet 2 (entry 6 tab. 3). 104
Figure III.7: UV analysis of the quinhydrone test solution for entry 6 Table III.1. 105
Figure III.8: Positive ESI-HRMS of the monomers couple forming alphabet 3. 109
Figure III.9: Negative ESI-HRMS recorded for entry 4 tab.4. 109
Figure III.10: Mass spectrometry analysis of the sequence-coded oligomer entry 5 tab 4. . 111
Figure IV.1: General concept of stepwise selective fragmentation. 116
Figure IV.2: Structures of the hydroxyl-functionalized nitroxides employed in his work. 117
Figure IV.3: ESI-HRMS analysis registered for monomer Pr_1. 118
Figure IV.4: Positive ESI-HRMS recorded for monomer Pr_2. 119
Figure IV.5: Negative ESI-HRMS registered for a polymer coding for 0_a-T_3-0_a-SG1-0_a-T_3. 120
Figure IV.6: Molecular structure of the fully deprotonated oligomer. 121
Figure IV.7: Breakdown curves measured for the oligomer coding for 0_a-T_3-0a-SG1-0_a-T_3. . 122
Figure IV.8: ESI-HRMS and ESI-MS/MS analysis of a sequence-defined polymer coding for 1_a-SG1-1_a-Pr_2. 124
Figure IV.9: ESI-HRMS recorded for a sequence-defined polymer coding for 1_a-Pr_2-1_a-SG1-0_a-Pr_2-1_a-SG1. 126
Figure IV.10: ESI-MS/MS of the precursor ion corresponding to 528. (grey labelled peak). 127
Figure IV.11: ESI-HRMS recorded for a sequence-defined polymer coding for1_a-Pr_2-1_a-SG1-0_a-Pr_2-1_a-SG1-P_0-SG1-P_0-Pr_2. 128
Figure IV.12: ESI-MS/MS of the precursor ion corresponding to 515.3. (grey labelled peak). 129
Figure IV.13: General concept proposed to obtain byte fragmentations and sequencing *via* stepwise selective alkoxyamine bond fragmentation. 130
Figure E.1: General strategy employed for the development of poly(alkoxyamine phosphodiester)s for fragmentation-free sequencing. 150
Figure E.2: Example of Sequencing of poly(alkoxyamine phosphodiester)s. 152
Figure E.3: Size exclusion chromatogram recorded in water/acetonitrile for oligomer 9 in Table E.1 153
Figure E.4: NMR recorded in $CDCl_3$ for the coding monomer 0_a. 154
Figure E.5: NMR spectra recorded in D_2O for entry 3 in Table E.1 corresponding to the sequence α-$0aT_30aT_30aT_30aT_3$. 155
Figure E. 6: Capping agent NMR recorded in $CDCl_3$. 156

Figure E.7: Mass spectrometry analysis of the sequence-coded oligomer entry 6 tab.4. 163
Figure E.8: Examples of NMR spectra obtained for oligomer 3 in Table E.2 recorded in D_2O. 164
Figure E.9: Size exclusion chromatograms recorded in water/acetonitrile for oligomers P6, P4 and P2 in Table E.2. 165
Figure E.10: Selective fragmentation of the SG1-based alkoxyamine bonds obtained at 10 eV for a polymer coding for1_a-SG1-1_a-Pr_2. 167
Figure E.11: NMR spectra recorded in D_2O for an oligomer coding for 1_a-SG1-1_a-Pr_2. 169

List of schemes

Scheme I.1: Protecting groups free approaches developed for the synthesis of sequence-defined polymers. 37
Scheme I.2: Phosphoramidite cycle for oligonucleotides synthesis. 48
Scheme II.1: General strategy for the coding monomers synthesis. 68
Scheme II.2: Acylation mechanism for the coding monomers synthesis. 69
Scheme II.3: Phosphoramidite coupling mechanism for the synthesis of the coding monomers. 69
Scheme II.4: General strategy employed for the synthesis of digitally encoded poly(alkoxyamine hosphodiester)s. 71
Scheme II.5: Synthetic strategy employed for the solid-support modification. 72
Scheme II.6: Mechanism of the phosphoramidite coupling on solid support. 73
Scheme II.7:Mechanism of phosphite oxidation into a phosphate. 74
Scheme II.8: Nitroxide radical coupling mechanism on solid support. 75
Scheme II.9: General strategy for the synthesis of linker T_2. 77
Scheme II.10: General strategy for the synthesis of linker T_3. 78
Scheme II.11: Capping agent synthesis. 89
Scheme II.12: Synthetic strategy employed for the synthesis of monomer L_1. 89
Scheme III.1: General strategy for the synthesis of monomers 0_d and 1_d. 99
Scheme III.2: Synthetic strategy employed for the synthesis of monomer 0'd. 101
Scheme III.3: First synthetic strategy for the synthesis of monomer 1'd. 101
Scheme III.4: Mechanism for the synthesis of intermediate 11. 102
Scheme III.5: Synthetic strategy employed for the synthesis of monomer 1'd. 102
Scheme III.6: First synthetic strategy employed for the synthesis of monomer 1"d. 106
Scheme III.7: Synthetic strategy employed for the synthesis of alphabet 3. 107
Scheme III.8: Protection mechanism using TIPS-Cl. 107
Scheme III.9: Mechanism of TEMPO catalyzed oxidation of an alcohol into a carboxylic acid. 108
Scheme III.10: Amidification mechanism using PyBop as activating agent. 108
Scheme IV.1: Synthetic strategy employed for the synthesis of monomer Pr_1. 118

Scheme IV.2: Synthetic strategy employed for the synthesis monomer Pr_2. 119
Scheme E.1: General strategy employed for the solid-support modification. 141
Scheme E.2: General strategies for the syntheses of the monomers employed in this work. 143
Scheme E.3: Synthetic strategy employed for the synthesis of linker L_1. 147
Scheme E.4: Synthetic strategy employed for the synthesis of the capping agent. 149
Scheme E.5: Synthetic strategy employed for the synthesis of alphabet 1. 156
Scheme E.6: Synthetic strategies employed for the synthesis of alphabet 2. 157
Scheme E.7: First synthetic strategy employed for the synthesis of alphabet 3. 159
Scheme E.8: Synthetic strategy employed for the synthesis of alphabet 3. 160
Scheme E.9: Synthetic strategy employed for the synthesis of Pr_1. 165
Scheme E.10: Synthetic strategy employed for the synthesis of Pr_2. 166

List of tables

Table I.1: Resins classification of according to the cleavage conditions. 33
Table II.1: ESI-HRMS Individual characterization of the digital polymers synthesized and studied in this work. 80
Table III.1: ESI-HRMS individual characterization of the digital polymers synthesized using alphabet 2. 103
Table III.2: Negative ESI-HRMS characterization of the sequence-coded polymers 110
Table E.1: ESI-HRMS characterization of the digital polymers synthesized and studied in this work. 151
Table E.2: Negative ESI-HRMS characterization of the sequence-coded polymers synthesized using alphabet 3. 162

List of abbreviations

BAIB	(Diacetoxyiodo)benzene
NMI	1- methylimidazole
TEMPO free radical	2,2,6,6-Tetramethylpiperidin-1-yl)oxyl
Phosphoramidite chloride	2-cyanoethyl-N,N-diisopropylchlorophosphoramidite
P-Cl	2-cyanoethyl-N,N-diisopropylchlorophosphoramidite
NPPOC	3‘-Nitrophenylpropyloxycarbonyl
3-Amino-PROXYL	3-amino-2,2,5,5-tetramethyl-1-pyrrolydinyloxy
DMT-Cl	4,4’-dimethoxytriphenylmethyl chloride
4-Amino-TEMPO	4-Amino-2,2,6,6-tetramethylpiperidine-1-oxyl
DMAP	4-Dimethylaminopyridine
4-Hydroxy-TEMPO	4-Hydroxy-2,2,6,6-tetramethylpiperidine
FMOC	9-Fluorenylmethoxycarbony
CH3CN	Acetonitrile
ADMET	Acyclic diene metathesis
AT-NRC	Atom transfer nitroxide radical coupling
ATRA	Atom-tranfer radical addition
ATRP	Atom-tranfer radical polymerization
PyBOP	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
CHCl3	Chloroform
CID	Collision-induced dissociation
CPG	Controlled pore glass
CuBr	Copper bromide
CuAAC	Copper(I)-catalyzed azide-alkyne cycloaddition
DBTL	Dibutyltin dilaurate
DMPA	2,2-Bis(hydroxymethyl)propionic acid
DP	Degree of polymerization
DNA	Deoxyribonucleic acid
CDCl3	Deuterated chloroform
DCM	Dichloromethane
Et2O	Diethyl ether
DMSO	Dimethyl sulfoxide
DMF	Dimethylformamide
Đ	Dispersity index
ESI	Electrospray ionization

EtOH	Ethanol
EtOAC	Ethyl acetate
F2S	fragmentation-free sequencing
HRMS	High resolutiom mass spectrometry
I2	Iodine
LDA	Lithium diisopropylamide
MgSO4	Magnesium sulphate anhydrous
MeOH	Methanol
CH3I	Methyl iodide
KH2PO4	Monopotassium phosphate
DCC	N,N'-Dicyclohexylcarbodiimide
DIPEA	N,N-diisopropylethylamine
Brine	NaCl saturated solution
Abbreviation	Name
n-BuLi	n-Butyllithium
NMP	Nitroxide mediate polymerization
NRC	Nitroxide radical coupling
SG1-OH	N-tert-butyl-1-diethoxyphosphoryl-N-$l^{1}-oxidanyl-2,2-dimethylpropan-1-amine
SG1	N-tert-butyl-1-diethoxyphosphoryl-N-$l^{1}-oxidanyl-2,2-dimethylpropan-1-amine
NMR	Nuclear magnetic resonance
PTSA	Para-toluenesulfonic acid
P-3C	Passerini three components reaction
PtBA	Poly(*tert-butyl* acrylate)
PEG	Polyethylene glycol
PEO	polyethylene oxide
RAFT	Reversible addition−fragmentation chain-transfer polymerization
RNA	Ribonucleic acid
ROP	Ring opening polymerization
rt	Room temperature
SET-NRC	Single electron transfer nitroxide radical coupling
SEC	Size exclusion chromatography
NaHCO3	Sodium bicarbonate
Na EDTA	Sodium ethylenediaminetetraacetic acid
NaH	Sodium hydride
NaHCO3	Sodium hydrogen carbonate
NaNO3	Sodium nitrate

Na2SO4	Sodium sulphate anhydrous
Na2S2O3	Sodium thiosulphate
MS/MS	Tandem mass spectrometry
Boc	*Tert-Butoxycarbonyle*
t-Bu	*Tert-Butyl*
TBAF	Tetrabutylammonium fluoride
THF	Tetrahydrofuran
TLC	Thin layer chromatography
TCA	trichloroacetic acid
TCA	Trichloroacetic acid
NEt3	Triethylamine
TIPS-Cl	Triisopropylsilyl chloride
Me6TREN	Tris[2-(dimethylamino)ethyl]amine
U-4C	UGI four components reaction
MsCl	Methanesulfonyl chloride
PMDETA	N,N,N’,N’’,N’’-Pentamethyldiethylenetriamine
DCPA	Dimethyl tetrachloroterephthalate
DIC	N,N'-Diisopropylcarbodiimide
HOBT	Hydroxybenzotriazole
Pd(PPH3)Cl2	Bis(triphenylphosphine)palladium(II) dichloride
NBS	N-Bromosuccinimide
NaOMe	Sodium methoxide

Introduction Générale

Dans la société moderne, le volume de données stockées augmente continuellement. Il a été estimé que 35 zétabytes (1000^7 bytes) de données digitales seront générées durant l'année 2020.[1] Cette grande quantité d'information est stockée dans des centres de données, où les conditions de chaleur et d'humidité sont contrôlées, ce qui entraine des consommations d'espace et d'énergie. Une alternative possible à ce mode de stockage de donnée serait d'encoder l'information dans des molécules. En effet, le stockage de données à l'échelle moléculaire permettrait d'accroitre considérablement la densité d'information, ce qui résulterait en une réduction de l'espace dédié aux centres de données. Ces dernières années, des efforts ont été réalisés afin de stocker de l'information à l'échelle moléculaire dans de biomolécules comme l'ADN.[2] En effet, différentes méthodes ont été développées afin de stocker différents formats comme l'ASCII, le PDF ou encore le JPEG dans de l'ADN.[3] Par ailleurs, une vidéo a aussi été encodée dans le génome d'une bactérie.[4] Néanmoins, des alternatives aux biopolymères ont aussi été proposées. Par exemple, notre équipe a suggéré l'utilisation de polymères à séquences définies synthétiques, aussi appelés polymères digitaux afin de stocker de l'information.[5] Il a déjà été montré que de l'information peut être contenue dans des polymères synthétiques.[6] Ainsi, de l'information binaire peut être introduite dans un copolymère en utilisant deux monomères arbitrairement définis comme « bit 0 » et « bit 1 ». Pour permettre une bonne fiabilité dans le déchiffrement de l'information, ces molécules doivent être monodispereses et parfaitement définies en séquence. De plus, elles doivent aussi permettre un séquençage facile et rapide. Le terme séquençage fait ici référence à une technique capable de retrouver l'ordre dans lequel des monomères sont organisés dans le polymère, et ainsi pouvoir remonter à l'information contenu dans ce dernier (Figure 0.1). L'accès à une grande variété de structures chimiques[7, 8] confère un avantage à l'utilisation de polymères synthétiques par rapport aux biopolymères. Il est ainsi possible de concevoir des polymères pouvant être séquencés de façon optimale.[9]

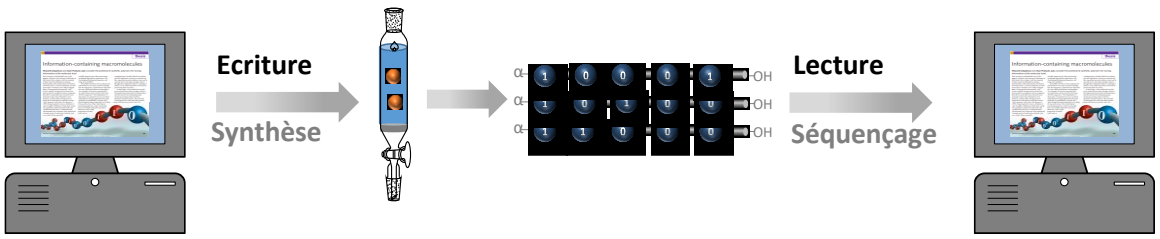


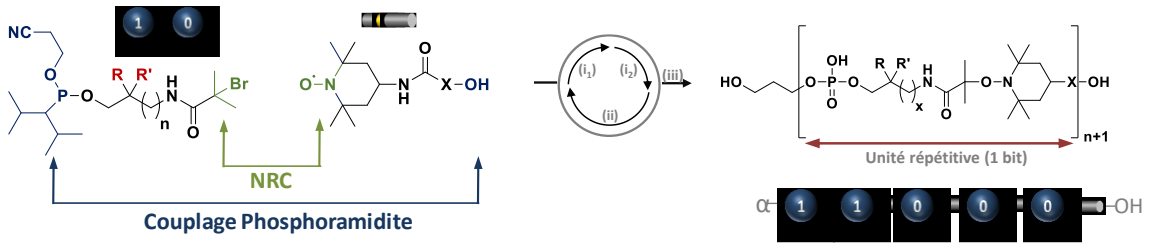
Figure 0.1: Concept général du codage et du décodage d'information dans des polymères synthétiques définis en séquence.

La synthèse de polymères digitaux est basée sur l'utilisation de la polymérisation multi-étapes dans laquelle les monomères sont ajoutés les uns après les autres sur le polymère croissant. Cette technique, développée par Merrifield pour la synthèse de polypeptides,[10] est aujourd'hui la technique la plus utilisée dans la synthèse de polymères à séquences définies. Cette approche permet le contrôle total de la structure primaire du polymère ainsi

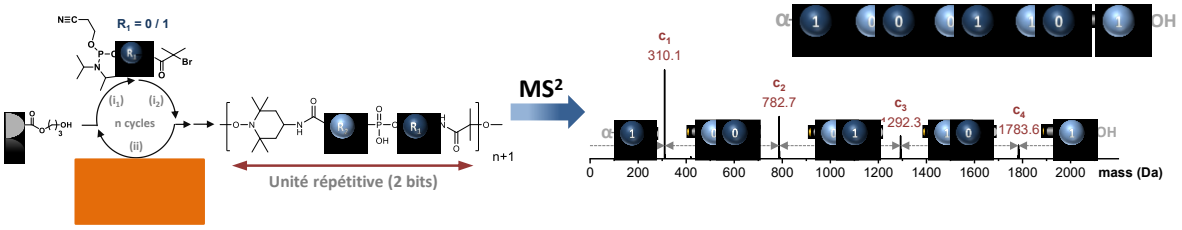
que de son degré de polymérisation, ce qui permet un codage fiable. Notre équipe a montré que l'introduction dans les monomères de groupements latéraux hydrogène (bit 0) ou méthyle (bit 1) était suffisant afin d'obtenir un alphabet binaire séquençable par spectrométrie de masse en tandem.[11] Parmi les polymères digitaux développés à ce jour, les poly(phosphodiester)s sont une des classes les plus prometteuses. En effet, de longues chaînes de poly(phosphates)s peuvent être synthétisées relativement rapidement grâce à la méthode phosphoramidite automatisée.[12] Néanmoins, ces molécules génèrent des spectres de fragmentation complexes, ce qui rend difficile leur séquençage par spectrométrie de masse en tandem[13]

La classe des poly(alkoxyamine amide)s quand a elle montre que le séquençage par MS/MS peut être simplifié par l'ajout de liaisons alkoxyamines dans la chaine principale.[14] En effet, la liaison alkoxyamine étant labile, elle est fragmentée prioritairement en MS/MS ce qui génère des spectres de fragmentation faciles à interpréter. La différence de stabilité des liaisons alkoxyamines des monomères 0 et 1 entraine néanmoins des spectres de fragmentation dont la prédiction dépend de la séquence des polymères. Dans ce travail de thèse, le développement d'un protocole chimiosélectif pour la synthèse d'une nouvelle classe de polymères digitaux, les poly(alkoxyamine phosphodiester)s, a été développé. Comme cela a déjà été mentionné, l'avantage d'utiliser des polymères définis en séquences synthétiques pour stocker des données est de pouvoir modifier la chaine principale du polymère pour lui donner des propriétés nouvelles. La classe des poly(alkoxyamine phosphodiester)s a été ainsi conçue afin de simplifier le séquençage des poly(phosphodiester)s via l'insertion de liaisons alkoxyamines dans la chaine principale. La synthèse de ces polymères digitaux est une stratégie AB+CD décrite dans la Figure 0.2 a. Cette stratégie utilise deux monomères portant chacun deux fonctions différentes. La fonction A ne pourra réagir qu'avec la fonction C, de même pour les fonctions B et D, ce qui empêche la polycondensation entre les monomères. Le protocole permet d'accélerer la synthèse et améliore l'économie des étapes.[15] La méthode phosphoramidite ainsi que le couplage radicalaire de nitroxide ont étés employés pour la synthèse de ces polymères digitaux. Ces réactions ont été étudiées intensivement par la communauté scientifique durant ces dernières années et permettent des rendements quantitatifs et peu de produits secondaires de réaction. Les liaisons labiles alkoxyamine insérées dans la chaine principale de phosphodiester permettent le décodage de ces molécules. En effet, dans des conditions de MS/MS, la liaison alkoxyamine est rompue préférentiellement ce qui génère des fragments faciles à interpréter. De plus, la chaine principale a été conçue afin d'éviter des phénomènes de fragmentation propres à la séquence, problème qui existe avec la classe des poly(alkoxyamine amide)s.

a) Stratégie (AB+CD) pour la synthèse de poly(alcoxyamine phosphodiester)s encodés numériquement



b) Synthèse de poly(alcoxyamine phosphodiester)s avec des alphabets de nitroxydes étendus



c) Des poly(alkoxyamine phosphodiester)s avec des liaisons alcoxyamines de stabilités différentes

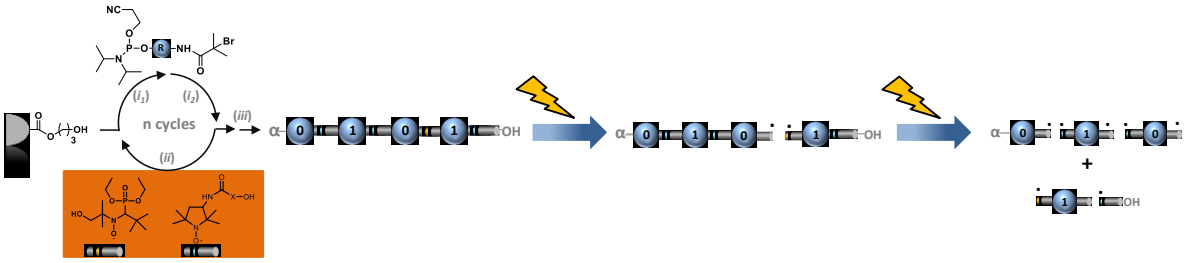


Figure 0.2:Résumé des projets présentés dans cette thèse. a) Chapitre II: Protocole itératif pour la synthèse de poly(alkoxyamine phosphodiester)s. Conditions expérimentales: (i1) couplage phosphoramidite: t.a., AcCN, tétrazole; (i2) oxydation: t.a., I2, 2,6-lutidine, THF/H2O; (ii) couplage radicalaire de nitroxydes: CuBr, Me6TREN, DMSO; (iii) clivage: pipéridine, AcCN, t.a., puis MeNH2, NH4OH, H2O, t.a.; b) Chapitre III: Conception, synthèse et séquençage de Poly(alkoxyamine phosphodiester)s générant des dyades en MS/MS; c) Chapitre IV :Poly(alkoxyamine phosphodiester)s modifiés pour l'obtention de clivage sélectif de liaisons alkoxyamine.

Dans le cadre de cette thèse, la chaine principale des poly(alkoxyamine phosphodiester)s a été modifiée afin de conférer de nouvelles propriétés aux polymères digitaux. Par exemple, la densité d'information de ces molécules a été augmentée via l'introduction de dyades clivables en MS/MS (Figure 0.2 b). La détection de dyades en spectrométrie de masse en tandem rend le décryptage de ces polymères digitaux relativement facile. Une autre modification a été conçue afin d'obtenir des poly(alkoxyamine phosphodiester)s à séquences contrôlées dont la structure primaire peut être clivée sélectivement en différents fragments (Figure 0.3 b). Cette thèse est composée de quatre chapitres. Le Chapitre I est un état de l'art de la recherche sur les polymères à séquences contrôlées. Dans une première partie, les biopolymères comme l'ADN et les protéines sont brièvement décrits. Une

deuxième partie discute le contrôle de la structure primaire dans les polymères synthétiques, et en particulier des différentes techniques de polymérisation employées afin d'obtenir des polymères à séquences contrôlées, et plus récemment des polymères à séquences définies. Dans cette partie, une importance particulière sera donnée à l'approche de polymérisation multi-étapes. La troisième partie est un rappel historique des différentes réactions employées pour le protocole de synthèse itératif présenté dans cette thèse. La dernière section décrit les polymères digitaux: leur développement, leur séquençage et leur lecture.

Le **Chapitre II** décrit dans une première partie la synthèse et le décryptage de poly(alkoxyamine phosphodiester)s digitaux. Différents monomères synthétiques ainsi qu'un bras espaceur ont été utilisées dans le protocole itératif décrit dans la Figure 0.2 a. Un chimiothèque de poly(alkoxyamine phosphodiester)s digitaux contenant différents alphabets a été synthétisée. Ces molécules présentent des spectres de fragmentation MS/MS faciles à interpréter grâce au clivage sélectif des liaisons alkoxyamines, ce qui permet décrypter l'information encodée. La deuxieme partie de ce chapitre décrit l'utilisation de ces polymères avec la technique de séquençage sans fragmentation (F^2S). Cette technique a été développée pour le criblage de chimiothèques de peptides biologiquement actifs par spectrométrie de masse.[16] Afin de pouvoir utiliser la méthode F^2S avec ces poly(alkoxyamine phosphodiester)s, un faible pourcentage d'agent de coiffage a été ajouté durant le couplage phosphoramidite. Apres n+1 cycles, le produit final est un mélange du polymère visé et des n polymères de plus faible masse moléculaire coiffés expérimentalement. En se basant sur la différence de masse entre chaque espèce, cette analyse permet la reconstruction de la séquence recherchée par spétrométrie de masse, sans utiliser la technique de MS/MS.

le **Chapitre III** décrit le développement de poly(alkoxyamine phosphodiester)s contenant des dyades binaires clivables en MS/MS. Une nouvelle architecture moléculaire a été développée pour cette classe de polymères à séquences définies. Elle permet l'insertion de deux bits d'information dans chaque unité répétitive. Comme décrit dans la Figure 0.2 b, quatre monomères codants différents (deux phosphoramidites et deux nitroxides) ont été utilisés comme unités codantes afin de permettre le stockage d'information lors du couplage radicalaire de nitroxide. Trois différent couples de nitroxides codants ont été synthétisés et testés pour la synthèse de polymères. Le nouveau concept permet d'augmenter la densité d'information des polymères digitaux tout en gardant un séquençage aisé. Dans des conditions de MS/MS, ces macromolécules génèrent des dyades de différentes masses moléculaires. Le processus de séquençage est permis par la détection de ces dyades (Figure 0.2 b) et un algorithme a été développé afin de décoder ces polymères digitaux en quelques millisecondes.

Dans le **Chapitre IV** est présenté le développement de poly(alkoxyamine phosphodiester)s comprenant des liaisons alkoxyamines sélectivement clivables. Comme présenté

précédemment, la liaison C-ON joue un rôle décisif pour le séquençage par MS/MS. L'énergie requise pour le clivage homolytique dépend de la stabilité des radicaux générés par les conditions de fragmentation. L'utilisation de différents nitroxides durant le couplage radicalaire rend possible l'insertion de liaisons alkoxyamines ayant des stabilités différentes dans la chaine principale du polymère. De plus, les différentes liaisons alkoxyamines peuvent être sélectivement fragmentées par spectrométrie de masse en tandem en utilisant des énergies d'ionisation différentes (Figure 0.2 c). Le chapitre décrit les différentes tentatives d'obtention d'un couple de monomères nitroxides ayant les caractéristiques recherchées, la synthèse des monomères et leur utilisation dans le protocole itératif préalablement décrit, ainsi que les études spectroscopiques préliminaires concernant l'énergie requise pour le clivage homolytique.

Ce travail de thèse présente le développement de poly(alkoxyamine phosphodiester)s en tant que polymères digitaux. De plus, la modification de la chaine principale afin d'obtenir une augmentation la densité d'information d'une part, et le clivage sélectif de la liaison alkoxyamine d'autre part y est discuté.

General introduction

In the modern society, the amount of stored data increases constantly; it is estimated that in 2020, 35 zettabytes (1000^7 bytes) of digital information will be generated.[1] This huge amount of information is stored in data centres in controlled temperature and humidity conditions, resulting in a high consumption of space and energy. A valid alternative to the solid-state memories may consist in storing data into molecules. Indeed, molecular data storage can considerably increase the information data density and be convenient in terms of space and energy consumption. In the past years, several efforts were made to store information at the molecular level using biomolecules as DNA.[2] Recently, different methodologies allowing to store different data format as ASCII, PDF, JPEG, etc. in DNA [3] or encoding a short movie into bacteria's genome were developed.[4] However, new approaches, alternative to biopolymers, were proposed in the last years.

For instance our group suggested using synthetic sequence-defined polymers, so called digital polymers, for this purpose.[5] As it was already demonstrated, it is possible to store information at the molecular level using synthetic polymers.[6] Indeed, every digital information is a binary sequence which can be encoded in a copolymer composed of two monomers intentionally defined as 0 and 1. These molecules must be monodisperse and perfectly defined to ensure a reliable molecular encryption. Moreover, they also have to allow fast and easy sequencing. The term sequencing refers to the use of techniques capable of revealing the monomers sequence in the polymer backbone allowing the information decryption (Figure 0.1). The advantages in using synthetic polymers for molecular data storage lies in the fact that these molecules allow to access a very large variety of chemical structures.[7, 8] Moreover, the backbone can be carefully designed in order to have an optimal read-out.[9]

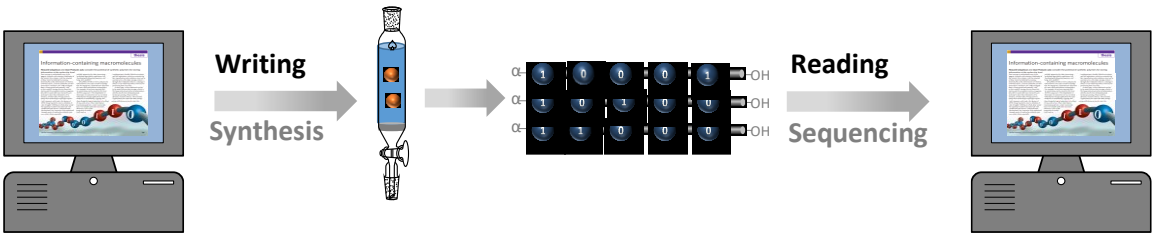


Figure 0.1: General concept of molecular encryption and decryption in synthetic sequence-defined polymers.

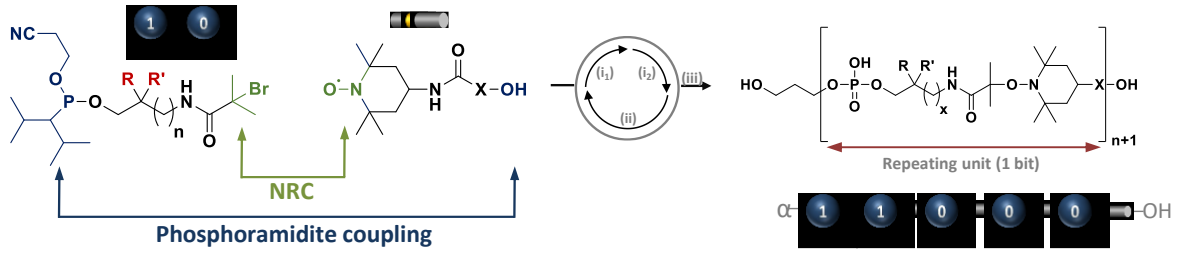
The digital polymers synthesis is usually done via a multi-step growth polymerization in which monomers are added one by one to a growing chain. This technique, initially developed by Merrifield for the polypeptide synthesis[10] is nowadays the most employed methodology for the synthesis of sequence-defined polymers. Indeed, this approach ensures the uniformity of the polymeric chains and the full control over the primary structure hence, a reliable molecular encryption. In addition, our group showed that the implementation of a binary alphabet relying on the methyl/hydrogen groups on the backbone side chain is sufficient to implement a MS/MS readable molecular binary alphabet.[11] Among the digital

polymers developed so far, poly(phosphodiester)s represent one of the most promising classes. In particular, long digitally encoded poly(phosphodiester) chains can be synthesized in a relatively short time using an oligonucleotide synthesizer.[12] Nevertheless, these molecules generate a complicated MS/MS spectrum making the sequencing difficult, in particular for long polymer chains.[13]

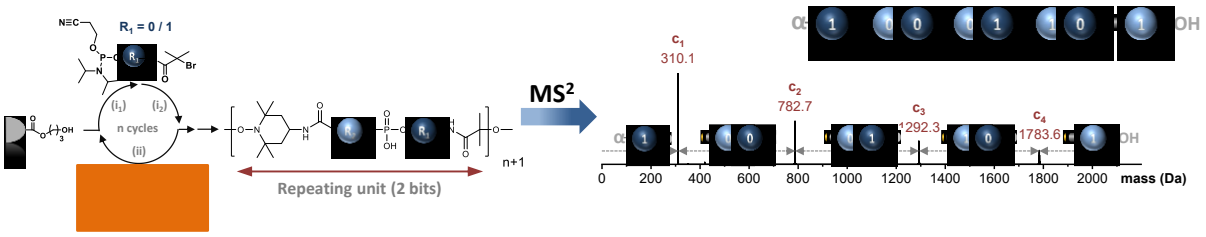
On the other hand, the poly(alkoxyamine amide)s class showed that the MS/MS read-out can be extremely simplified by adding an alkoxyamine bond in the polymer backbone.[14] Indeed, the labile alkoxyamine bond is preferentially cleaved in MS/MS conditions generating an easy to interpret fragmentation pattern. However, because of the different stabilities of the alkoxyamine bond generate by monomer 0 and monomer 1, the read-out of these molecules may depend on the sequence.

In the frame of the PhD project, the development of a chemoselective protocol for the synthesis of a new class of digital polymers namely the poly(alkoxyamine phosphodiester)s was developed. As said before, the advantage in using synthetic sequence-defined polymers for data storage application lies in the fact that it is possible to modify the backbone structure to improve some features. In this context, the poly(alkoxyamine phosphodiester)s class was designed with the aim to improve the poly(phosphodiester)s reading by alkoxyamine bond insertion in the polymer backbone. The synthesis of such digitally encoded polymers relies on an (AB+CD) approach depicted in Figure 0.2 a. This strategy employs two heterobifunctional monomers in which A can react only with C and B can react only with D avoiding the polycondensation of the monomers. The protocol allows to speed the synthesis up and improving the step economy.[15] In particular, phosphoramidite coupling and nitroxide-radical coupling (NRC) were employed for the synthesis of these digital polymers. These reactions were intensively studied in the past years and both lead to quantitative yields in a short time without the generation of side-products. The labile alkoxyamine bond inserted into the phosphodiester backbone enables the easy read-out of these molecules. Indeed, in MS/MS conditions the alkoxyamine bond is preferentially cleaved, generating fragments which are easy to sequence. Moreover, the backbone was designed in order to avoid sequence-dependent fragmentation phenomena occurring in the poly(alkoxyamine amide)s class. In the frame of the PhD studies, the poly(alkoxyamine phosphodiester) backbone was further modified in order to confer additional features to these digital polymers. For instance, the information density of these molecules has been increased by encoding dyads insertion in the polymer backbone (Figure 0.2 b). The dyad detection in tandem MS makes the data extraction from these digital polymers relatively easy.

a) (AB+CD) strategy for the synthesis of digitally encoded poly(alkoxyamine phosphodiester)s



b) Synthesis of poly(alkoxyamine phosphodiester)s using extended nitroxides alphabets



c) Poly(alkoxyamine phosphodiester)s having alkoxyamine bonds of different strength

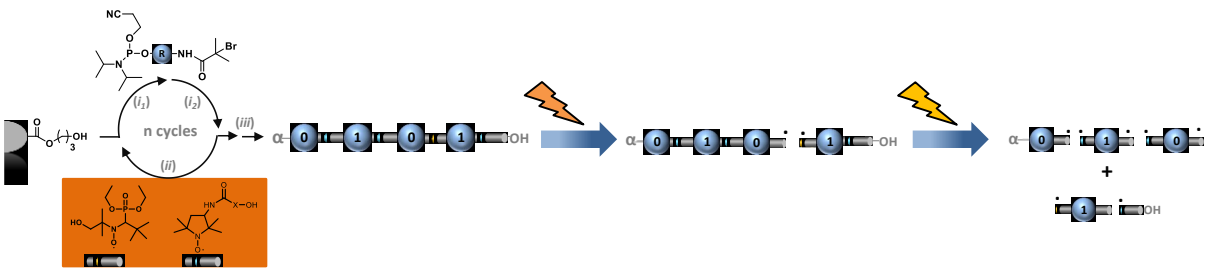


Figure 0.2: Summary of the projects presented in the thesis. **a) Chapter II:** Iterative protocol for the synthesis of poly(alkoxyamine phosphodiester)s. **(i_1)** phosphoramidite coupling: rt, CH_3CN, tetrazole; **(i_2)** oxidation: rt, I_2, 2,6-lutidine, THF/H_2O; **(ii)** nitroxide–radical coupling: CuBr, Me_6TREN, DMSO; (iii) cleavage: piperidine, CH_3CN, rt, then $MeNH_2$, NH_4OH, H_2O, rt; **b) Chapter III:** Design, synthesis and sequencing of poly(alkoxyamine phosphodiester)s assessed for encoding dyads detection; **c) Chapter IV:**Poly(alkoxyamine phosphodiester)s modified to obtain a selective alkoxyamine bond cleavage.

Moreover, a third modification was designed in order to obtain sequence-defined poly(alkoxyamine phosphodiester)s in which the primary structure can be selectively cleaved in defined parts (Figure 0.2 c). The following thesis is formed by four chapters. The **Chapter I** reports the state-of-the-art in the field of sequence-controlled polymers. In the first part, biopolymers as DNA and proteins are briefly described. The second part is about the primary structure control in synthetic polymers. In particular the different polymerization techniques which allow to obtain sequence-controlled and more recently sequence-defined polymers are discussed. In this part, particular importance is given to the multi-step growth polymerization approach. The third part is a historical background on the reactions

employed in the iterative protocol presented in this thesis. The last section describes the digital polymers: their development, sequencing and reading.

The **Chapter II** describes the synthesis and the read-out of digitally encoded poly(alkoxyamine phosphodiester)s. A library of different coding monomers was synthesized and employed in combination with the linker for the implementation of the iterative protocol depicted in Figure 0.2 a. A library of different digitally encoded poly(alkoxyamine phosphodiester)s containing different alphabets was synthesized. It was found that these molecules are relatively easy to sequence in tandem MS. Indeed, the selective alkoxyamine cleavage originates easy-to-interpret fragmentation patterns. The second part of the chapter describes the assessment of sequence-defined poly(alkoxyamine phosphodiester)s for fragmentation free sequencing approach (F^2S). This technique was developed for the screening of biologically active peptide libraries via MS analysis.[16] In order to assess poly(alkoxyamine phosphodiester)s for F^2S, a small percentage of a capping agent was added during the phosphoramidite coupling. After *n*+1 cycles, the final product is a mixture of the original polymer and *n* sequence specific truncated fragments. The MS analysis allows the reconstruction of the original sequence relying on the MS differences between the truncated fragments avoiding MS/MS fragmentation.

In **Chapter III,** the development of encoding cleavable binary dyads for poly(alkoxyamine phosphodiester)s is described. A new molecular architecture was developed for this class of sequence-defined polymer, allowing the insertion of two bits of information in each repeating unit. As depicted in Figure 0.2 b, in order to enable the information encryption during the nitroxide radical coupling, a set of four different coding monomers (two phosphoramidites and two nitroxides) was employed as coding units. In particular, three different couples of coding nitroxides were synthesized and tested for the polymer synthesis. The new design allows to increase the information density of digital polymers maintaining the sequencing easy. Indeed, in MS/MS conditions, these macromolecules generate encoding dyads containing different molecular weights. The sequencing, which relies on the dyads detection in MS/MS, is greatly simplified as depicted in Figure 0.2 b. Moreover, an algorithm was developed for the dyad detection allowing the decoding of these digital polymers in a few milliseconds.

In **Chapter IV** the development of a selective alkoxyamine bond cleavage for poly(alkoxyamine phosphodiester)s is presented. As discussed before, the C-ON bond plays a decisive role for the sequencing by MS/MS. The energy required for the homolytic cleavage depends on the stability of the resulting radicals generated in MS/MS conditions. The use of different nitroxides during the nitroxide-radical coupling makes the insertion of alkoxyamine bonds with different stability in the polymer backbone possible. Hence, the different alkoxyamine bonds can be selectively fragmented in tandem MS using different ionization energies (Figure 0.2 c). The chapter describes the attempts to find a couple of nitroxide

monomers suitable for this purpose; the synthesis of the different building blocks and their implementation in the iterative protocol. Moreover, the preliminary MS studies concerning the energy required for the selective homolytic cleavage are reported.

In the following thesis the development of poly(alkoxyamine phosphodiester)s as digital polymers is presented. Moreover, the backbone modifications designed for this class of digital polymers, which allows to increase the information density and obtain a selective alkoxyamine bond cleavage are discussed.

Chapter I

Sequence-defined polymers: state-of-the-art

1 Introduction

In the past century, the development of polymerization reactions and the resulting use of synthetic polymers in the everyday life have changed forever our society. Indeed, since 1950 the production of these molecules has constantly increased until reaching the annual production of 280 Mio.t of plastics in 2016.[17] Polymers are molecules composed of several repeating subunits.[18] Because of their different properties and their relatively easy synthetic access, they were intensively studied and employed over the last 150 years. Just like synthetic polymers, biopolymers are macromolecules fromed by natural subunits. Because of their involvement in all the biochemical processes, molecules as DNA, RNA, proteins, etc. were deeply investigated in the past decades. Biopolymers contain a fully defined monomer sequence which is the feature enabling them to carry on all the fundamental biochemical reactions. On the other hand, synthetic polymers generally contain a statistical monomer composition and different chain lengths. The control of the monomer sequence in the synthetic polymer backbones and the weight distribution control have always been one of the main goals for polymer chemists. In the past decades, the efforts made to regulate these parameters *via* standard polymerization enabled the synthesis polymers having a certain degree of control. For instance, synthesis of alternating, gradient, periodic and block copolymers containing narrow chain length distribution were described.[19] Nevertheless, full primary structure control is still far from being achieved with these techniques. In the last years, polymer chemists importing tools and techniques from other fields as biochemistry, biology and engineering, achieved the sequence regulation for synthetic polymers. For instance, techniques as solid-state assisted synthesis, template-assisted synthesis, etc. are extensively employed for the production of sequence-defined polymers. This new class of molecules gives the great possibility to finely tune the primary structure and thus control the macroscopic properties,[20] opening the field toward new polymer designs and applications. In this thesis, the development of a new class of sequence-defined polymers for data storage application is presented.

2 Sequence-Defined Polymers

2.1 Biopolymers

As written in section 1, several examples of sequence-defined polymers are present in nature. The fully defined primary structures of these molecules define their functions. This section is a general introduction about structures, biosyntheses and functions of natural sequence-defined polymers such as proteins and DNA.

2.1.1 DNA

DNA is a biomolecule that stores the genetic information and is involved in all the biochemical processes of all living organisms and many viruses. This molecule is composed of two antiparallel polynucleotides chains (single strands). Each strand is a sequence-defined polymer containing four nucleobases (adenine, cytosine, thymine and guanine) as repeating

units. The bases are linked to the backbone which is composed of deoxyribose units and phosphate groups. The characteristic double helix motif formed by two single strands is due to the selective hydrogen bonding between the four different nucleobases (adenine with thymine and guanine with cytosine) called base pairing rule [21] depicted in Figure I.1. According to this rule, only complementary sequences can interact each other.

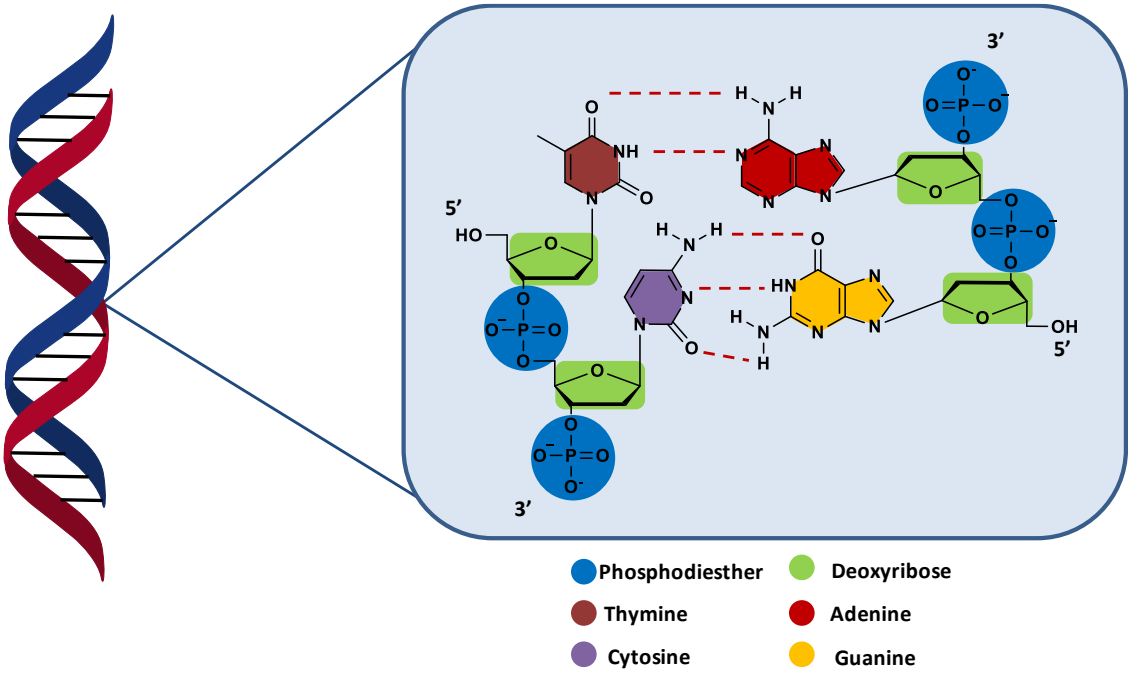


Figure I.1: DNA double helix, chemical structure and base pairing rule.

DNA is extremely important for life because it carries the biological information in all the living species. In 1958, Crick postulated the central dogma of molecular biology, depicted in Figure I.2, defining the flow of the genetic information in biological systems: "*The central dogma deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred back from protein to either protein or nucleic acid*".[22]

Figure I.2: Schematic representation of the central dogma of molecular biology.

Replication, transcription and translation are fundamental steps for the control of cell growth, differentiation and gene expression.[23, 24] The precise control of the primary structure during the DNA and RNA biosynthesis ensures the reliable information transfer between biomolecules. DNA replication is a process that allows the formation of two identical DNA copies starting from one original DNA molecule. The DNA synthesis is 5' to 3' directed and is catalyzed by DNA polymerase.[25] This enzyme enables the formation of

phosphodiester bonds between the growing chain and the deoxynucleoside 5'triphosphates. To ensure the control over the primary structure, the whole process relies on the base pairing rule between the growing chain and a single strand DNA template. The process is a semi-conservative synthesis. Indeed, the two new DNA molecules are formed by a template strand belonging to the original molecule and a new strand synthesized during the process.[26] According to the central dogma, during the transcription, the information contained in the DNA flows toward the RNA. The RNA polymerase is the enzyme catalyzing the reactions involved in the RNA synthesis.[23] The control of the primary structure relies on the base pairing rule between the growing RNA chain and the single strand DNA template. RNA plays a central role in the translation process and thus in gene regulation. Indeed, this biomolecule is capable of transfer the information stored into the DNA to the ribosome and thus activate the protein biosyntheses.[27] Ribosomes are natural molecular machines that translate the four-letter alphabet encoded in mRNA into the twenty-letter alphabet forming the polypeptides.

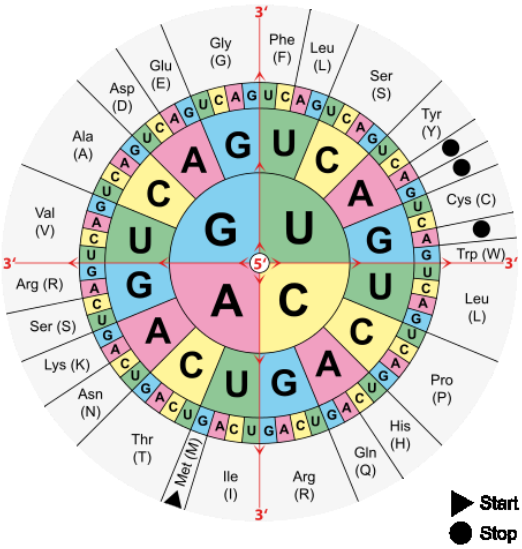


Figure I.3: Translation table from codons to amino acids.

During the translation, series of three bases called codons translate for a single amino acid[28] (Figure I.3). Also in this case, base pairing between the three bases forming the codon and the anticodon in the aminoacyl-tRNA[24] is the key step to obtain the sequence regulation.

2.1.2 Proteins

Proteins are the biopolymers that allow the expression of genetic information. The sequences of 22 amino acids linked *via* peptidic bonds, define the properties of the resulting macromolecules. The control of the amino acids sequence during their biosynthesis enables the production of enzymes, antibodies, hormones and several other molecules, that play a decisive role in the biochemical processes.[24] To better understand the relationship between function and structure in proteins, a hierarchal classification relying on four complexity levels

is applied. The lower level, called primary structure, takes into account the amino acid sequence in the backbone (Figure I.4). The monomer sequence determines the folding and the higher order self-assembling interactions of the resulting protein.[29] The secondary structure originates from weak interactions as hydrogen bonding or hydrophobic interactions. For instance, the alpha helix formation is due to the hydrogen bonding between an amine group and a carboxylic function located three or four residues earlier in the backbone.[30] The tertiary structure describes the three dimensional folding of the whole polypeptide sequence which is also due to weak interactions between amino acids involved in different types of secondary structures.[24] The quaternary structure defines the spatial organization of the different subunit, forming the protein. According to their architecture, these macromolecules can carry on several biochemical processes. For instance, many proteins are enzymes that catalyse the biochemical reactions. Other polypeptides have mechanical functions as actin and myosin or are involved in the cell signalling, cell growth mechanisms, etc.[24]

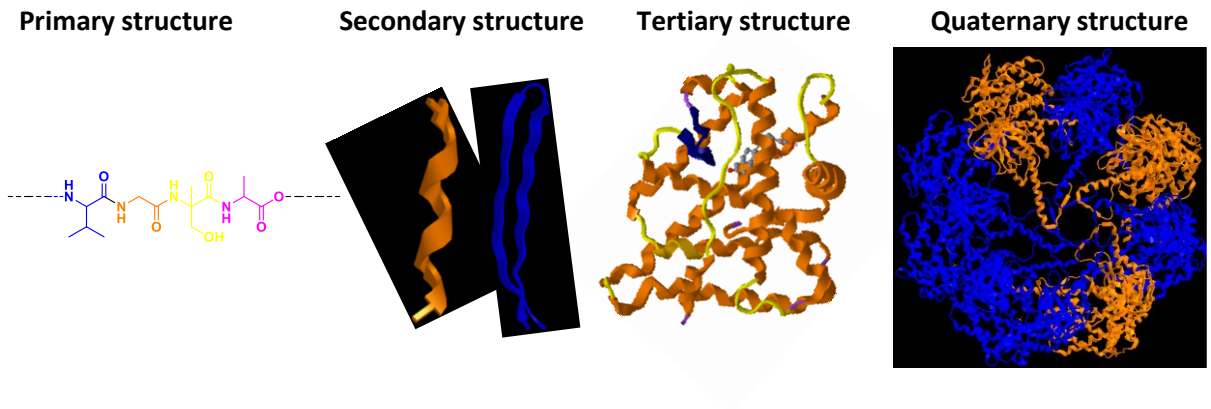


Figure I.4: Representation of primary, secondary, tertiary and quaternary structures of a protein.

As said before, proteins are involved in many biochemical processes in all living organisms. Moreover, some polypeptides show pharmacological activity, and in many cases they are drug targets. For all these reasons, scientists and in particular biochemists, were deeply interested in developing methodologies for the chemical synthesis of proteins. However, before 1963, because of complicated purification processes, polypeptide synthesis was only possible for small molecules.[31] Bruce Merrifield developed a revolutionary approach for polypeptide synthesis based on solid state iterative chemistry,[32] simplifying and making convenient the synthesis of long sequences. Nowadays, solid-state assisted techniques are the standard method for the synthesis of polypeptide. This approach and other solid-state methodologies applied for the production of different classes of sequence-defined polymers is discussed in the next sections.

2.2 Synthetic polymers

Nowadays, synthetic polymers are produced using three main techniques depicted in Figure I.5: (i) step-growth polymerization, (ii) chain-growth polymerization, (iii) multi-step growth polymerization. The development of new polymerization techniques as living polymerization [33] and controlled radical polymerization[34-36] improved the control on the primary structure. Indeed, these techniques enabled the synthesis of sequence-controlled polymers as block or gradient copolymers containing a narrow molecular weight distribution. The development of multi-step growth polymerization *via* solid state chemistry[20, 31] enabled the synthesis of molecules containing fully defined primary structures and uniform chain length distributions called sequence-defined polymers. In this section the most relevant polymerization techniques are discussed highlighting their strengths and weaknesses.

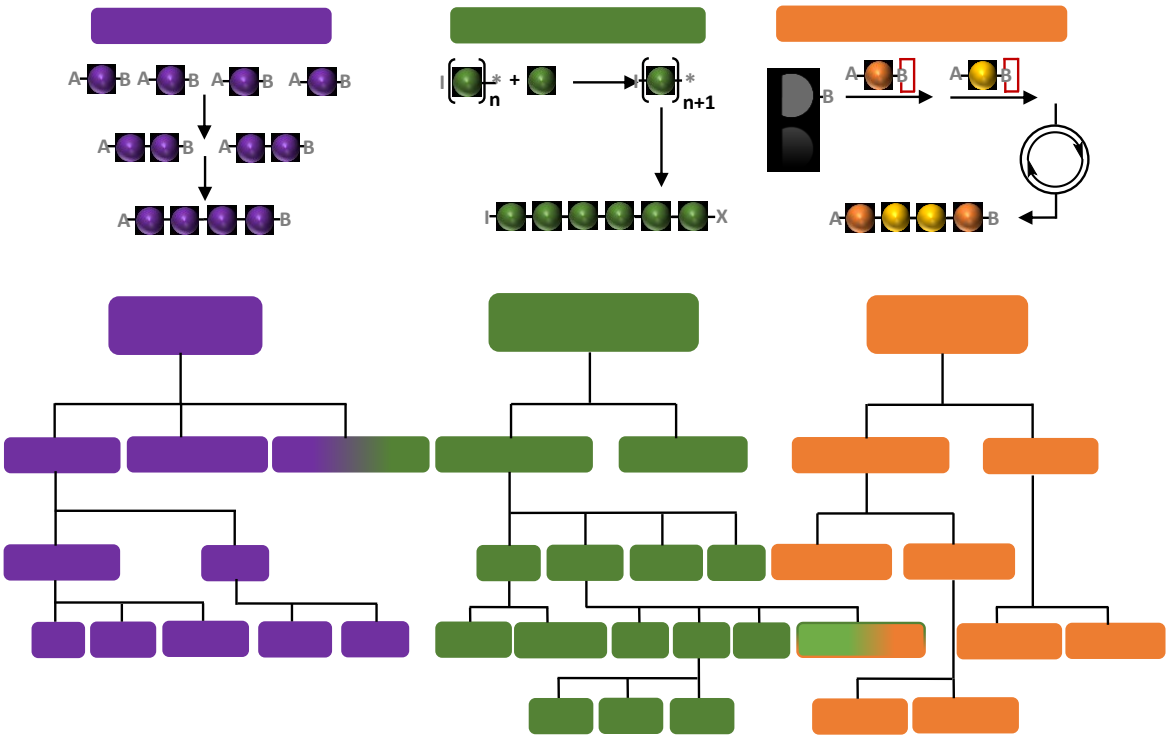


Figure I.5: General classification of the polymerization techniques.

2.2.1 Step-growth polymerization

Step-growth polymerization employs bifunctional monomers.[37] As depicted in Figure I.5, after the first reaction the dimers may react with another dimer or a monomer forming respectively, a tetramer or a trimer. The process continues until the formation of high molecular weight polymers. The pioneer in the field of step-growth polymerization is Carothers who first studied these reactions.[38, 39] Step-growth polymerization is largely applied in industry for the synthesis of polyesters, polyurethanes and polyamides. However, the low control over the primary structure and the broad weight distribution obtained employing standard step-growth methodologies make this technique not appealing for the

synthesis of sequence-defined polymers. However, in the last years, new toolboxes improving the control over the primary structure *via* step-growth polymerization were developed. For instance ADMET,[40, 41] CuAAC[42] or the metal-catalyzed step-growth radical polymerization[43] allow the synthesis of periodic copolymers. ADMET polymerization, described by Wagener in 1991,[41] relies on the olefin metathesis mechanism and employs α,ω-dienes as monomers. Usually, a side chain group is present in the monomer backbone, after the polymerization and hydrogenation, uniform side chain distribution is obtained. This technique was employed for the synthesis of ionomers,[44] block copolymers, soft materials[45] and for many other applications.[46] Step growth polymerization *via* CuAAC is a powerful tool enabling the synthesis of high molecular weight polymers in a relatively short time. Several syntheses of block copolymers have been reported in the past years using this technique.[47] The general procedure for step-growth CuAAC polymerization employs building blocks functionalized with an alkyne and an azide at the two chain ends, called α-alkyne, ω-azido heterotelechelic polymers. After the building block synthesis, the CuAAC reaction enables the production of high molecular weight block copolymers in which the different blocks are connected *via* 1,2,3-triazole linkage. Lutz for example, applied this methodology for the synthesis of polymers containing periodic functional motif.[42] In this easy concept depicted in Figure I.6, different maleimides bearing functional units were added at precise positions of polystyrene chains. The following building block modification into heterotelechelic polymers made possible the CuAAC reactions between the different units, enabling the synthesis of high molecular weight polymers containing periodical functional monomers.

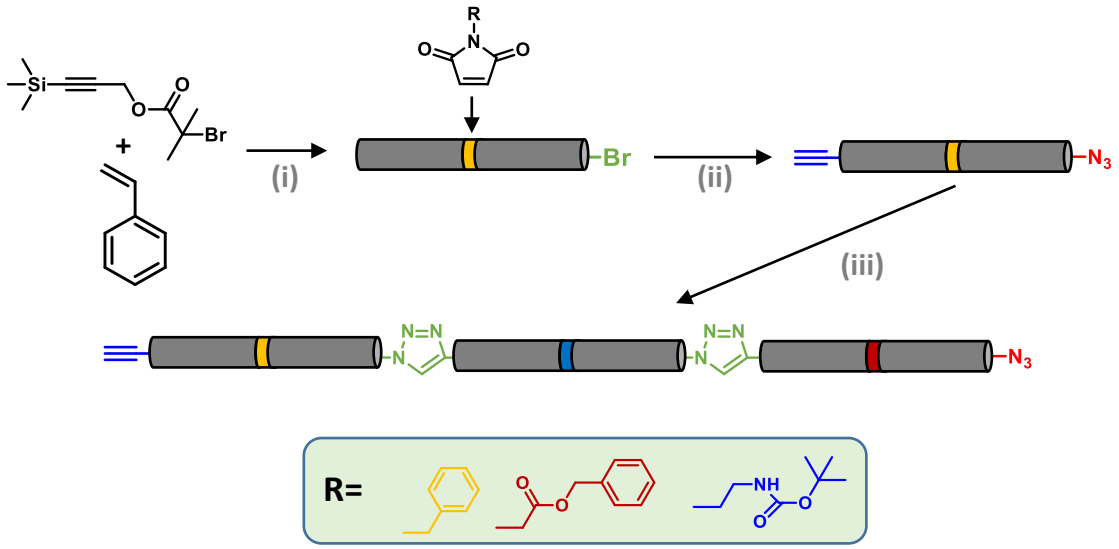


Figure I.6: CuAAC step-growth polymerization. **(i)** maleimide single insertion: CuBr, PMDETA, 110°C; **(ii)** deprotection and azide formation: various possible pathways; **(iii)** CuAAC: CuBr, PMDETA, THF, rt.

In 2010, Kamigaito described the synthesis of ABCC-sequence-regulated copolymers *via* metal-catalyzed step-growth radical polymerization (Figure I.7).[43] In this work, the ABC or

ABCC building blocks were polymerized using different metal catalyst *via* step-growth radical polymerization. These reaction proceeds through the formation of a C-C bond between the radical species generated via metal catalyst activation of the C-Cl bond and the unconjugated C=C bond of another molecule (Figure I.7). Even though this technique allows the synthesis of block copolymers, the main drawback is the poor control on the weight distribution.

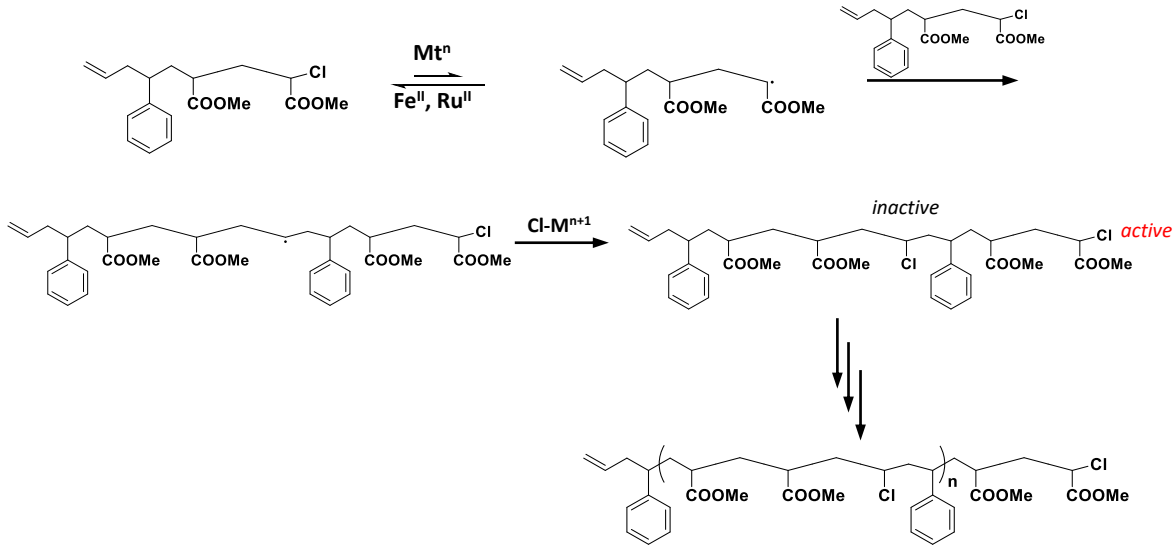


Figure I.7: Synthesis of ABCC sequence regulated vinyl copolymer *via* metal-catalyzed step-growth radical polymerization.

Recently, it has been demonstrated that the synthesis of high molecular weight block copolymers can be achieved *via* radical addition-coupling polymerization (RACP).[48] This approach relies on the alternating reaction of carbon-centred radical with N-O bond of a nitroso compound. The following radical-radical coupling between the nitroxide, generated in the previous step and a second carbon-centred radical, allows the synthesis alternating and periodic copolymers (Figure I.8).[49]

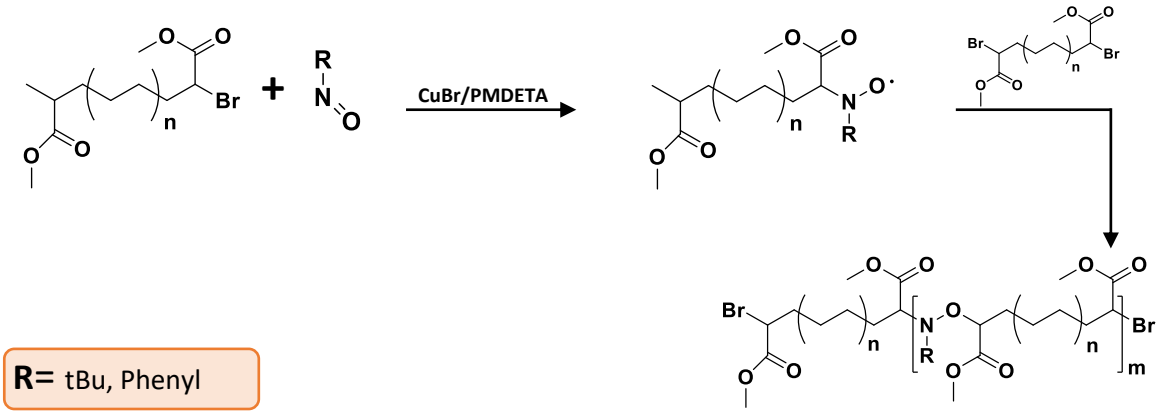


Figure I.8: RACP general strategy for the polymer synthesis.

Relying on this strategy, the synthesis of periodic copolymers containing [ABAC], [ABCD], and [ABCDBCAD] repeating units was described.[49] One more recent step-growth polymerization technique allowing uniform molecular weight distribution is living step-growth polymerization. In principle, the transition from step-growth to living step-growth is achieved making the polymer end group more reactive than the monomer; this condition can be achieved in two manners. For a model polymer X-AB-Y for instance, the reactivity of Y end group can be increased *via* monomer design or *via* catalytic chain end activation. A more detailed explanation of the process is described in a recent review.[50]

2.2.2 Chain-growth polymerization

In chain-growth polymerization, the monomers react one by one with a growing chain containing one or more reactive centres. The monomers can thus react *via* anionic, cationic, radical or ring opening mechanisms, as depicted in Figure I.5. The primary structure of the resulting polymers relies on the kinetics of propagation and on the reactivity of the monomers employed. In most of the cases, the comparable reactivity of the building blocks involved leads to statistical polymers with broad chain length distribution. The development of living polymerization was found to be extremely useful to improve the primary structure control. The first example of living polymerization is the synthesis of a styrene isoprene block copolymer *via* living anionic polymerization described by Szwarc.[33] The peculiarity of this reaction is the absence of termination step; hence the chains are always capable of receiving monomers. Moreover, all the chains are initiated simultaneously and the resulting polymers contain a narrow weight distribution. Another milestone in the sequence regulation of synthetic polymers is the development of controlled radical polymerization (CRP). These reactions are chain-growth polymerizations that undergo without chain transfer and termination events. Hence, the creation of an equilibrium between active and dormant species during the polymerization enables the simultaneous growing chains propagation.[8, 51] These features allow the synthesis of sequence-controlled polymers containing narrow chain length distributions. One of the first examples of controlled radical polymerization is the synthesis of gradient copolymers described by Matyjaszewski.[52] Compare to living polymerization, the tolerance to different functional groups and the easy reaction procedures make CRP a robust tool for synthesis of polymers containing controlled molecular architectures. Among the controlled radical polymerization methods, NMP,[53] ATRP[34, 35] and RAFT[36] are the most employed techniques.[54] During the last years, it has been demonstrated that the synthesis *via* living polymerization of longer multiblock polymers is possible [55, 56] and promising for the design of soft materials[57]. Haddleton for instance, described the synthesis of multiblock polymers *via* Cu-mediated photopolymerization.[56] In this approach, a dodecablock polymer containing four methacrylate repeating units with dispersity lower than 1.2 was synthesized trough the consecutive additions of monomers, $CuBr_2$, ME_6TREN and irradiation at 360 nm. Perrier also described the synthesis of multiblock copolymers containing precisely defined structures *via* RAFT.[55] The synthesis of such precise copolymers relies on a radical polymerization which

proceeds by degenerative transfer. Exploiting this strategy the synthesis of a dodecablock copolymer with a degree of polymerization (DP) of 10 for each block and an icosablock copolymer (20 blocks) were described. In 2015, a study concerning the precision of the monomer insertion in two ABACADACAB decablock copolymers, synthesized using this technique was performed by Perrier and Harrisson. Interestingly, they found out that the uncertainty of the position of a monomer in the chain depends on the polymer length and on the length of the blocks [58] (Figure I.9).

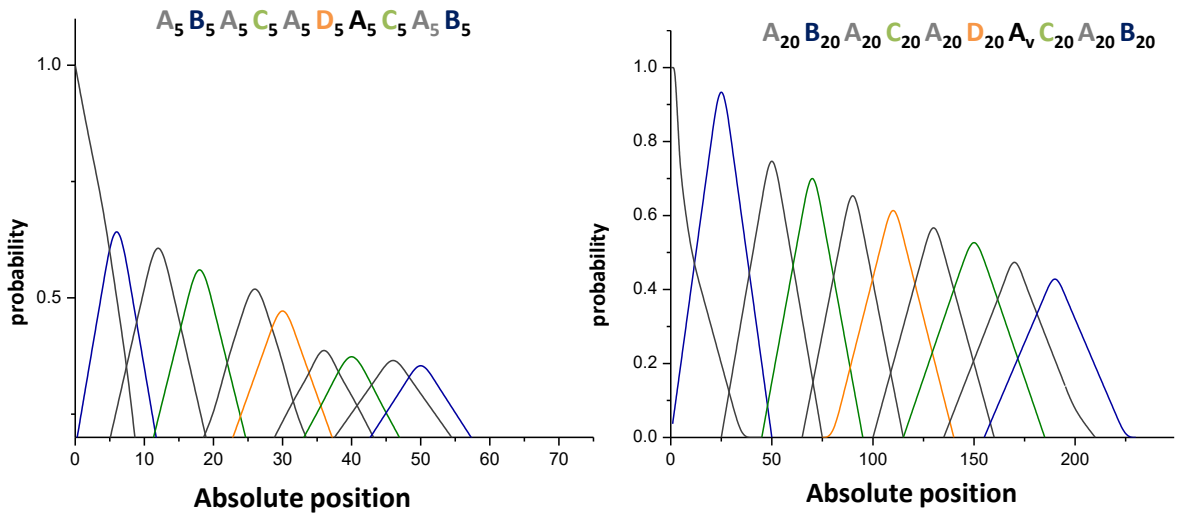


Figure I.9: Probability of finding monomers from each block of of an $A_5B_5A_5C_5A_5D_5A_5C_5A_5B_5$ (left) and $A_{20}B_{20}A_{20}C_{20}A_{20}D_{20}A_{20}C_{20}A_{20}B_{20}$ (right). In both schemes the probability is in function of the absolute position.

Recently, Sawamoto and co-workers described two methodologies for the selective monomer incorporation into polymer chains. The incorporation can be achieved *via a* template-assisted mechanism[59] or using bulky side chain groups.[60] In the latter example, the bulky side group restraining the propagation process allows single monomer insertion in vinyl polymer segments. As depicted in Figure I.10, the protocol relies on three chemical steps: radical addition, cleavage and functionalization. This approach was found to be very promising for the synthesis of block copolymer carrying sequence-controlled segments.

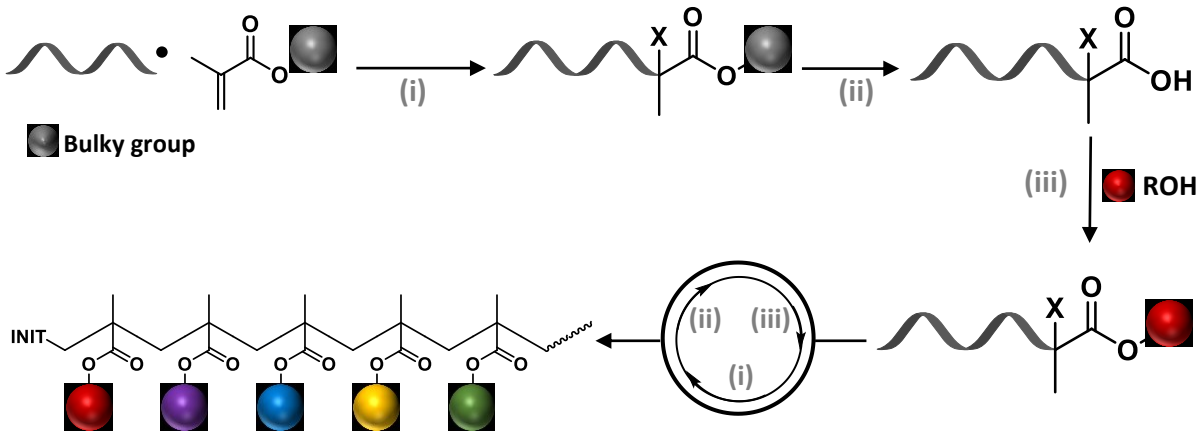


Figure I.10: Template-assisted single insertion. **(i):** radical addition: [[Cp*Ru(μ_3-Cl)]$_4$], BPMO, IPAMA, tetralin, DCPA, toluene, 4h, 100˚C; **(ii)** acidolysis: TFA. 0˚C .30 min; **(iii)** esterification: alcohol, DMAP, DIC, DCM, MeOH, 0˚C, 1h.

Hawker and Russell described in 2000 a methodology based on living polymerization of polystyrene and maleic anhydride for the synthesis of block polymers composed of a heterogeneous mixture of homopolystyrene and various narrow regions of styrene/maleic anhydride copolymers.[61] Inspired by this concept, Lutz et al. developed a protocol for the precise insertion of different N-substituted maleimides in the polystyrene chain *via* atom transfer radical polymerization (ATRP).[62] In this approach, small amounts of acceptor, usually different N-substituted maleimides, copolymerize in polystyrene excess. The maleimide insertion in narrow regions of the polystyrene chain relies on the high consumption rate of these molecules

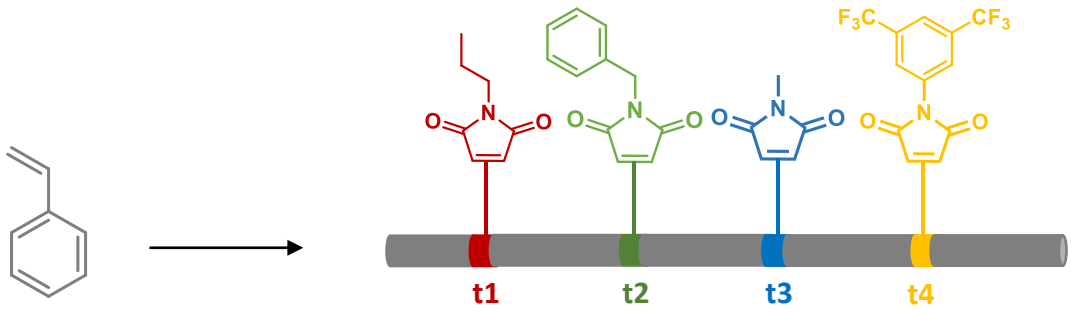


Figure I.11: General concept of ultra-precise maleimide insertion.

The same group, relying on this principle, described the synthesis of a single sugar chain array *via* CuAAC post-polymerization.[63] In particular, three different protected maleimides were inserted into narrow regions of the polystyrene backbone and after selective deprotection, each of the resulting alkynes reacts *via* CuAAC with different azide-functionalized sugars, allowing the synthesis of high molecular weight glycopolymers.

2.2.3 Multi-step growth polymerization

Even though, several improvements toward sequence regulation *via* controlled radical polymerization were achieved in the past decades, uniform polymers having a fully defined primary structures can still not be obtained using this methodology.[8] Indeed, only syntheses of sequence-defined trimers and tetramers are reported so far using the chain-growth polymerization technique. Multi-step growth polymerization is a powerful tool for the synthesis of sequence-defined polymers. The technique relies on the iterative monomer addition to a growing chain and it allows to obtain fully defined primary structures and uniform chain lengths distribution.[64] Furthermore, this approach offers the opportunity to perform the reactions in liquid [65-67] or solid phase.[31, 68] All these features made the technique the most common methodology employed for the synthesis of sequence-defined polymers. The solid-state synthesis concept relies on insoluble supports connected to the growing chain *via* a cleavable bond. The use of these matrices allows to simplify the purification steps, which is done by filtration. As depicted in Figure I.12, standard solid-support assisted protocol is composed of two steps: (i) a reaction between the growing chain and the desired monomer, (ii) a filtration of the unreacted monomers. Moreover, to remove the polymer chain from the solid-support an additional cleavage step is done.

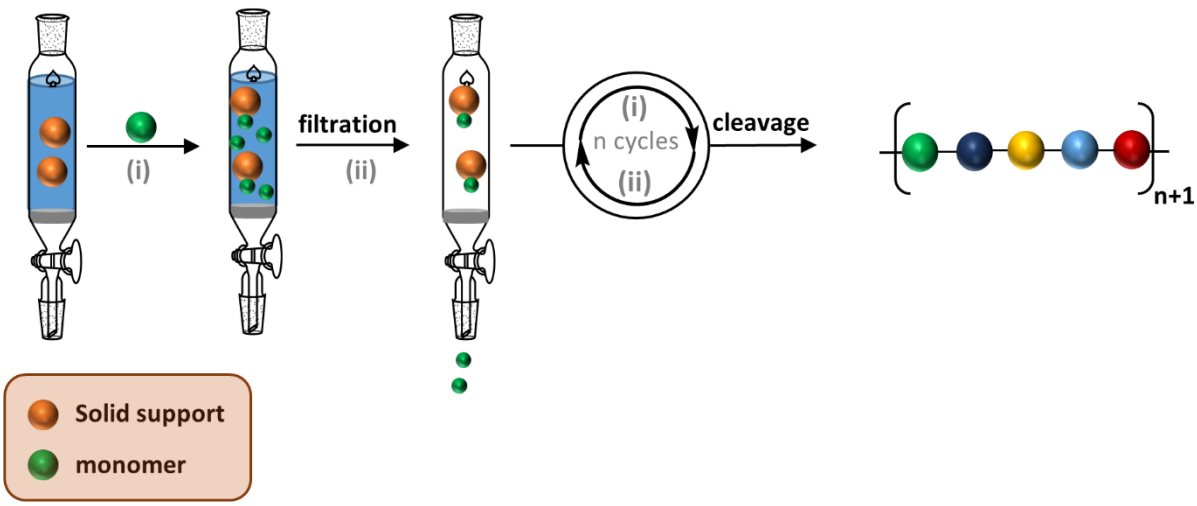


Figure I.12: Solid phase-assisted synthesis concept.

As said in section 2.1.2, this technique was firstly described for protein synthesis in 1963.[32] At that time, polypeptide synthesis in solution was extremely complicated and time consuming because of growing chain solubility issues and complicated purification processes required after each coupling.[69] The solid-state protocol developed by Merrifield incredibly simplified the synthesis of these molecules,[31] making the synthesis of long polypeptides in a short time possible. The Merrifield protocol is a two-step protocol, at each cycle an amino acid is covalently attached to the polymer chain. As depicted in Figure I.13, the first step (i) is an amidification coupling between the carboxylic moiety of N-protected amino acid and the amino group on the growing chain. The second step (ii) is the deprotection of the N-

protected terminal residue attached during the previous reaction. The final step (iii) is the chain cleavage from the solid support. In 1965, the protocol automation and the following replacement of Cbz with Boc[68] made the polypeptide synthesis more efficient. Chang and Meienhofer further optimized the protocol, employing Fmoc as protecting group and making chemoselective the deprotection step.[70] However, performing reactions on a solid support decreases the reaction rate. Indeed, a monomers excess is required to push the reaction to completion. In 1986, was discovered that microwaves could reduce the reaction time for organic reactions in solution.[71, 72] Inspired by this concept, Wang applied this methodology to solid support-assisted synthesis,[73] reducing the reaction time and making the synthesis of higher molecular weight polymers faster and easier.[74]

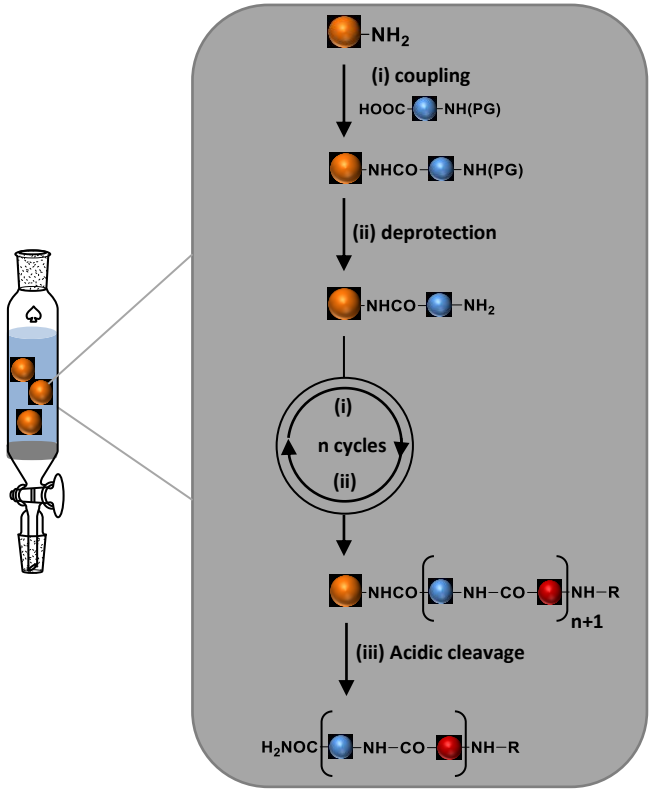


Figure I.13: Standard protocol for the polypeptide synthesis on solid support.

During the last 50 years, many optimizations to expand the synthetic possibilities and increase the efficiency were achieved. Usually, during the design of a solid-state protocol, several parameters need to be taken into account like the nature of the reactions involved, the solid support stability in the cleavage conditions, etc. Nowadays, a huge number of supports and reaction sets are available allowing a potentially endless number of combinations. The reactions involved in a multi-step protocol need to be fast and quantitative.[31] Indeed, the high reaction rate is a crucial parameter because it enables the access to high molecular weight molecules in a short time while the coupling efficiency determines the uniform chain length distribution. Solid supports are formed by two elements: the matrix and the linker. Matrices are usually formed by different kinds of

polystyrene cross-linked with divinylbenzene in different percentages.[75] The reactions are carried out in low-polar solvents in order to maintain the matrix swollen and guarantee the accessibility of the growing chains. Other polystyrene-based matrices like Tentagel and Agrogel were found to be particularly useful to perform reactions in polar solvents. The linker connects the matrix to the anchoring point, which is involved in the reaction with the first monomer. The stability in the reaction conditions, orthogonality and high cleavage yields are the features required for this moiety.[75] According to the cleavage conditions, the solid supports can be classified in three categories: acid-cleavable, base-cleavable and photocleavable supports.

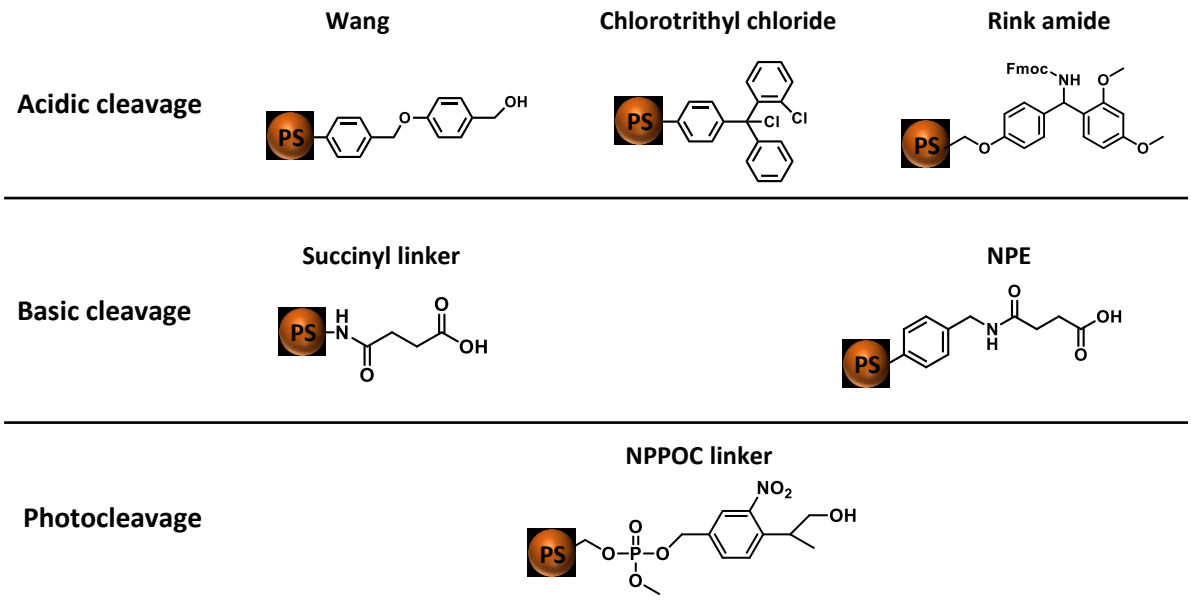


Table I.1: Resins classification of according to the cleavage conditions.

Nowadays, solid-state assisted protocols are applied for the synthesis of a wide range of molecules such as synthetic polymers, [5, 76, 77] oligonucleotides,[78-81] bioconjugates[82, 83] and many others. However, the drawbacks of this technique are the slower reaction rate compared to the liquid phase reactions and the difficulties to scale up. Liquid-state multi-step polymerization enables the synthesis of big amounts and high molecular weight sequence-defined or periodic polymers in a relatively short time. Indeed, performing the reactions in solution may speed the reaction rate up. However, long purification procedures as precipitation or chromatographic columns are required after each coupling. The Iterative exponential protocol is a liquid phase approach for the synthesis of sequence-controlled polymers.[65] In this strategy, a bifunctional XY monomer protected in X reacts with a bifunctional XY monomer protected in Y. After the first step, the resulting dimers are split in two parts, deprotected on one side and reacted again. The big limitation of this technique is the time-consuming purification step required after each coupling. Furthermore, primary structure control is limited to periodic polymers. However, the possibility to access high

molecular weights polymers in relatively few steps makes this technique appealing in particular for industrial applications. The iterative synthesis on soluble supports is a technique which allows to obtain sequence-controlled polymers in multi-gram scale.[67] The soluble supports employed are usually PEG or linear polystyrene-based polymers exposing the anchoring point. Standard multi-step growth polymerization on soluble support is performed in two steps: coupling and purification. The first step involves the reaction between the growing chain anchored to the support and the monomer. This reaction is performed in a solvent in which the support is soluble. The purification step is done by soluble support precipitation in a non-solvent and usually requires several hours. As for other support-assisted synthesis, cleavage is the last step allowing to separate the polymers from the support. Soluble supports are largely employed for the synthesis of several non-natural[84, 85] and natural sequence-defined polymers such as polypeptides,[86-88] oligonucleotides,[89] bioconjugates[90] and polysaccharides.[67, 91] Several approaches for the synthesis of sequence-defined polymers using multi-step growth polymerization were developed during the past years; the different strategies are deeply discussed in the next section.

3 Synthetic strategies for the synthesis of sequence defined polymers

3.1 Introduction

As described in the previous section, the multi-step iterative protocol is the standard procedure for the synthesis of sequence-defined polymers. In the last decades, several protecting group and orthogonal strategies leading to the synthesis of diverse classes of natural and non-natural sequence-defined polymers[92] such as oligonucleotides, oligosaccharides, bioconjugates etc. were developed. In this chapter, these strategies are described highlighting the pros and cons for each one.

3.2 Protecting group approach (AB+AB)

The (AB+AB) approach was the first strategy applied for the polypeptide synthesis *via* an iterative solid state protocol.[32] As depicted in Figure I.14, the monomers employed contain two reactive groups. One of those is protected in order to avoid oligomerization *via* polycondensation mechanisms. Moreover, cleavage from the solid support is usually orthogonal to the protecting group deprotection.

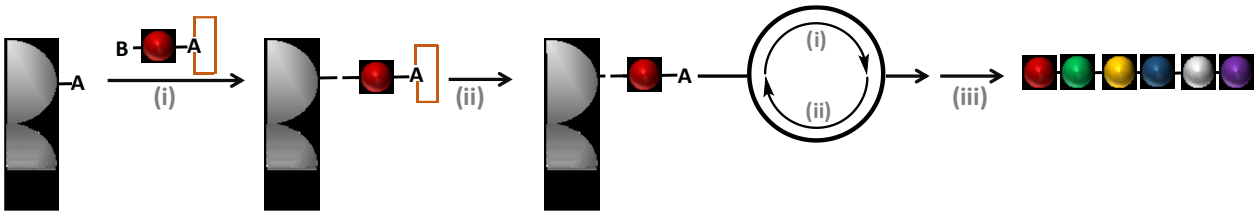


Figure I.14: General strategy for (AB+AB) approach.

The (AB+AB) strategy was also applied for the synthesis of different classes of natural and non-natural macromolecules. Letsinger and Mahadevan, for instance, introduced the solid-support assisted protocol for oligonucleotide synthesis in 1965. [78] In 1981 Beaucage and Caruthers, exploiting the better stability of the phosphoramidite nucleosides improved the protocol efficiency. [80] Nowadays, the synthesis of oligonucleotides containing 200 units is possible on solid support in a relatively short time.[76] Recently, Häner highlighted the possibility to produce responsive two dimensional polymers using sequence-defined self-assembled DNA networks (Figure I.15).[93] In this work, a building block containing a benzene 1,3,5 functionalized with three self-complementary oligonucleotides was synthesized using the phosphoramidite protocol. The oligonucleotides contain an anthracene placed in the middle of each sequence. The building block, at high concentration, leads to the formation of a 2D supramolecular network. In this system, the anthracene moieties are close enough to react *via* UV-induced dimerization which leads to the formation of a 2D polymer.

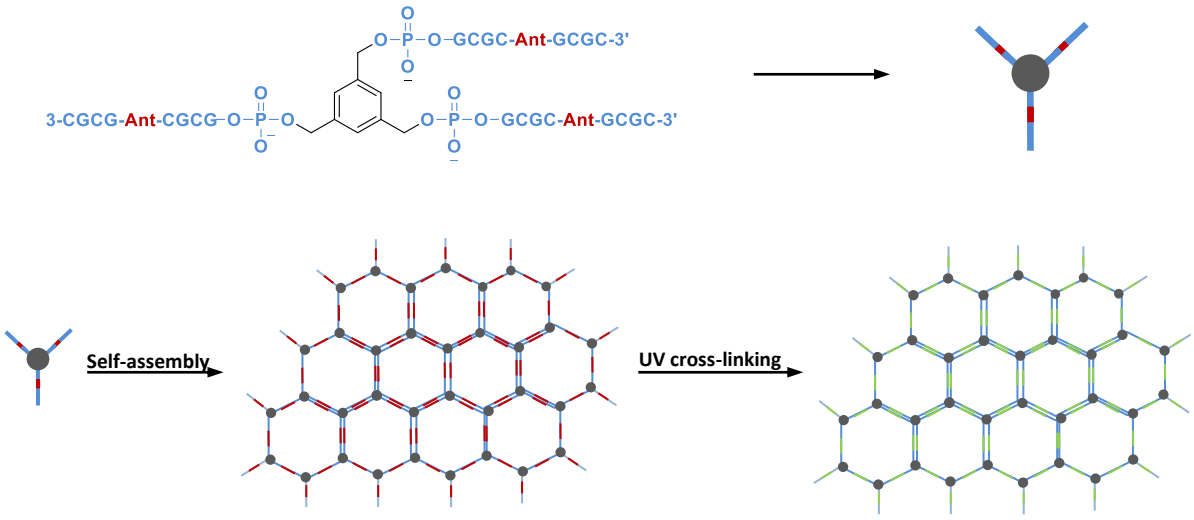


Figure I.15: General strategy for the formation of responsive 2D polymers.

In 2014, Sleiman developed a methodology for the synthesis of sequence-defined polymer-oligonucleotide hybrids.[94] As depicted in Figure I.16, the synthesis of a short single strand DNA was done using the automated phosphoramidite protocol. Afterwards, the bioconjugate was obtained adding DMT protected monomers to the 5' alcohol of the oligonucleotide segment. Dodecanediol phosphoramidite and hexaethylene glycol phosphoramidite were the monomers involved in these iterative additions. Interestingly, it

was found that it is possible to tune the hydrophobicity, the self-assembly and the guest encapsulation of the hybrids regulating the monomers insertion in the backbone.

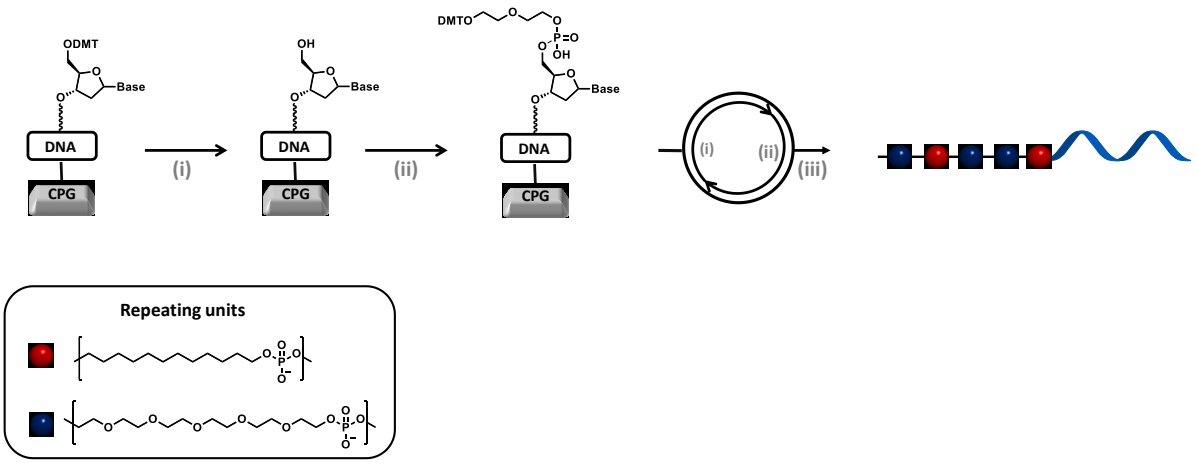


Figure I.16: General strategy or the synthesis of sequence-defined polymers oligonucleotide hybrids. **(i)** DMT deprotection: H^+; **(ii)** monomer addition and oxidation: tetrazole, monomer, CH_3CN, I_2, water, 2-6 lutidine; **(iii)** cleavage: OH^-.

The phosphoramidite chemistry was also employed for the synthesis of non-natural sequence-defined polymers for data storage applications.[95] The iterative insertion of two different phosphoramidite monomers, intentionally defined as 0 and 1, allows the synthesis of non-natural sequence defined poly(phosphodiester)s encoding information. This class is deeply discussed in the digital polymers section. The (AB+AB) strategy was also employed for the synthesis of non-natural sequence-defined polymers. One of the first example is the synthesis of sequence-defined oligocarbamates described by Shultz in 1993[96] (Figure I.17). The different N-protected p-nitrophenyl carbonate monomers react with the amino group of the growing chain *via* nucleophilic substitution at the carbonyl group, producing a carbamate function. The following protecting group removal releases the amine group involved in the next coupling.

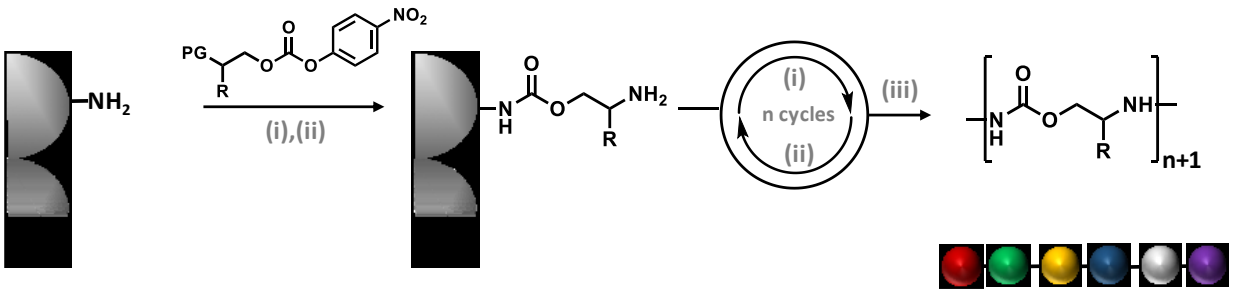


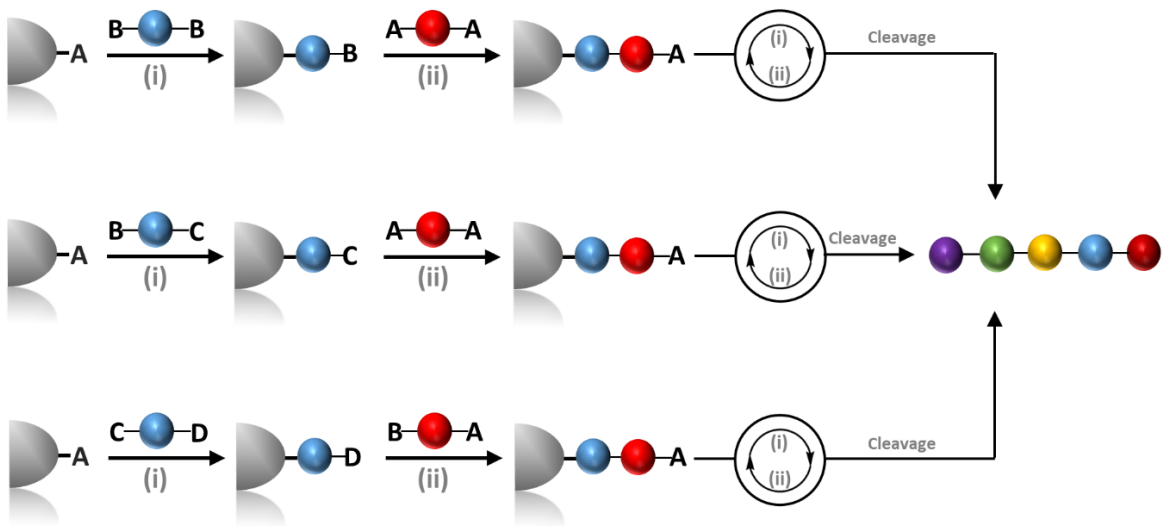
Figure I.17: Strategy for the synthesis of sequence-defined oligocarbamates. **(i)** coupling: Fmoc, N-protected monomer, DIPEA, HOBt, NMP, 4h; **(ii)** Fmoc deprotection: 20% piperidine, NMP; **(iii)** cleavage: H^+.

Nowadays, several unnatural sequence-defined polymers are synthesized using protecting group approaches, for example polycarbamates,[96, 97] oligoureas,[98] polyamines,[99] etc. Angelo and Arora described in 2005 the synthesis of non-peptidic foldamers based on 1,3-substituted triazole structures.[100] The protocol relies on three steps. In the first step, the

amine group of different amino acids is transformed into an azide. The second step is the CuAAC coupling of the azide with N-protected amino alkyne. The final step is the amine deprotection. The resulting macromolecules fold with a motif reminding the beta-strands. Recently Barner-Kowollik described an iterative synthesis *via* photoligation[101] in which different building blocks, bearing photoreactive benzaldehyde and a furan protected maleimide are involved in photochemically induced Diels-Alder reactions, allowing the production of high molecular weight sequence-defined polymers in few steps.

3.3 Iterative synthesis using latent reactive groups

The protecting group strategy was the first methodology applied for sequence-defined polymers synthesis. However, the use of protected monomers leads to a mandatory deprotection step after each cycle, this additional step is usually time consuming and inconvenient in particular for the synthesis of long sequences. As depicted in Scheme I.1, several strategies were developed to avoid the use of protected monomers.



Scheme I.1: Protecting groups free approaches developed for the synthesis of sequence-defined polymers.

A valid alternative to avoid the deprotection steps is represented by the use of latent reactive groups. This approach relies on the activation *via* chemical or physical stimuli of "dormant" species contained in the sequence. One example is the oligothiophene synthesis described in 1998 by Fréchet.[102] Polythiophenes are conductive materials and are largely studied for solar cell applications.[103] As depicted in Figure I.18, different building blocks are linked *via* Stille reaction between bromine and a trimethyl-stannyl group. The dormant group is activated *via* a α tiophene bromination (ii), allowing the successive Stille coupling.

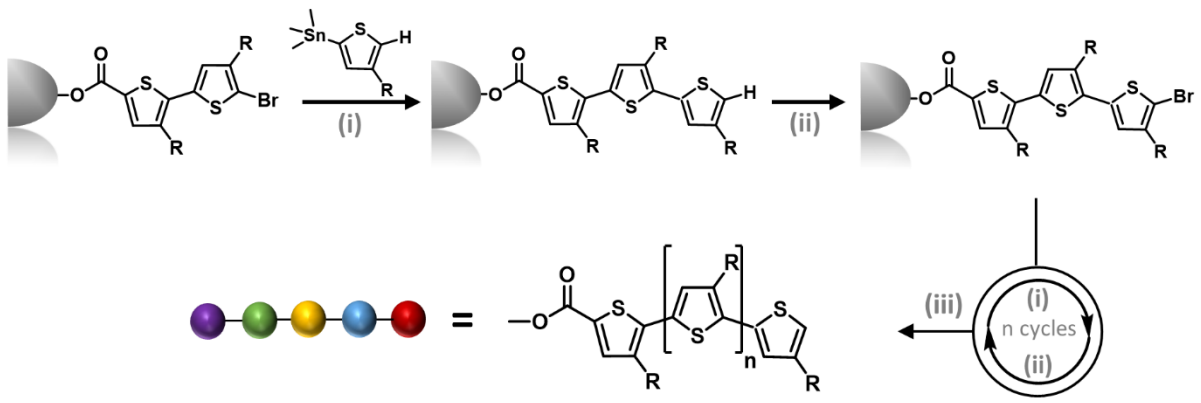


Figure I.18: General strategy for sequence-defined oligothiophenes. **(i)** stille coupling: $Pd(PPh_3)Cl_2$, DMF, 80°C; **(ii)** bromination: NBS, DMF; **(iii)** cleavage: NaOMe, THF, reflux, MeI, 18°C, 6h, reflux.

The synthesis of ether oligomers described by Reymond in 2003 is another example of iterative synthesis which employs latent groups.[104] In this approach, an ether linkage is formed *via* Mitsunobu coupling between a benzyl alcohol and hydroxy aromatic aldehydes or ketones. The latent group activation consists in the carboxylate reduction into a benzylic alcohol moiety which is involved in the following step. This approach was employed for the synthesis of a library of different uniform macromolecules showing inhibitory activity for trypsin, chymotrypsin, and subtilisin. In 2016, Lutz *et al*. described the synthesis of sequence-defined oligocarbamates using a latent reactive group approach on a solid support. Interestingly, this class of non-natural sequence-defined oligomers was found to be easy to sequence by MS/MS. For this reason, they found application as molecular barcodes.[105, 106]

3.4 (AA+BB) approach

The (AA+BB) approach employs two homobifunctional monomers in large excess to guarantee that only one of the two functional groups reacts with the growing chain. The big disadvantage is the low atom economy mainly due to the massive monomer excess needed for each coupling. One of the first (AB+AB) approaches was described by Rose and Vizzavona[107] using a diacid and a diamine. The protocol consists in an amidification reaction between an amino group and a succinic anhydride, carboxylic acid activation with carbonyldiimidazole and following coupling with an excess of different diamines. The iterative repetitions of this cycle enables the synthesis of uniform sequence-defined polyamides. Börner employed the same protocol, replacing the PEG-based diamines with spermine. The resulting sequence-defined polymers were successively linked to a polypeptide or to a PEO chain enabling the synthesis of bioconjugates showing peptidomimetic activity.[108] Another example is the triazine-based polymers (Figure I.19) in which different cyanuric chlorides derivatives react with diamines to afford uniform rigid structures macromolecules.[109] Moreover, regulating the primary structure and thus the hydrogen bonding is possible tuning the backbone-backbone interactions.

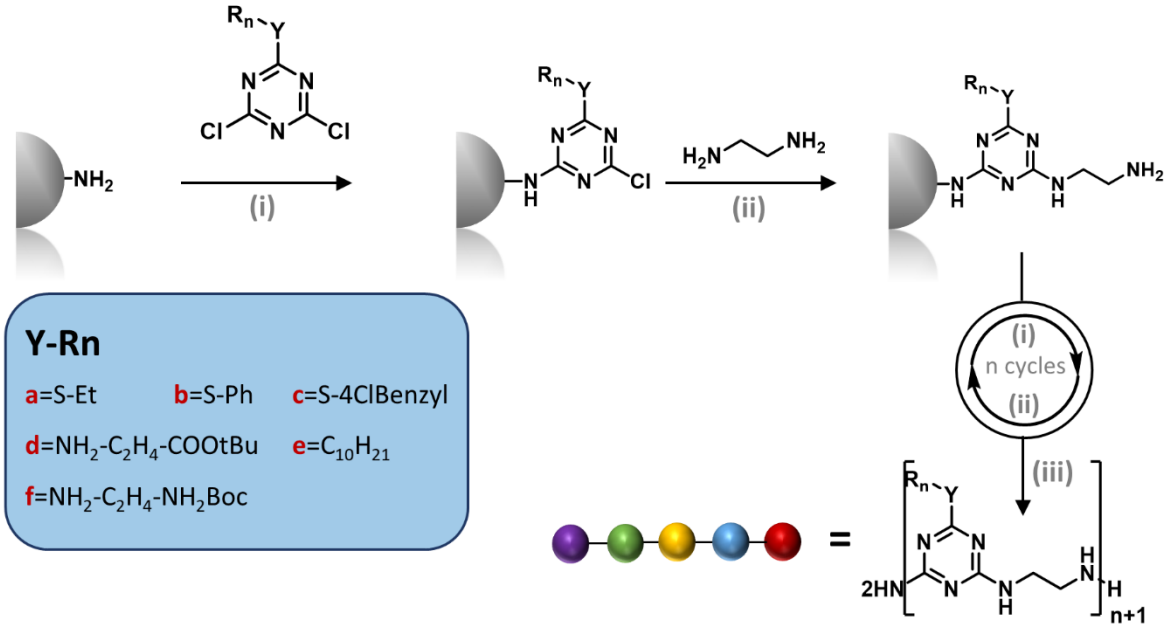


Figure I.19: (AB+AB) solid support-assisted synthesis of triazine-based polymers. **(i)** triazine coupling**:** DIPEA, THF 35°C; **(ii)** diamine coupling: DIPEA, NMP 80°C; **(iii)** cleavage: TFA.

3.5 Chemoselective (AB+CD) approach

Another elegant strategy to bypass protecting group in solid-state-assisted multi-step synthesis is employing chemoselective strategies. This concept was coined by Torst in 1983: "*The ability to discriminate among the reactive sites is referred to as chemoselectivity*".[110] This feature can be obtained *via* two different mechanisms[111] (Figure I.20) or *via* a single mechanism. The latter case for instance, can involve the acidic deprotection of two acetal groups, but due to the graduated reactivity, increasing acidity fully removes one and then the other group.[112]

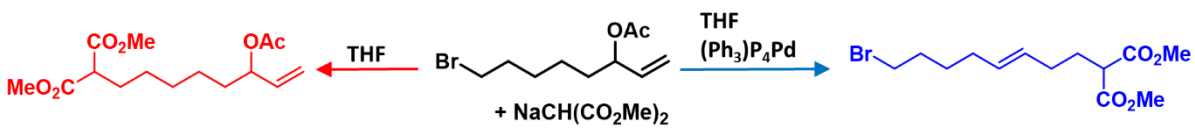


Figure I.20: Chemoselectivity concept: the different conditions employed lead to the insertion of the same functional group in different parts of the molecule *via* two different mechanisms.

As depicted in Scheme I.1, chemoselective (AB+CD) approaches employ two heterobifunctional monomers in which A can react only with C and B can react only with D, thus avoiding monomer polycondensation reactions. One of the biggest advantages to employ protecting-group-free strategies lies in the improved step economy.[15] The peptoid synthesis, reported by Zuckermann in 1992,[113] is the first example of chemoselective solid phase supported synthesis. Peptoids exhibit biological activity flexibility, higher proteolytic stability, compatibility and can fold in secondary and tertiary structures.[114] For this reason, these molecule were found to be interesting in material science and for pharmaceutical

applications.[115, 116] The strategy developed by Zuckermann relies on two "submonomers" which are involved in two steps (Figure I.21). The step (i) is the reaction between a secondary amine and bromoacetic acid. The resulting α-bromoacetamide reacts with a primary amine *via* nucleophilic displacement (ii), generating the secondary amine moiety involved in step (i). The cycle is iteratively repeated allowing the full control of the primary structure and uniform weight distribution.

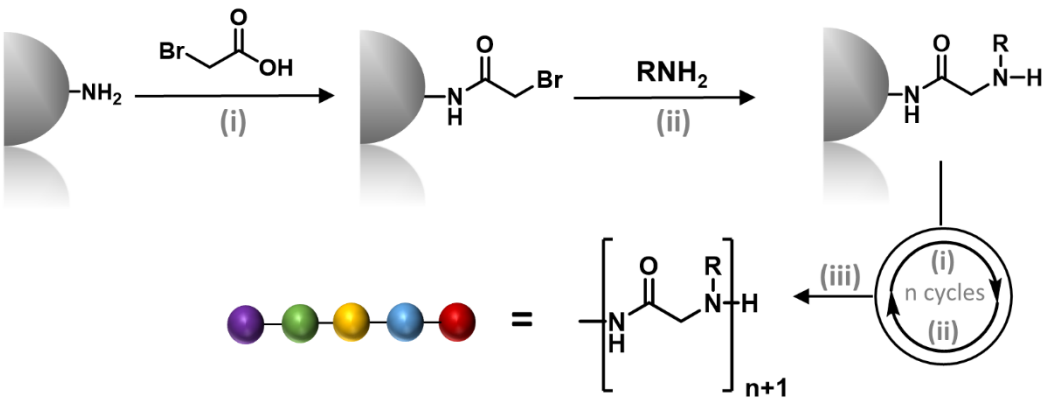


Figure I.21: Peptoids synthesis. **(i)** amidification: haloacetic acid, DCC, DMF; **(ii)** nucleophilic displacement: amine, DMF; **(iii)** cleavage: H^+.

The protocol was followingly automatized employing a polypeptide synthesizer.[117] Moreover, the use of a combinatorial chemistry approach[114, 118, 119] made the synthesis of long peptoid libraries[120] possible. In 2009 Lutz proposed an (AB+CD) strategy on solid and soluble support for the synthesis of sequence-defined poly(triazole amide)s.[84] In this approach the iterative repetitions of CuAAC and amidification reactions lead to the formation of uniform sequence-defined polymers. A more recent orthogonal approach is the thiolactone protocol developed by Du Prez based on aminolysis/chain extension steps.[121] As shown in Figure I.22 a, the aminolysis step (i) is performed using different amines. The following chain extension *via* thiol-ene reaction allows the reinstallation of the thiolactone moiety (ii). Nevertheless, the protocol still showed some big limitation mainly due to the formation of disulfide bridges during step (i). Indeed, the treatment with phosphines as reducing agents was required after this step. This treatment, however, led to the formation of side products, making the synthesis of long polymers impossible. In 2016, the protocol was improved[122] (Figure I.22 b) replacing the amine with an amino alcohol. The thiol generated during the first step, reacts with an acrylate or an acrylamide *via a* thiol-ene reaction. The chain extension (ii), performed between the hydroxy function and α-isocyanato-γ-thiolactone, allows the reinstallation of the thiolactone moiety. The milder conditions employed in the first step, avoid the disulfide bridge formation and improve the protocol efficiency. Moreover, the protocol was performed using a modified polypeptide synthesizer that enabled the parallel synthesis of decamers in a relatively short time.

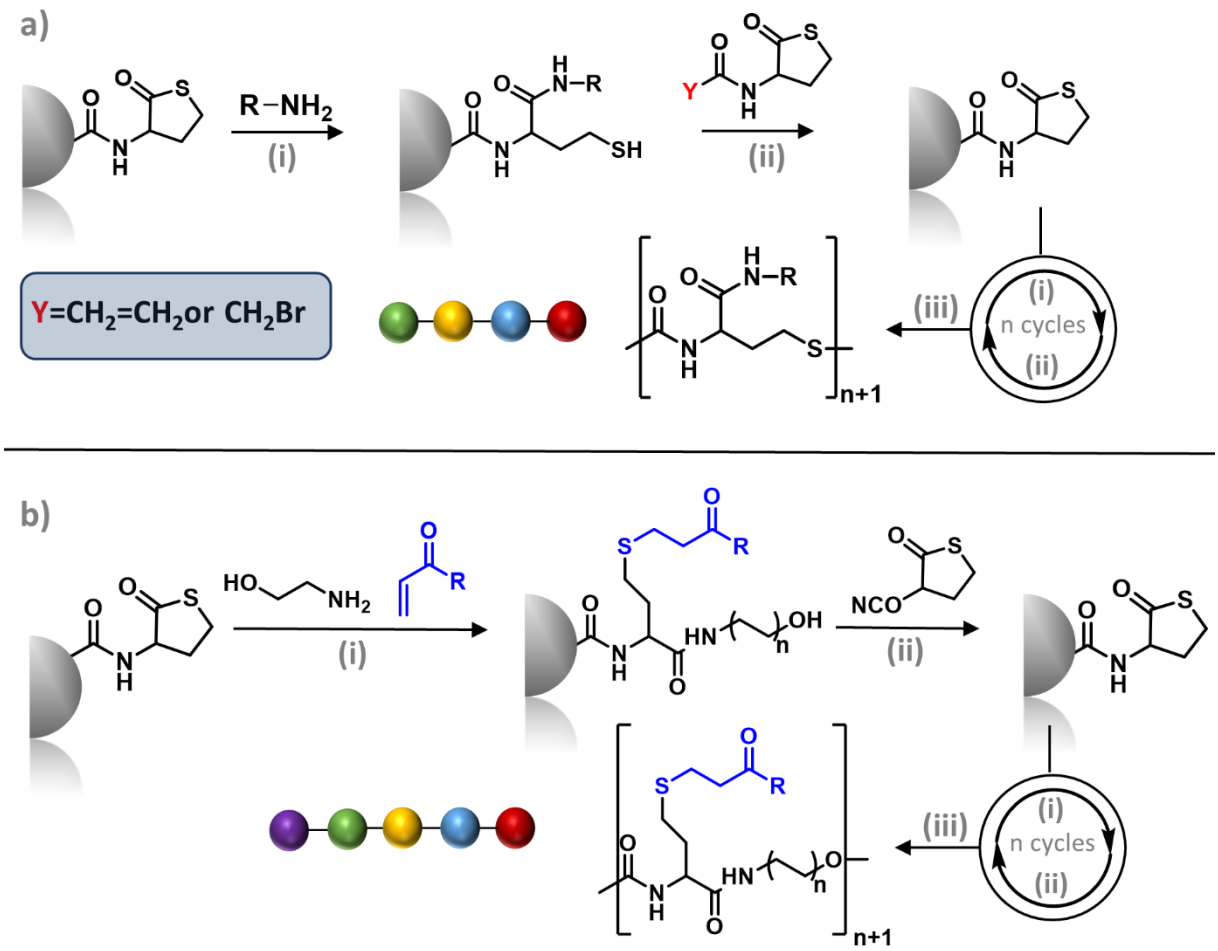


Figure I.22: Development of sequence-defined polymers based on thiolactone chemistry. **a)** First strategy employed for sequence-defined polymers based on thiolactone chemistry. **(i)** aminolysis: CH_2Cl_2; **(ii)** chain elongation: $PhMe_2P$, DMF; **(iii)** cleavage: TFA; **b)** Improved strategy for the synthesis of sequence-defined polymers based on thiolactone chemistry. **(i)** aminolysis, chain functionalization: $CHCl_3$, 15 min; **(ii)** chain extension: $CHCl_3$, DBTDL; **(iii)** cleavage: TFA.

Lutz developed in 2016 an (AB+CD) protocol for the synthesis of xeno nucleic acids. Because of their ability to hybridize with RNA and DNA single strands these molecules were found to be very interesting.[123] In this strategy, two monomers are employed: the first one bears a carboxylic acid and an azide function, the second one contains alkyne and amine moieties; both monomers are covalently linked to different nucleobases *via* an amide bond. The protocol relies on successive amidifications and CuAAC steps enabling the synthesis of uniform sequence-defined PTzNAs.[124] Other interesting examples of sequence defined polymers based on chemoselective (AB+CD) approach are poly(alkoxyamine amide)s[11] and oligo(triazole amide)s[125] which are further discussed in the digital polymers section.

3.6 (AA+BC) approach

This approach can be described as an intermediate strategy between (AA+BB) and (AB+CD). As depicted in Scheme I.1, the first step involves the chemoselective reaction between the functional group A on the growing chain and the B moiety of a BC monomer. The second step is the reaction between an excess of homobifunctional monomer AA and C. Alabi

employed this approach for the synthesis of sequence-defined synthetic polymers.[126] As depicted in Figure I.23, the iterative protocol involves two steps: the first one is the thiol-ene reaction between the dithiol excess and the allylic moiety of the hetero-bifunctional monomer. The second step is the phosphine-catalyzed Michael addition between the acrylamide moiety and the thiol function on the growing chain.

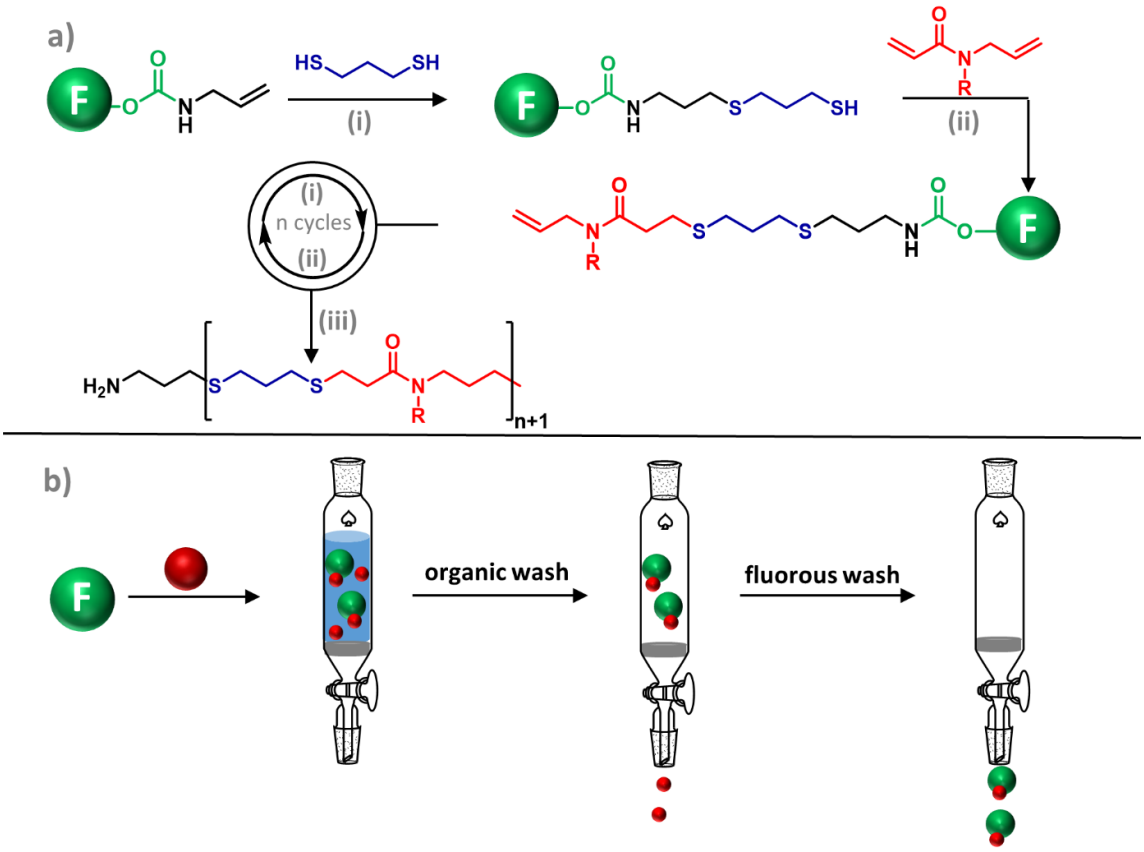


Figure I.23: General scheme for the synthesis of sequence-defined polymers *via* phosphine-catalyzed Michael and thiol-ene "click" addition using fluorous tags. **a) (i)** thiol-ene "click": DMPA, hv, 90s; **(ii)** phosphine-catalyzed Michael addition: Me_2PhP, 300s; **(iii)** cleavage: TFA; **b)** Fluorous solid phase extraction concept.

Interestingly, the use of fluorous tags as a liquid phase reaction support[127] enables the monomer addition in solution, decreasing the reaction time. Fluorous tags are removable perfluorocarbon alkyl chains, usually employed to obtain faster and easier purification processes. As depicted in Figure I.23 b, the reaction mixture is purified employing a fluorous silica solid phase column; the successive organic and fluorous washes allow the fast separation of the fluorous-tagged growing chain from the reaction mixture. In 2016, the same group described a methodology for the synthesis of sequence-defined macrocycles. After the macromolecule synthesis, an acid-catalyzed cascade reaction enables the one pot cleavage of the linear precursor from the fluorous tag and its macrocyclization.[128]

3.7 Multicomponent strategies

Multicomponent reactions can connect three or more starting materials in a single step.[129] The Passerini-three component reaction (P-3C) for instance employs an isocyanide, an

aldehyde (or ketone) and a carboxylic acid to afford an α-acyloxy amide.[130] Recently, these reactions were implemented in iterative protocols for the synthesis of sequence-defined polymers. Meier and co-workers, described the synthesis of this class of molecules employing the P-3C and thiol-ene reaction.[131] Figure I.24, outlines the strategy for the sequence-defined polymers synthesis employing a multicomponent reaction. in the first step, a carboxylic acid moiety (blue) reacts *via* P-3C with an aldehyde bearing a double bond (red) and different substituted isocyanides (green) forming an α-acyloxy carboxamide. The second step is the thiol-ene reaction between the α-acyloxy carboxamide and the thiol moiety of the mercaptopropionic acid. The iterative repetition of these two steps enables the synthesis of sequence-defined tetramers and pentamers. In order to increase the yield and speeds the synthesis up, the whole protocol was also performed employing a PEG-based soluble support.

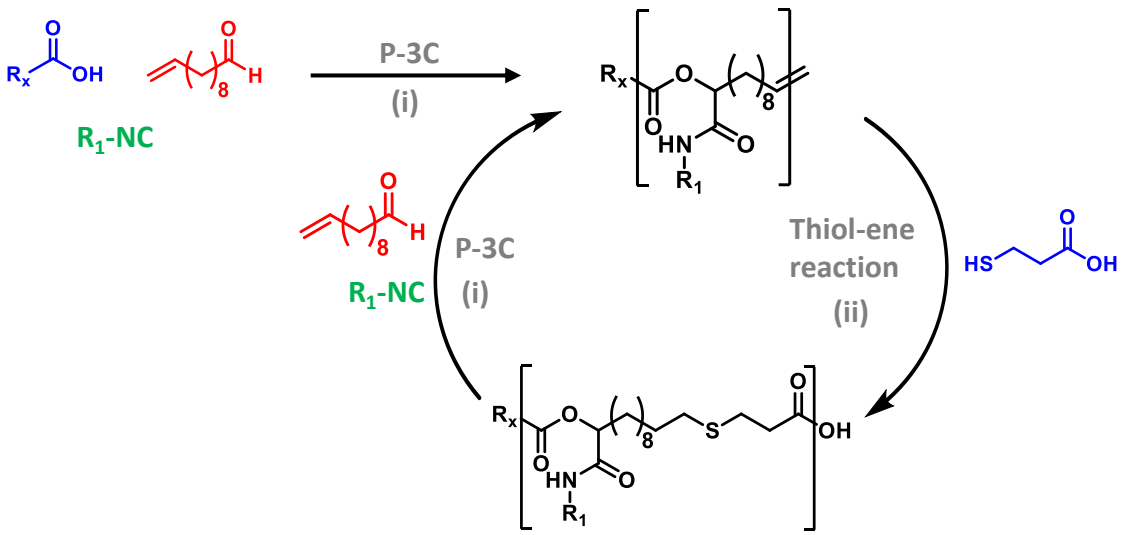


Figure I.24: Strategy for sequence defined polymer synthesis *via* multicomponent approach. **(i)** P3-C: DCM, 2h; **(ii)** thiol-ene "click": DMPA, 320nm, 1h.

Recently, a convergent protocol in which thiolactone chemistry is combined with P-3C reaction was described.[132] In this approach, a thiolactone-carboxylic acid reacts *via* P-3C with 10-undecenal and different isocyanides in the first step. In the second step, the resulting compound reacts with a mercaptopropionic acid *via* a thiol-ene reaction. The resulting macromolecules are synthesized in high yields and in multi-gram scale. UGI reaction (U-4C) is a four component reaction which employs the same three reagents involved in the Passerini and an amine, leading to the formation of α-aminoacylamides.[133] This reaction plays an important role in combinatorial chemistry[134] and is also employed for the synthesis of cyclic peptoids.[135] The multicomponent reaction was recently combined with thiol-ene reaction for the iterative synthesis of sequence-defined polymers offering the possibility to insert two side groups in the backbone in one step and increase the structural complexity of the resulting macromolecules.[132].

4 Chemical tools for the synthesis of poly(alkoxyamine phosphodiester)s

4.1 Introduction

In this section, the phosphoramidite coupling and the nitroxide-radical coupling (NRC) are discussed. These reactions were extensively employed for the synthesis of sequence-defined poly(alkoxyamine phosphodiester)s which are presented in chapter 2.

4.2 Phosphoramidite chemistry historical background and development

After the discovery of the DNA structure in 1953 by Watson and Crick,[21] several efforts were made in order to develop tools allowing the access to synthetic DNA in a facile and fast manner. In this section, the most significant discoveries leading to the development of the modern phosphoramidite protocol are summarized.

The first dinucleotide chemical synthesis was described by Todd in 1955.[136] The phosphodiester linkage between two deoxynucleotides was obtained *via* a single condensation between 3'-OAc-protected deoxythymidine and 5'-OAc-protected deoxythymidine bearing phosphorochloridate in 3'position (Figure I.25).

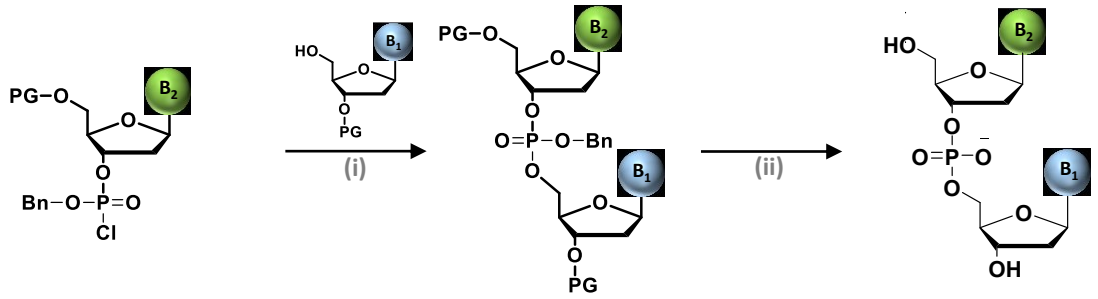


Figure I.25: First dinucleotide synthesis by Todd **(i)** coupling; **(ii)** protecting groups removal.

In 1956 Kohrana developed another procedure for the dinucleotide synthesis,[137] In this methodology, a phosphomonoester 5'-phosphorylated deoxynucleoside reacts with the 3'-hydroxyl group of another deoxynucleoside. Subsequently the same group developed a strategy for the synthesis of DNA and RNA using this approach. Contrary to the first method, the synthesis is performed in 3' 5' and exploit the more nucleophile 5'-OH that reacts with a DCC activated 3'-phosphate monoester (Figure I.26).[138, 139]

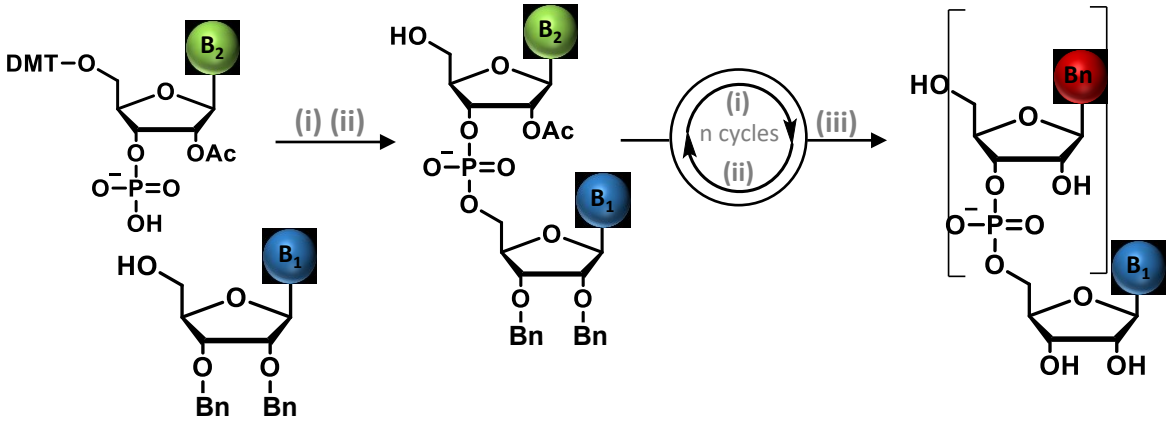


Figure I.26: Phosphodiester method. **(i)** phosphodiester synthesis: DCC, **(ii)** DMT protecting group release: H^+, **(iii)** benzyl groups deprotection: NH_3.

This method was referred to as phosphodiester method that allowed to perform the two step cycle several times and to produce short oligonucleotides. However, the increasing amount of phosphate groups, generated after each cycle, could still react with the monomer in a competitive manner. This hyperbranching made the purification process tedious.

In 1965, Letsinger solved this problem developing a new approach called phosphotriester method (Figure I.27).[78]

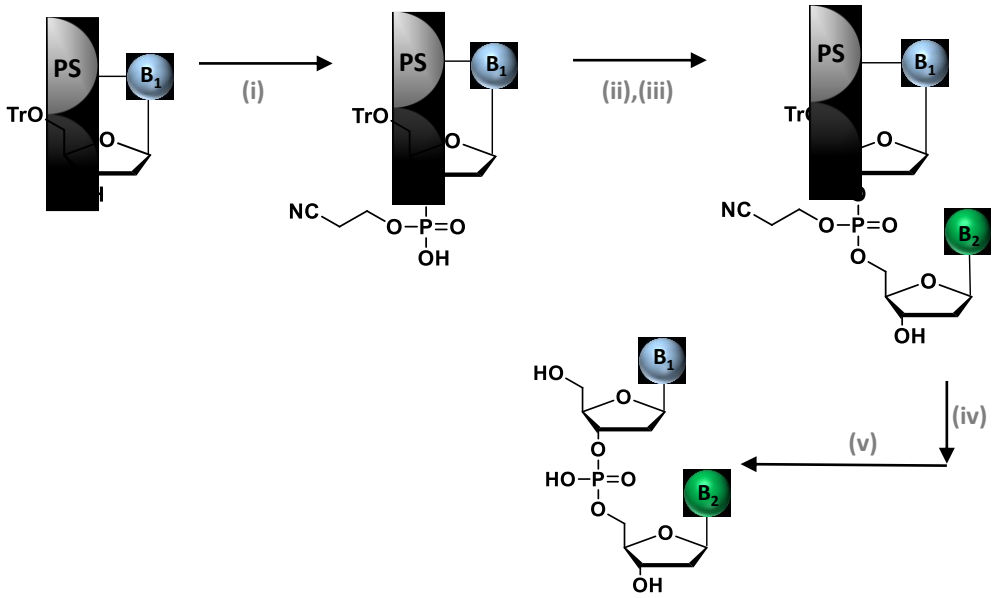


Figure I.27: Phosphotriester method. **(i)** phosphorilation: 2-cyanoethyl phosphate, DCC; **(ii)** phosphate activation: MsCl; **(iii)** coupling: deoxynucleoside 5'- OH; **(iv)** cyanoethyl deprotection: thiophenol in pyridine; **(v)** trityl deprotection and cleavage: H^+ then OH^-.

In this strategy, a 5'-OTr-protected deoxynucleoside is transformed into the corresponding deoxynucleotide with a 2-cyanoethyl-protected phosphoester at the 3' position (Figure

I.27(i)).[140] The following condensation of the activated nucleotide with a 5'-OH deoxynucleoside produces a phosphate triester linkage (Figure I.27). The formation of the phosphotriester intermediate precludes hyperbranching. The following phosphate triester deprotection, done under slightly basic conditions, gives the phosphodiester moieties after the desired oligonucleotide sequence has been synthesized. Furthermore, the protocol was performed on solid support, making the synthesis faster and the purification step easier.[78] Thanks to these two new tools, the iterative synthesis of short oligonucleotides became relatively easy and fast.[141] In 1975, Letsinger discovered that phosphorodichloridites react in a fast manner with hydroxy groups, generating a phosphite triester moiety. The resulting phosphite triester is easily oxidized into a phosphate group using iodine and water.[142] Relying on this observation, a new strategy, called phosphite trimester method, was developed. This new approach was in the beginning applied in solution [143, 144] and then on solid-state.[145] The iterative cycles were performed in three steps: (i) the coupling, (ii) the phosphite triester oxidation, (iii) the 5'OH deprotection (Figure I.28). Even though the oligonucleotide synthesis on solid supports drastically simplified the purification procedures, the use of polymers as supports had some drawbacks. For instance, the slow solution diffusion rate into the matrix or the irreversible reagent adsorption were big limitation for these materials.[146] In 1980, Caruthers described the synthesis of nonanucleotides *via* the phosphite triester approach[145] replacing the standard polymers, usually employed as solid supports, with a macroporous silica gel-based support.

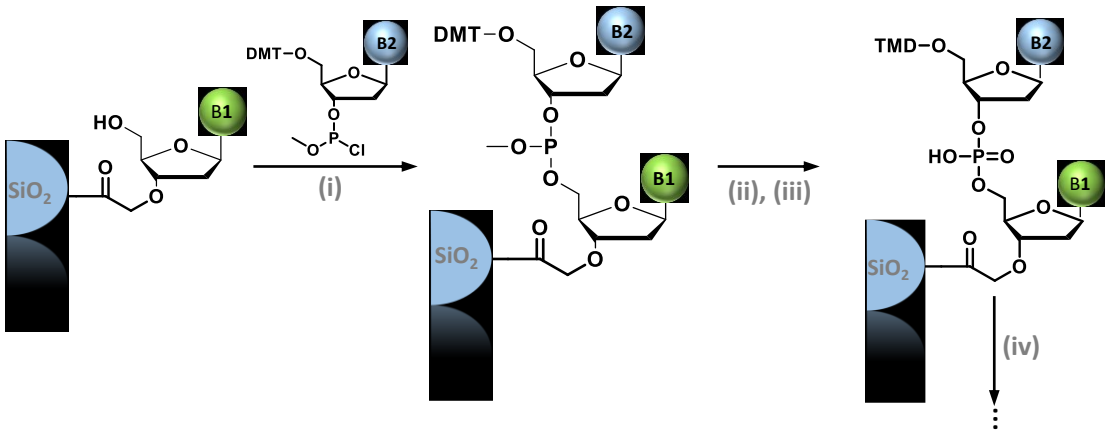


Figure I.28: Phosphite triester method on silica matrix **(i)** coupling: CH_3CN, tetrazole, **(ii)** oxidation: iodine, H_2O, nucleobase; **(iii)** protecting groups removal; **(iv)** cleavage: triethylammonium thiophenoxide, then ammonia.

The development of this new class of matrices and the following protocol automation[79] made the DNA synthesis much faster and more reliable. Further optimization of the silica gel support led to the development of the modern CPG-based supports. These porous materials have a constant volume and solvent-dependent swelling was not an issue anymore. Moreover, the application of high flow rates is possible under all the reaction conditions.[147] Nevertheless, the phosphite triester methodology still had a big limitation as the low

monomer stability toward hydrolysis and air oxidation made the molecules difficult to store and handle. These complications made the DNA synthesis *via* phosphite triester hardly suitable for biologists and biochemists. In 1981, Caruthers described the synthesis of phosphoramidite nucleosides or nucleotides and their use as synthons for the DNA synthesis.[79, 80] Phosphoramidite are not prone to oxidation and thus easy to store. Furthermore in mild acidic conditions, the monomers are activated toward the formation of the phosphite triester linkage that could be easily oxidized into the corresponding P(V). Due to the high stability of phosphoramidite monomers and the selective activation with mild acids, the coupling is performed at room temperature. Further protocol optimizations developed during the last years [148, 149] led to the implementation of the modern phosphoramidite protocol described in section 4.3.

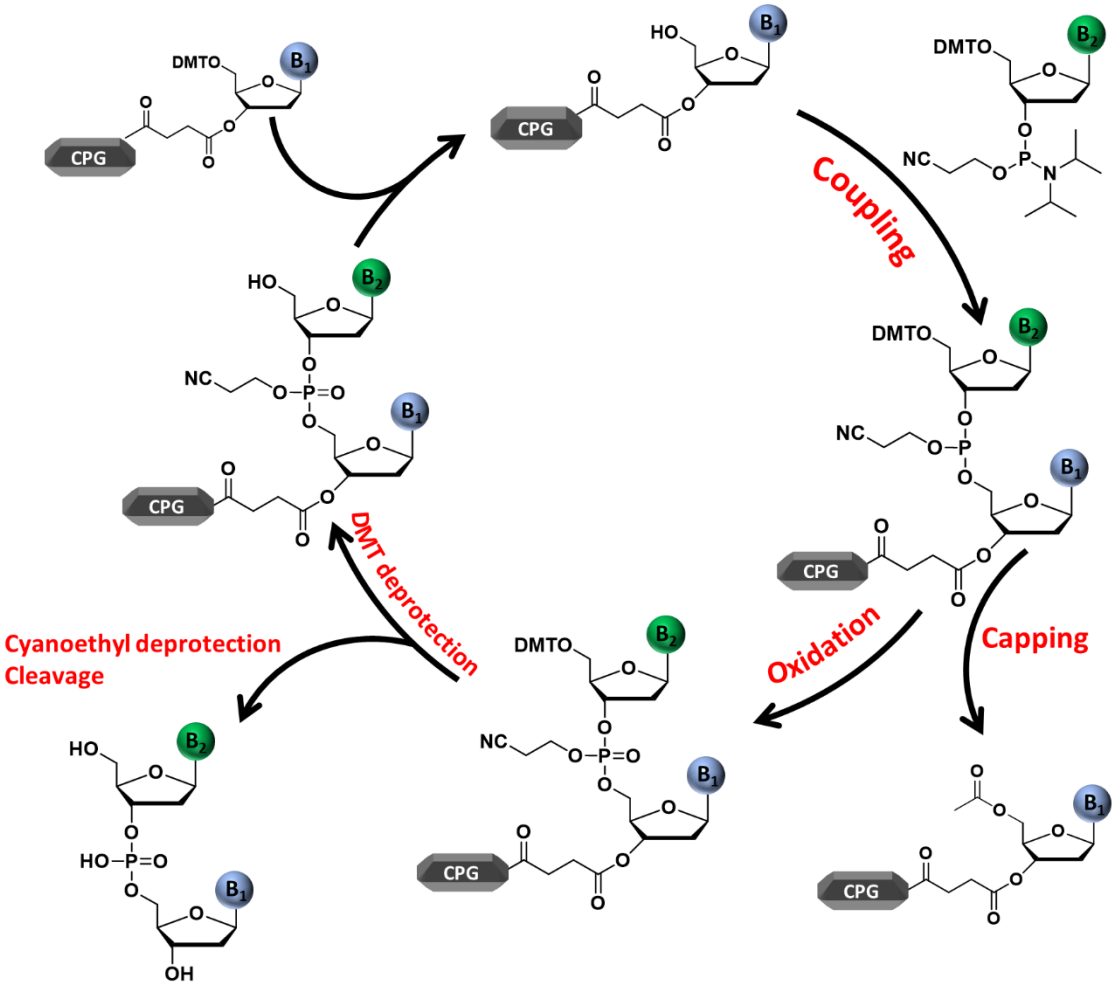
4.3 Phosphoramidite chemistry protocol

Nowadays, the standard procedure for DNA synthesis is performed using oligonucleotide synthesizers. These machines inject the different reagents into a column containing the solid support, usually CPG. A software allows the choice of the different conditions (the reaction time for each step, the synthesis direction, the primary structure composition, etc.) The automated phosphoramidite protocol is depicted in Scheme I.2.

Phosphoramidite coupling: the reaction between the 5'OH on the growing chain and a phosphoramidite activated with tetrazole takes place inside the column and reagents are delivered from the machine in the right amount. Usually the coupling time is less than 30 seconds with almost quantitative yields.

Capping: as explained before, the coupling yield is around 98%. In order to ensure uniform chain length distribution, the statistical 2% of unreacted growing chains need to be capped before performing the next phosphoramidite coupling. This step is done using a group highly reactive toward the free 5'-OH of the unreacted growing chains, for instance acetic anhydride.

Oxidation: the solution of iodine and water oxidizes the phosphite formed during the coupling step into a more stable phosphate. The reaction is completed within few seconds.



Scheme I.2: Phosphoramidite cycle for oligonucleotide synthesis.

DMT deprotection: the growing chains are 5'DMT protected. Before performing the following phosphoramidite coupling, it is necessary to deblock this functionality. DMT deprotection is performed in mild acidic conditions, usually using a diluted solution of TCA in DCM. The amount of released DMT cation can be quantified measuring the absorbance of the resulting solution at 500 nm and can be correlated to the coupling yield.

2-Cyanoethyl deprotection: before cleaving the growing chains from the solid support, it is necessary to remove the cyanoethyl group.

Cleavage: the synthesized poly(phosphodiester) chains are cleaved from the solid support using a strong basic solution usually ammonia or a mixture of ammonia and methylamine 1/1. After cleavage, the sequence is usually recovered by filtration and following solvent evaporation.

4.4 Nitroxide radical coupling

The nitroxide radical coupling (NRC) is a powerful tool for the synthesis of polymers and materials.[51] In this reaction, the combination of copper and ligand generates a macroradical from molecules bearing halogen terminal groups. The resulting radical is trapped by a nitroxide in a fast reaction leading to the formation of stable alkoxyamine bonds. The high tolerance to several functional groups,[150] the possibility to modulate the reaction rate,[151] and the reversibility at high temperatures[152] make the NRC a powerful tool for the synthesis of polymers containing complex architectures, for the polymer chain end functionalization and for many other applications.[153] As depicted in Figure I.29 a, the macroradical can be achieved either *via* atom transfer radical coupling (AT-NRC), catalyzed by Cu(I), or single electron transfer radical coupling (SET-NRC) Cu(0)-catalyzed. The NRC was initially developed by Salomon for spin trapping[154] and was following employed by Matyjaszewski in 1998 for the synthesis of alkoxyamines.[155] In the latter work, the different alkoxyamines were produced from the reaction between macroradicals, generated from acrylates, methacrylates and acrylonitriles in presence of Cu(I), PMDETA with TEMPO (Figure I.29 b). Interestingly, the combination of ligand and Cu(0) powder (1.05 eq.), added in the reaction mixture led to the regeneration of Cu(II) in Cu(I) allowing the complete conversion of the halide in a macroradical.

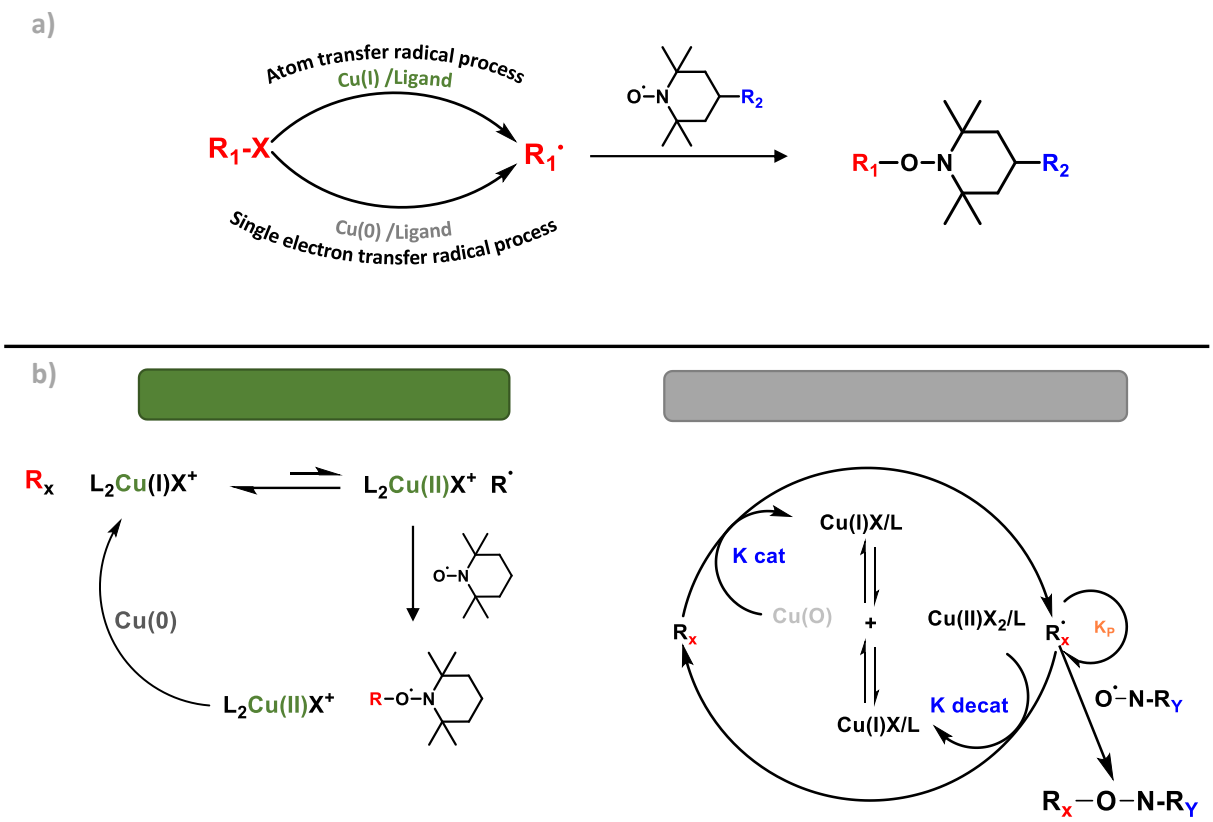


Figure I.29: Mechanisms involved in NRC and radical formation. **a)** General mechanisms of NRC; **b)** Mechanisms of radical formation.

In 2006, Percec described the living radical polymerization *via* single electron transfer mechanism (SET-LRP). In this methodology the macroradical was obtained using a Cu(0)-based catalyst.[156] The process is based on the disproportionation phenomena of Cu(I) into Cu(0) and Cu(II) in DMSO. As depicted in Figure I.29 b, the radical is generated by reactive Cu(0) species whereas the radical deactivation is regulated by $Cu(II)X_2$. Furthermore, the disproportionation rate constant is influenced by the solvent and the ligands employed. From the mechanistic point of view, the radical formation *via* single electron transfer can occur *via* inner sphere electron transfer (ISET) or outer sphere electron transfer (OSET).[157] The main difference is that for ISET a transition state consisting of a bridge between the halide and the metal is formed whereas OSET is a two-step process in which the species are not covalently connected hence the electron is forced to move through the space. In the last years, several examples of nitroxide radical coupling *via* atom transfer or single electron transfer mechanisms were described for the design of polymer architectures and materials.[153] One example is the poly (GTEMPO-co-EO)-g-PS (or PtBA) synthesis.[158] In this strategy, the halogenated chain ends of PS or PtBA, produced *via* ATRP, are transformed in the macroradical *via* atom transfer radical mechanism. Afterward, the radicals generated in the previous step react with the nitroxide moieties of poly(GTEMPO-co-EO) forming the grafted copolymers. NRC is widely employed for the preparation of terminal functional polymers. Indeed, functional group insertion at the polymers chain end is extremely useful for the synthesis of molecule with complex architectures as bioconjugates and organic-inorganic composites.[153] Monteiro *et al*., in 2011, described the end functionalization of different polystyrenes and poly(*tert*-butyl acrylate)s *via* NRC coupling. Polymeric chains containing terminal Br moiety were synthesized *via* ATRP; end modification was obtained *via* SET-NRC using different substituted TEMPO monomers (Figure I.30 a). Furthermore, the monomers employed for the chain end modification bearing clickable groups, activated esters, biomolecules and biosensors [150] allow the furthr chain end post modification (Figure I.30 b).

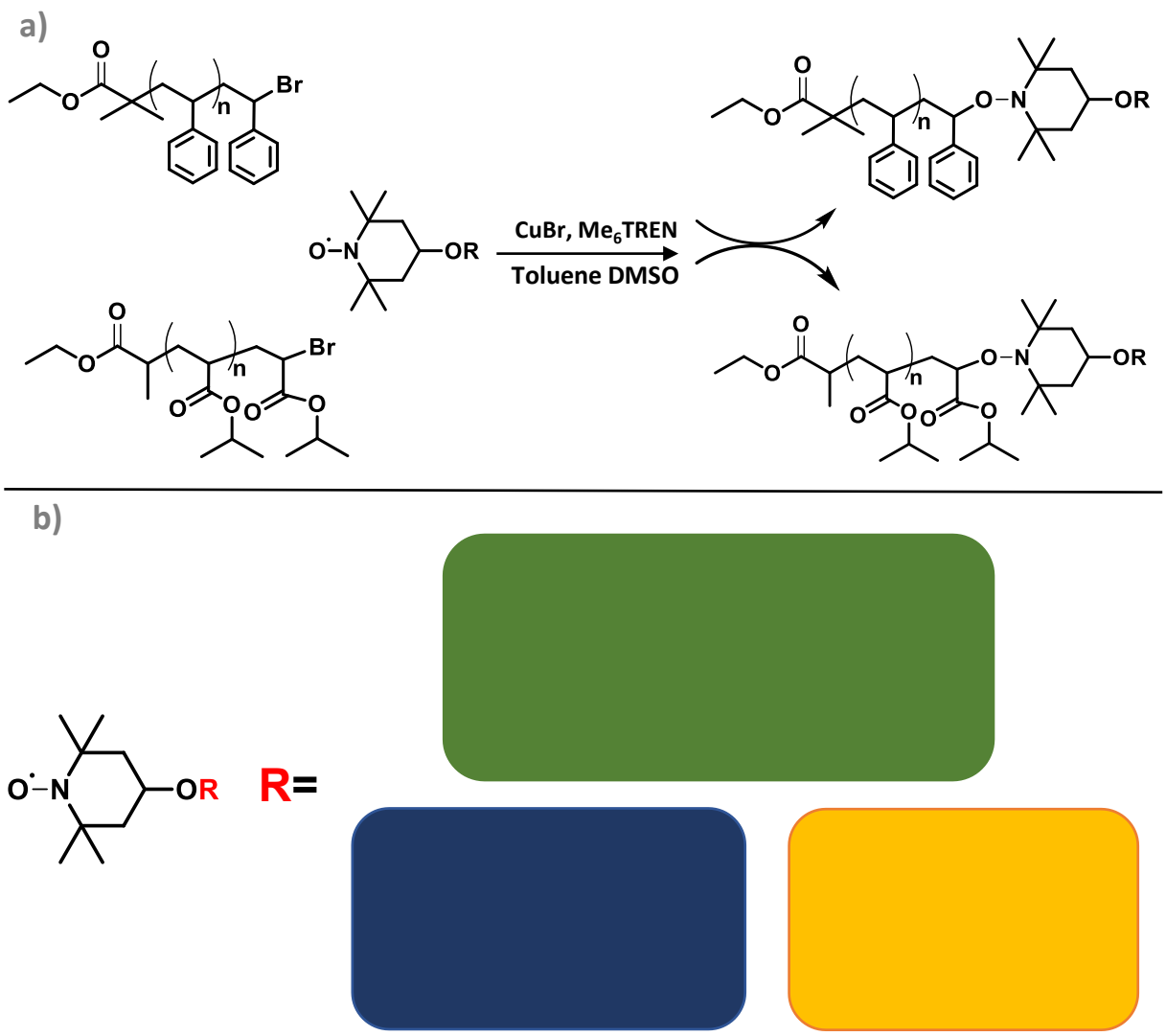


Figure I.30: Chain end functionalization strategy. **a)** Chain end modification strategy *via* NRC coupling; **b)** Molecules employed for the chain end modification.

NRC was also employed for the synthesis of sequence controlled polymers as block copolymers. In 2009 the synthesis of block copolymers using SET-NRC was described.[159] Different bromine and TEMPO end polymers were synthesized employing respectively ATRP and ROP. The two blocks were then linked one another using SET-NRC conditions. Interestingly, the macroradical generation, achieved at room temperature, prevents the side reactions as β-hydrogen transfer or thermal cross-linking. Huang reported several examples of triblock copolymers using nitroxide radical coupling.[160, 161] The strategy for the synthesis of these sequence-controlled polymers, depicted in Figure I.31, relies on the one-pot NRC and CuAAC reactions of the three blocks previously synthesized. The monomer structures namely azide poly(*tert*-butyl acrylate) PtBA-N_3 (I), alkyne-PSBr (II) and the PEO TEMPO or PCL TEMPO blocks (III) are depicted in Figure I.31. Mixing the three building blocks in DMF with CuBr and PMDETA for 1 day at 90°C, the formation of triazole and alkoxyamine linkages enable the triblock copolymer formation.[161]

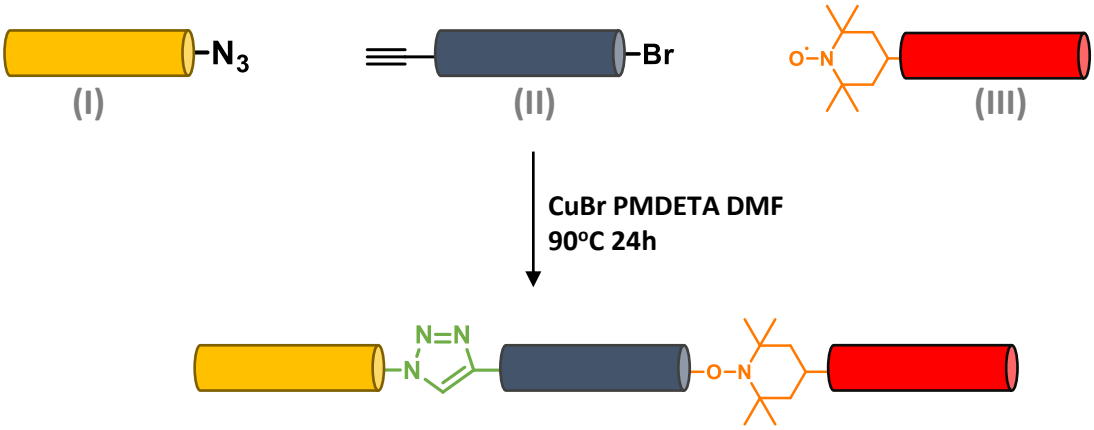


Figure I.31: General strategy for the synthesis of triblock copolymers *via* one-pot CuAAC and NRC reactions.

Another interesting example of NRC application for macromolecular design was reported in 2009.[162] A polymer chain synthesized *via* ATRP reacts with *tert*-butyl-phenylnitrone generating the nitroxide moiety which is involved in a reaction with a second macroradicalic polymer chain. As a structural feature, the resulting conjugated polymer shows an alkoxyamine moiety at the mid-chain. Lutz and co-workers incorporated the NRC into an iterative chemoselective solid-state protocol for the synthesis of digitally encoded poly(alkoxyamine amide)s.[11] As described in the digital polymers section, the alkoxyamine bond formed during this reaction simplifies the sequencing by tandem MS of the resulting sequence-defined polymers.

5 Digital polymers

5.1 Introduction

Nowadays, the amount of stored data increases constantly. In 2012 for instance, about 2.5 exabytes of data were created every day, this number was doubling every 40 months.[163] This huge amount of information needs to be stored in enormous data centres in controlled temperature and humidity conditions, resulting in a high consumption of space and energy. A valid alternative to the modern solid-state memories consists in storing data into molecules. Indeed, molecular data storage can considerably increase the information data density, and may be convenient in terms of space and energy consumption. One of the first molecules studied for data storage applications is DNA.[164] The four letter alphabet employed for storing the genetic code in DNA and RNA makes these molecules a natural highly dense data storage system. Theoretically, using DNA it is possible to store 455 exabytes per gram per single strand.[165] Last year, Zielinski described the storage of 215 petabytes per gram of DNA using the DNA fountain technique,[166] creating the most dense data storage DNA-based system reported so far. Recently, different methodologies allowing to store different data format as ASCII, PDF, JPEG, etc. in DNA[3] or encoding a short movie into a bacteria's genome[4] were developed. Nevertheless, DNA is not the only molecule that

allows data storage. In principle, every sequence-defined polymer containing two different monomers intentionally defined as 0 and 1 can encrypt information relying on the binary alphabet.[5] Obviously, the molecules employed for data storage applications must allow fast and easy synthesis and sequencing. The term sequencing refers to the use of analytical techniques capable of revealing the monomeric sequence in the polymer backbone.[167] In the last decades, a huge variety of techniques for the synthesis of sequence-defined polymers have been developed allowing the easy synthetic access to a large variety of structures (see section 3).[8, 20] Furthermore, synthetic polymers give the unique possibility to carefully design the backbone structure to obtain a fast and facile sequencing.[6] Nowadays, several techniques as tandem Mass spectrometry,[168] nanopore techniques,[169] etc. are employed for the sequencing of synthetic polymers. In this chapter, the synthesis and the sequencing of the different classes of sequence-defined polymers employed for data storage applications, called digital polymers will be discussed.

5.2 Oligo(triazole amide)s

Synthetic sequence-defined polymers give the possibility to access a large variety of chemical structures and tuning them in order to obtain an easier sequencing. The first example of digital polymer is represented by the oligo(triazole amide)s described by Lutz *et al*. in 2014.[125] These molecules are synthesized on a solid support using an (AB+CD) approach.

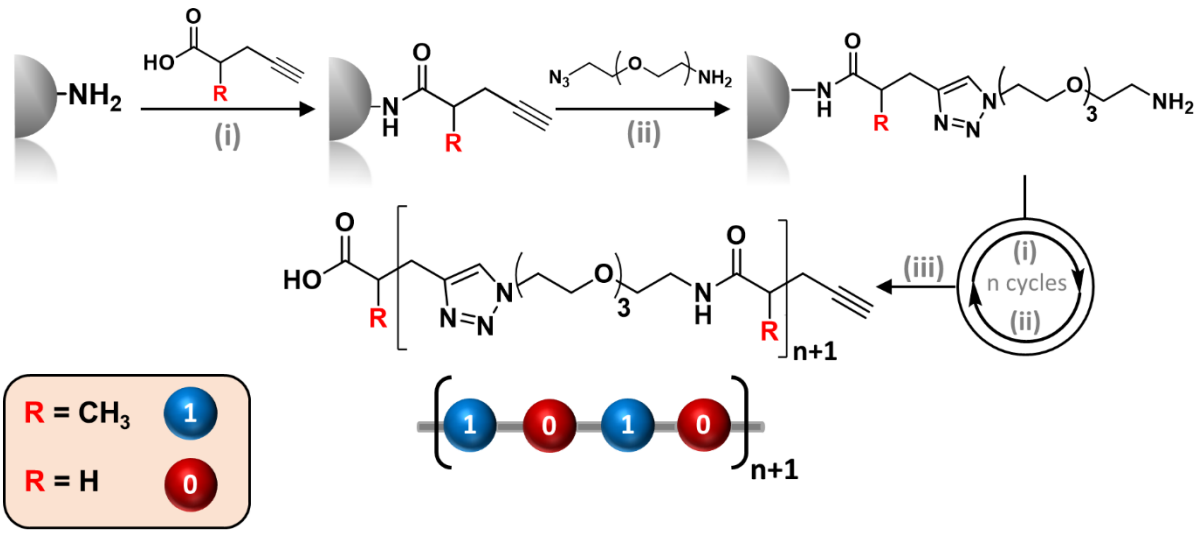


Figure I.32: Oligo(triazole amide) synthesis. **(i)** amidification: PyBOP, DIPEA, DCM; **(ii)** CuAAC: CuBr, dNbipy, THF; **(iii)** cleavage: TFA.

The iterative protocol relies on two iterative coupling reactions, namely amidification and CuAAC steps. For this class of digital polymers 4-pentynoic acid and 2-methyl-4-pentynoic acid are respectively defined as bit 0 and 1 whereas amino-11-azido-3,6,9-trioxaundecane is employed as a spacer (Figure I.32). Once the desired sequence is obtained, the resulting

macromolecules are cleaved from the Wang support in acidic conditions. After the synthesis, the digitally encoded sequence-defined polymers were analyzed *via* MALDI-TOF, NMR and SEC showing uniform chain length distributions and fully controlled primary structures. The sequencing *via* tandem MS analysis evidences multiple cleavage mechanisms resulting from the amide bond fragmentation and the last ether bond cleavage in each unit (Figure I.33).

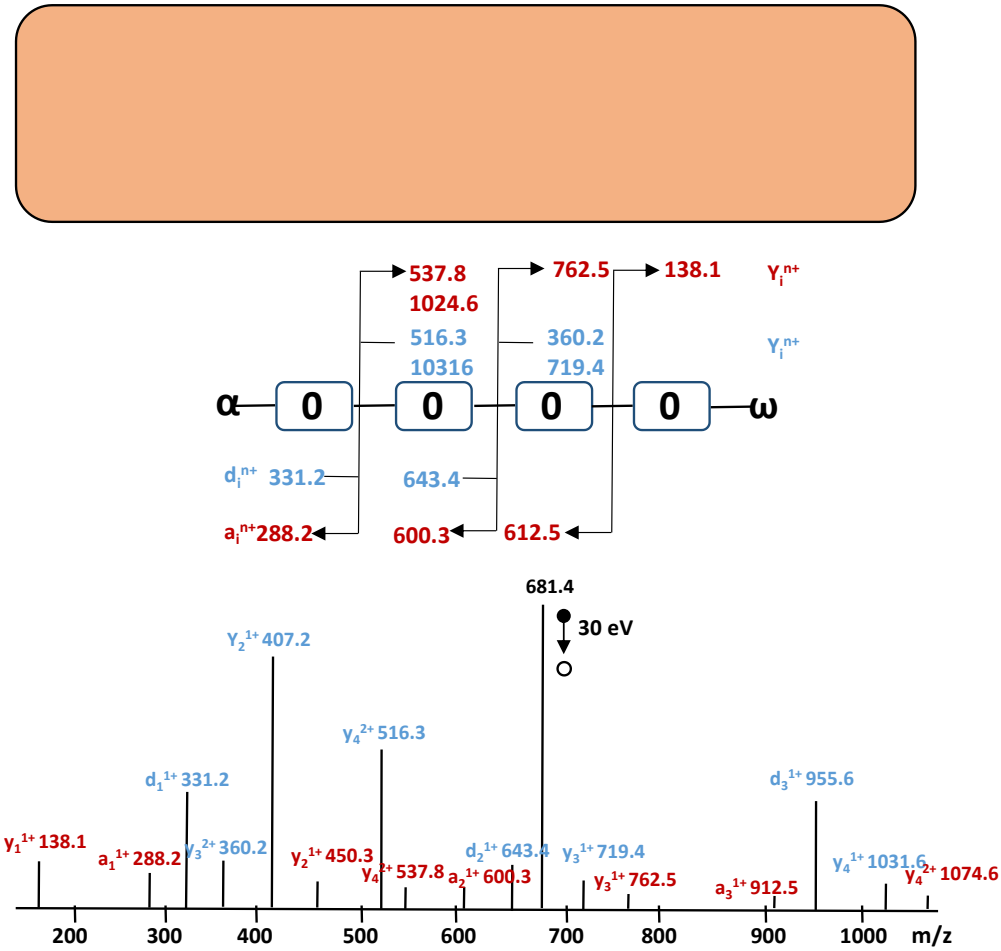


Figure I.33: Fragmentation pattern for oligo(triazole amide)s and sequencing *via* tandem MS of a sequence coding for 0-0-0-0.

Moreover, the multi charge state of some fragments makes the sequencing possible, but very complicated even for short and simple sequences.[170] Even though, the sequencing of these molecules by MS/MS was not optimal, the oligo(triazole amide)s represent the first proof of concept of non-natural sequence-defined polymer encoding information. One of the main lacks of the solid support assisted synthesis is the long time required for the synthesis of long chains. In order to obtain high molecular weight polymers in a relatively short time, a dyad ligation strategy was developed for the synthesis of oligo(triazole amide)s.[171] In this work, a library containing all the possible combinations of two bits and one spacer was synthesized. The encoded dyads were employed in the standard solid support protocol, reducing considerably the number of steps and the time required for the synthesis of one byte.

5.3 Poly(alkoxyamine amide)s

The problem of the complicated fragmentation pattern observed for the oligo(triazole amide) class was solved in 2015 by replacing the triazole moiety with a labile alkoxyamine bond, generated by nitroxide radical coupling. The development of the poly(alkoxyamine amide) class results one of the breakthrough in the field of the digital polymers. Indeed, this class allows the easy and fast reading by tandem mass spectrometry.[11] Two different building blocks are employed in the protocol: the first one contains a symmetric acid anhydride and alkyl bromide groups, whereas the other one contains a nitroxyl radical and a primary amine. The anhydrides represent the coding monomers, in particular the 2-bromo-isobutyric acid anhydride and the 2-bromopropionic acid anhydride define respectively the bit 1 and the bit 0 (Figure I.34). The chemoselective reactions involved in the protocol are the amidification reaction between amine and anhydride and the NRC between the carbon-centred radical and the amino-TEMPO (Figure I.34). Relying on these quantitative and fast reactions, the solid support protocol was employed for the synthesis of a library of different poly(alkoxyamine amide)s. After acidic cleavage, the resulting SEC, NMR and ESI-HRMS analysis highlight the uniform weight distribution and high purity of the resulting digital polymers.

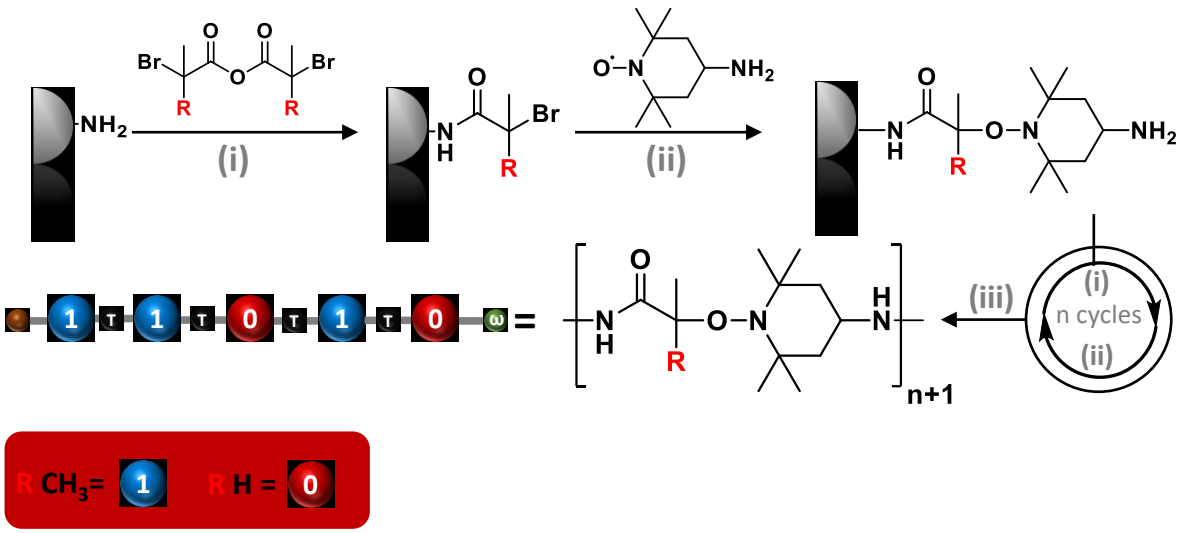


Figure I.34: General strategy for the synthesis of sequence-defined poly(alkoxyamine amide)s. **(i)** amidification: K_2CO_3 or DIPEA, THF, 50min; **(ii)** radical-radical coupling: CuBr, Me_6TREN, DMSO, 25min; **(iii)** cleavage: TFA.

The introduction of the labile alkoxyamine bond in the backbone enables the fast and easy sequencing. Indeed, at low activation energy, the homolytic alkoxyamine bond cleavage is the only occurring event. This selective fragmentation leads to a single bond cleavage for each unit in the backbone, generating extremely easy to interpret fragmentation patterns. As depicted in Figure I.35, in the tandem MS spectrum, it is possible to identify four different fragments corresponding to the radical motifs: α-0 (130 Da), α-1 (144 Da), ω-0 (305 Da) and ω-1 (319 Da).

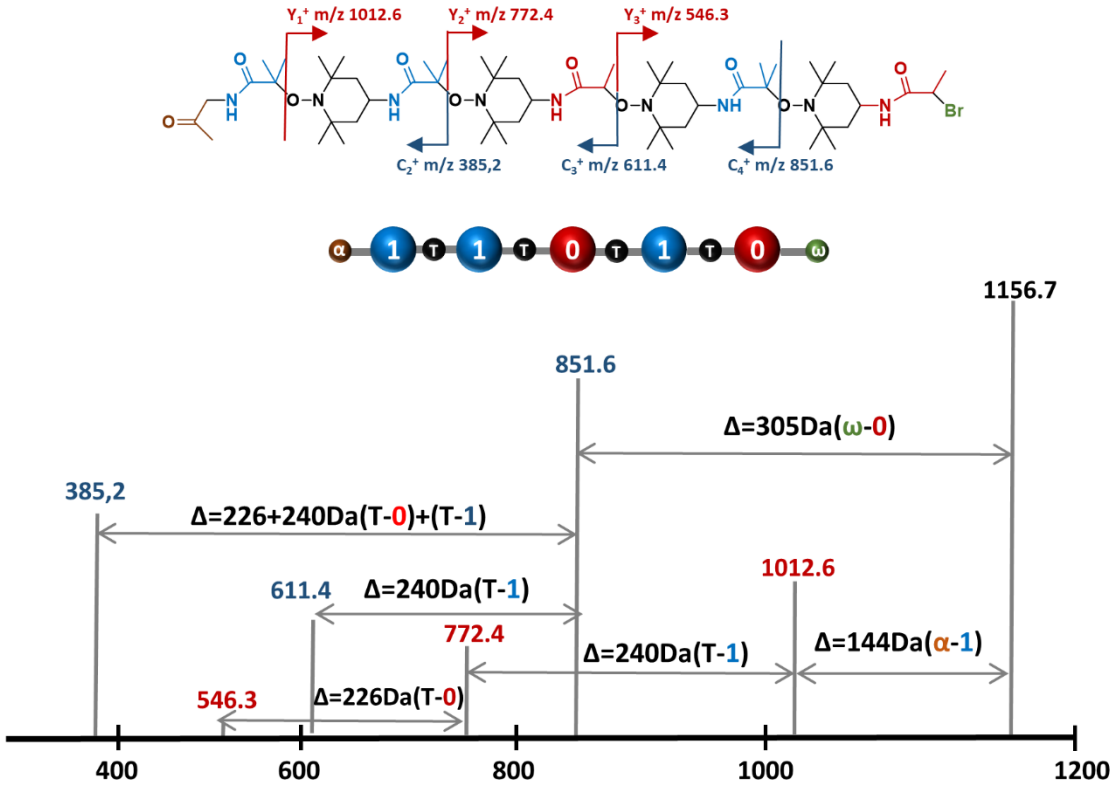


Figure I.35: MS/MS sequencing of a poly(alkoxyamine amide) sequence coding for 1-1-0-1-0.

The sequencing of the whole primary structure relies on the intervals of 226 and 240 between α and ω corresponding respectively to TEMPO-0 and TEMPO-1.[14] In 2015, it was shown that it is possible to produce long digitally encoded poly(alkoxyamine amide)s in few steps *via* a convergent synthesis.[172] Indeed, a library containing all the possible combinations of two coding units and one spacer was synthesized on a solid support. The trimers were iteratively added to a tetramer chain, anchored to a Wang resin, following the standard protocol. This strategy allowed the synthesis of a sequence-controlled 10-mer in few steps. Another interesting feature of this class of molecules is the possibility to be "erased". Indeed, these polymers start to degrade above 60°C and the degradation rate can be regulated according to the temperature. This mechanism relies on the reversible nature of the alkoxyamine bond[173, 174] producing a mixture of non-uniform fragments from the original sequence. However, it was noticed for this class of digital polymers that the coding could influence the yields. Indeed, the low stability of the radical, produced during the NRC of monomer 0 result in lower yields compare to the one originated from the NRC of monomer 1. Moreover, as a result of the different radical stabilities, in CID conditions the tertiary radical generated from the alkoxyamine bond cleavage of bit 1 is more abundant in tandem MS/MS[14], hence also the coding influences the sequencing.

5.4 Sequence-coded polyurethanes

Polyurethanes are one of the most employed and studied polymer classes. The first polyurethanes synthesis was reported in 1947 by Bayer.[175] Nowadays, these molecules are largely employed in industry as coatings, adhesives, synthetic fibres, etc.[176] However, the standard step-growth polymerization process employed for their synthesis doesn't ensure primary structure control and narrow molecular weight distribution. In the past decades, several strategies for the synthesis of sequence-defined polyurethanes relying on protecting group approaches were reported.[96, 177] In 2015, the first protecting group-free protocol of the synthesis these molecules was described by Lutz et al. Furthermore, the implementation of a binary alphabet relying on this strategy opened the possibility to use them for data storage applications.[178] The described iterative protocol relies on two steps. In the first step, the hydroxy group anchored to the growing chain reacts with N,N'-disuccinimidyl carbonate (DSC) forming an unsymmetrical active carbonate. In the second step, the resulting carbonate selectively reacts with the amino group of different substituted amino alcohols (Figure 1.36). The binary alphabet implementation relies, as for the other digital polymers, on the presence of a methyl group on the amino alcohol side chains. Indeed, monomers containing methyl groups define the bit 1 while monomers without functional groups on the side chains define the bit 0 (Figure I.36) In order to show the reproducibility of the method, a library with different couples of amino alcohols was synthesized. The iterative repetition of these two steps enables the synthesis of uniform digitally encoded polyurethanes.

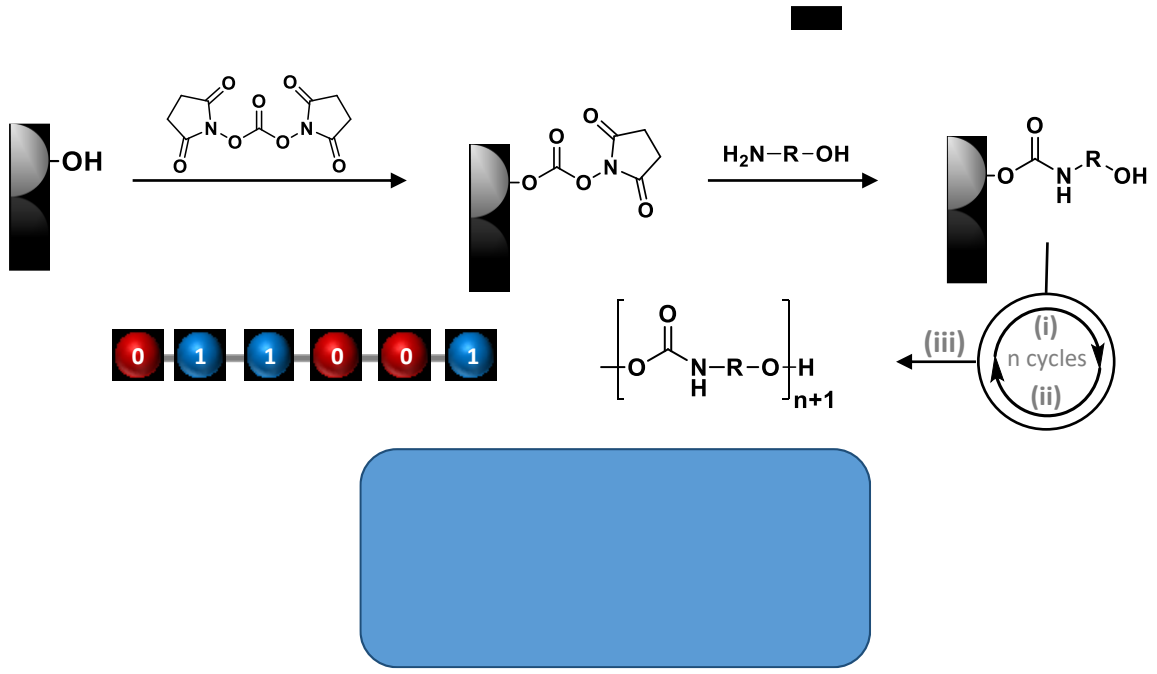


Figure I.36: Monomer structures and iterative protocol for the synthesis of sequence-defined oligocarbamates.

The selective C-O carbamate bond fragmentations, obtained in negative tandem MS conditions, lead to the formation of easy-to-interpret spectra, allowing the easy and fast read-out of the molecues. The discrimination of isobaric sequences is possible relying on the molecular weight differences between two peaks in the tandem MS which correspond to the molecular weight of bit 0 or bit 1 in the sequence (Figure I.37).[179]

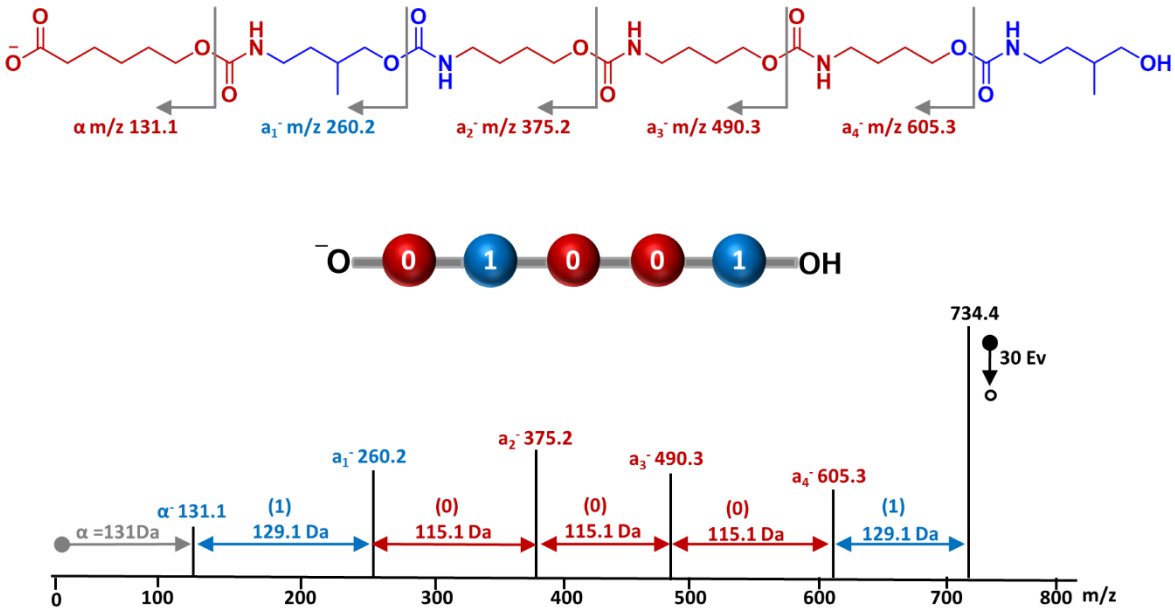


Figure I.37: Tandem MS sequencing of digitally encoded oligocarbamates.

Thanks to their extremely easy sequencing ,their biocompatibility[180] and to the possibility to be blended with other polymers, these molecules have found application as molecular tags.[105, 106] Indeed, several techniques were shown to be suitable for tagging and extracting the oligocarbamate-based tags from different polymeric materials as lenses, films, etc. After the extraction, the tags are easily sequenced by negative mode tandem MS seems to be very promising molecules for anti-counterfeiting applications. Moreover, because of their absence of toxicity in living organisms, this class of sequence-defined polymers was recently employed as in-*vivo* molecular taggants for the identification of implants.[181]

5.5 Poly(phosphodiester)s

The biochemical methodologies developed during the past decades can also be employed for the synthesis of non-natural sequence-defined polymers.[20] In this context, phosphoramidite chemistry is a promising toolbox. Indeed, the intensive studies conducted in the past decades (see section 4.2) make this coupling reaction extremely fast, nearly quantitative and thus very appealing for the synthesis of sequence-controlled polymers. Based on these preliminary observations, in 2015, the synthesis of non-natural sequence-defined poly(phosphodiester)s using the phosphoramidite chemistry protocol was described.[95] The monomers employed for the synthesis contain a hydroxy group DMT-protected and a phosphoramidite moiety (Figure I.38). In order to encrypt binary information into these molecules, the two different building blocks differ for two methyl groups defining respectively the bit 1 for the dimethyl derivative and the bit 0 for the other

one. The protocol employed is an (AB+AB) approach relying on three steps, namely phosphoramidite coupling, oxidation and DMT deprotection. Phosphoramidite coupling is performed between the OH group on the growing chain and the phosphoramidite moiety of the coding monomer. The coupling is extremely fast and the reaction is completed in approximately 15 minutes, leading to the formation of a phosphite group. Phosphite oxidation into a phosphate group is done using an oxidizer solution containing iodine, water and lutidine; this oxidation step takes few minutes. The DMT deprotection releases the OH group involved in the following coupling and is performed using a TCA 3% solution in DCM. The iterative repetition of these three steps, cyanoethyl deprotection and final basic cleavage from the solid support enable the synthesis of digitally encoded sequence-defined poly(phosphodiester)s in a short time.

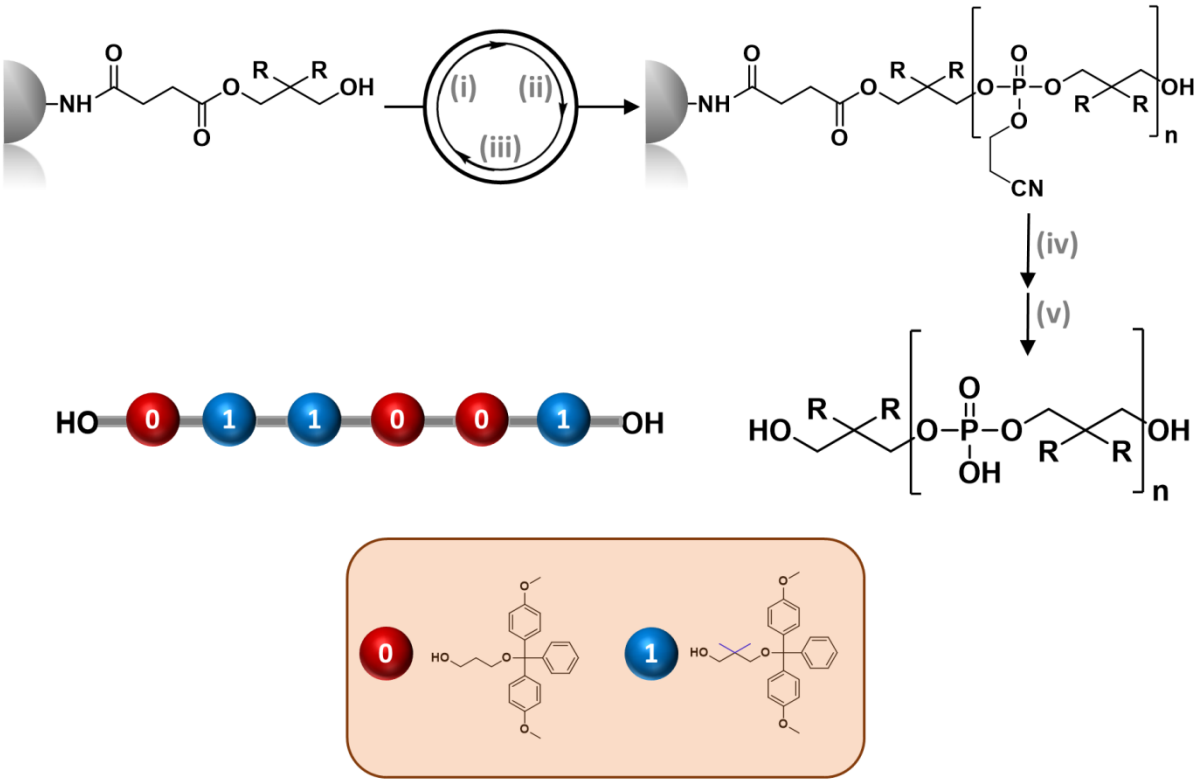


Figure I.38: Monomer structures and general strategy for the synthesis of poly(phosphodiester)s.

The fast synthesis of long digital polymer sequences is nowadays one of the big challenges in the field. Indeed, handmade iterative solid-state syntheses are usually time-consuming processes, requiring several days for the synthesis of high molecular weight macromolecules. However, it is known from biochemistry that iterative synthesis can be drastically accelerated using solid phase synthesizers.[182] These machines are largely employed for the production of synthetic biomolecules as polypeptides and oligonucleotides. In 2015, relying on the protocol previously described, the automated synthesis of non-natural sequence-defined poly(phosphodiester)s was developed using an

oligonucleotide synthesizer.[12] All the reagents needed for the synthesis were charged into the machine and the iterative cycle was performed automatically. As for DNA, the synthesis of long chains requires an additional capping reaction before performing the following phosphoramidite coupling. Indeed, the phosphoramidite reaction is nearly quantitative, thus the unreacted growing chains need to be capped to ensure uniform chain length distribution. Nevertheless, the synthesis of digitally encoded poly(phosphodiester)s with chain lengths above DP-100 was completed within a few hours. Moreover, to show that message encryption in these molecules is possible, the words "Macromolecules" and "Maurice" were ASCII-encoded using the binary alphabet. The resulting macromolecules were analyzed by MS and subsequently sequenced *via* MS/MS. The conditions employed for the sequencing lead to the dissociation of all phosphate bonds in the backbone, making the reading relatively easy only for short sequences. For long chains, the expected ion series are often detected at multiple charge states. Moreover, a big number of secondary fragments, originated from the dissociation of highest charge state fragments leads to complicated tandem MS spectra.[183] As said before, synthetic polymers offer the possibility to tune the backbone structure in order to obtain an optimal read-out.[6] Indeed, the poly(phosphodiester)s backbone was modified adding two structural features namely a labile alkoxyamine bond between the bytes and mass tags for bytes labelling (Figure 1.39 a). These new motifs allow the creation of a series of tagged bytes after selective alkoxyamine bond fragmentation in tandem MS (Figure 1.39 b).[184] The bytes produced in MS^2 can be selectively activated and decrypted by MS^3 in a fast and facile manner. Furthermore, the natural and non-natural nucleobases, employed as mass tags, allow knowing the position of the byte in the original sequence. In order to implement this new architecture, the polymer synthesis was achieved *via* an iterative phosphoramidite protocol using an oligonucleotide synthesizer and inserting the mass tags and the cleavable units between the bytes. The synthesis of a library of different polymers and their following sequencing *via* MS^2 and MS^3 highlighted the possibility to encode and decode long non-natural polymers chains in a fast and facile manner. The poly(phosphodiester) backbone was also modified in order to make this class suitable for nanopore sequencing. The nanopore analysis is a promising non-destructive read-out technique already applied for DNA.[185] The different interactions of the monomeric sequence with the pore, generate a characteristic current variation that can be related to the monomer sequence. In order to have a facile sequencing *via* nanopore analysis, the poly(phosphodiester) backbone was post-modified inserting two different PEG-based groups in the monomers side chain via CuAAC.[186] In principle, the different side chains connected to the coding monomers, having different interactions with the pore, should allow the polymer sequencing. Recently, a new protocol for the synthesis of poly(phosphodiester)s was developed. In this approach the DMT protecting group was replaced by a photoremovable protecting group (NPPOC). The use of NPPOC allows the photocontrolled synthesis of sequence-defined poly(phosphodiester). The synthesis was performed on a soluble support enabling the scale-up of the resulting digital polymers.[187]

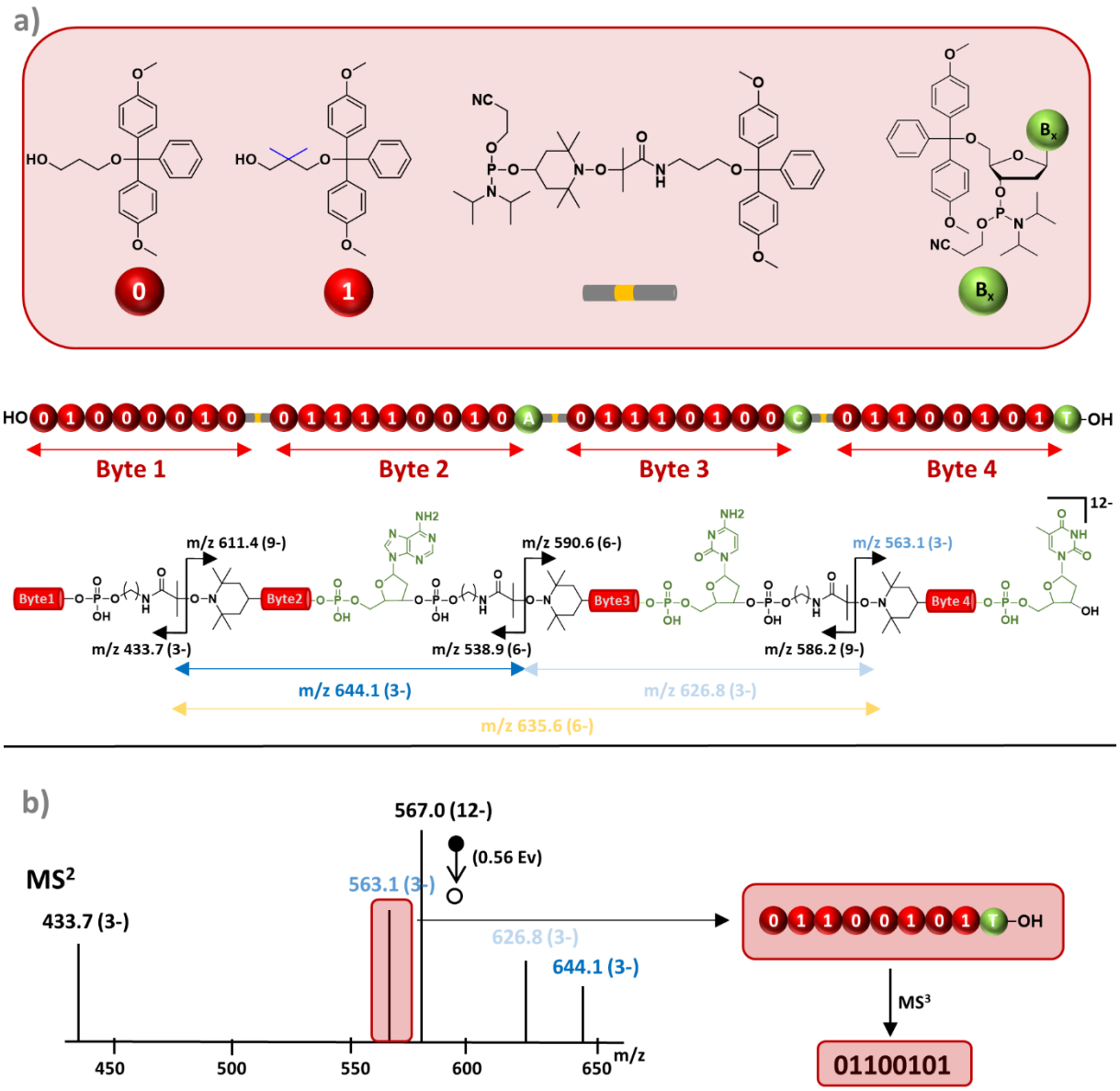


Figure I.39: General concept of molecular byte tags applied for poly(phosphodiester)s. **a)** Monomers structures and chemical structure of the poly(phosphodiester)s with mass tags; **b)** MS^2 of the structure depicted in a) and MS^3 and decoding of the byte 4.

6. Conclusions

In the last years, the efforts made toward the sequence regulation in synthetic polymers provided tools enabling the synthesis of macromolecules containing fully defined primary structures. These tools, mainly imported from other fields, transformed the polymer chemistry area in a broad interdisciplinary subject. The sequence-defined polymers produced with these techniques show interesting innovative features. Nevertheless, many progresses are still to be done, for instance increasing the attainable length, providing robust protocols and techniques for the scale-up and develop fast and efficient non-destructive sequencing techniques. Moreover, the possibility to tune the backbone offered by these

molecules, allows to obtain new and useful features. Even if this class of molecules has not yet found industrial applications, the unique possibility to tune the primary structure offered by these polymers, enable the synthesis of macromolecules with finely defined new properties. For all these reasons, these new molecules are promising candidates for the implementation of next-generation polymers.

Chapter II

Orthogonal synthesis of poly(alkoxyamine phosphodiester)s

1 Introduction

Digital polymers are synthetic sequence-defined polymers capable of store information at the molecular level.[188] As discussed in the chapter 1, every sequence-defined polymer containing two different monomers, intentionally defined as 0 and 1, can encrypt information at the molecular level.[5] The synthesis of such digital polymers is done *via* multi-step growth polymerization.[64] This technique, developed by Merrifield for polypeptide synthesis,[10] allows to obtain molecules containing a fully defined primary structure and uniform weight distribution, ensuring thus a reliable molecular encryption.[6] The different multi-step growth polymerization approaches developed in the past years (see chapter 1), allow obtaining uniform sequence-defined polymers in a relatively short time.

Moreover, the backbone of the resulting molecules can be designed to obtain an optimal sequencing. The term sequencing refers to the use of analytical techniques revealing the monomer sequence in the polymer backbone. Indeed, after the information has been encrypted at the molecular level, reliable sequencing techniques are needed for the decryption of the encoded message. In the past years, different techniques were developed for the sequencing of natural and non-natural sequence-defined polymers. For instance, MS/MS has been intensively employed and nowadays is the most employed methodology for the sequencing of digital polymers.[13] Interestingly, the backbone of these molecules may be modified not only to improve the MS/MS reading but also to obtain tailor-made features; for instance, increasing the information density, or obtaining digital polymers which can be read using non-destructive sequencing techniques.

Among the different classes of digital polymers developed so far the poly(phosphodiester)s are one of the most promising. The synthesis of these molecules relies on the phosphoramidite chemistry protocol described in chapter 1. Moreover, long digitally encoded poly(phosphodiester)s chains can be synthesized using an oligonucleotide synthesizer in a relatively short time.[12] Nevertheless, these molecules follow the typical dissociation pattern of deprotonated oligonucleotides.[189] Hence, they originate a complex tandem MS spectra which makes difficult the sequencing of long chains.

On the other hand, the development of the poly(alkoxyamine amides)s demonstrated that the read-out by MS can be extremely simplified by adding an alkoxyamine bond in the polymer backbone.[11] Indeed, this bond is selectively cleaved in CID conditions generating predictable and easy to interpret fragmentation patterns.[14] However, because of the different stability of the alkoxyamine bond generated from monomer 0 and monomer 1, the read-out of these molecules may be influenced by the sequence.

As explained before, one of the big advantages in using synthetic molecules for data storage is the possibility to carefully design the backbone structure to achieve an optimal sequencing. In this context, the poly(phosphodiester)s readability may be improved *via* alkoxyamine bond insertion into the polymer backbone. Based on this observations, a new class of digital polymers namely poly(alkoxyamine phosphodiester)s was developed and is presented in this chapter. These sequence-defined polymers generate easy to interpret MS spectra; as a result they are easier to sequence compared to poly(phosphodiester)s. Moreover, the backbone of these molecules was designed in order to avoid sequence dependent fragmentations phenomena.

An iterative solid-state protecting-group-free approach was designed for the synthesis of these digital polymers. The use of orthogonal reactions allows to bypass the deprotection step simplifying the iterative synthesis[190] and improving the step economy.[15] The developed (AB+CD) strategy relies on the phosphoramidite coupling and the nitroxide-radical coupling (NRC); both reactions proceed in a short time, at room temperature and in quantitative yields. The first reaction is based on the phosphoramidite chemistry that was developed for the DNA synthesis over the past 30 years (see chapter 1). The second one is the nitroxide-radical coupling (NRC) that enables the alkoxyamine bond insertion into the polymer backbone. This reaction, initially developed for the spin trapping,[191, 192] was following employed for the synthesis of alkoxyamines. The alkoxyamines synthesis *via* NRC proceeds without the generation of side products at room temperature, [155, 191] making the reaction suitable for the iterative strategy.

In this chapter, the synthesis of the building blocks and the optimization of the conditions which led to the development of the iterative protocol for the synthesis of digitally encoded poly(alkoxyamine phosphodiester)s is described. Moreover, the dissociation behaviour of these molecules in MS/MS conditions and the sequencing theories will be described.

2 Result and discussion

2.1 Monomers design and synthesis

The solid-state protocol developed for the synthesis of poly(alkoxyamine phosphodiester)s relies on the phosphoramidite coupling and on the NRC. These reactions allow the successive insertion of a phosphoramidite coding monomer and a hydroxy-functionalized nitroxide to the growing chain. The building blocks involved in the iterative protocol were designed in order to allow the consecutive repetition of these two reactions. The synthesis of the building blocks has to be fast, simple and it should allow to obtain pure molecules in high yields. The coding monomers are different bromo-functional phosphoramidites, the general structures of these building blocks are depicted in Figure II.1. The synthesis of bromo-functional phosphoramidites nonomers was already described by Matjazewsky in

another contest;[193] the chemical structure of this molecule resulted convenient to be implemented in the iterative protocol described in Scheme II.4.

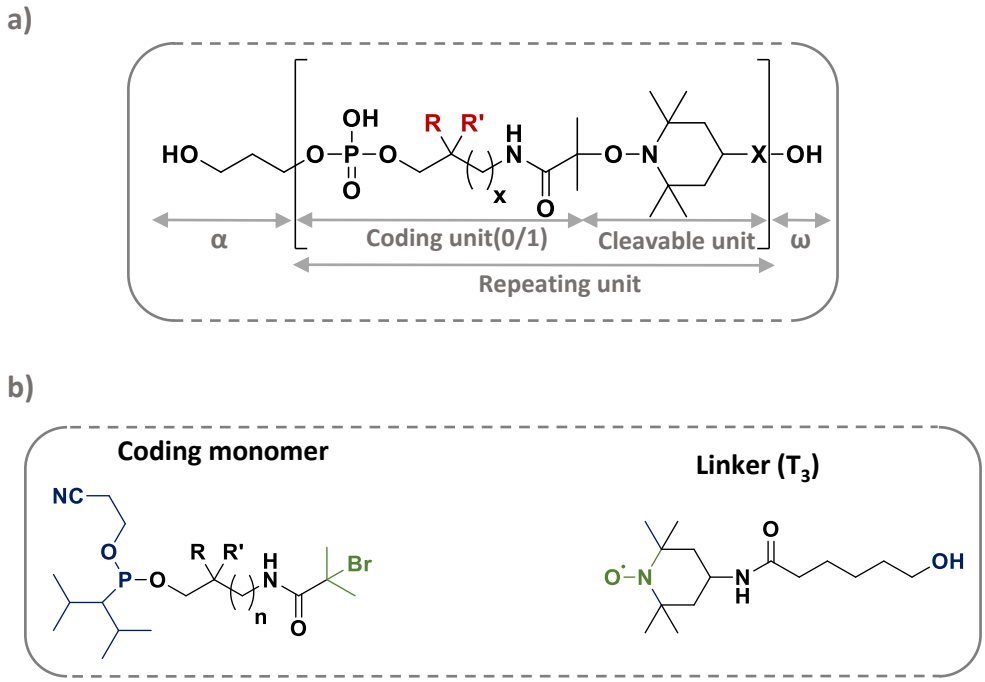


Figure II.1: a) General structure of poly(alkoxyamine phosphodiester)s. **b)** General structures of the monomers employed in the iterative protocol, the blue functional group represent the A and D functionalities while the green ones represent the B and C functionalities.

However, in order to build a MS/MS readable binary alphabet, the backbone of the molecule was slightly modified. Indeed, it has already demonstrated, that is possible to implement a MS/MS readable binary alphabet using building blocks with hydrogen/methyl couples on the side chain.[11] For this reason, the coding monomers employed in this work contain alkyl spacer for bit zero and methyl side chains for bit one (Figure II.2). In addition, the tertiary bromo-functionalized moiety allow to obtain quantitative yields during the NRC and to avoid sequence dependent fragmentation phenomena.[170]

Phosphoramidite monomers

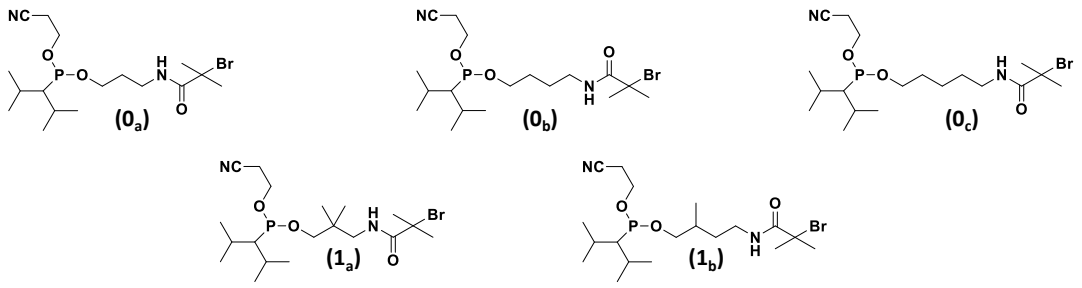
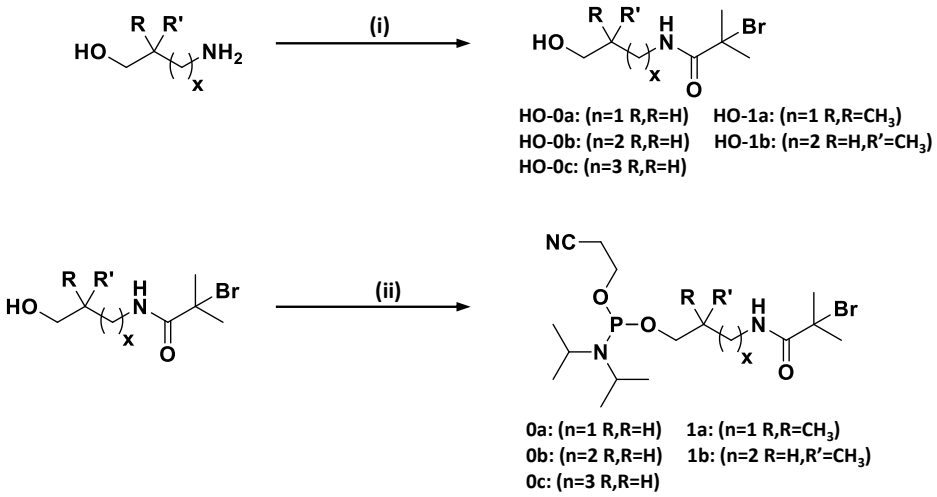


Figure II.2: Structures of the different coding monomers employed for the synthesis of poly(alkoxyamine phosphodiester)s.

In order to test if different monomers could influence the yields and to prove the reproducibility of the developed protocol, the poly(alkoxyamine phosphodiester)s synthesis was performed using different sets of coding monomers. Indeed, in the frame of this project, three different alphabets were synthesized. As depicted in Figure II.2, monomers zero contain several alkyl spacer (*n-propyl, n-butyl and n pentyl*) whereas the monomers one bear methyl groups on the side chain. The synthetic strategy applied for the synthesis of these building blocks is a two steps synthesis, depicted in Scheme II.1.

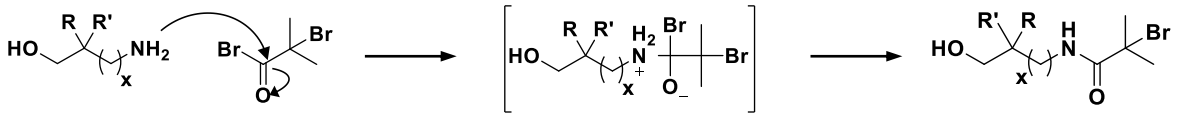


Scheme II.1: General strategy for the coding monomers synthesis. **(i)** acylation: α-Bromoisobutyryl bromide, NEt_3, DCM, 0°C then rt; **(ii)** phosphoramidite coupling: 2-Cyanoethyl N,N-diisopropylchlorophosphoramidite, DIPEA, NMI DCM 0°C then rt.

As previously discussed, the bromo-functional phosphoramidites synthesis was already described from Matyjaszewski.[193] In this work a bromo-functional phosphoramidite monomer was attached to the chain end of a DNA single strand *via* a phosphoramidite coupling. Following polymerization using the tertiary bromide carbon as initiator allowed the synthesis of polymers-DNA bio-hybrids.

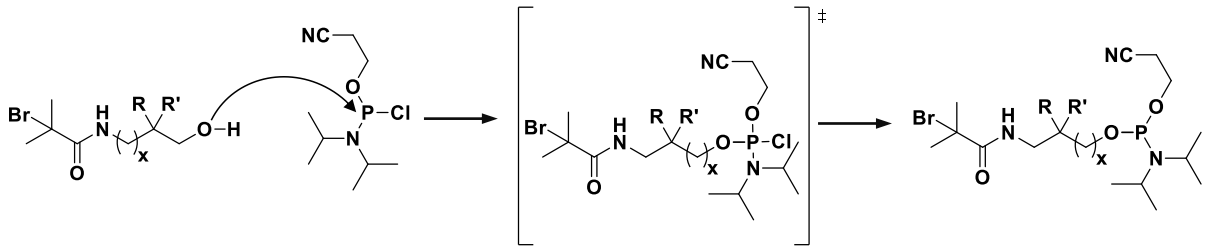
The first step in the building block synthesis is an acylation between the different amino alcohols and α-bromoisobutyryl bromide (Scheme II.2); the reaction is carried out in

anhydrous conditions for 1.5 hours. Afterwards, treatment with 5% NaOH solution is done to hydrolyse the small amount of ester formed in the reaction conditions.



Scheme II.2: Acylation mechanism for the coding monomers synthesis.

The length of the amino alcohols employed and the number of methyl groups doesn't influence the yields which are all between 75% and 85%. The second step is the nucleophilic substitution between the hydroxy function of the synthesized intermediates and the 2-cyanoethyl N,N-diisopropylchlorophosphoramidite. The addition elimination reaction at the phosphorous (III) involves the vacant d orbital of this heteroatom and it occurs *via* the formation of a pentacoordinated transition state (Scheme II.3). [194]



Scheme II.3: Phosphoramidite coupling mechanism for the synthesis of the coding monomers.

The coupling is carried out for two hours in anhydrous conditions and after purification, the final products were recovered in good yields (between 75 and 92%). All the coding monomers were following analyzed by ^{1}H-NMR and ^{13}C-NMR (Figure II.3) and ^{31}P-NMR (see experimental part).

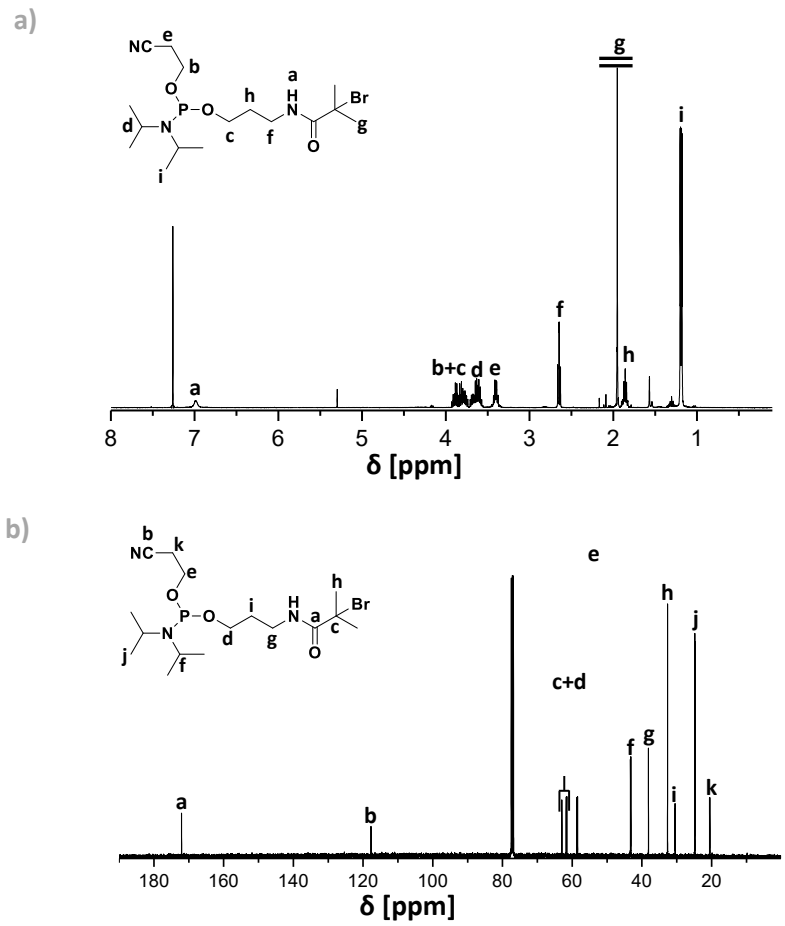
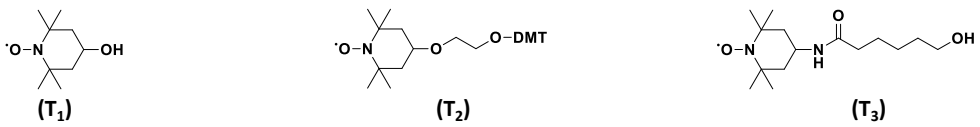


Figure II.3: Example of ^{1}H-NMR (a) and ^{13}C-NMR (b) for monomer 0_a registered in $CDCl_3$.

As discussed previously, the hydroxy-functionalized nitroxide allows the alkoxyamine bond insertion in the polymer backbone. This bond has been found to be extremely important for the MS/MS sequencing. In the frame of this project, three generations of hydroxy-functionalized nitroxides were tested for the development of the iterative protocol. The structures of these building blocks are depicted in

Nitroxide monomers



(T_1) **(T_2)** **(T_3)**

Figure II.4.

Nitroxide monomers

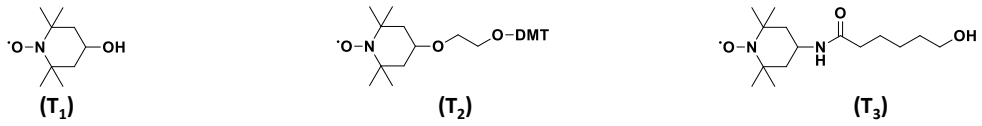
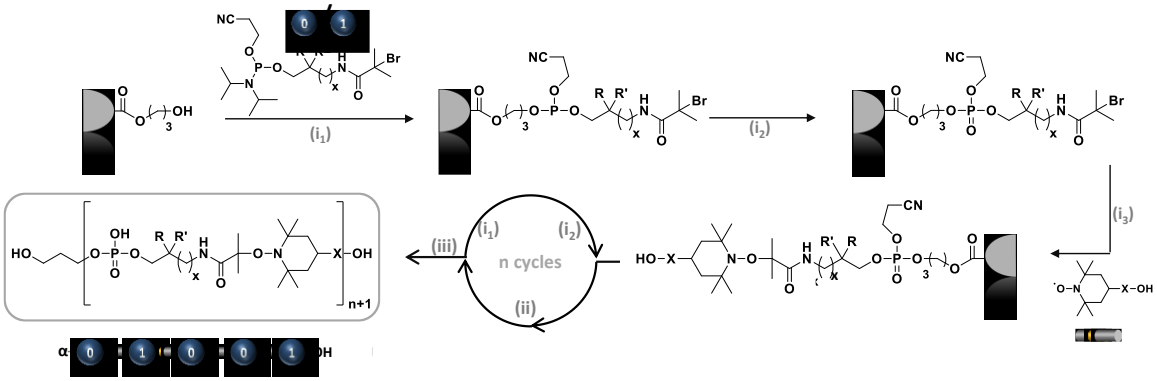


Figure II.4: Hydroxy-functionalized nitroxide monomers tested in this work.

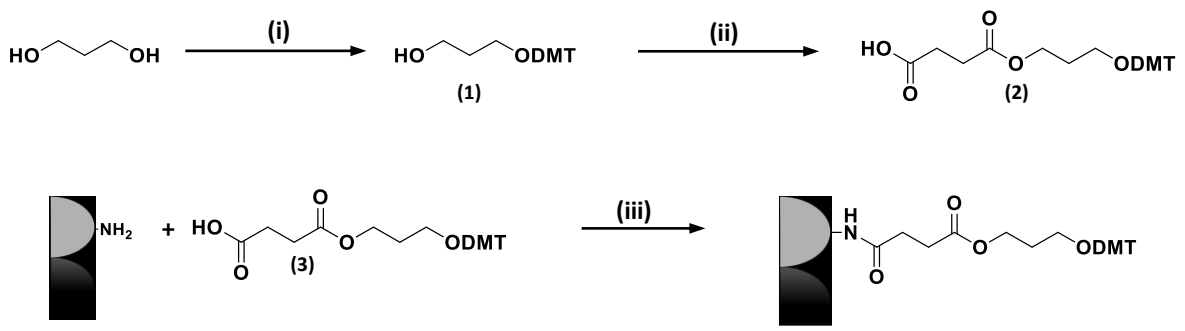
2.2 Solid-state iterative protocol

The solid-state protocol designed for the poly(alkoxyamine phosphodiester)s synthesis relies on phosphoramidite coupling and nitroxide radical coupling (Scheme II.4). The iterative repetition of these two reactions allows to obtain uniform sequence-defined poly(alkoxyamine phosphodiester)s encoding information.



Scheme II.4: General strategy employed for the synthesis of digitally encoded poly(alkoxyamine phosphodiester)s. **(i_1)** phosphoramidite coupling: CH_3CN, tetrazole, rt; **(i_2)** oxidation: I_2, 2,6-lutidine, THF/H_2O, rt; **(ii)** radical–radical coupling: CuBr, Me_6TREN, DMSO; **(iii)** cleavage: piperidine, CH_3CN, rt, then $MeNH_2$, NH_4OH, H_2O, rt.

The multi-step iterative synthesis is performed on a modified amino polystyrene resin cross-linked with divinylbenzene. The solid-support modification is a three steps synthesis which enables the insertion of a basic-cleavable DMT protected succinyl linker (Scheme II.5). The first step is the protection of one alcoholic function of the propanediol with DMT-Cl. Following reaction of the unprotected alcoholic function with succinic anhydride gives the corresponding succinate ester which is coupled with the aminopolystyrene resin to obtain a highly functionalized resin. Subsequently, the loading is calculated via UV spectroscopy. The term loading refers to the number of functional groups attached to the solid support and is determined in mmol/g. Hence, in order to calculate this parameter, a small amount of resin is deprotected using a solution 3% TCA in DCM.



Scheme II.5: Synthetic strategy employed for the solid-support modification. **(i):** DMT protection: DMT-Cl, Pyr, THF, rt; **(ii)**: coupling: succinic anhydride, DMAP, pyridine 37°C; **(iii):** amidification: DCC, DMAP, DCM, rt, overnight.

The concentration of DMT in the solution is calculated using the Lambert-Beer's law. According to this law the absorbance is proportional to the concentration of the DMT in solution. This law is valid in the absorbance range between 0 and 1.[195]

$$A = \varepsilon * l * c$$

A= Absorbance, Ɛ= extinction coefficient ($L \cdot mol^{-1}\ cm^{-1}$), l= length of the sample (1 cm), C= DMT concentration in DCM ($mol \cdot L^{-1}$)

Hence, the concentration of DMT is calculated using the absorbance value at 504 nm, at this wavelength the extinction coefficient is 67000.[196] This parameter is related to the number of functional groups attached to the resin. The UV-spectrometer of the deprotection solution is depicted in Figure II.5.

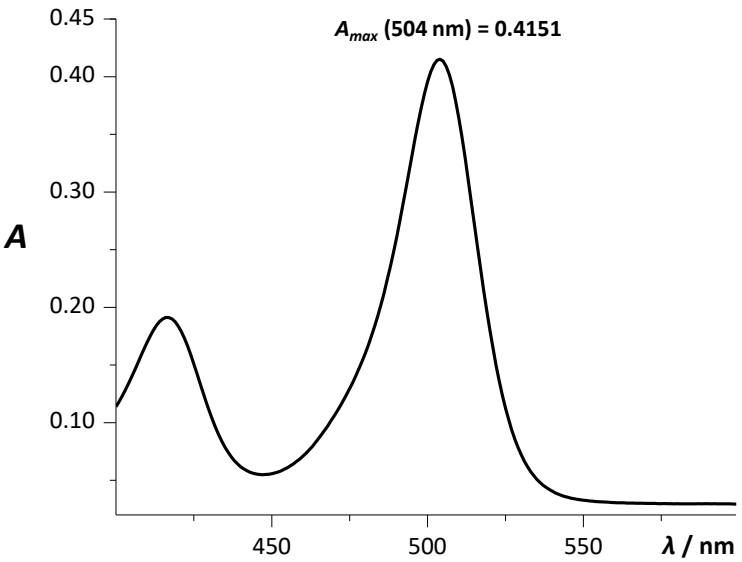
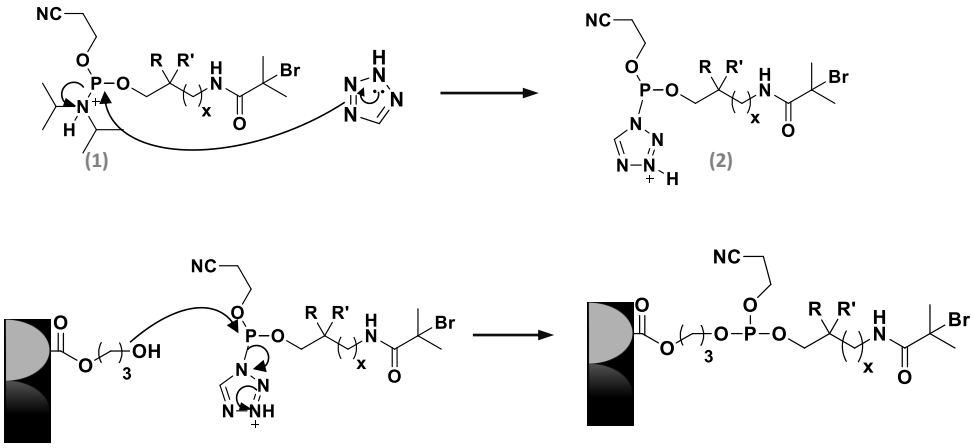


Figure II.5: UV-Spectra obtained for the deprotection solution of the modified resin using a solution of TCA 3 % in DCM. In the example depicted in herein, 78 mg of modified resin were cleaved using a solution of TCA 3 % in DCM. Afterwards, the filtrate was filled to 50 mL. An aliquot was then diluted of a factor of 200 and the absorbance at 504 nm was determined at 0.415. This value was implemented into the Lambert's Beer law to calculate the concentration of DMT cation in the initial solution which was 1.09 mM. Afterwards, the number of mmol in 50 ml were calculated (n=c V)=56.4mmol of DMT were cleaved from 78mg of resin hence the final loading was found to be 0.69mmol/g.

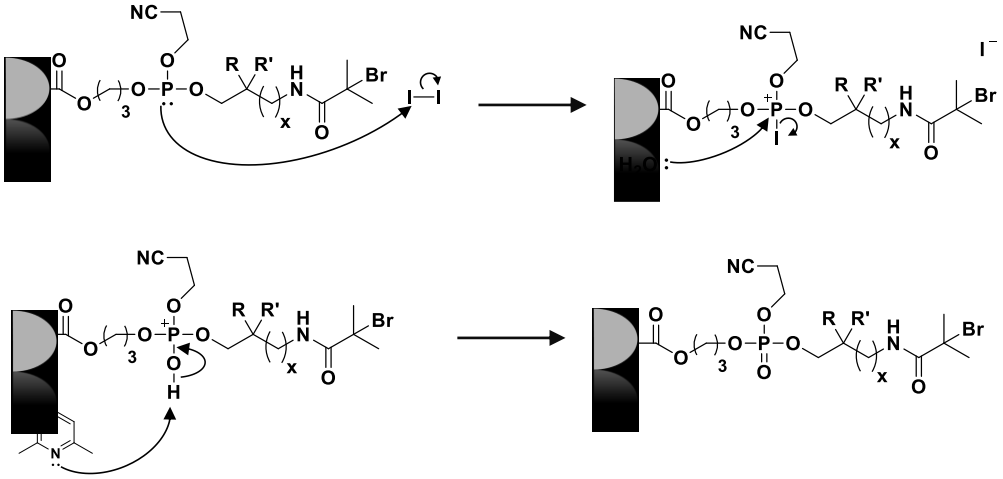
After DMT deprotection in acid conditions, the resulting hydroxy group, attached to the linker, reacts with the phosphoramidite moiety of the selected coding monomer (0/1) (step i_1). The reaction is carried out in anhydrous acetonitrile using tetrazole as catalyst.[197] During the reaction an equilibrium is established between the intermediate (1) and (2) in Scheme II.6, the latter one reacts with the alcoholic function generating the corresponding phosphite.



Scheme II.6: Mechanism of the phosphoramidite coupling on solid support.

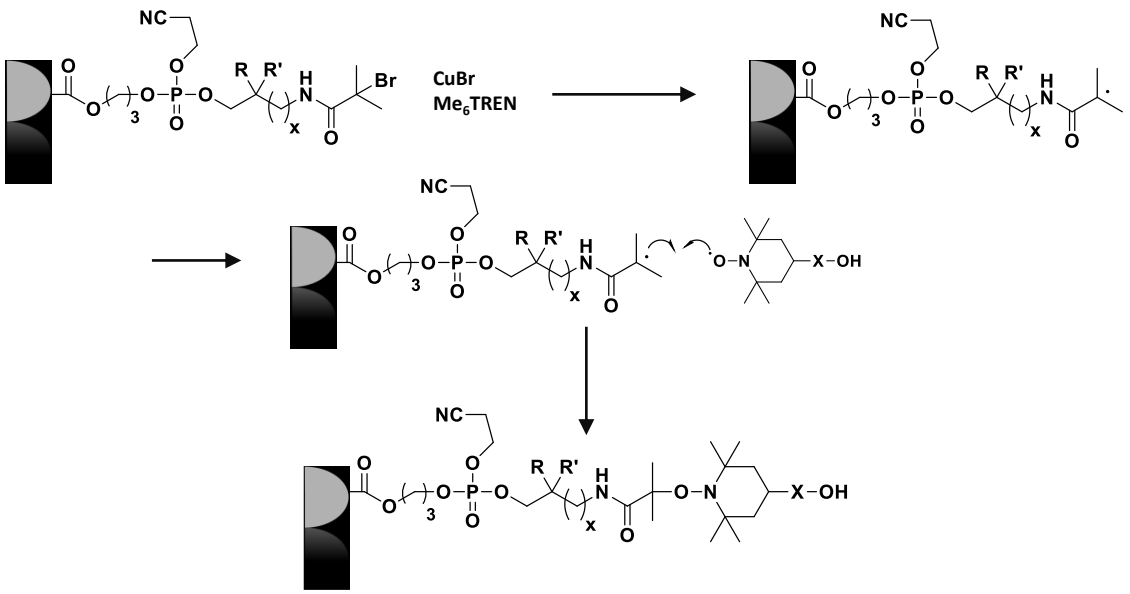
The resulting phosphite is following oxidized into a more stable phosphate using a solution of iodine, water and lutidine in THF (i_2). The reaction of I_2 with the phosphite generates a

tetrahedral intermediate. Following iodine displacement gives the corresponding phosphate (Scheme II.7).



Scheme II.7:Mechanism of phosphite oxidation into a phosphate.

The step (*ii*) is the nitroxide-radical coupling (NRC). In this reaction, the bromo-functional resin generates a carbon-centered radical in presence of copper bromide and a ligand (Me_6TREN). The radical is following spin trapped by the nitroxide moiety of the TEMPO-based monomer, generating the corresponding alkoxyamine bond. The reaction mechanism is depicted in Scheme II.8. All the nitroxides employed in this work contain a hydroxy terminal group which can react with the phosphoramidite moiety of the following coding monomer. Hence, after the first cycle, the growing chain can be involved in several other cycles until the desired sequence is achieved.



Scheme II.8: Nitroxide radical coupling mechanism on solid support.

Afterwards, Cyanoethyl deprotection is done using a solution 10% piperidine in acetonitrile. The cleavage is performed using a solution of ammonia/methylamine (1/1). After, resin filtration and lyophilisation allow to recover the digital polymers in god yields.

2.3 Synthesis of sequence-defined poly(alkoxyamine phosphodiester)s using 4-Hydroxy-TEMPO

The first synthesis of digitally encoded poly(alkoxyamine phosphodiester)s was done using the coding monomer 0_a and the commercially available 4-Hydroxy-TEMPO in step (*ii*). The phosphoramidite coupling and the nitroxide radical coupling were performed using the same conditions previously developed in our group, [11, 95] increasing the reaction times to one hour for both reactions. Moreover, after step (*ii*) the copper excess was removed with a 2% solution of sodium EDTA in water. For the first test, the iterative cycle was repeated 5 times in order to obtain a sequence-defined decamer. The resulting product was analyzed by ESI-HRMS. As depicted in Figure II.6, the targeted molecule was detected at different charge states ranging from 2^- to 4^- (peaks labelled in green). The formation of these series is due to the negative charge carried by the phosphate groups.

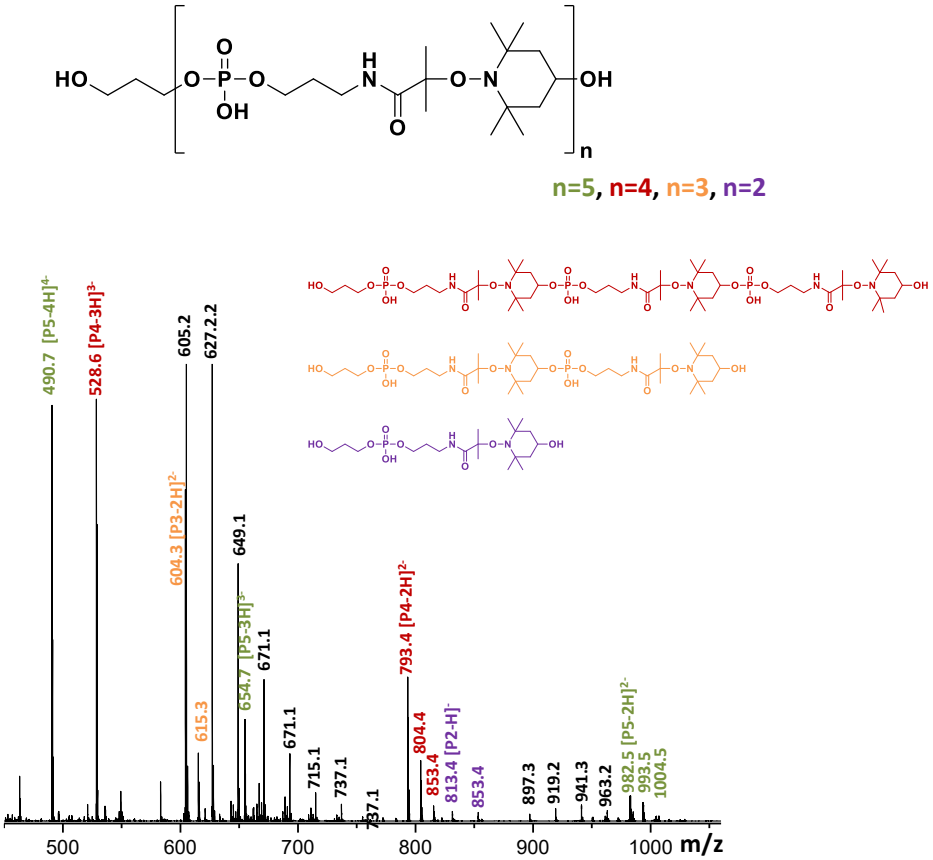
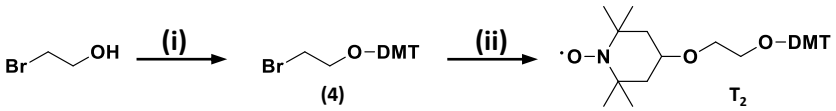


Figure II.6: ESI-MS recorded for a decamer synthesized using monomer 0_a in step (*i*) and 4-Hydoxy-TEMPO in step (*ii*). Peaks labelled in green correspond to the desired product; the peaks labelled in other colours correspond to products generated form incomplete phosphoramidite coupling. The structures in red orange and purple correspond to the structures of the lower congeners.

The analysis showed also the presence of lower congeners which were originated from incomplete phosphoramidite coupling (i_1) (red, orange and purple labelled peaks in Figure II.6). Obviously, the *non*-quantitative yields in this step led to the synthesis of *non*-uniform polymers. Hence, these molecules were not suitable for storing information. Indeed, as already discussed, only sequence-defined polymers having uniform molecular weight distribution and fully defined primary structure ensure a reliable molecular encryption. In addition, the high rate of H^+/Na^+ exchanges originated from the polymers and from the sodium EDTA (black labelled peaks) made the MS spectra difficult to interpret.

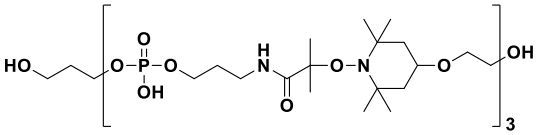
2.4 Design of optimized hydroxy-functionalized nitroxides

The *non*-quantitative yields obtained in step (i_1) led to the synthesis of heterogeneous polymers unsuitable for storing information. The incomplete yields obtained during the phosphoramidite coupling were probably due to the low reactivity of the secondary alcoholic function of the 4-Hydroxy-TEMPO. In order to confirm this hypothesis, the monomer T_2 was synthesized. This building block, containing a protected primary alcoholic function, was synthesized following the synthetic strategy depicted in Scheme II.9. The first step is the protection of the 2-bromo ethanol with DMT-chloride. Williamson etherification between the intermediate 4 and the 4-Hhydroxy-TEMPO gives the desired monomer.



Scheme II.9: General strategy for the synthesis of linker T_2. **(i)** DMT protection: DMT-Chloride, THF, pyridine, rt; **(ii)** etherification: NaH, THF 0°C, 4-hydroxy TEMPO 0°C then rt.

Subsequently, the monomer 0_a and the linker T_2 were tested for the synthesis of a hexamer (3 repeating units). After step (*ii*), the DMT deprotection allowed to release the primary alcoholic function involved in the following phosphoramidite coupling. Moreover, a 2% solution of 4,4'-di-n-nonyl-2,2'-dipyridyl in dichloromethane was used to wash the copper excess. Indeed, is well known from the literature that bipirydines can complex copper and other metals efficiently [198, 199] moreover the DCM, keeping the resin swollen, improves the copper extraction. After performing three iterative cycles using these conditions, the resulting compound was analyzed *via* negative ESI-HRMS that confirmed the homogeneity of the resulting sequence-defined polymer (



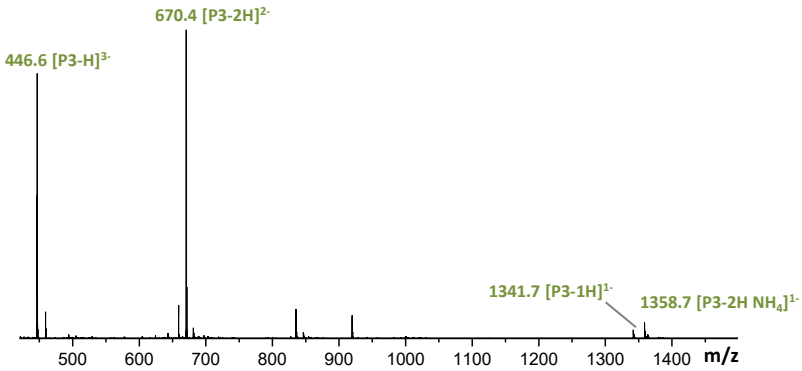


Figure II.7). Indeed, the targeted product was detected at different charge states moreover, impurities corresponding to *non*-quantitative coupling yields or side reactions were not detected in ESI-HRMS. Interestingly, the use of the bypiridine solution as copper scavenger reduces the amount of H/Na^+ exchange making the resulting MS spectra less complicated.

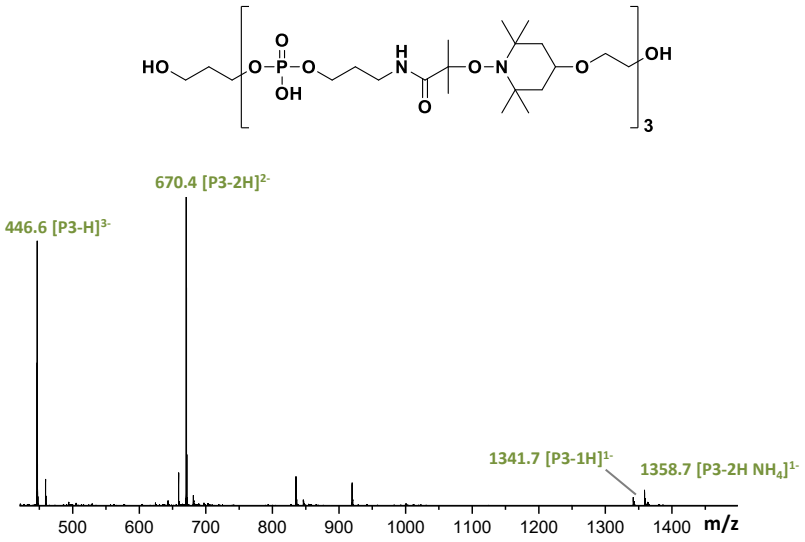
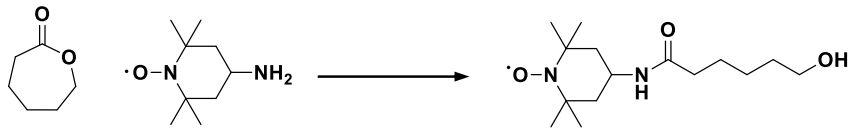


Figure II.7: ESI-HRMS registered for a hexamer synthesized using linker T_2 and coding monomer 0_a.

The iterative synthesis performed using nitroxide T_2 confirmed that the primary alcohol leads the phosphoramidite coupling to quantitative yields. Even though, the replacement of 4-Hydroxy-TEMPO with linker T_2 allowed the synthesis of uniform digital polymers, this compound was not suitable for the development of the iterative protocol for several reasons. First of all, the aim of the project was to obtain sequence-defined poly(alkoxyamine phosphodiester)s *via* a chemoselective approach in order to avoid deprotection steps and make the iterative synthesis faster. Moreover, this monomer was recovered in low yields after two steps. As said before, the building blocks synthesis should be achieved in few steps and the final product should be recovered in high yields. For all these reasons, the monomer T_2 was replaced with the hydroxy-functionalized nitroxide T_3. This building block was obtained opening the Ɛ-caprolactone with 4-Amino-TEMPO.



Scheme II.10: General strategy for the synthesis of monomer T_3. Experimental conditions: 130°C reflux 2.5 h.

The reaction doesn't need anhydrous conditions and the product is recovered in 73% yield in a single step. The monomer was following analyzed by ESI-HRMS (Figure II.8) which shows the targeted monomer as predominant species. Moreover side products were not detected in ESI-HRMS. The easier synthetic access made the monomer T_3 a convenient building block for the implementation of the solid-state protocol. The iterative protocol was then performed using the reaction conditions employed in the previous experiment using the monomer 0_a in step (*i*) and the linker T_3 in step (*ii*). The resulting tetramer was analyzed by

negative ESI-HRMS that confirmed also in this case the homogeneity of the resulting sequence-defined polymer.

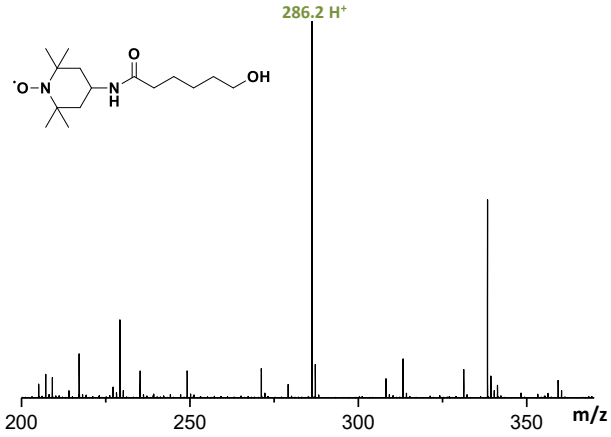


Figure II.8: Positive ESI-HRMS recorded for the linker T_3.

Hence, the complete yields for both steps in the iterative synthesis and the efficiency of the strategy developed in this work. Indeed, as shown in Figure II.9, the product was detected at different charge states. A small amount of the sodium adduct of the polymer containing only a deprotonated phosphate group was also detected at low intensity. Moreover, peaks corresponding to side products were not detected.

Interestingly, it was found that the optimized nitroxide T_3 improves the MS analysis and the following read-out. Indeed. as will be discussed in the sequencing section, the longer hexanamide chain provided by nitroxide T_3, keeping the phosphate groups at an optimal distance, enables their simultaneous deprotonation during the MS analysis facilitating the sequencing.[170]

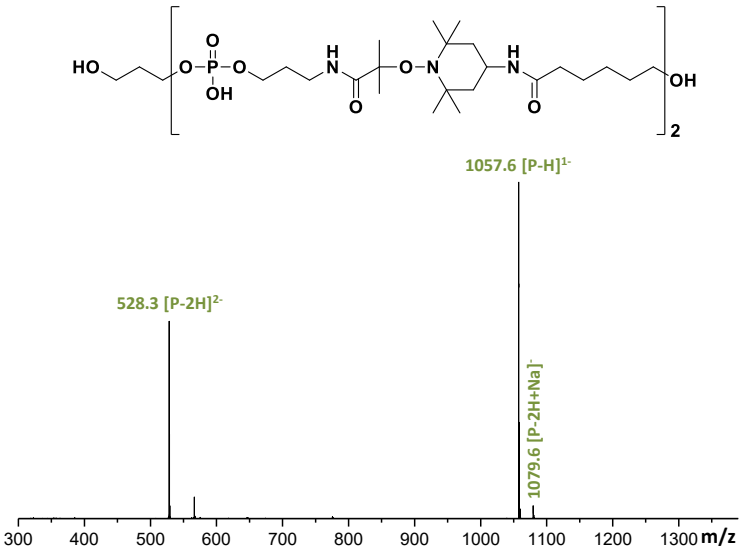


Figure II.9: Negative ESI-HRMS recorded for entry 3 Table II.1

In order to show the reproducibility of the method with different monomers, three different coding alphabets (a, b and c in Figure II.2), were used in combination with the linker T_3 for the synthesis of molecularly-encoded poly(alkoxyamine phosphodiester)s (Table II.1). Moreover, the optimization of the reactions conditions allowed to perform the phosphoramidite coupling using 3 eq. of coding monomers (0 or 1) for 30 minutes and the radical-radical coupling using 5 eq. of linker (T_3) for 40 minutes. In all the cases, the resulting digital polymers were recovered in good yields (between 40 and 60%). Interestingly, the coding for this class of molecules doesn't influence the yields or the sequencing as it has been noticed for other digital polymer classes.[11]

Table II.1: ESI-HRMS Individual characterization of the digital polymers synthesized and studied in this work.

	sequence	*mass*	*m/z$_{th}$*	*m/z$_{exp}$*
1	α-$0_aT_10_aT_10_aT_10_aT_1$	1967.0	490.7458[b]	490.7465
2	α-$0_aT_20_aT_20_aT_2$	1339.7	1339.6923[a]	1339.6910
3	α-$0_aT_30_aT_30_aT_3$	1549.9	515.6196[a]	515.6187
4	α-$0_aT_30_aT_30_aT_30_aT_3$	2041.2	509.2819[b]	509.2814
5	α-$1_aT_30_aT_30_aT_31_aT_3$	2097.2	523.2975[b]	523.2977
6	α-$(1_aT_3)_2(0_aT_3)_31_aT_31_aT_3$	4118.4	514.0417[c]	514.0429
7	α-$0_bT_30_bT_30_bT_31_bT_3$	2097.2	523.2975[b]	523.2970
8	α-$0_bT_30_bT_31_bT_31_bT_3$	2125.3	530.3035[b]	530.3060
9	α-$0_bT_31_bT_3$	4174.4	521.0496[c]	521.0495
10	α-$0_cT_30_cT_30_bT_30_cT_3$	2153.3	537.3132[b]	537.3123

a $[M - 3H]^{3-}$, b $[M - 4H]^{4-}$ c $[M - 8H]^{8-}$, measured at isotopic maximum.

Moreover, side products originated from disproportionation reaction during the NRC were not detected. These observations are in agreement with the literature which indicates that the disproportionation reactions for a tertiary radical in the NRC are usually observed at high

temperatures.[155, 200] The disproprtionation reactions are due to the hydrogen abstraction in γ position of the alkyl radical. The side product originated from this process are a hydroxylamine generated from the nitroxide and an alkene originated from the alkyl radical.[201]. In Figure II.10, the products originated from a hypothetical disproportionation occurring during step (ii) are depicted.

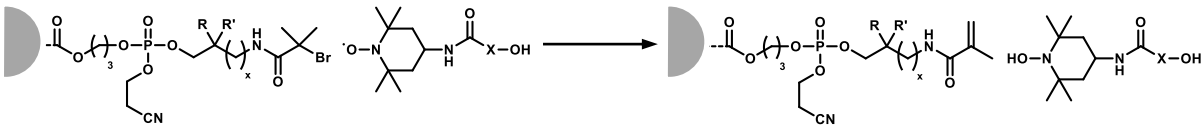


Figure II.10: Products originated from a hypothetical disproportionation reaction in step (*ii*).

In addition, the combination of copper and ligand doesn't lead to chain breaks as it has been already reported for DNA.[202] All the sequences in Table II.1 were analyzed by ESI-HRMS which show in all the cases that the measured molecular weights are in line with the theoretical masses. Moreover, the presence of side products was never detected. All these elements prove the reliability of the molecular encryption. The digital polymers were analyzed by ^{1}H-NMR and ^{31}P-NMR (see experimental section). The uniformity of the sequence-defined hexadecamer is further confirmed by SEC experiments in water/acetonitrile (Figure E.3). Thanks to its efficiency and reproducibility, the developed solid state strategy result in a powerful tool for the synthesis of digitally encoded sequence-defined polymers.

3 Conventional sequencing

As discussed in the introduction of this chapter, the poly(alkoxyamine phosphodiester)s were designed with the aim to improve the tandem MS readability of poly(phosphodiester)s The MS and the MS/MS experiments presented in the first part of this chapter were performed by Jean-Arthur Amalian at Marseille University. During the MS/MS experiments a precursor ion, generated in ESI-HRMS, is selected and fragmented in a collision cell by impact with an inert gas. The conversion of kinetic energy in internal energy leads to the fragmentation of the molecular ions in smaller fragments.

In these conditions, the labile alkoxyamine bond of poly(alkoxyamine phosphodiester)s is preferentially cleaved; the rupture occurs at low ionization energies (between 15 to 40 eV). This homolytic breaking produces a predictable and easy to interpret fragmentation pattern that allows deciphering the original sequence. During the tandem MS process two series of fragments are produced according to the chain end-groups (α or ω) (Figure II.11). Thus, relying on the mass of the different building blocks, generated in these conditions, the sequence in the polymer chain can be determined from the m/z values of the obtained fragments (Figure II.11 b). The first MS/MS sequencing of poly(alkoxyamine

phosphodiester)s was done on a homopolymer containing $0_a/T_3$ (entry 4 Table II.1). As shown in Figure II.11, the precursor ion selected for the fragmentation in MS/MS is the one containing all the phosphate deprotonated, $[M-4]^4$. The fully deprotonated species are selected as precursor ions because their fragmentation in MS/MS generates two fragmentation series in which each fragment contains a single charge signature. Hence, the resulting fragmentation patterns are easier to correlate to the original structure. On the other hand, partially ionized polymers, containing randomly deprotonated phosphate groups, generate complicated MS/MS spectra because of the formation of fragments with different ionization states. Interestingly, it was found that the long side chain of the nitroxide T_3 is extremely useful for the sequencing. Indeed, the hexanamide chain, placing the phosphates at an optimal distance, facilitates the simultaneous deprotonation of these groups increasing the abundance of the fully deprotonated species.[170] Moreover, the tertiary radical, generated after homolytic cleavage of bit 0 and 1 avoids sequence-dependent fragmentations as it was already observed for other classes of digital polymers.[11] In Figure II.11, the MS/MS sequencing of a short homopolymer coding for 0_a-T_3-0_a-T_3-0_a-T_3-0_a-T_3 is depicted. After the sequencing of short chains containing three different alphabets, the read-out of longer polymers was performed. Indeed, the read-out of a hexadecamer, coding for 8 bits was tested in MS/MS. Also in this experiment the fully deprotonated species was selected as precursor ion for the MS/MS experiment (red labeled peak in

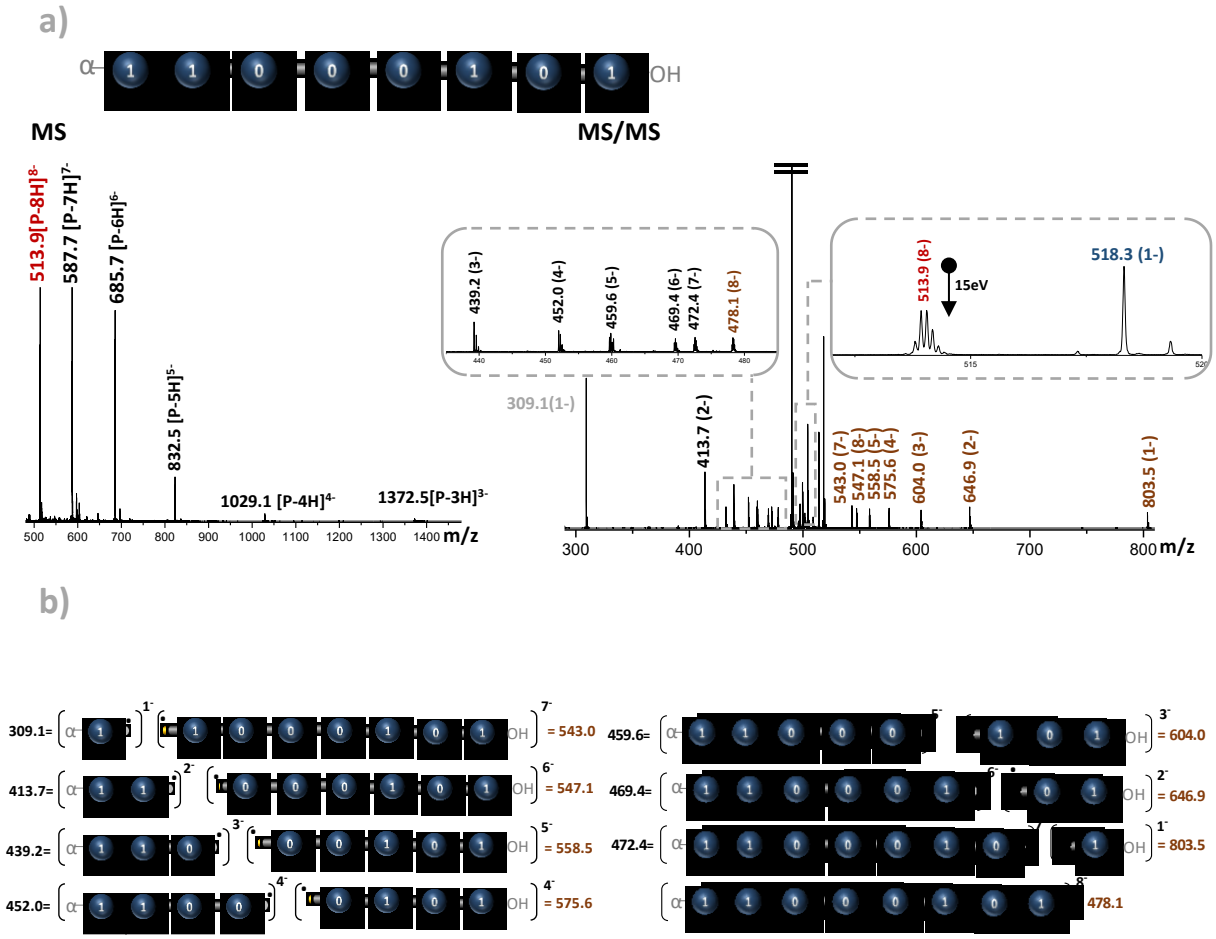


Figure II.12). As shown in

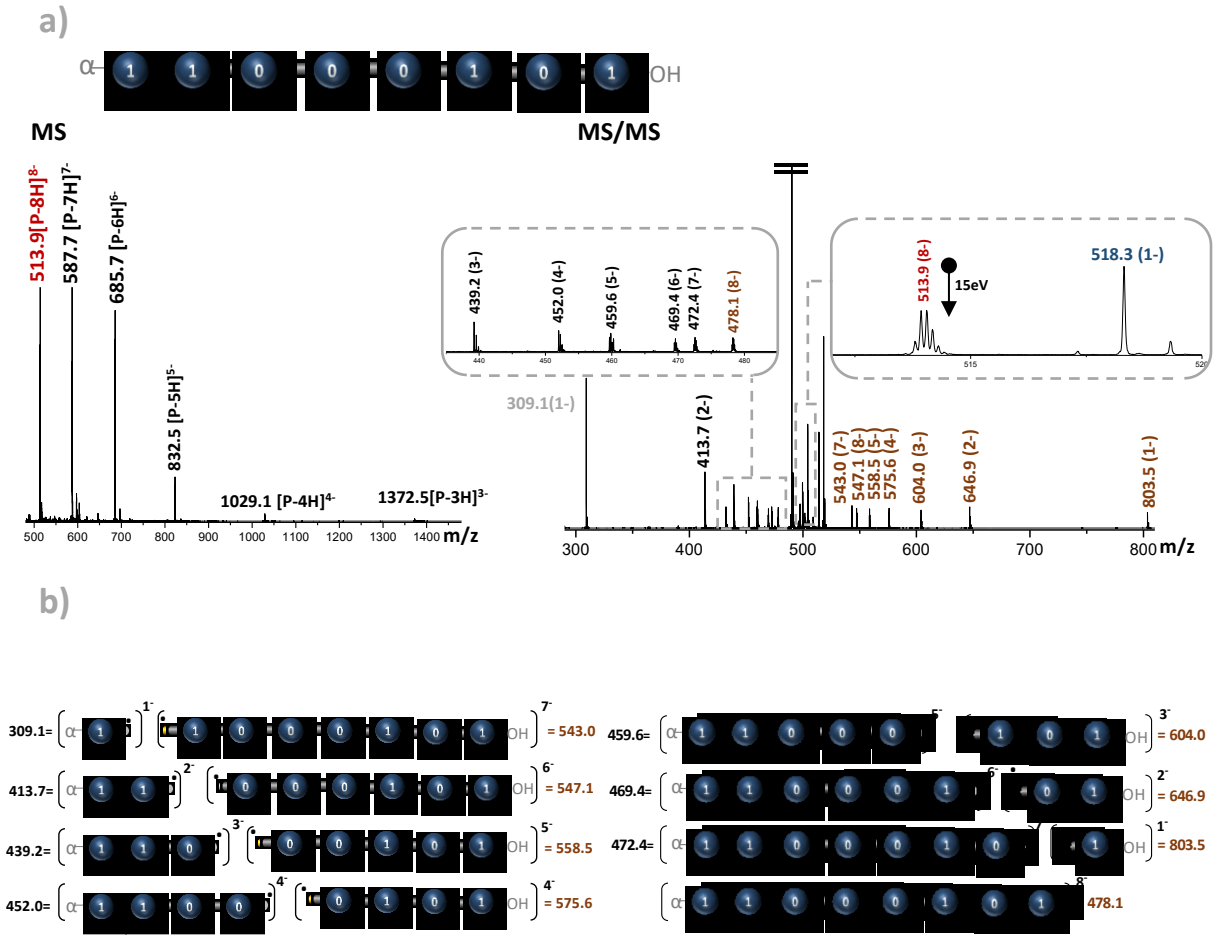


Figure II.12 b, all the fragments originated from the alkoxyamine bond cleavage were detected in MS/MS, allowing the easy sequencing even of longer polymer chains.

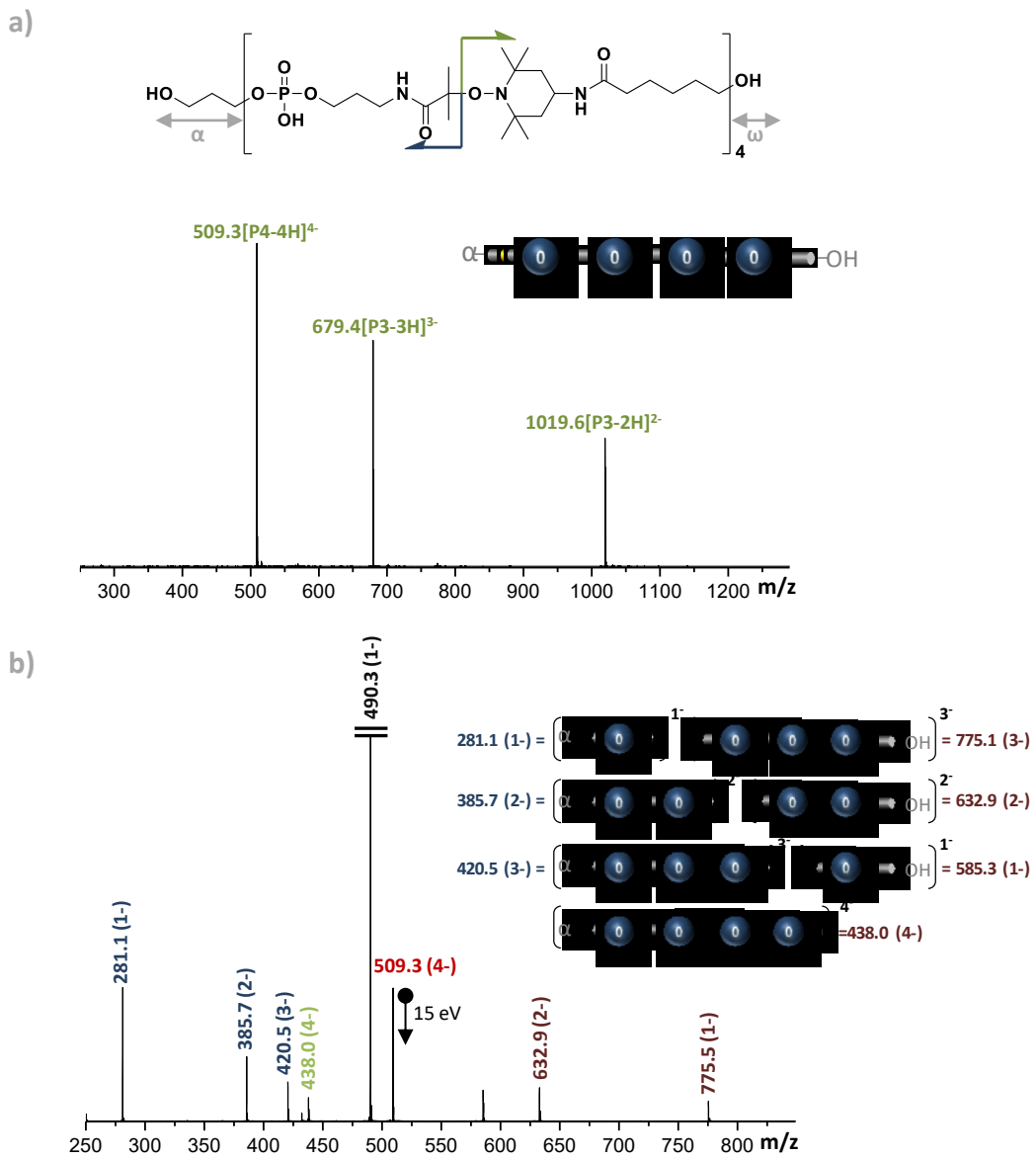


Figure II.11: Example of Sequencing of poly(alkoxyamine phosphodiester)s. **a)** Negative ion mode ESI mass spectrum of a poylmer coding for, α-$0_aT_30_aT_30_aT_30_aT_3$ (entry 6 tab. 2). **b)** MS/MS sequencing of [P – 4H]$^{4-}$ at m/z 509.3 (red labelled) (collision energy : 15 eV) and corresponding dissociation scheme.

This last example shows that even for longer sequences it is possible to decipher the original structure relying on the different fragments generated in MS/MS conditions. All the synthesized digital polymers were successfully sequenced by MS/MS highlighting the efficiency and the reproducibility of this sequencing technique for this class of digital polymers.

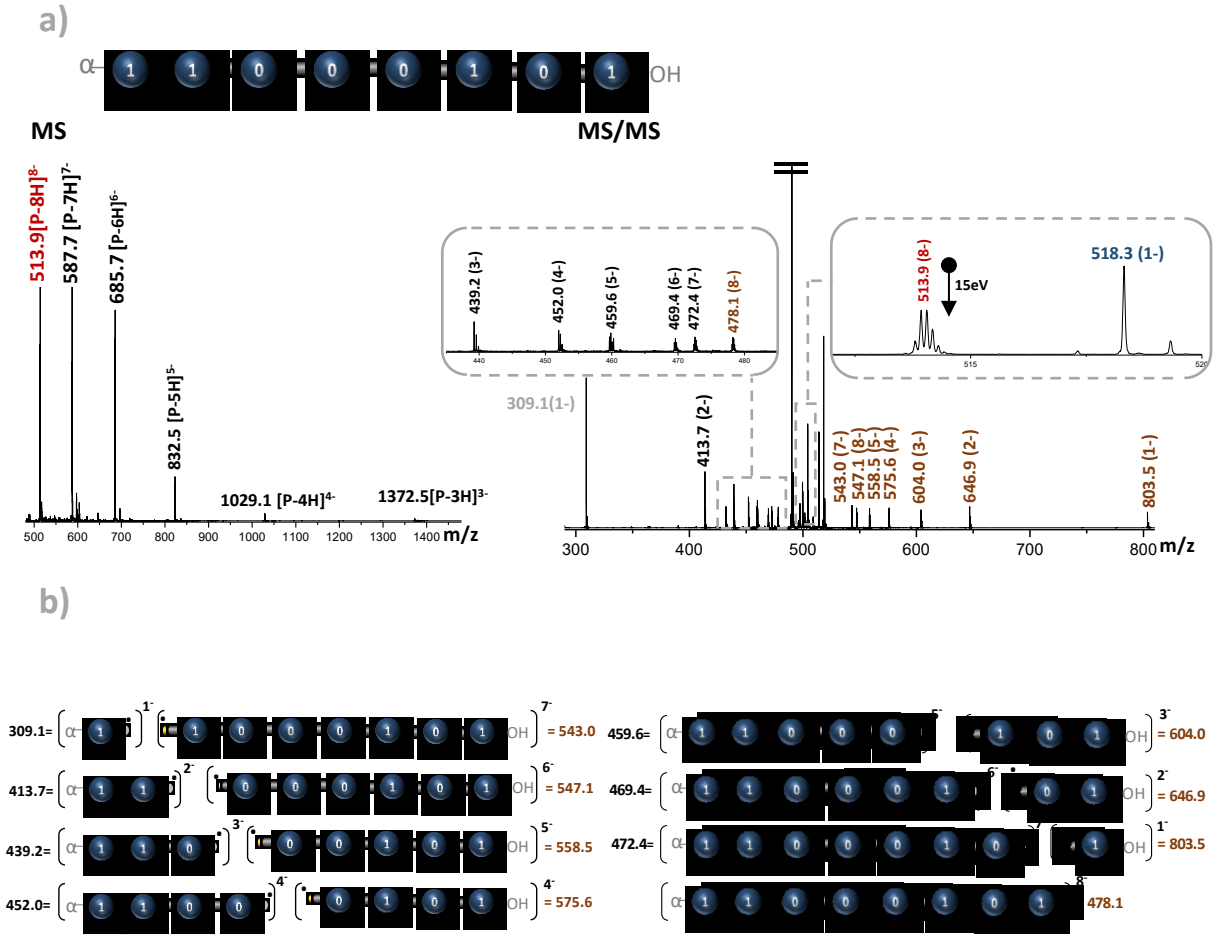


Figure II.12: Example of sequencing for a polymer coding for one byte. **a):** Negative mode ESI-MS characterization of a encode copolymer synthesized using the building-blocks 0a/1a/T_3 (Table II.1, entry 5) (right), MS/MS sequencing of the same polymer after CID of precursor $[M - 8H]^{8-}$ at m/z 513.9 (red labeled peak). **b):** Representation of the chain fragmentation recovered in MS/MS for the same copolymer.

4 Fragmentation-free sequencing approach

4.1 Introduction

Thanks to its efficiency and to the possibility to sequence digital polymers in a relatively short time, tandem MS is nowadays the standard procedure for the sequencing of digital polymers.[13] However, in the last years other sequencing techniques were developed for the read-out of digital polymers. Another convenient approach to facilitate the sequencing could be the fragmentation-free sequencing (F^2S) developed by Keough and co-workers.[203] This strategy, was initially applied for the screening of biologically active peptides libraries *via* MS analysis.[16] The concept relies on the generation of small percentages of specific truncated sequences and the full-length polymer during the iterative synthesis. The addition of a small percentage of a capping agent during each monomer coupling, allows the formation of sequence-specific termination products. Hence, after *n*+1 cycles, each resin bead contains a mixture of the original polymer and *n* sequence specific truncated fragments (Figure II.13).

The sequencing relies on the weight difference between the termination products produced during the iterative synthesis. The advantage in using F^2S lies in the fact that the sequencing by MS is faster and easier as it has been already demonstrated for natural products.[16] Indeed, this method allows the sequencing bypassing chemical or MS/MS fragmentations which usually lead to the formation of complicated MS/MS spectra.

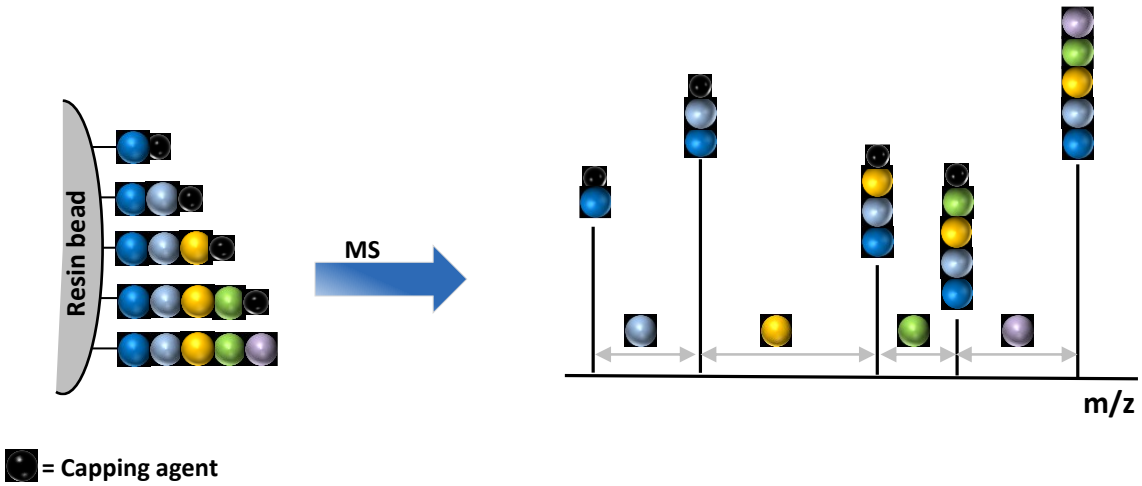


Figure II.13: General concept of fragmentation free sequencing.

In the frame of the PhD thesis, the poly(alkoxyamine phosphodiester) class was assessed for F^2S approach. This project was developed in collaboration with the group of Prof. Hans Börner at Humboldt university of Berlin and is discussed in this section.

4.2 General strategy for F^2S using poly(alkoxyamine phosphodiester)s

The general solid-state strategy developed in this work is depicted in Figure II.14. The solid support employed for the iterative synthesis is a polystyrene-based resin, modified with a base cleavable linker to whom is attached an amino acidic sequence. The polypeptide is coupled to the resin in order to ensure the MS analysis in a convenient m/z range also for short fragments and to improve the polymers ionization during the MS experiments. The poly(alkoxyamine phosphodiester) chains are connected to the polypeptide *via* an hydroxy carboxylic acid linker. The truncated sequences are produced during the phosphoramidite coupling. Indeed, a small percentage of a specifically designed capping agent is employed in this step in combination with the desired coding monomer (0_a/1_a).

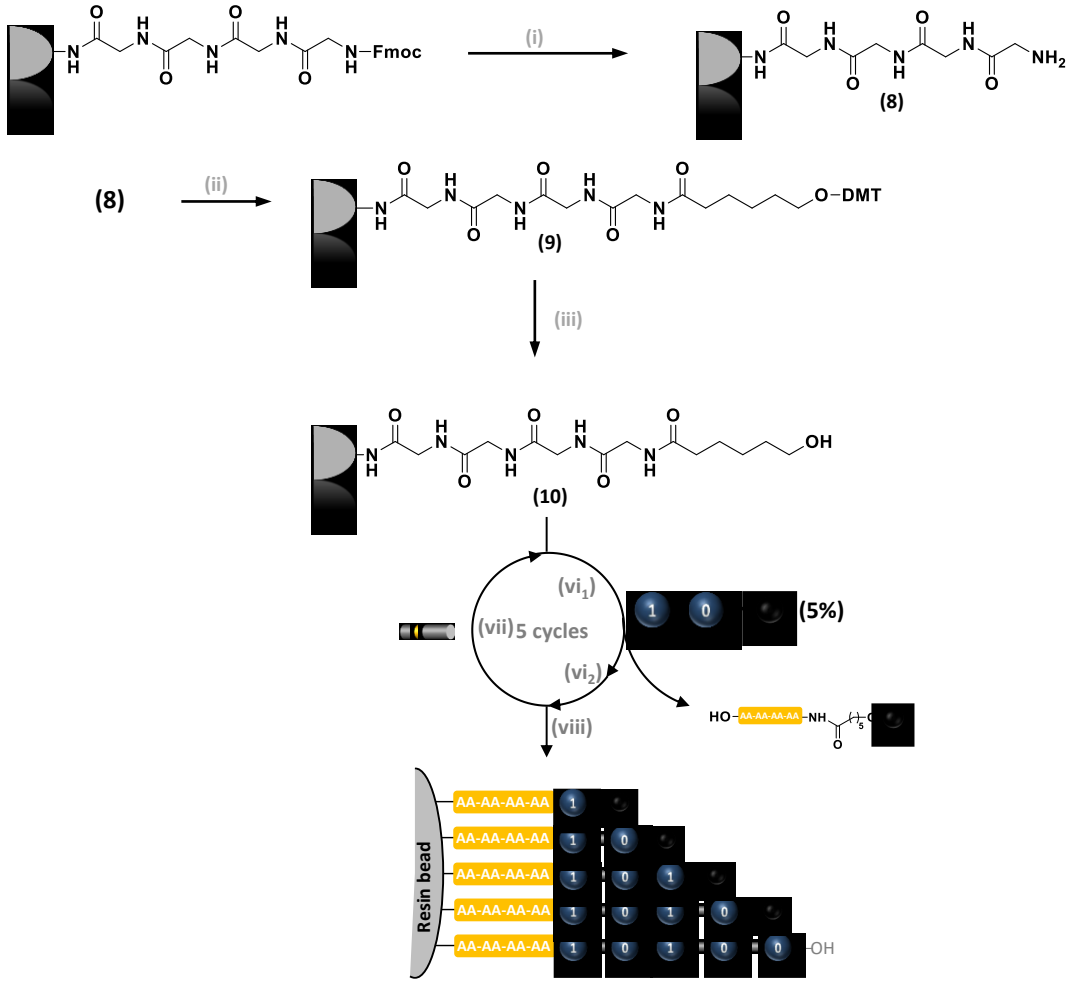


Figure II.14: General strategy employed in this work. **(i)** Fmoc deprotection: 25% pyperidine in DMF, 20 min 2 times; **(ii)** linker coupling: L_1, PyBop, DIPEA,DCM, overnight; **(iii)** DMT deprotection: TCA 3% in DCM, 10 min. 2 times; **(iv_1)** phosphoramidite coupling:, Acetonitrile, tetrazole, rt; **(iv_2)** oxidation: I_2, 2,6-lutidine, THF/H_2O, rt; **(v)** radical–radical coupling: CuBr, Me_6TREN, DMSO; **(vi)** cleavage: piperidine, acetonitrile, rt, then NH_4OH, H_2O, rt, overnight.

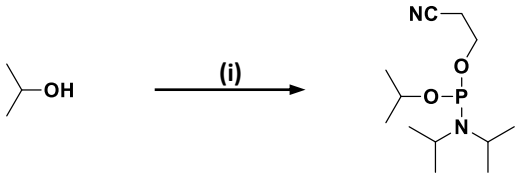
The iterative protocol for the synthesis of poly(alkoxyamine phosphodiester)s is performed using a mixture of the selected coding monomers (0_a or 1_a) and the capping agent in step (*iv*) whereas, the nitroxide T_3 is employed in step (*v*). The synthesis of the coding monomers and the linker was done following the same conditions described previously in this chapter, whereas the synthesis of the capping agent is discussed in the following section. The resin modification, the polypeptides synthesis and all the MS analyses presented in this section were performed by the group of Prof. Hans Börner at Humboldt university of Berlin.

4.3 Results and discussion

4.3.1 Synthesis of the capping agent

The formation of the truncated fragments relies on the reaction of the growing chain with the 5% of capping agent added in the reaction mixture during the phosphoramidite coupling. The synthesis of this monomer was done in a single step *via* phosphoramidite coupling

between 2-isopropnaol and the 2-Cyanoethyl N,N-diisopropylchlorophosphoramidite (Scheme II.11).

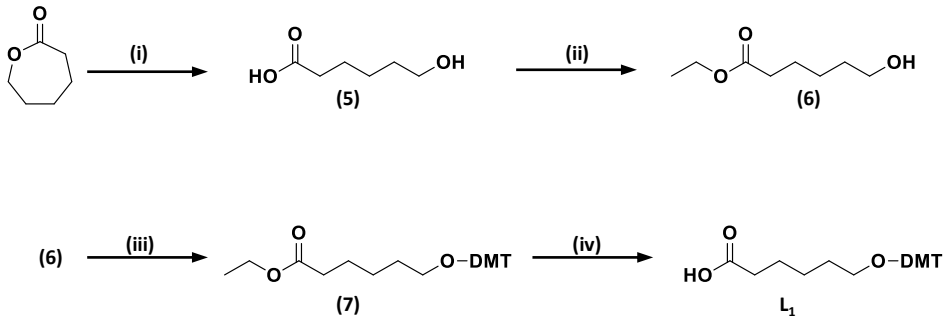


Scheme II.11: Capping agent synthesis. **(i):** 2-Cyanoethyl N,N-diisopropylchlorophosphoramidite, DIPEA, dichloromethane 2.5h, rt.

The reaction mixture is stirred for 2.5 hours and after purification the final product is obtained in 85% yields. The product was following analyzed by ^{1}H-NMR, ^{13}C-NMR and ^{31}P-NMR (see experimental part).

4.3.2 Iterative protocol optimization and F^{2}S sequencing

The first test to asses poly(alkoxyamine phosphodiester)s for F^{2}S was done synthesizing an amino acidic sequence containing two MTT protected lysines and two glycines. The lysines fragments were selected in order to obtain a better ionization in positive ESI-MS. As depicted in Figure II.14, intermediate 16 was coupled to the growing chain using PyBop as activating agent. The silyl ether was then deprotected using a solution of TBAF in THF. Subsequently, the hydroxy terminal growing chain was involved in the iterative protocol described in Figure II.14. For the first test, the capping agent was not employed during step (*iv*) and only one cycle was performed. Following protecting groups removal and cleavage in ammonia gave a colourless oil. The product was analyzed in positive and negative ESI-MS. The MS analysis evidenced the non-quantitative yield during the, phosphoramidite coupling. Indeed, the predominant species in the spectra was represented by the polypeptide connected to the linker (intermediate 10 in Figure II.14). The low yield in step (*iv*) was probably due to the presence of fluoride anions originated during the TIPS deprotection. Indeed, these ions increase the PH, reducing the activity of the tetrazole as catalyst. In order to bypass this problem, a new monomer containing an acid cleavable protecting group was employed as linker. The synthesis of this monomer (L_1) is depicted in Scheme II.12



Scheme II.12: Synthetic strategy employed for the synthesis of monomer L_1. **(i)** lactone opening: NaOH, H_2O 3h; **(ii)** esterification: PTSA, EtOH, 65°C, 4h; **(iii)** DMT protection: DMT-Cl, THF, pyridine, 3h, rt; **(iv)** ester hydrolysis: NaOH, EtOH, 6h, rt.

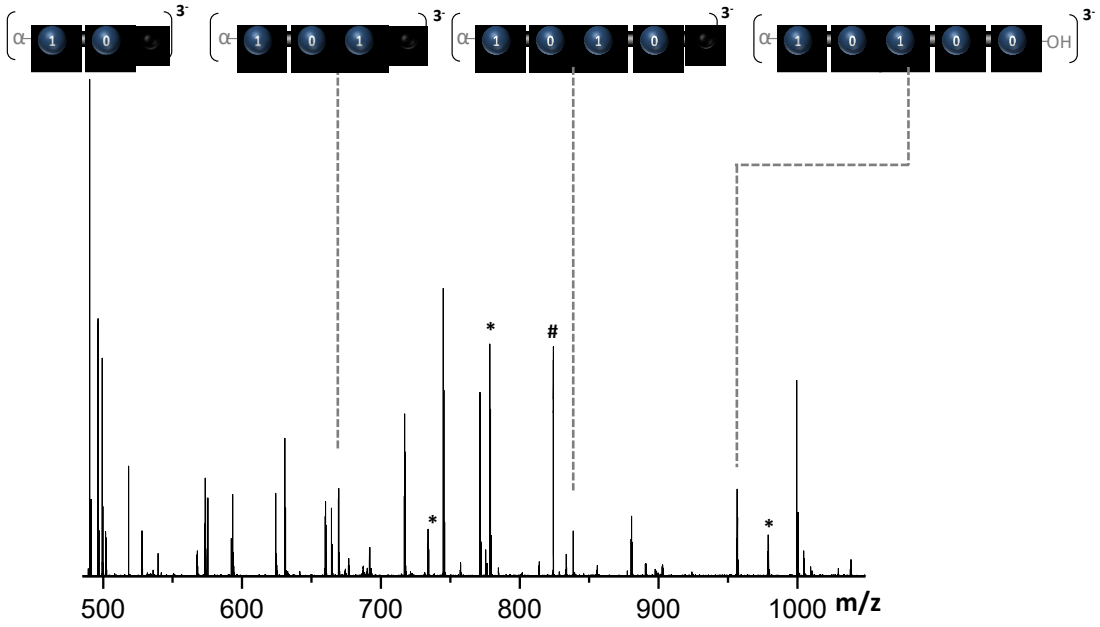
Moreover, the tetraeptide was also modified. Indeed, it was found that the phosphate groups were sufficient to ensure good polymer ionization in negative ESI-MS. Hence, the second test was done using a tertrapeptyde containing glycine and monomer L_1 as linker. Also in this experiment, only one iterative cycle was performed and the capping agent was not employed in step (*iv*). As depicted in

Figure II.15, the resulting negative ESI-MS show the presence of the targeted molecule as predominant species. These results demonstrated that the new conditions allow to obtain quantitative yields in both couplings hence, uniform sequence-defined polymers.

Figure II.15: ESI-MS registered for a polymer synthesized using 4 glycines, L_1 as linker and performing one iterative cycle.

After the first successful test, the synthesis of a polymer containing four glycines and the sequence 1_a-T_3-0_a-T_3-1_a-T_3-0_a-T_3-0_a-T_3 was done in F^2S conditions. Hence, in this experiment, 5% of capping agent was employed in step (*iv)*. After cleavage in ammonia and filtration, the resulting solid was analyzed by negative ESI-HRMS showing the presence of the targeted sequence-defined polymer at different charge states. Moreover, as depicted in

a)

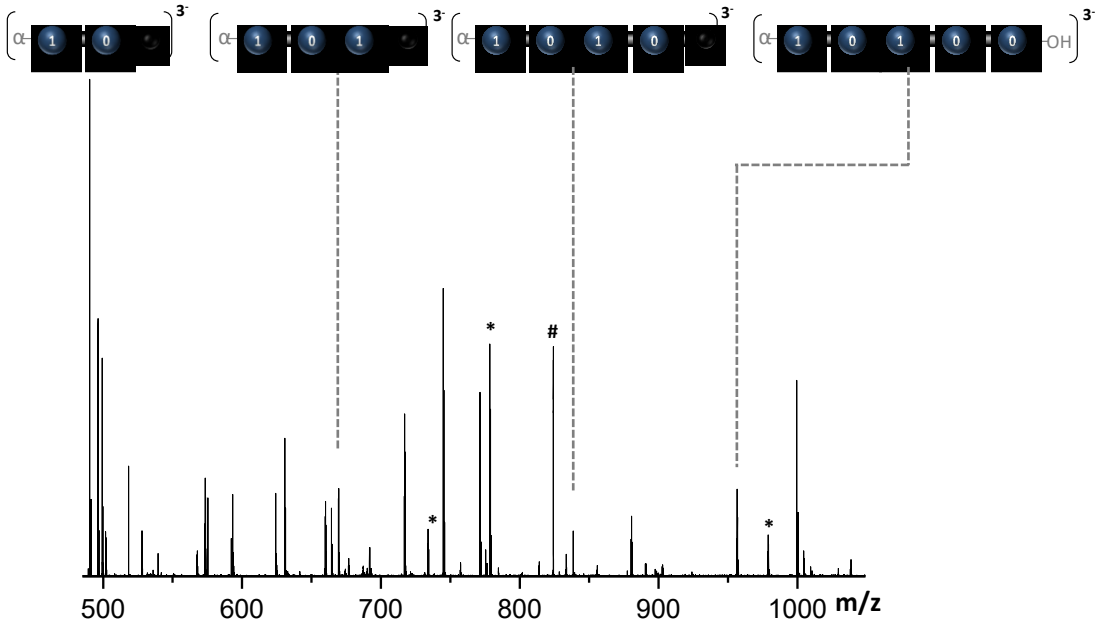


b)

Sequence	$[M]^{-}_{th}$	$[M]^{-}_{exp}$	$[M]^{2-}_{th}$	$[M]^{2-}_{exp}$	$[M]^{3-}_{th}$	$[M]^{3-}_{exp}$	$[M]^{4-}_{th}$	$[M]^{4-}_{exp}$
α-$1_a T_3 0_a T_3 1_a T_3 0_a T_3 0_a T_3$	2870.61	–	1434.81	–	956.20	956.20	716.90	716.90
α-$1_a T_3 0_a T_3 1_a T_3 0_a$**X**	2501.34	–	1250.17	–	833.11	833.11	624.59	624.58
α-$1_a T_3 0_a T_3 1_a$**X**	2010.07	–	1004.54	1004.53	669.35	669.35	501.77	501.76
α-$1_a T_3 0_a$**X**	1490.76	–	744.88	744.88	496.25	496.25	371.94	–
α-1_a**X**	999.49	999.49	499.25	499.24	332.50	–	249.19	–
α-**X**	480.18	480.19	239.59	–	159.39	–	119.30	–

Figure II.16b, the sequence-terminated fragments, originated from the reaction between the growing chain and the capping agent, were found at different charge states, according to the number of phosphate groups they still bear. However, the detection of the short sequence-terminated fragments was not optimal. In

a)

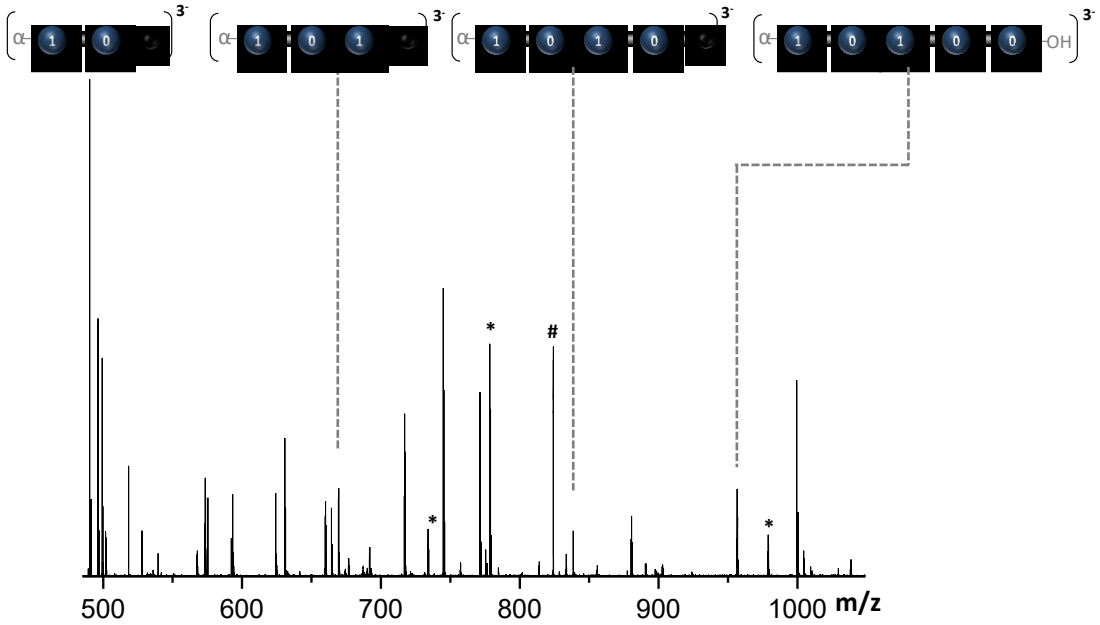


b)

Sequence	$[M]^{-}_{th}$	$[M]^{-}_{exp}$	$[M]^{2-}_{th}$	$[M]^{2-}_{exp}$	$[M]^{3-}_{th}$	$[M]^{3-}_{exp}$	$[M]^{4-}_{th}$	$[M]^{4-}_{exp}$
α-$1_{\alpha}T_3 0_{\alpha}T_3 1_{\alpha}T_3 0_{\alpha}T_3 0_{\alpha}T_3$	2870.61	–	1434.81	–	956.20	956.20	716.90	716.90
α-$1_{\alpha}T_3 0_{\alpha}T_3 1_{\alpha}T_3 0_{\alpha}$**X**	2501.34	–	1250.17	–	833.11	833.11	624.59	624.58
α-$1_{\alpha}T_3 0_{\alpha}T_3 1_{\alpha}$**X**	2010.07	–	1004.54	1004.53	669.35	669.35	501.77	501.76
α-$1_{\alpha}T_3 0_{\alpha}$**X**	1490.76	–	744.88	744.88	496.25	496.25	371.94	–
α-1_{α}**X**	999.49	999.49	499.25	499.24	332.50	–	249.19	–
α-**X**	480.18	480.19	239.59	–	159.39	–	119.30	–

Figure II.16 a, only the detected peaks corresponding to the 3^{-} are depicted.

a)



b)

Sequence	$[M]^{-}_{th}$	$[M]^{-}_{exp}$	$[M]^{2-}_{th}$	$[M]^{2-}_{exp}$	$[M]^{3-}_{th}$	$[M]^{3-}_{exp}$	$[M]^{4-}_{th}$	$[M]^{4-}_{exp}$
$\alpha\text{-}1_{\alpha}T_3 0_{\alpha}T_3 1_{\alpha}T_3 0_{\alpha}T_3 0_{\alpha}T_3$	2870.61	–	1434.81	–	956.20	956.20	716.90	716.90
$\alpha\text{-}1_{\alpha}T_3 0_{\alpha}T_3 1_{\alpha}T_3 0_{\alpha}$**X**	2501.34	–	1250.17	–	833.11	833.11	624.59	624.58
$\alpha\text{-}1_{\alpha}T_3 0_{\alpha}T_3 1_{\alpha}$**X**	2010.07	–	1004.54	1004.53	669.35	669.35	501.77	501.76
$\alpha\text{-}1_{\alpha}T_3 0_{\alpha}$**X**	1490.76	–	744.88	744.88	496.25	496.25	371.94	–
$\alpha\text{-}1_{\alpha}$**X**	999.49	999.49	499.25	499.24	332.50	–	249.19	–
α-**X**	480.18	480.19	239.59	–	159.39	–	119.30	–

Figure II.16: Sequencing of the polymer synthesized using F^2s approach. **a)** ESI-MS analysis registered for the polymer asses for F^2S coding for 1-0-1-0-0. The star corresponds to the Na adducts, the hash correspond to artefacts; **b)** Table of the fragments calculated and found corresponding to the sequence-terminated fragments and the original sequence.

In order to displace the shorter sequence-terminated fragments in a more convenient m/z range and improve their detection, a longer peptide may be employed for the synthesis of the sequence-defined polymers. Moreover, charge state MS deconvolution could facilitate the interpretation of the MS spectra obtained in F^2S conditions. Even though, some parameters need still to be optimized, the F^2S approach seems promising for the sequencing of synthetic sequence-defined polymers. Indeed, truncated sequences generated from the reaction between the growing chain and the capping agent allow the polymer sequencing avoiding MS/MS or chemical fragmentations.

5 Conclusions

In conclusion, a new iterative protocol based on two chemoselective steps was developed for the synthesis of digitally encoded sequence-defined poly(alkoxyamine phosphodiester)s. The building blocks were designed in order to obtain quantitative yields and to avoid side reactions during the iterative synthesis. Moreover, the monomers design avoids sequence-dependent fragmentations, ensuring an optimal reading by MS/MS. The chemoselective solid-state protocol, optimized in this work allows obtaining uniform digital polymers in a relatively short time. The molecular encryption for these molecules relies on the H/Me couple as for other digital polymers. Moreover, the multi-step growth polymerization approach guarantees the uniformity of the resulting sequence-defined polymers. A library of digitally encoded sequence-defined polymers was synthesized in order to show the feasibility of the method.

Interestingly, the labile alkoxyamine bonds, inserted in the polymers backbone during the nitroxide radical coupling, are selectively fragmented in MS/MS conditions originating predictable and easy to interpret fragmentation patterns. The simple read-out and the relatively easy synthetic access, makes this class of digital polymers promising candidates for long-term data storage applications.

Moreover, after the conventional cleavage by MS/MS, the sequencing *via* F^2S approach was tested for poly(alkoxyamine phosphodiester)s. This technique was largely employed in the past years for the sequencing of natural sequence-defined polymers as polypeptides. The iterative strategy presented in this chapter enables the formation of sequence-terminated fragments during the phosphoramidite coupling. The truncated sequences, generated in this step, enable the fast and reliable sequencing bypassing MS/MS or chemical fragmentations. To the best of our knowledge, the described work is one of the first examples of fragmentation-free sequencing applied to synthetic sequence-defined polymers. The promising results, obtained using this sequencing technique allows to obtain a fast sequencing. However, some parameters have still to be optimized.

Even though, the first generation of poly(alkoxyamine phosphodiester)s ensure a reliable molecular encryption only one bit of information is inserted in each repeating unit. As said in the introduction, the structure of synthetic sequence-defined polymers can be modified in order to obtain tailor-made features. In this context, the development of a new strategy for the insertion of two bits of information for each repeating unit is discussed in the next chapter. As is explained in chapter 2, this design allow to increase the information density keeping the read-out *via* MS/MS easy.

Chapter III

Increasing the storage capacity of poly(alkoxyamine phosphodiester)s using extended nitroxide alphabets

1 Introduction

One of the big advantages of molecular data storage consists in the higher data density of the resulting molecules compare to modern flash memories. In 2012, Church calculated the data density of several commercial data storage media and compared it with the one calculated for DNA. Interestingly, it was found that DNA was six orders of magnitude more dense compare to flash memories. More details about the storage density calculation are in the figure and data section of the same article.[204] The DNA could theoretically store 455 exabytes (10^{18} bytes) per gram for single strand.[164] As already discussed in this thesis, a valid alternative to DNA for data storage applications, could be represented by synthetic polymers. Indeed, these molecules give the unique possibility to design the backbone in order to obtain tailor-made features. In this context, the digital polymers backbone could be designed in order to enhance the data density. In 2015, Lutz suggested that the use of enhanced alphabets (i.e. tertiary, quaternary etc.) for the code implementation may drastically improve the storage capacity and could allow the encryption of a big number of combinations into short chains.[6] However, the introduction of these alphabets may increase the time required for the monomer synthesis and the resulting digital polymers could not be compatible with the existing information technology which nowadays is based on the binary encryption. Moreover, the complicated molecular structures of these sequence-defined polymers may lead to multiple fragmentation mechanisms in CID conditions, making more complicated the resulting sequencing by MS/MS.

In this context, an easy concept allowing to increase the information density, maintaining the sequencing by tandem MS easy was tested for poly(alkoxyamine phosphodiester)s and is presented in this chapter. As explained in chapter 2, the iterative protocol developed for the synthesis of these digital polymers relies on two chemoselective steps. During the phosphoramidite coupling, a coding monomer (0/1) is added to the polymer chain whereas the nitroxide-radical coupling provides the insertion of the labile alkoxyamine bond in the backbone. Even though, the insertion of this bond is extremely useful to simplify the MS/MS sequencing, no information is encrypted during this step. Hence, the resulting digital polymers contain one bit for each repeating unit (21 backbone atoms) showing a lower data density compare to other digital polymers classes. For this reason, the first generation of poly(alkoxyamine phosphodiester)s is not suitable for the molecular encryption of long messages. In order to increase the data density, a new molecular architecture based on encoding dyads was designed. The iterative synthesis based on dyads was already employed from our group in the context of convergent synthesis of poly(alkoxyamine amide)s. In this work, the dyads employed were trimers containing two coding units and one spacer. A library containing all the possible dyad combinations was synthesized on solid support. Afterwards, the trimers were iteratively added to the growing chain in order obtain long sequence-defined polymers in few steps. However, the use of dyads didn’t increase the storage capacity of the resulting digital polymers. In the work presented herein, four

different monomers (two phosphoramidites and two nitroxides) are employed as coding units and are implemented into the iterative protocol depicted in

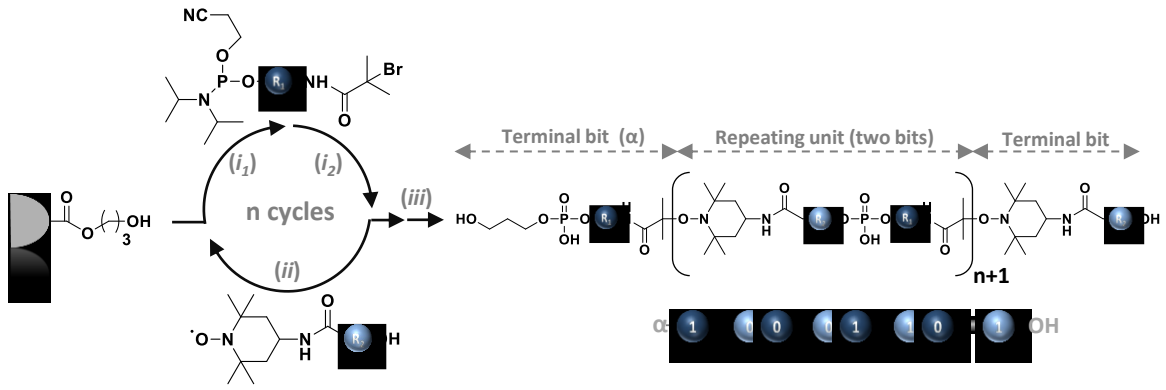


Figure III.1. In order to encrypt information during the NRC, the hydroxy-functionalized nitroxide monomers contain hydrogen/methyl or methyl/dimethyl couples on the side chain (Figure III.2). The new molecular architecture allows the encryption of one bit for each iterative step. Hence, the backbone contains two bits for each repeating unit (20 backbone atoms). Moreover, as discussed in the sequencing section, the binary dyads, generated after alkoxyamine homolytic cleavage in MS/MS allow the easy sequencing also for long polymeric chains.

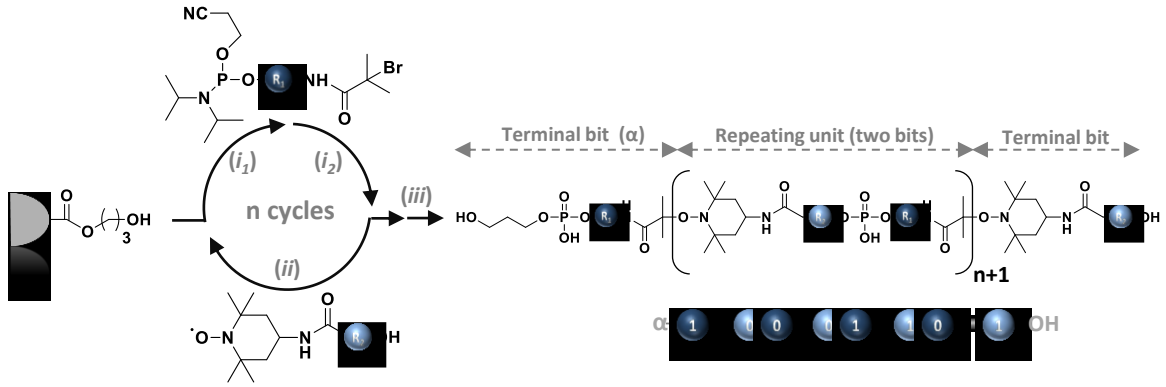


Figure III.1: General strategy employed for the synthesis of poly(alkoxyamine phosphodiester)s encoding dyads. **(i_1)** phosphoramidite coupling: acetonitrile, tetrazole, rt; **(i_2)** oxidation: I_2, 2,6-lutidine, THF/H_2O, rt; **(ii)** radical–radical coupling: CuBr, Me_6TREN, DMSO; **(iii)** cleavage: piperidine, AcCN, rt, then $MeNH_2$, NH_4OH, H_2O, rt.

The nitroxides employed in this project gave the possibility to place the coding units in different positions of the alkylic chain. In order to study if the placement of this groups could influence the yields, three different alphabets were synthesized and tested for the synthesis of digitally encoded poly(alkoxyamine phosphodiester)s encoding dyads. As shown in Figure III.2, all the hydroxy-functionalized nitroxides bear H/Me or methyl/dimethyl couples in different positions of the backbone. In this chapter, the synthesis and the optimization of the different building blocks which led to the synthesis of sequence-defined poly(alkoxyamine phosphodiester)s encoding dyads is presented. Moreover, the dissociation behaviour in MS/MS conditions and the sequencing theories for this class of molecules are discussed.

Results and discussion

For the synthesis of poly(alkoxyamine phosphodiester)s encoding dyads two sets of coding monomer were employed as depicted in Figure III.2. The phosphoramidite monomers 0_a and 1_b were synthesized in two steps, following the synthetic strategy described in chapter 2. For the hydroxy-functionlized nitroxides three different alphabets were tested: alphabet 1 ($0_d/1_d$), alphabet 2 ($0'_d/1'_d$) and alphabet 3 ($0''_d/1''_d$). The structures of all the building blocks synthesized in this work are depicted in Figure III.2.

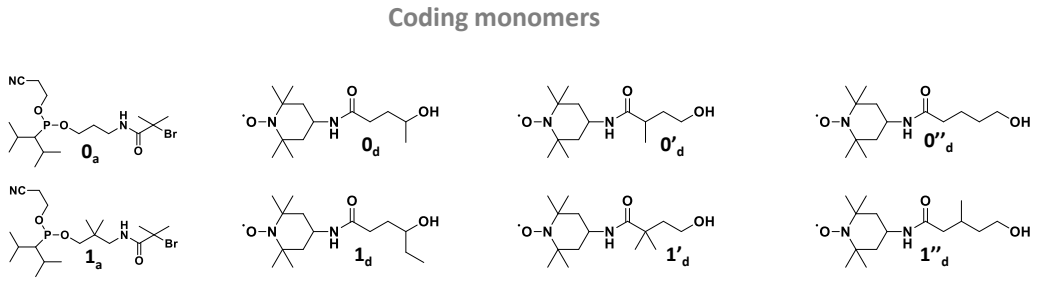
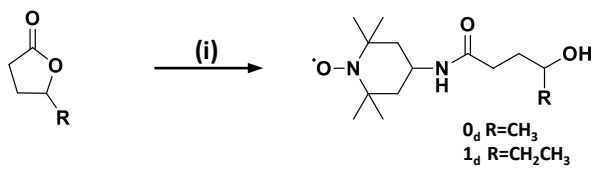


Figure III.2: Structures of the building blocks synthesized and tested for the development of poly(alkoxyamine phosphodiester)s encoding dyads.

The solid-support modification and the loading calculation were done following the procedure described in chapter 2. All the digital polymers presented in this chapter were synthesized *via* a multi-step growth polymerization approach, using the same conditions described in the previous chapter.

2.1 First generation of poly(akoxyamine phosphodiester)s encoding dyads using alphabet 1

The first alphabet which was investigated for the synthesis of poly(alkoxyamine phosphodiester)s encoding dyads is represented by the couple $0_d/1_d$. These monomers were the first tested because of their easy and high yield synthesis. Indeed, both molecules were recovered in satisfactory yields in one step. These building blocks were synthesized opening the corresponding lactones with 4-Amino-TEMPO. In particular, the monomer 0_d was synthesized opening the γ-valerolactone and the monomer 1_d was synthesized opening the γ-hexanolactone (Scheme III.1). For both monomers, the conditions employed were the same described for the synthesis of linker T_3.



Scheme III.1: General strategy for the synthesis of monomers 0_d and 1_d. **(i)** 4-Amino-TEMPO, 130°C

The resulting compounds were analyzed by positive ESI-HRMS. For both molecules, the theoretical weights are in line with the analytical ones. Moreover, peaks corresponding to

products originated from side reactions were not detected as shown in Figure III.3. The alphabet 1 was subsequently employed in combination with the phosphoramidite monomer 0_a for the iterative synthesis of digitally encoded poly(alkoxyamine phosphodiester)s encoding dyads.

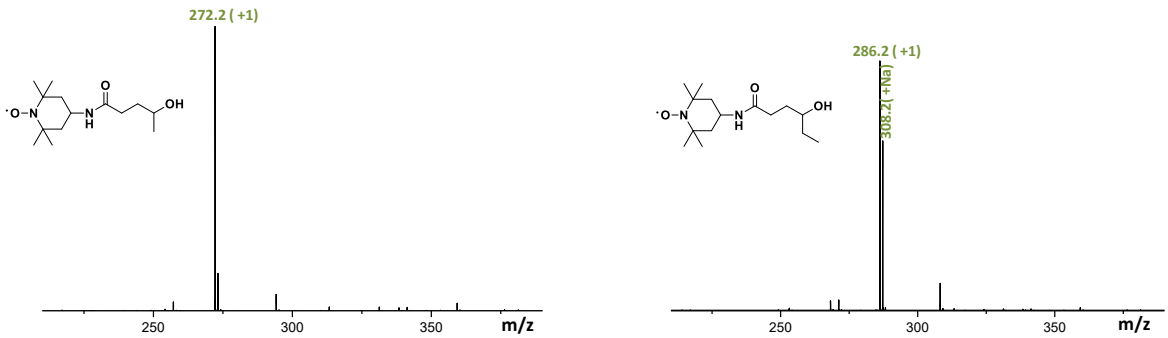


Figure III.3: Positive ESI-HRMS recorder for the monomers couple 0_d and 1_d (alphabet 1).

The resulting tetramer coding for 0_a-1_d-0_a-0_d was synthesized *via* an iterative solid-state protocol using the conditions described in chapter 2. After cleavage, the product was recovered in only 37% yields and was following analyzed by ESI-HRMS. As for the first generation of poly(alkoxyamine phosphodiester)s, the targeted product was detected as preponderant species at different charge states ranging from 1^- to 2^- (green labelled peaks). However, the presence of lower congeners, originated from the non-quantitative yields during the second phosphoramidite coupling (red labelled peak in Figure III.4) was detected at lower intensity. The structure of the lower congeners is depicted in the same picture.

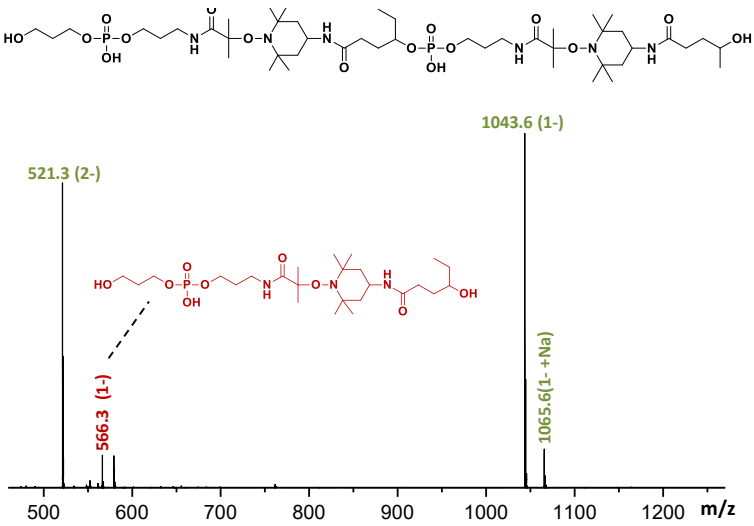


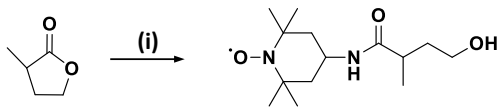
Figure III.4: Negative ESI-HRMS analysis of polymer synthesized using alphabet 1 coding for 0_a-1_d-0_a-0_d red labelled peak corresponds to the lower congeners originated from non-quantitative phsophoramidite coupling.

As already observed for the first generation of poly(alkoxyamine phosphodiester)s, the sterically hindered secondary alcohol leads to incomplete yields during the phosphoramidite coupling (step (i)). The presence of the lower congener, made the digital polymers

synthesized using the alphabet one non-uniform hence, unsuitable for storing data at the molecular level. For this reason, alphabet 1 was abandoned and alphabet 2 was tested.

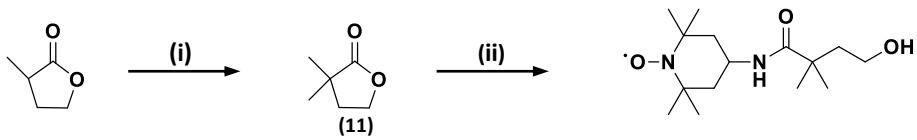
2.2 Second generation of poly(alkoxyamine phosphodiester)s encoding dyads using alphabet 2

The second generation of hydroxy-functionalzed nitroxides tested for the synthesis of uniform sequence-defined poly(alkoxyamine phosphodiester)s encoding dyads is represented from the monomers $0'_d$ and $1'_d$; the structures of these building blocks are depicted in Figure III.2. These monomers were designed to bypass the problem of the *non*-quantitative yields obtained during the phosphoramidite coupling using alphabet one. Indeed, the displacement of the methyl/dimethyl couple in α position allows to have a primary alcoholic function which ensures the quantitative yields in step (i_1). As depicted in Scheme III.2, the synthesis of the monomer $0'_d$ was achieved in a single step, opening the corresponding α-methyl-γ-butyrolactone with 4-Amino-TEMPO. Also in this case, the conditions employed were the same described in chapter 2. After purification, the desired product was recovered in good yields.



Scheme III.2: Synthetic strategy employed for the synthesis of monomer 0'd. **(i)** 4-amino-TEMPO, 130°C reflux, 2.5h.

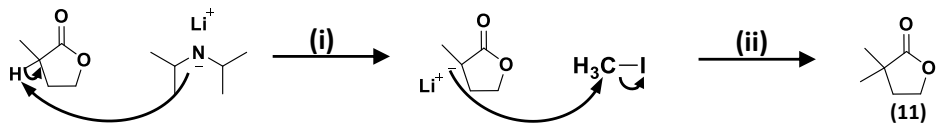
The first synthetic strategy employed for the synthesis of monomer $1'_d$ is depicted in Scheme III.3. The α-methyl-γ-butyrolactone was methylated using LDA and methyl iodide to obtain the intermediate 11. Following reaction between the dimethyl lactone with 4-Amino-TEMPO gave the desired product in two steps. Hence, the synthetic pathway proposed would be extremely convenient.



Scheme III.3: First synthetic strategy for the synthesis of monomer 1'd. **(i)** lactone methylation: n-BuLi, diisopropylamine, THF 0°C, then -78°C; CH_3I, HMPA, overnight; **(ii)** lactone opening**:** 4-Amino-TEMPO, different reaction time and temperatures, the final product was not obtained.

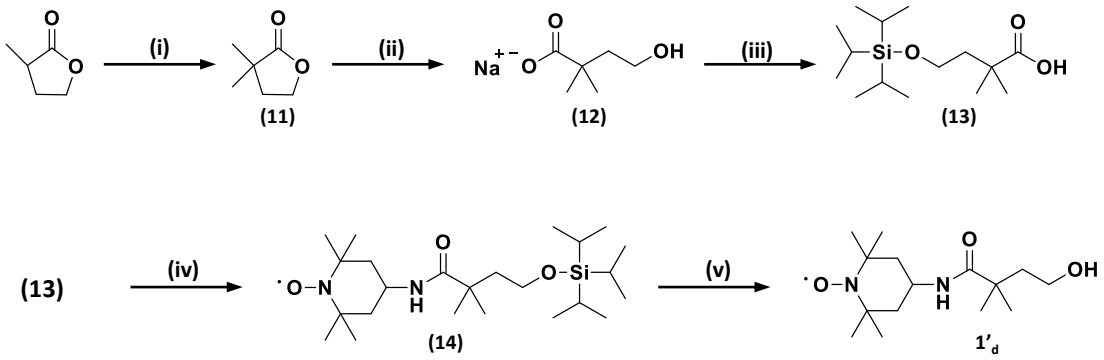
As depicted in Scheme III.4, the intermediate 11 was synthesized using LDA, methyl iodide and HMPA. The LDA was produced *in situ* reacting the distilled N-N diisopropylethylamine with a solution of n-butyllithium in THF for 30 minutes at 0°C. The anion generated from the deprotonation of the hydrogen in α position attacks the methyl iodide to afford the intermediate (11). The HMPA was employed to accelerate the nucleophilic substitution.

Indeed, this molecule solvates the lithium cation and creates a more naked anion. The reaction was carried out at -78°C to ensure the anion stability. The following reaction between the intermediate 11 and the 4-Amino-TEMPO to afford the monomer 1'$_d$ have been tried in different conditions but the targeted product was not recovered.



Scheme III.4: Mechanism for the synthesis of intermediate 11.

This was probably due to the lower reactivity of intermediate 11 compare to the α-methyl-γ-butyrolactone. Indeed, the insertion of a second methyl group in alpha increases the steric hindrance reducing the accessibility of the carboxylic function. For this reason, a new synthetic strategy for the synthesis of the monomer 1'$_d$ was designed. This new approach, depicted in Scheme III.5, allowed to obtain the desired monomer in five steps. In this strategy, the intermediate 11 was transformed in the corresponding hydroxy acid after reaction with sodium hydroxide. The alcoholic function of the intermediate 12 was then protected using TIPS-Cl. This step was done in order to avoid the formation of side products that could originate during the following reaction. Indeed, in the conditions employed for the amidification, the alcohol could react with the carboxylic function to form the corresponding ester. The last step was the TIPS deprotection of the intermediate 14 using TBAF to obtain the monomer 1'$_d$.



Scheme III.5: Synthetic strategy employed for the synthesis of monomer 1'd. **(i)** lacone methylation: n-BuLi, diisopropylamine, THF 0°C, then -78°C; CH_3I, HMPA, overnight; **(ii)** lactone opening: NaOH, water 3h, rt; **(iii)** TIPS protection: TIPS-Cl, DMF overnight; **(iv)** 4-Amino-TEMPO coupling: 4-Amino-TEMPO, DIPEA, PyBOP, dichloromethane, 0°C then rt, overnight, **(v)** TIPS deprotection: TBAF, THF, 2.5h, rt.

After the synthesis, the building blocks forming the alphabet 2 were analyzed by positive ESI-HRMS. As shown in Figure III.5, for both monomers the preponderant peaks were the ones corresponding to the targeted products, the sodium adducts were also detected at lower

intensities. Moreover, peaks corresponding to side products were not detected in positive ESI-HRMS in both samples.

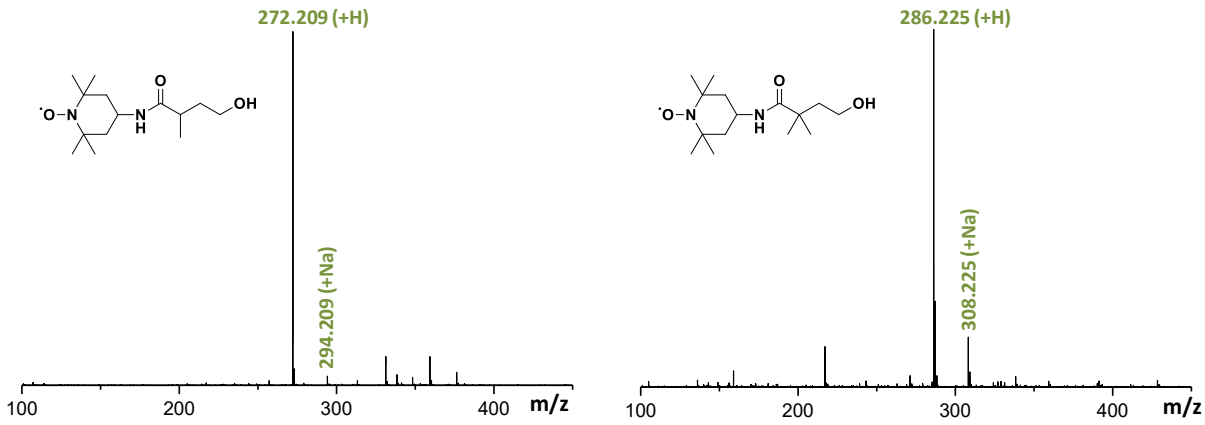


Figure III.5: Positive ESI-HRMS recorder of monomer 0'$_d$ and monomer 1'$_d$ (alphabet2).

The alphabet 2 was then employed in step (*ii*) of the iterative protocol described in

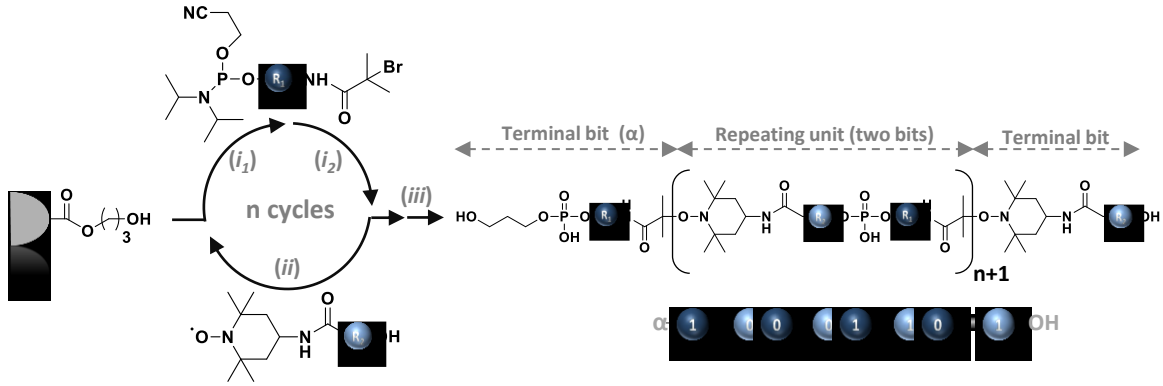


Figure III.1 for the synthesis of a library of six different digitally encoded poly(alkoxyamine phosphodiester)s. As shown in Table III.1, after cleavage, the digital polymers were recovered in moderate yields. Furthermore, impurities corresponding to primary amine-terminated oligomers were detected in ESI-HRMS (purple, orange and red peaks in Figure III.6).

Table III.1: ESI-HRMS individual characterization of the digital polymers synthesized using alphabet 2.

	sequence	*yield*	*m/z$_{th}$*	*m/z$_{exp}$*
1	α-0$_a$-1'$_d$-0$_a$-0'$_d$	47%[a]	1044.5894	1044.5899[a]
2	α-1$_a$-0'$_d$-0$_a$-0'$_d$	50%[a]	1057.5972	1044.5923[b]
3	α-0$_a$-1'$_d$-1$_a$-0'$_d$	48%[a]	535.3028	535.3034[a]
4	α-1$_a$-0'$_d$-0$_a$-1'$_d$	52%[a]	535.3028	535.3034[a]
5	α-0$_a$-0'$_d$-1$_a$-0'$_d$-1$_a$-1'$_d$-0$_a$-1'$_d$	45%[a]	1033.5867	1033.5863[a]
6	α-1$_a$-0'$_d$-0$_a$-0'$_d$-1$_a$-0'$_d$-0$_a$-0'$_d$	55%[a]	1019.5710	1019.5743[a]

[a] $[M-2H]^{2-}$, [b] $[M-H]^-$.

In Figure III.6 the negative ESI-HRMS recorded for one of the sequence-defined polymers corresponding to entry 6 Table III.1 is depicted. As shown in the picture, the targeted

product was detected at different charge states ranging from 2^- to 4^- (green labelled peaks). However, the side products corresponding to amine-terminated oligomers were also found at lower intensities (red labelled peaks). These impurities were detected at different charge states according to the number of phosphate groups that they still bear. The chemical structure of the amine-terminated fragments is depicted in Figure III.6. Analogues ESI-HRMS were recorded for all the sequence-defined polymers in Table III.1.

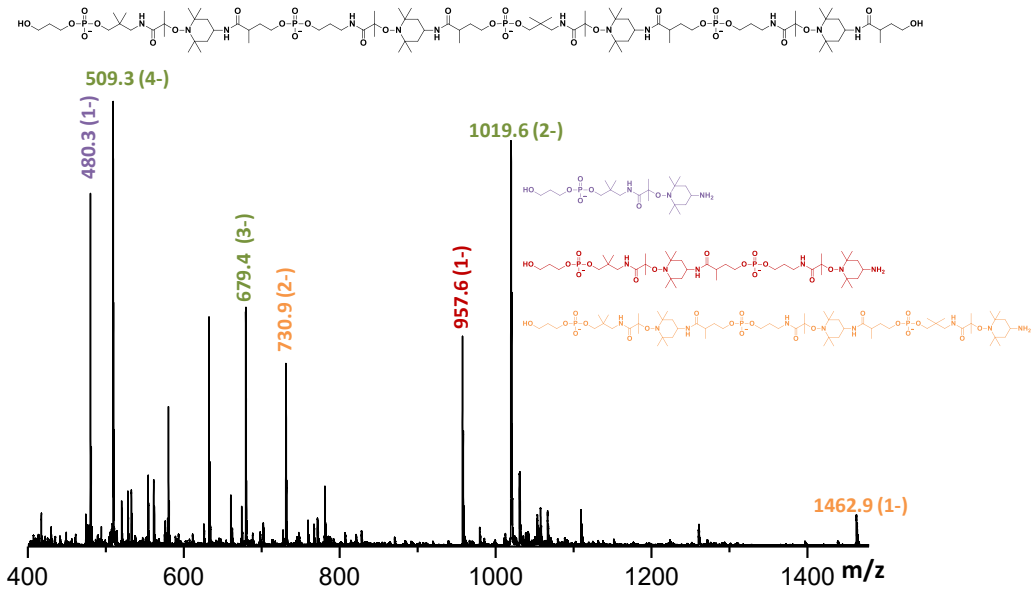


Figure III.6: Negative ESI-HRMS analysis of oligomer synthesized using alphabet 2 (entry 6 tab. 3). The green labelled peaks correspond to the desired oligomer whereas the red labelled peaks correspond to the amine truncated sequences.

These contaminations were obviously not originated from any residue of 4-Amino-TEMPO contained in the coding monomers employed for the iterative synthesis. Indeed, these building blocks were analyzed by ESI-HRMS and these impurities were not detected even several weeks after the synthesis. The positive ESI-MS analysis is suitable for the detection of 4-Amino-TEMPO traces in the samples. Indeed, the amine moiety bear by this molecule are highly ionized in positive ESI conditions, generating intense peaks. To confirm the absence of contaminations in the building blocks, the monomers were anlayzed in positive ESI-HRMS in combination with different percentages of 4-Amino-TEMPO which was detected in all the cases. Afterwards, it was assumed that the detection of these fragments in ESI-HRMS could be due to the amide cleavage *via* a Mclafferty rearrangement in MS conditions.[205] However, this mechanism relies on the abstraction of a gamma proton in ionization conditions hence; the same reaction pathway should have occurred also for the hydroxy-functionalized nitroxides employed previously. Moreover, the ionization energies employed for the analysis are too low to observe this reaction. To confirm that the amine-terminated sequences detected in ESI-HRMS were not generated during the MS analysis, the digital polymers in Table III.1 were submitted to the quinhydrone test. The quinhydrone is a complex of hydroquinone and benzoquinone (1:1) and is usually employed for the amines detection.[206] The reaction of this complex with the amines generates a change of colour in the resulting

solution. This test, performed at room temperature, is particularly convenient for the amine detection in the poly(alkoxyamine phosphodiester) samples. Indeed, in these conditions the stability of the alkoxyamine bond is ensured. The test was performed dissolving the samples in water and treating the aqueous solution with a methanolic solution of quinhydrone. The change of the solution colour suggested the presence of amines. Moreover, the resulting test solution was analyzed by UV spectroscopy. As depicted in Figure III.7, the shift of the quinhydrone bands in the UV spectra (red labelled UV spectra) further suggested the presence of free amines in the samples.

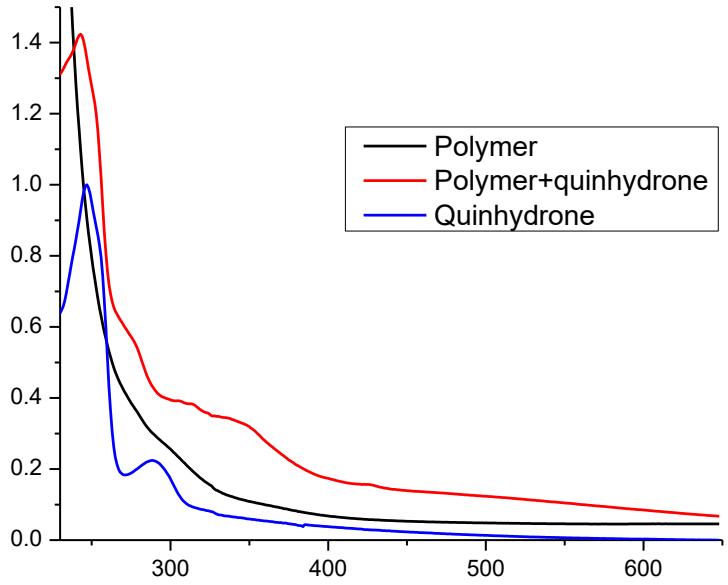


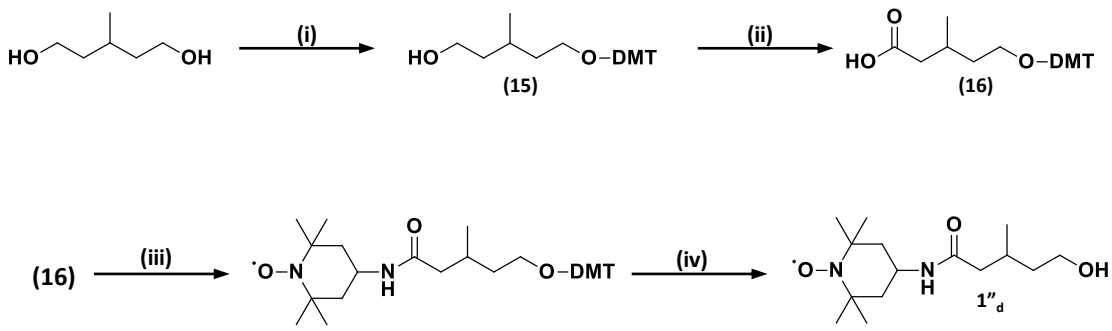
Figure III.7: UV analysis of the quinhydrone test solution for entry 6 Table III.1. UV spectrum for polymer 6 Table III.1 (black) UV spectra of quinhydrone (blue), UV spectra of the test solution (red).

Finally, it was assumed that these fragments could be originated from a copper-induced amide cleavage mechanism.[207] However, in the frame of this project the mechanism, which led to the formation of amine-terminated fragments using alphabet 2 was not investigated. The presence of impurities in the sequence-defined polymers, synthesized using the monomer couple $0'_d$ /$1'_d$ made the alphabet 2 unsuitable for the synthesis of uniform poly(alkoxyamine phosphodiester)s encoding dyads, hence this alphabet was abandoned and alphabet 3 was tested.

2.3 Synthesis of uniform poly(alkoxyamine phosphodiester)s encoding dyads using alphabet 3

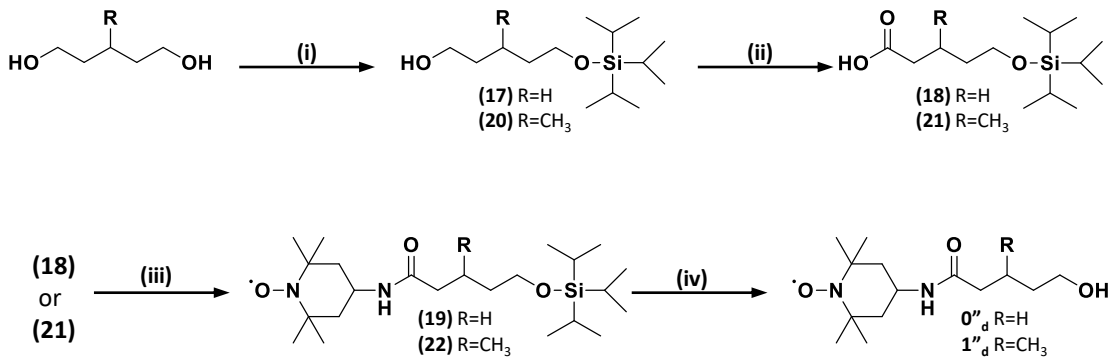
The quinhydrone test, the moderate yields obtained for the polymers synthesized using alphabet 2 and the MS studies performed on the monomers, suggested that the amine truncated fragments, detected in ESI-HRMS, were originated during the iterative synthesis. In order to overcome this problem, the use of alphabet 3 for the synthesis of uniform poly(akoxyamine phosphodiester)s encoding dyads was investigated. The monomers forming alphabet 3 were designed displacing the coding units (H/Me) in β position (Figure

III.2). The first attempt, to synthesize monomer 1”$_d$ was done following the synthetic strategy described in Scheme III.6.



Scheme III.6: First synthetic strategy employed for the synthesis of monomer 1“d. **(i)** DMT protection: DMT-Cl, pyr, THF 5h, rt, **(ii)** oxidation: PDC, DMF different reaction times and temperatures however the intermediate 116 was not recovered.

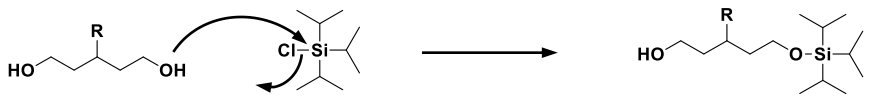
The first step was the protection of one alcoholic of the 3 methyl 1-5 pentanediol using DMT-Cl. The alcohol monoprotection was done in order to avoid the formation of a mixture of products originating from the oxidation of both alcoholic functions in the following step. The reaction was carried out in a mixture of anhydrous THF and pyridine and the resulting intermediate was recovered in good yields after purification. The second step was the oxidation of the corresponding protected diol into a carboxylic acid using pyridinium dichromate (PDC). The oxidation of alcohols bearing acid labile moieties using PDC was described by Corey and Shmidt in 1979.[208] Indeed, this reagent is less acidic compared to other chromate salts and prevents the DMT deprotection in the oxidation conditions. The reaction was performed stirring the intermediate 15 and the PDC in DMF. The dimethylformamide was chosen as solvent because as reported in literature, this compound speeds the oxidation up.[209] However, the intermediate 16 was not isolated even tough harsh conditions were employed. For this reason, another synthetic strategy, depicted in Scheme III.7 was designed.



Scheme III.7: Synthetic strategy employed for the synthesis of alphabet 3. **(i)** TIPS protection: TIPS-Cl, imidazole, DMF overnight; **(ii)** oxidation: TEMPO, BAIB, dichloromethane, water, 3h, rt; **(iii)** 4-Amino-TEMPO: 4-Amino-TEMPO, DIPEA, PyBOP, dichloromethane, 0°C then rt, overnight; **(iv)** TIPS deprotection: TBAF, THF, 2.5h, rt.

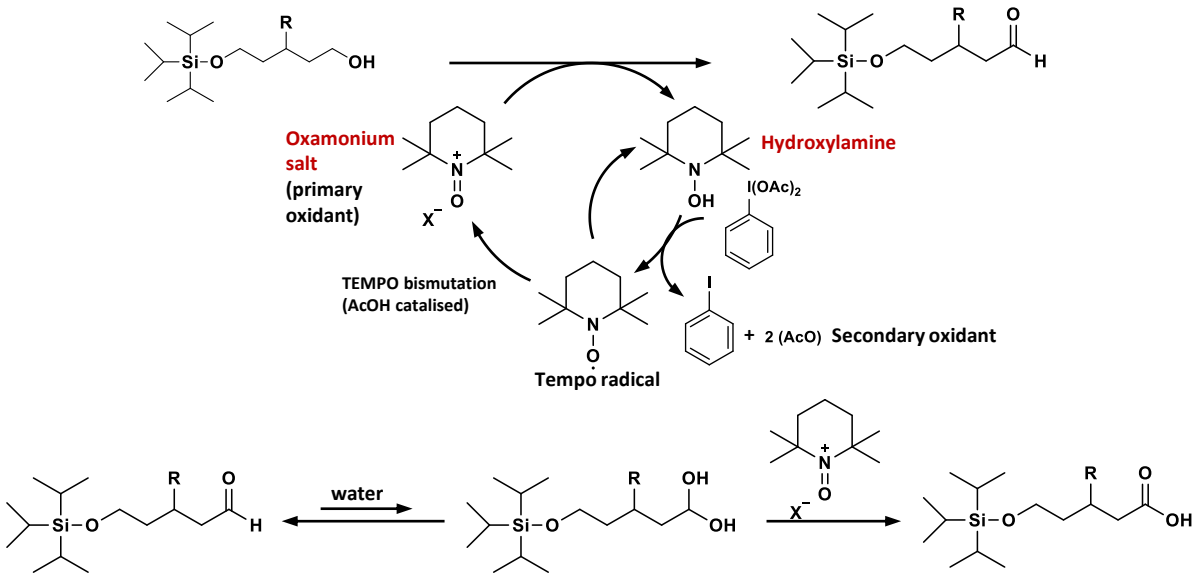
The new synthetic pathway employs TIPS chloride as protecting group and the couple TEMPO/(Diacetoxyiodo)benzene (BAIB) as oxidants. Another advantage of using the couple TEMPO/BAIB for the oxidation is their absence of toxicity compare to chromate salts. Hence, the second attempt to synthesize monomer 0“$_d$ and 1”$_d$ was done following the synthetic strategy depicted in Scheme III.7. The desired building blocks were synthesized in four steps and were recovered in satisfactory yields.

As discussed before, the first step of the synthetic pathway was the diols monoprotection with TIPS-Cl, performed in DMF at 0°C using imidazole as chloride scavenger. The TIPS was selected as protecting group because it is stable to the oxidation conditions employed in the following reaction. The mechanism of the TIPS protection is depicted in Scheme III.8. The silyl ether formation proceeds in few hours and the desired product was obtained in satisfactory yields.



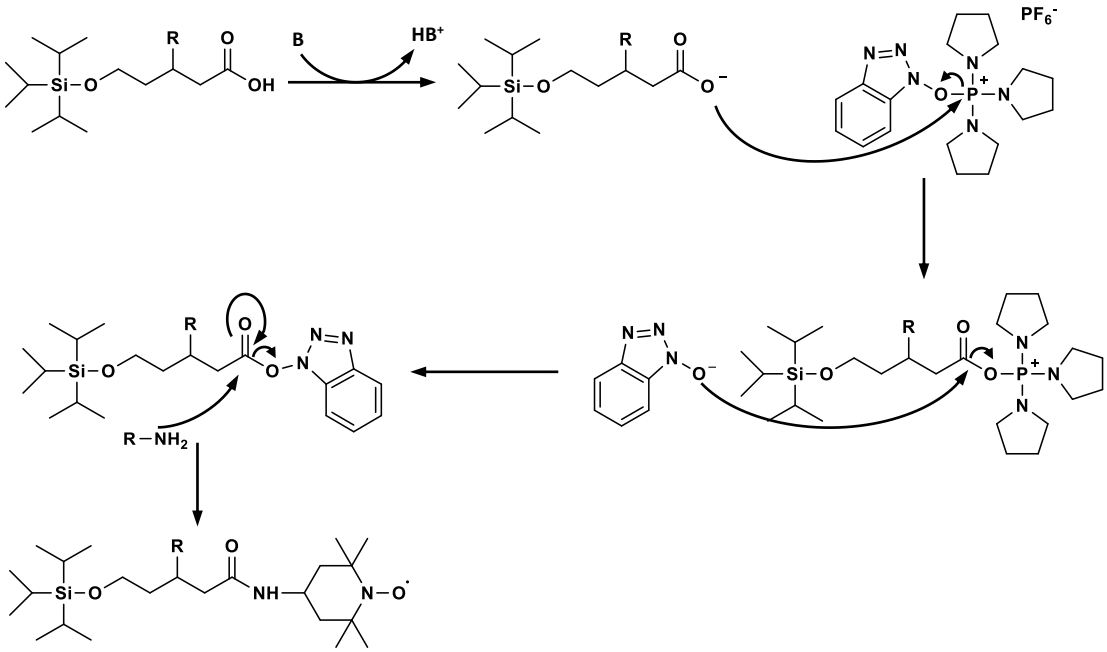
Scheme III.8: Protection mechanism using TIPS-Cl.

The reaction is a nucleophilic substitution in which the oxygen of the hydroxy function attacks the silicon atom with release of a chloride anion which reacts with the imidazole forming the resulting salt. The intermediates 17 and 20 were following oxidized into the corresponding carboxylic acids (18 and 21) using TEMPO, in catalytic amounts, and BAIB. The oxidation was carried out at 0°C in order to ensure the stability of the TEMPO and to increase the yields. The reaction mechanism is depicted in Scheme III.9. The first step is the AcOH-catalyzed bismutation of TEMPO into an oxoammonium salt and hydroxylamine, the oxamonium salt oxidizes the alcohol in the corresponding aldehyde. Afterwards, the hydroxylamine, generated during the alcohol oxidation reacts with the BAIB, regenerating the TEMPO and producing AcOH. In presence of water, the aldehyde formed during the first oxidation, is further oxidized into carboxylic acid following the same reaction pathway described previously.[210] In literature there are several examples of transformation of 1-4 and 1-5 hydroxy acids into the corresponding lactones using these oxidation conditions.[211] In order to avoid the hydroxy acid cyclization, the reaction was performed on TIPS monoprotected diols.



Scheme III.9: Mechanism of TEMPO catalyzed oxidation of an alcohol into a carboxylic acid.

After purification, the structure of the resulting intermediates was confirmed by ^{13}C-NMR analysis (see experimental part). The reaction between the intermediates 18 and 21 with 4-Amino-TEMPO gave the corresponding amides. The reaction mechanism is depicted in Scheme III.10. The amidification was carried out in anhydrous dichloromethane using PyBOP. The use of PyBOP as coupling agent leads to the formation of an activated ester which is attached by the 4-Amino-TEMPO forming the resulting amide.



Scheme III.10: Amidification mechanism using PyBop as activating agent.

Following TIPS deprotection using TBAF gave the corresponding coding nitroxides in good yields. After the synthesis, the monomers 0”$_d$ and 1”$_d$.were analyzed by positive ESI-HRMS which indicated for both monomers the presence of the targeted products and the

corresponding sodium adducts at lower intensities. Moreover, peaks corresponding to side products were not detected as depicted in Figure III.8.

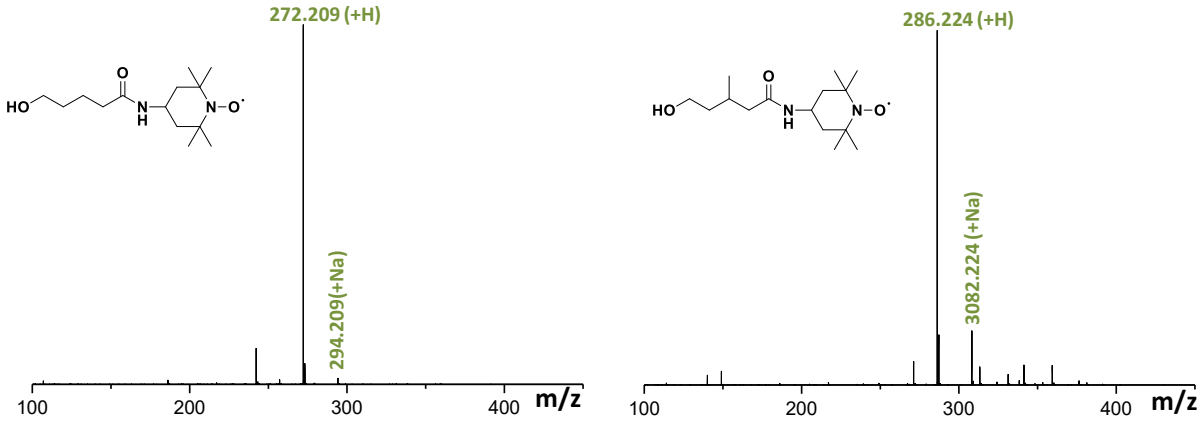


Figure III.8: Positive ESI-HRMS registered for the monomers couple forming alphabet 3.

Alphabet 3 was following implemented into the iterative protocol for the synthesis of a digitally encoded poly(alkoxyamine phosphodiester) coding for 1_a-0”$_d$-1_a-1”$_d$-0_a-0”$_d$ (3 repeating units) employing the conditions described in chapter 2. After purification and cleavage, the resulting powder was analyzed by negative ESI-HRMS and size exclusion chromatography in water acetonitrile (Figure E.9). As evidenced by negative ESI-HRMS, (Figure III.9), the targeted product was detected at different charge states ranging 2^- to 3^-. Moreover, products corresponding to side reactions, non-quantitative coupling yields or amine-terminated oligomers were not detected.

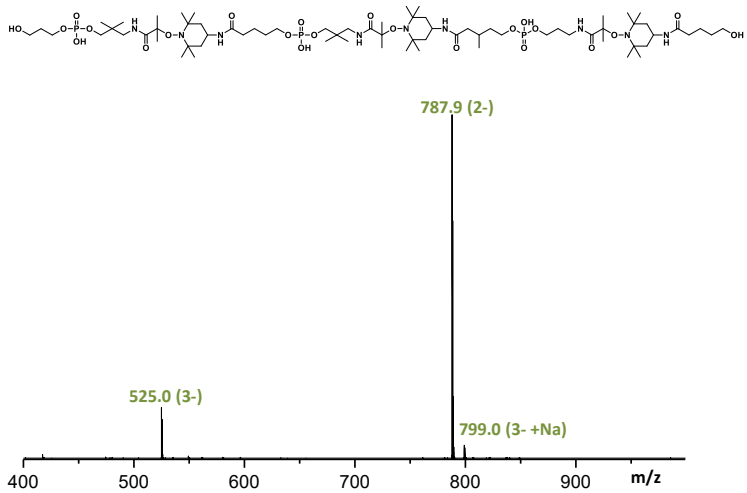


Figure III.9: Negative ESI-HRMS recorded for entry 4 tab.4. The peaks labeled in green correspond to the synthesized sequence defined polymer.

Size exclusion chromatography in water/acetonitrile further confirmed the uniformity of the macromolecule (see experimental part). These results confirmed the uniformity of the digital polymer synthesized using alphabet 3. Moreover, the product was analyzed by ^{1}H-NMR and ^{31}P-NMR (experimental part Figure E.8). Based on these promising results, the alphabet 3

was employed in step (*ii*) for the synthesis of a library of six sequence-defined poly(alkoxyamine phosphodiester)s encoding dyads (Table III.2). The iterative synthesis was performed using the conditions described in chapter 2.

Table III.2: Negative ESI-HRMS characterization of the sequence-coded polymers

	sequence	*yield*	*m/z$_{th}$*	*m/z$_{exp}$*
1	α-1$_a$-1”$_d$-1$_a$-1”$_d$	90%	556.3263	556.3261[a]
2	α-0$_a$-1”$_d$-0$_a$-1”$_d$-1$_a$-0”$_d$	73%	780.9408	780.9429[a]
3	α-1$_a$-0”$_d$-0$_a$-1”$_d$-1$_a$-0”$_d$	75%	787.9486	787.9510[a]
4	α-1$_a$-0”$_d$-1$_a$-1”$_d$-0$_a$-0”$_d$	78%	787.9486	787.9482[a]
5	α-0$_a$-1”$_d$-0$_a$-0”$_d$-1$_a$-0”$_d$-0$_a$-0”$_d$	93%	1012.5632	1012.5669[a]
6	α-1$_a$-0”$_d$-0$_a$-0”$_d$-1$_a$-1”$_d$-0$_a$-1”$_d$	85%	1033.5867	1033.5895[a]

[a] $[M-2H]^{2-}$, measured at isotopic maximum.

All the digital polymers synthesized using alphabet 3 were recovered in good yields. Afterwards, the samples were analyzed by negative ESI-HRMS and sequenced by ESI-MS/MS. The ESI-HRMS confirmed the uniformity of the sequence-defined polymers in all the cases. Indeed, the targeted products were detected at different charge states according to the number of phosphate groups they bear. Furthermore, the ESI-MS/MS allowed the reliable sequencing for all the digital polymers in Table III.2. More detail about the behaviour of these molecules in MS/MS and the sequencing theory is discussed within the next section. The results obtained suggest that the monomers 0”$_d$ and 1”$_d$ lead to quantitative yields in step *(i)* and *(ii).* In particular, the use of alphabet 3 in the iterative protocol leads to the synthesis of uniform digital polymers ensuring a reliable molecular encryption. This new molecular architecture enables the insertion of a second bit of information during the nitroxide-radical coupling increasing drastically the data density of the resulting digital polymers.

3 Sequencing *via* dyads extraction

As explained in the introduction of this chapter, the backbone modification, designed for this new class of digital polymers, allow to increase the information density maintaining the sequencing by MS/MS easy. The dissociation behaviour of the poly(alkoxyamine phosphodiester)s encoding dyads in MS/MS conditions is analogue to the previous generation. Indeed, also for this class of molecules the labile alkoxyamine bond is preferentially cleaved in CID conditions at relatively low activation energies. The homolytic cleavage originates two series of fragments: $C_i^{\cdot z-}$ with α on the left or $y_j^{\cdot z}$ with OH moiety on the right.[170] As shown in Figure III.10, the reconstruction of the original sequence can be done analysing the $c_i^{\cdot z-}$ series in which every members contains one dyad more compare to the previous one. As for the first generation, the precursor ions selected for the fragmentation in the collision cell are the fully deprotonated species. Indeed, as already explained, the fragmentation patterns originated from these species, are easier to correlate

to the original structure because of the absence of fragments having different ionization states in the series. The feature that makes the sequencing easy is the formation of encoding dyads in MS/MS. Indeed, in this class of digital polymers, the alkoxyamine bond cleavage generates encoding dyads that can be easily distinguished in MS/MS because of their different mass signature. The MS/MS sequencing was successfully achieved for all the digital polymers synthesized using alphabet 3.

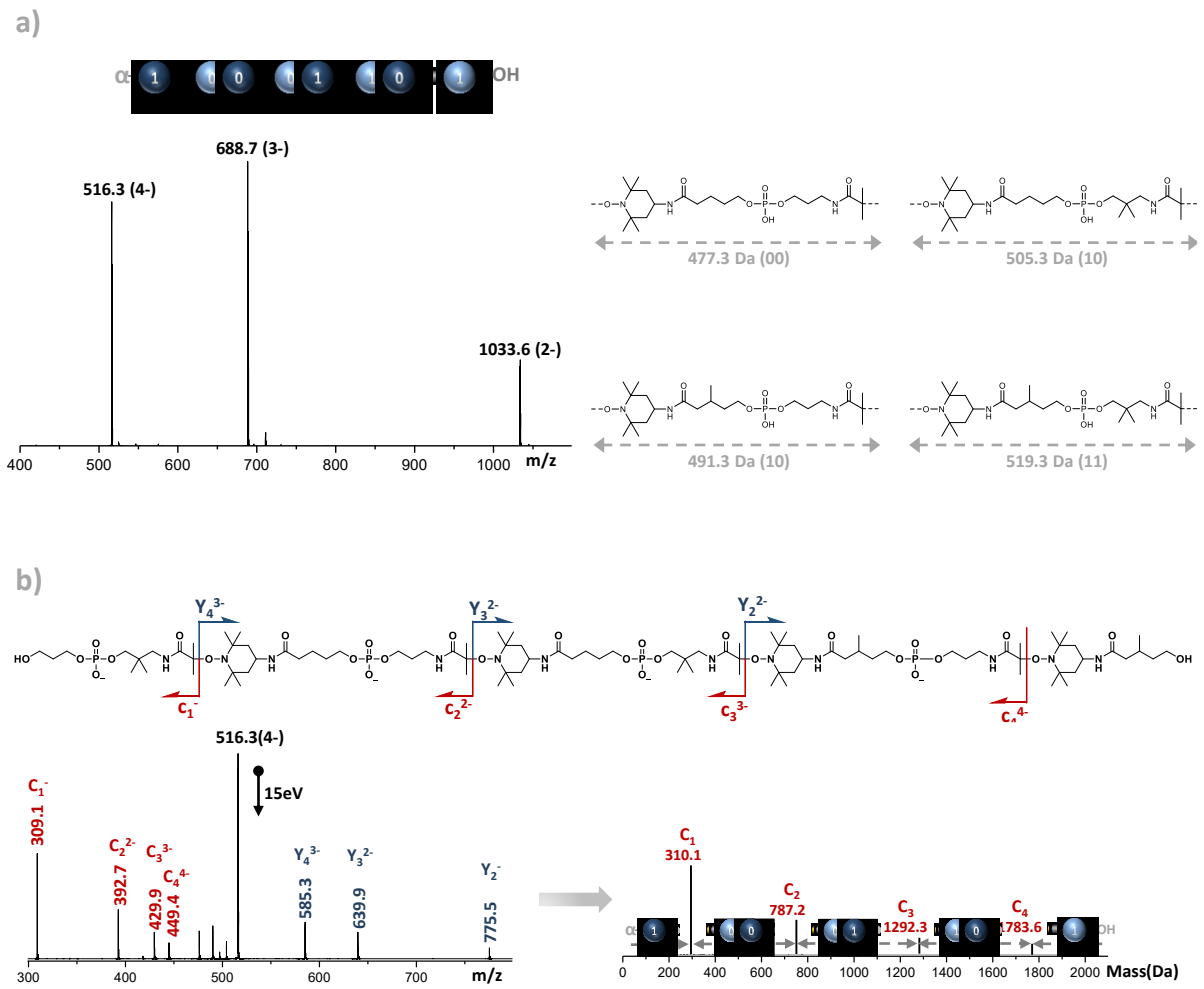


Figure III.10: Mass spectrometry analysis of the sequence-coded oligomer entry 5 tab 4: **a)** Negative ion mode ESI mass spectrum and molecular structure with the molar mass of the coded dyads produced in tandem MS (left); **b)** Molecular structure of the fully deprotonated oligomer with fragments produced in tandem MS; ESI-MS/MS obtained for this oligomer (left); and charge state deconvolution (right) of the $c_i^{\bullet z-}$ ion series (blue labeled in MS/MS),the peaks not labeled in tandem MS correspond to Internal fragment generated during secondary dissociation reactions.

Interestingly, the sequencing becomes even easier after charge sate MS/MS deconvolution. Indeed, as depicted in Figure III.10 b, two bits of information can be easily extracted from the single peak-to-peak distance in the ion series, making the resulting sequencing faster and easier. MS-DECODER is a software developed for the decoding of synthetic sequence-defined polymers.[212] In the past decades different software for data extraction from biopolymers as proteins and DNA were developed.[213, 214] The use of these algorithms can drastically reduce the time required for the sequencing of long polymeric chains. In this context, a new algorithm was implemented for MS-DECODER supporting dyads detection of

sequence-defined poly(alkoxyamine phosphodiester)s. The software, using the MS/MS spectra as input, enables the automatic sequencing, reducing drastically the time required for the data extraction from these digital polymers. MS-DECODER was tested for the six digital polymers synthesized using alphabet 3, allowing in all the cases the reliable decoding of the sequences in a few milliseconds.

3 Conclusions and perspectives

In conclusion, the sequence defined poly(alkoxyamine phosphodiester) backbone was modified in order to increase the data density for this digital polymers class. In the frame of this project, three different alphabets were tested in step (*ii*) of the iterative protocol. The new molecular architecture allows to increase the information density, keeping a predictable and "easy to interpret" dissociation pattern in MS/MS. The new molecular architecture brings several advantages to this class of digital polymers. First of all, the use of four coding monomers allows the insertion of two bits in each repeating unit, increasing drastically the information density. Moreover, this new architecture allows to reduce the steps required for the polymer synthesis. In addition, the binary encryption at the molecular level makes these molecules still compatible with the existing information technology. The dyads, generated in CID conditions after alkoxyamine homolytic cleavage, simplifies the data extraction from digital polymers, making the sequencing fast and efficient. In addition, the MS-DECODER software, developed for the dyads detection of these digital polymers allows the efficient decoding of long polymeric chains in a few milliseconds. This concept, which in this work was assessed for the poly(alkoxyamine phosphodiester) class, in principle could be applied for every digital polymer. Indeed, the only features required for the dyad encoding is the employment of quaternary alphabet which generates four cleavable units with different mass signatures. It is possible achieving this result using four different monomers in two different reaction steps but also using four different coding monomers containing different mass in a single step. Moreover, in order to produce longer digital polymers chains in a relatively short time, all the possible dyads combination could be synthesized in solution. Afterward, the dyads could be applied to an automated synthesizer machine in order to speed the iterative synthesis up. To conclude, the work presented in this chapter highlights the unique possibility to tune the digital polymer backbone in order to obtain tailor-made features. Indeed, in this example, a simple modification of the monomers, allowed to enhance the storage capacity of the resulting digital polymers and to keep the sequencing by MS/MS easy.

Chapter IV

Poly(alkoxyamine phosphodiester)s containing alkoxyamine bonds of different strength

1 Introduction

In literature there are several examples reporting how the molecular architecture of synthetic sequence-defined polymers has been tuned in order to obtain an optimal reading using different sequencing techniques.[184, 215] A recent proof-of-concept in which the backbone of sequence-defined poly(alkoxyamine phosphodiester)s was modified in order to obtain an enhanced storage capacity was described in the previous chapter. The iterative protocol developed for the synthesis of these sequence-defined polymers relies on phosphoramidite coupling and the NRC (see Chapter 2). The insertion of a labile alkoxyamine bond in the polymer backbone enables the easy sequencing by MS/MS of these molecules. The alkoxyamine bond is generated during the nitroxide-radical coupling between a carbon-centered radical and a nitroxide. This reaction, initially developed for spin-trapping,[192] was intensively studied in the past years because of its importance in the nitroxide-mediated polymerization (NMP).[51] More recently, the reversibility of the alkoxyamine bond was exploited for the synthesis of responsive polymers[216], to implement dynamic reaction networks.[217] and for the synthesis of self-healing polymers.[218] The nitroxides stability and the factors which influence alkoxyamine bond dissociation were intensively investigated in the past years to better understand the role of these molecules in NMP.[219] Moad and Rizzardo discovered that the alkoxyamine bond stability is influenced from polar, electronic and steric factors; among all these parameters, the steric hindrance is the most relevant.[220] Hence, the structure of the nitroxides employed in the NRC could influence the stability of the resulting alkoxyamine bonds.

In this context, the use of two different nitroxides in step (*ii)* of the iterative protocol described in Figure IV.1 allows the insertion of bonds having different strength in the poly(alkoxyamine phosphodiester) backbone. This feature could be exploited to obtain a stepwise selective fragmentation. The stability of the alkoxyamine bond in the digital polymers backbone was already investigated for the poly(alkoxyamine amide) class.[14] In this study short poly(alkoxyamine amide)s were synthesized using different combinations of coding monomer and nitroxide. The resulting MS/MS studies demonstrated that the PROXYL nitroxide generates more stable alkoxyamine bonds compare to the ones originated from the TEMPO. These results are in line with the existing literature as is discussed within the next sections.

The molecular architecture described in this chapter could potentially facilitate the sequencing making more efficient the data extraction from long polymeric chains. In order to develop a stepwise selective cleavage for poly(alkoxyamine phosphodiester)s, a new molecular architecture, depicted in Figure IV.1 was designed. This backbone modification enables the selective alkoxyamine bond cleavage in specific positions of the polymer and could result extremely useful to simplify the sequencing by MS/MS. Moreover, in order to obtain encoding dyads and to increase the information density, the hydroxy-functionalized

nitroxides could bear coding moieties (see chapter 2). Hence, the resulting digital polymers could show an improved storage capacity and readability.

The alkoxyamine bond may be fragmented not only in MS conditions. Indeed, is well known from literature that the homolysis of this bond may occur also *via* a thermal degradation. [221, 222] For instance, it was already observed for poly(alkoxyamine amide)s that the alkoxyamine bonds start to degrade at 65°C.[11]

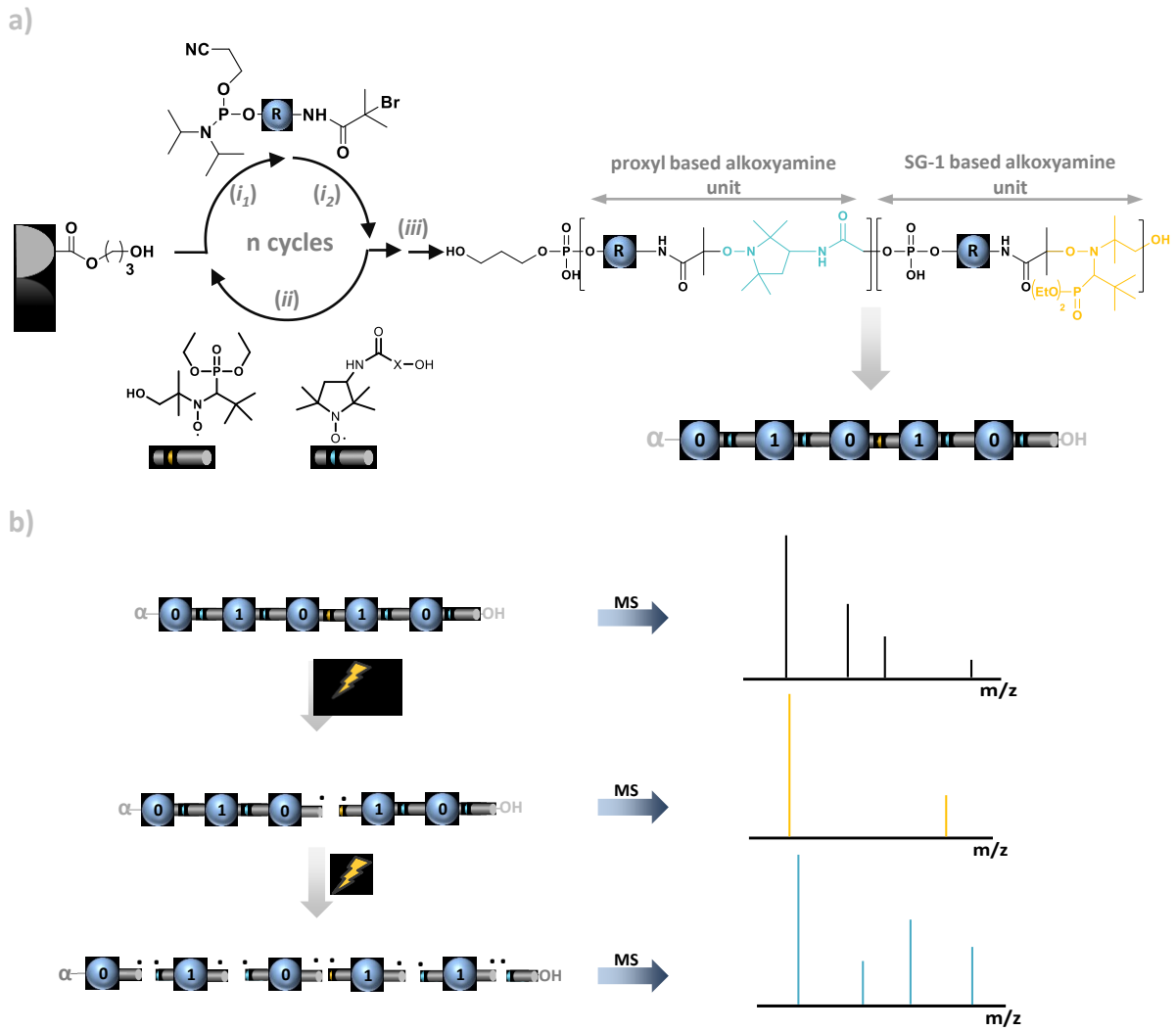


Figure IV.1: General concept of stepwise selective fragmentation. **a)** Iterative cycle for the synthesis of poly(alkoxyamine phosphodiester)s containing selective cleavable units and chemical structure of the resulting sequence-defined polymers. **b)** General concept of stepwise selective cleavage.

In this chapter, the development of the new molecular architecture allowing obtaining stepwise selective alkoxyamine bond fragmentation in MS/MS is presented. In particular, the screening of different combinations of nitroxide will be described. Furthermore, the synthesis of the monomers employed in this work and the preliminary MS/MS experiments, focused on the research of the optimal MS/MS conditions to obtain a stepwise selective fragmentation will be discussed.

2 Result and discussion

2.1 Monomers design and synthesis

2.1.1 Monomers design

The poly(alkoxyamine phosphodiester)s containing selective cleavable units were synthesized using an iterative solid-state approach, following the conditions described in chapter 2. In this work, the monomers 0_a and 1_a were employed in the phosphoramidite coupling. These building blocks were synthesized following the synthetic strategy described in chapter 2. In the frame of the project, a library of three different hydroxy-functionalized nitroxides (T_3, SG1-OH and Pr_2) was investigated in order to obtain the selective alkoxyamine cleavage in MS. As depicted in Figure IV.2, these monomers have linear and cyclic structures.

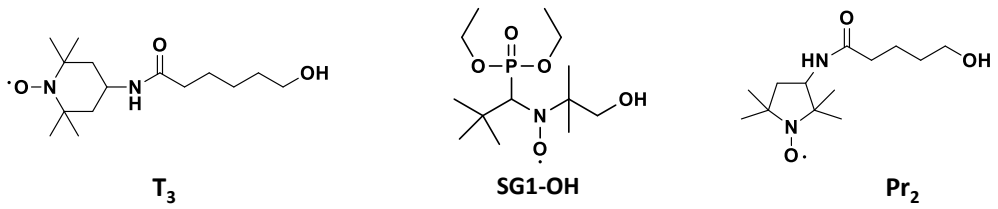
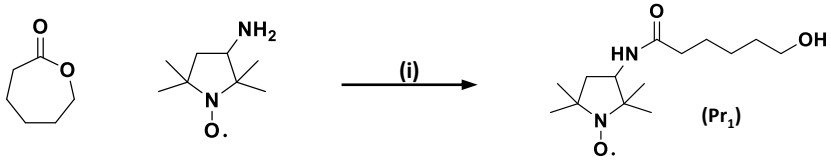


Figure IV.2: Structures of the hydroxyl-functionalized nitroxides employed in his work.

The monomer T_3, was synthesized following the synthetic strategy described in chapter 2. The SG1-OH was supplied by Jean-Louis Clément from Marseille University and the monomer Pr_2 was synthesized following the synthetic strategy depicted in Scheme IV.2. The different nitroxides employed in this work, allow to obtain alkoxyamine bonds with different strength in the polymer backbone. Indeed, as demonstrated form Rizzardo and Moad, the steric hindrance, the polarity of the substituents, the ring size, etc. influence the stability of the alkoxyamine bond.[220] The nitroxides forming the library, generate alkoxyamine bonds having an increasing stability in the order SG1-OH→T_3→Pr_2. Hence, the SG1-OH was tested in combination with T_3 and Pr_2 in order to find a couple of nitroxides suitable for obtaining a selective alkoxyamine bond cleavage.

2.1.2 Synthesis of monomer Pr_2

The synthesis of monomer Pr_2 was performed following the synthetic strategy depicted in Scheme IV.2. The choice of this nitroxide was done in order to obtain a more stable alkoxyamine bond compare to the T_3-based ones. Indeed, according to the study of Rizzardo and Moad, the stability of the alkoxyamine bond increases when the size of the ring decreases.[220] For this reason, a five-membered ring nitroxide was chosen. The first synthetic pathway employed for the synthesis of the hydroxy-functionalized PROXYL derivative (Pr_1) is depicted in Scheme IV.1.



Scheme IV.1: Synthetic strategy employed for the synthesis of monomer Pr_1.

The strategy investigated would have enabled the easy synthetic access to nitroxide Pr_1, making this molecule extremely convenient for the protocol implementation. Indeed, the synthesis of Pr_1 was performed following the procedure employed for the synthesis of nitroxide T_3, which is described in chapter 2. However, in these conditions the final product was not recovered. The reaction was then tried using harsh conditions, after purification the recovered compound was analyzed by positive ESI-HRMS.

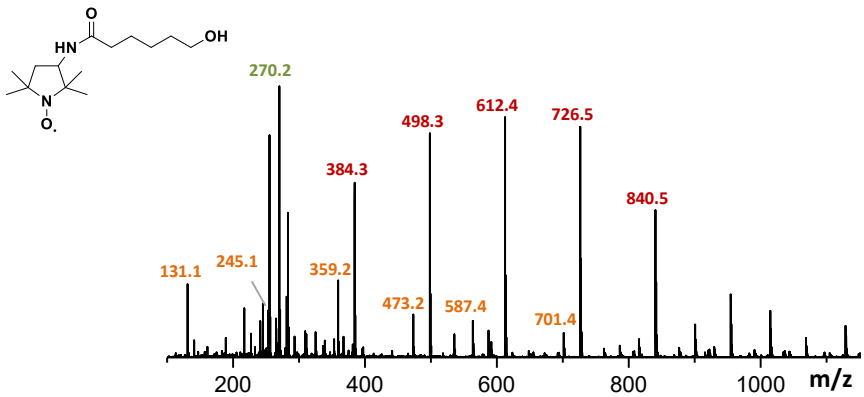
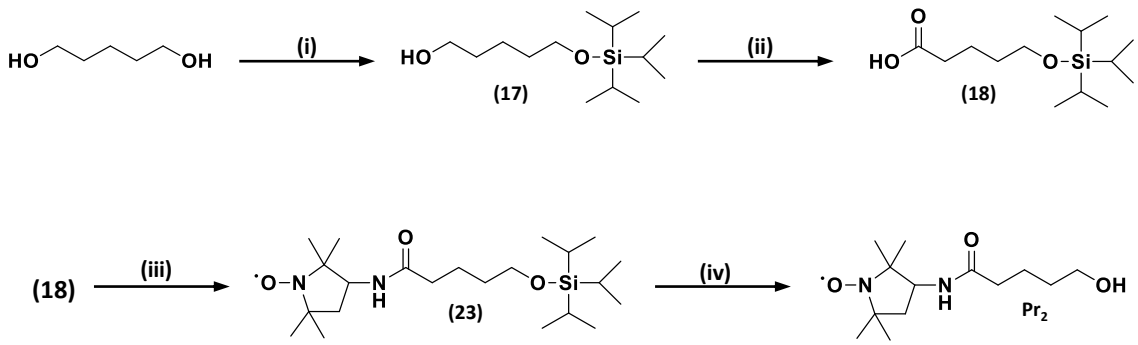


Figure IV.3: ESI-HRMS registered for monomer Pr_1, the peaks labeled in green and in red there correspond to the two series generated from the Ɛ caprolactone aggregation in ESI source. The green labeled peak corresponds to the desired product.

A shown in Figure IV.3, the ESI-HRMS exhibits two peaks distributions spaced by 114 Da (orange and red labeled peaks). This indicates that the sample contains large amount of Ɛ caprolactone that aggregates in ESI conditions. In particular, the red labeled peaks correspond to the targeted product which aggregates with the Ɛ-caprolactone, whereas the orange labeled series correspond to the lactone aggregation. The desired product was also found in the mixture (green labeled peak) however, the low yields obtained and the difficulties to purify the reaction mixture led to synthesis of a second hydroxy-functionalized PROXYL-based monomer (Pr_2). This molecule was obtained in four steps, following the synthetic strategy depicted in Scheme IV.2.



Scheme IV.2: Synthetic strategy employed for the synthesis monomer Pr_2. **(i)** TIPS protection: TIPS-Cl, imidazole, DMF overnight; **(ii)** oxidation: TEMPO, BAIB, dichloromethane, water, 3h, rt; **(iii)** 3-Amino-PROXYL coupling: 3-Amino-PROXYL, DIPEA, PyBOP, dichloromethane, 0°C then rt, overnight; **(iv)** TIPS deprotection: TBAF, THF, 2.5h, rt.

The first step was the protection of one alcoholic function of the 1-5 pentanediol using TIPS-Cl to afford the intermediate 17. The monoprotected diol was oxidized then *via* TEMPO-mediated oxidation to afford the corresponding carboxylic acid. As already described in chapter 3, the diol was monoprotected using TIPS-Cl to avoid the formation of the corresponding lactone in the oxidation conditions. Following amidification of the intermediate 18 with 3-Amino-PROXYL, allowed to obtain the intermediate 18. The final TIPS deprotection using TBA, gave the monomer Pr_2 in good yields. The resulting compound was analyzed by positive ESI-HRMS.

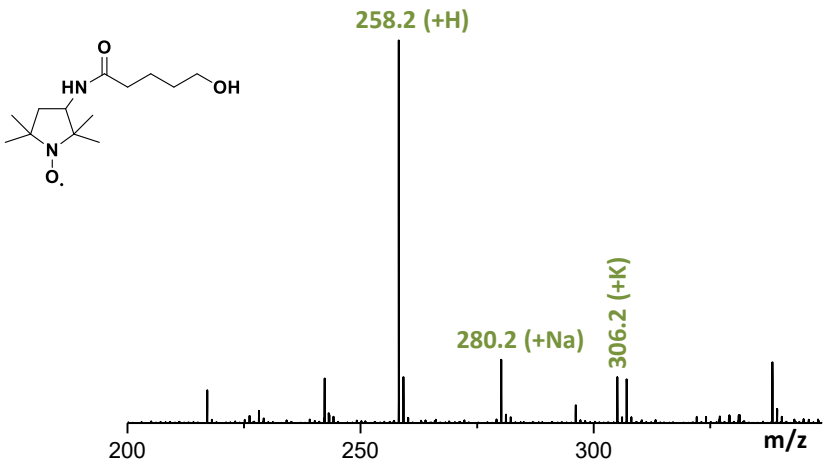


Figure IV.4: Positive ESI-HRMS recorded for monomer Pr_2.

As shown in Figure IV.4, the targeted product was detected as main peak moreover, the corresponding sodium and potassium adducts were also detected at lower intensities. In addition, peaks originated from side reactions were not found in ESI-HRMS. The building block was following employed in combination with SG1-OH for the synthesis of sequence-defined poly(alkoxyamine phosphodiester)s containing alkoxyamine bond of different strength. The results obtained using this nitroxide in combination with SG1-OH are described in section 2.3 of this chapter.

2.2 Synthesis of uniform sequence-defined poly(alkoxyamine phosphodiester)s using SG1-OH/T_3

The first hydroxy-functionalized nitroxides set which was tested to obtain selective alkoxyamine bond cleavage was formed by SG1-OH and the monomer T_3. The nitroxide T_3 was chosen because of its easy synthetic access. Indeed, the molecule was synthesized in one step following the synthetic procedure described in chapter 2. The second building employed was the β phosphorylated nitroxide SG1-OH (Figure IV.2). This monomer was found to be one of the most useful and versatile nitroxides for NMP.[223] The steric hindrance of the SG1-OH increases the dissociation constant[224] making the alkoxyamine bond, generated during the NRC, more labile compare to the cyclic T_3-based alkoxyamine bonds. In the first test, the nitroxides SG1-OH and T_3 were employed in combination with 0_a for the synthesis of a hexamer coding for 0_a-T_3-0_a-SG1-0_a-T_3. The sequence-defined polymer was synthesized using the conditions described in chapter 2, increasing the reaction time to one hour for the SG1-OH coupling. After cleavage, the sequence-defined poly(alkoxyamine phosphodiester) was analyzed by negative ESI-HRMS.

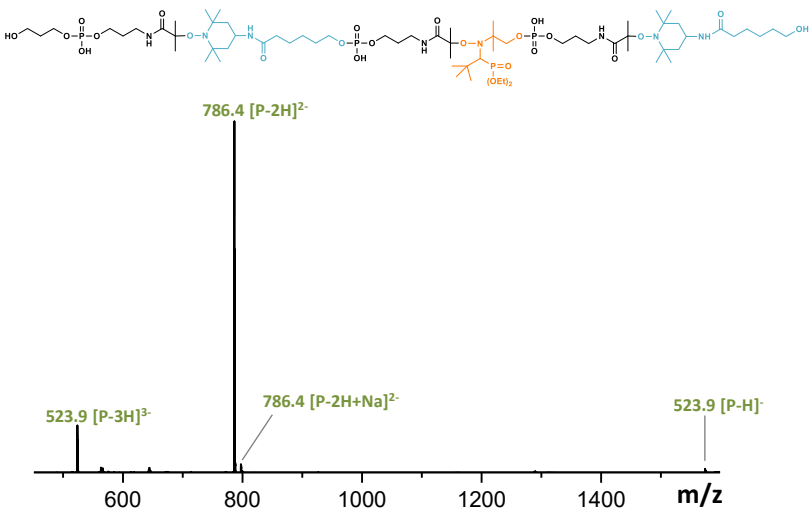


Figure IV.5: Negative ESI-HRMS registered for a polymer coding for 0_a-T_3-0_a-SG1-0_a-T_3.

As depicted in Figure IV.5, the targeted product was detected at different charge states ranging from 1^- to 3. The generation of these series is due to the phosphate groups contained in each repeating unit. In this experiment, a small percentage of the sodium adduct of the two times deprotonated species was also detected. Moreover, side products originated from incomplete coupling yields or disproportionation reactions were not detected. Afterwards, the selective alkoxyamine cleavage was tested in MS/MS at 10 eV. As for all the MS/MS experiments performed on poly(alkoxyamine phosphodiester)s, the precursor ion selected was the fully deprotonated species. Indeed, as it was already discussed in chapter 2, these species originates easy to interpret fragmentations patterns.

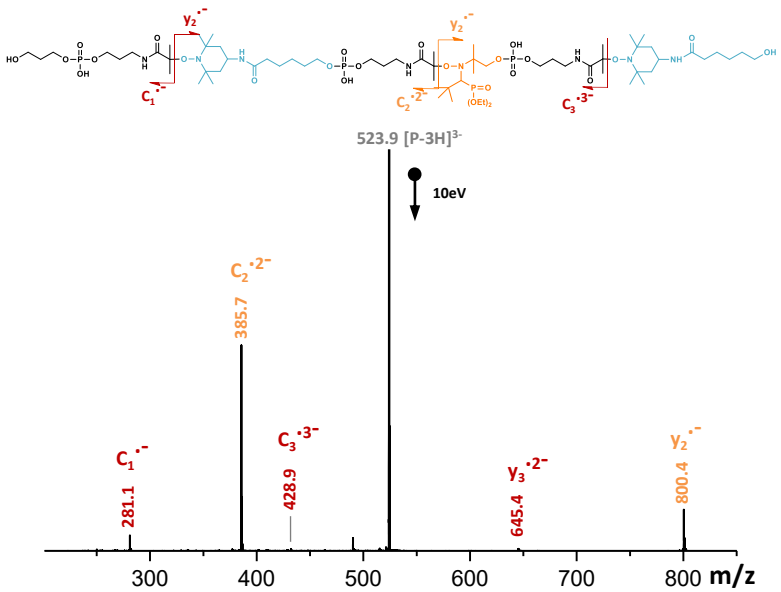


Figure IV.6: Molecular structure of the fully deprotonated oligomer with the fragments produced in tandem MS originated from the cleavage of all the alkoxyamine bonds and ESI-MS/MS of the same oligomer.

The peaks originated from the homolytic rupture of SG1-based alkoxyamine bond were detected at high intensities (orange labeled peaks in Figure IV.6), indicating that the rupture of these bonds in favored in MS/MS. However at 10eV, fragments corresponding to the homolytic cleavage of T_3-based alkoxyamine bonds were also detected at lower intensities (red labeled peak in Figure IV.6). In order to clarify if these nitroxides could be selectively cleaved using lower ionization energies, the “breakdown curves" was calculated by the group of Prof. Laurence Charles. In this study, the homolytic rupture is put in relation with the ionization energy. As shown in Figure IV.7, the breakdown curves show that even at lower ionization energies, the cleavage of the SG1-based alkoxyamine bond is preponderant. However, the rupture of the T_3-based alkoxyamine bond was always detected at lower intensities. This experiment confirmed that the SG1-OH/T_3 combination allowed to obtain the preferential fragmentation of the SG1-based alkoxyamine bonds in the polymer backbone. However, small traces of the products originated from the homolytic rupture of the T_3-based alkoxyamine bond were always detectable. For this reason, the possibility to employ a more stable alkoxyamine bond in combination with the SG1-OH in order to obtain a fully selective fragmentation was taken into account.

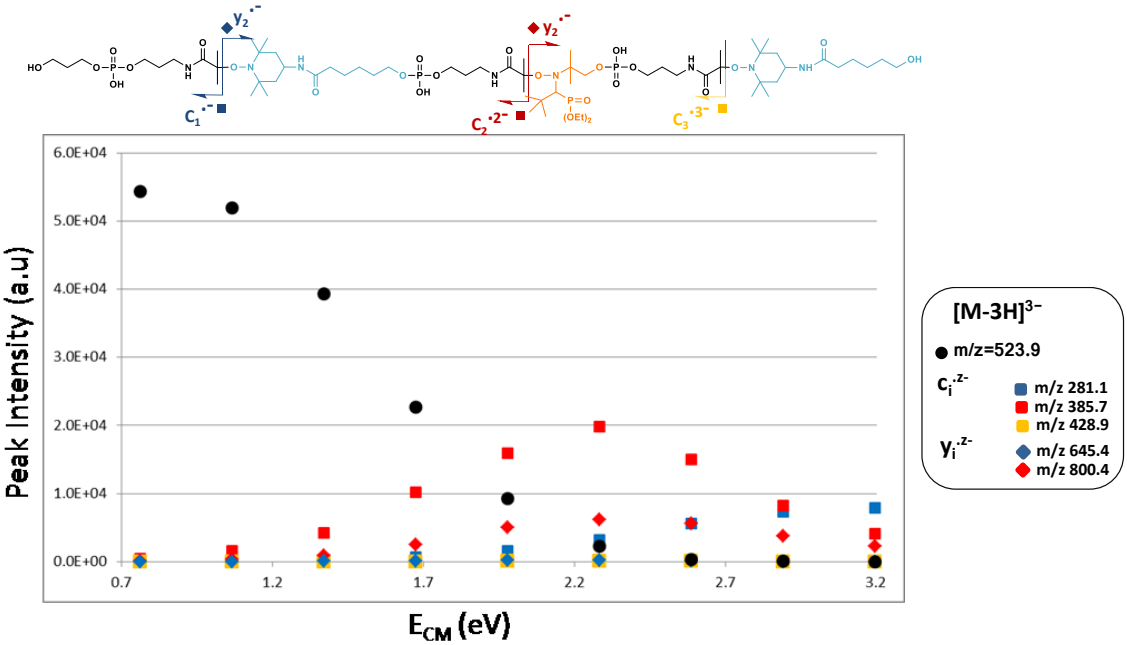


Figure IV.7: Breakdown curves measured for the oligomer coding for 0_a-T_3-0a-SG1-0_a-T_3.

2.3 Synthesis of uniform sequence-defined poly(alkoxyamine phosphodiester)s using SG1-OH/PROXYL

After the first test, the possibility to use a five membered ring nitroxide in combination with the SG1-OH, to obtain a selective fragmentation was investigated. Indeed, as it was explained previously, the PROXYL-based alkoxyamine bonds are more stable compared to the TEMPO-based ones because of the smaller ring size. Hence, the monomer Pr_2 was employed with SG1-OH and the monomer 1_a for the synthesis of a sequence-defined tetramer coding for 1_a-SG1-1_a-Pr_2. The tetramer was synthesized following the conditions described in Chapter 2. Moreover, the reaction times for the NRC were increased to one hour. After purification and cleavage, the resulting product was analyzed by negative ESI-HRMS. The MS analysis confirmed the uniformity of sequence-defined polymer synthesized using these monomers. Indeed, the targeted product was detected at different charge states ranging from 1^- to 2^-, the sodium adduct of the mono-deprotonated species was also detected at low intensities. Moreover, peaks corresponding to side reactions or *non*-quantitative coupling yields were not found. The fully deprotonated species (red labeled peak in

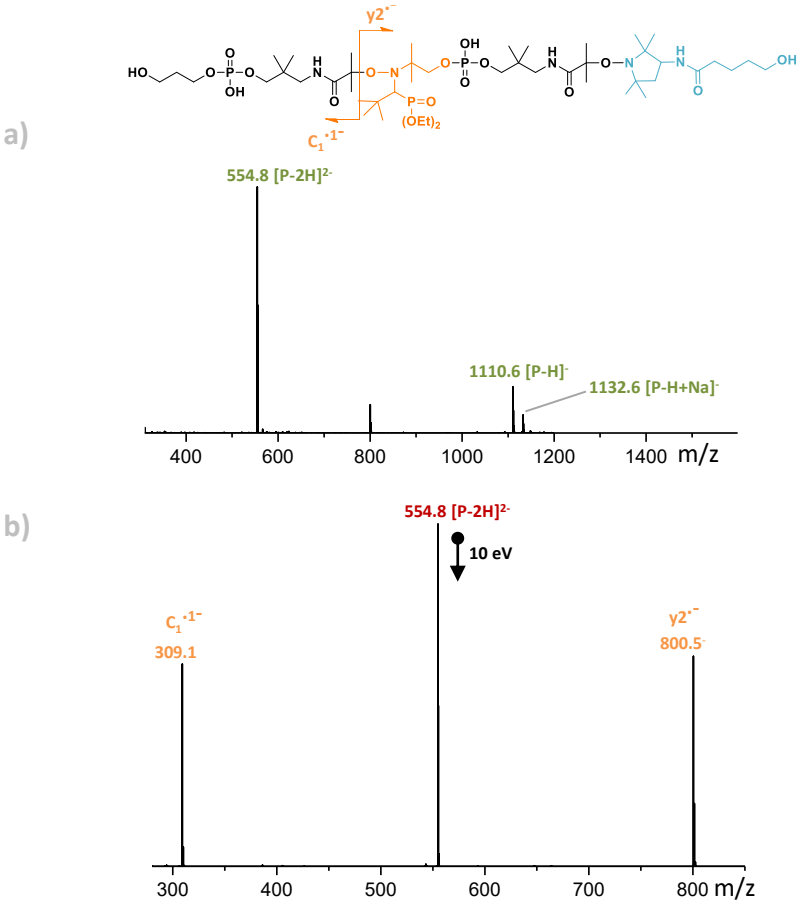


Figure IV.8 b), was then selected as precursor ion and it was fragmented in MS/MS at 10 eV.

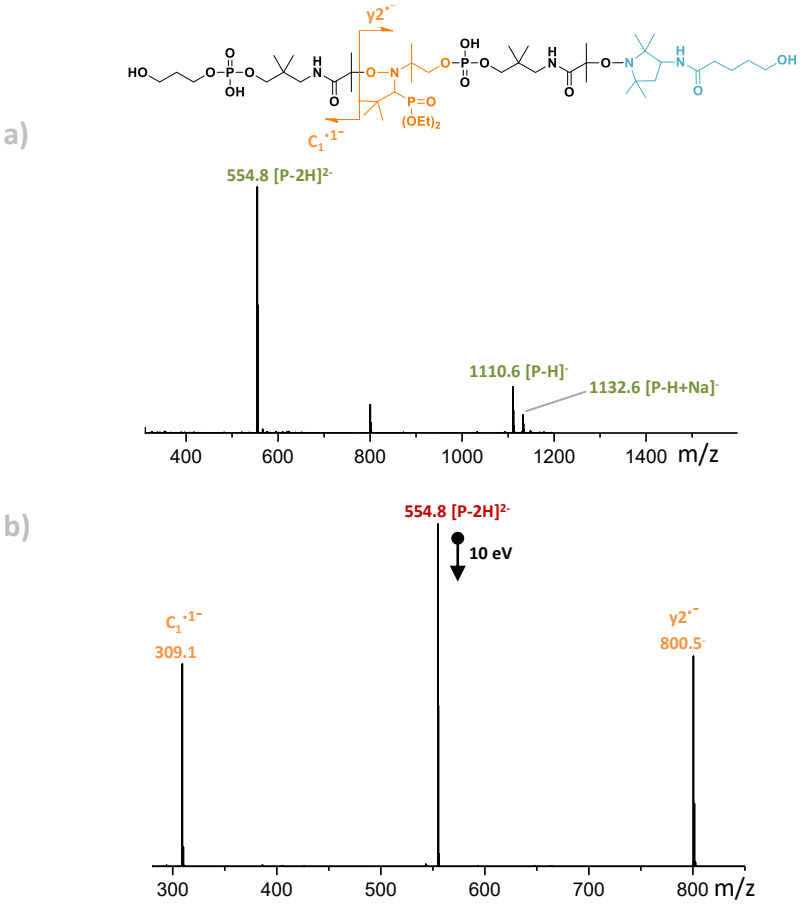


Figure IV.8: ESI-HRMS and ESI-MS/MS analysis of a sequence-defined polymer coding for 1_a-SG1-1_a-Pr_2. **a)** ESI-HRMS of a polymer coding for 1_a-SG1-1_a-Pr_2. **b)** ESI-MS/MS of the same polymer using the fully deprotonated species (grey labeled peak) as precursor ion. The orange labeled peaks originate from the selective SG-1 based alkoxyamine bond cleavage.

In these conditions is possible to obtain the selective fragmentation of the SG1-based alkoxyamine bond. Indeed, as depicted in

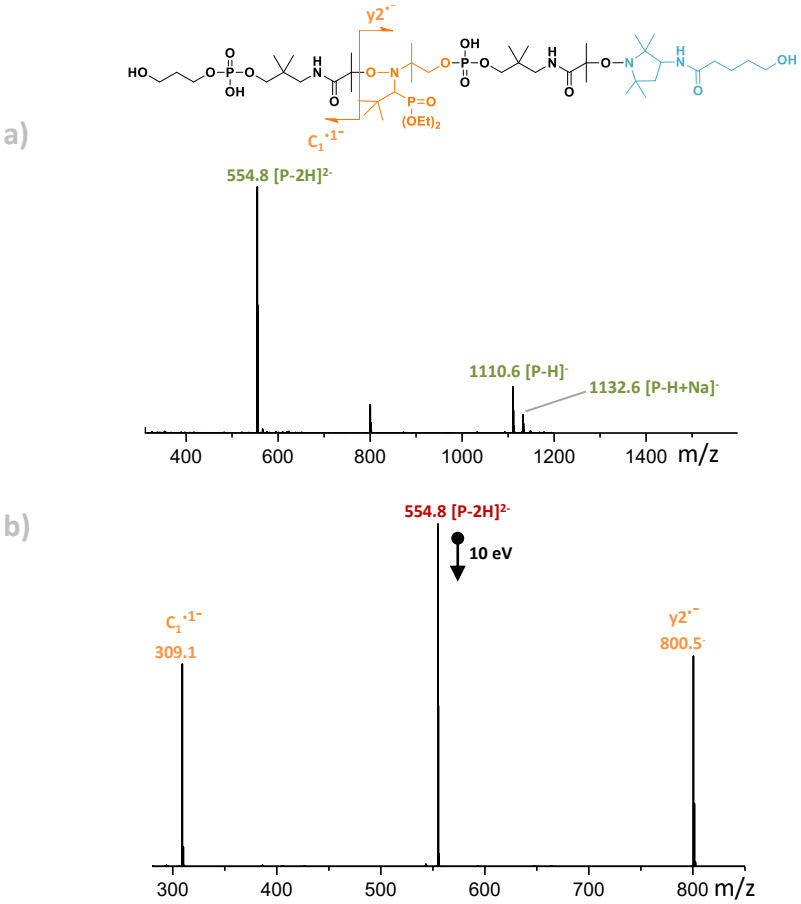


Figure IV.8b only the peaks corresponding to the fragments originated from the cleavage of SG1-based alkoxyamine bond were detected (orange labeled peaks). The sequence-defined tetramer was then analyzed by ^{1}H and ^{31}P-NMR. Interestingly, the phosphorous NMR showed one peak a t 0ppm, corresponding to the phosphate group and one peak at 28.4 ppm corresponding to the phosphonate moiety[225] (see Figure E.12). After this first encouraging result, the syntheses of an octamer and a dodecamer coding respectively for 1_a-Pr_2-1_a-SG1-0_a-Pr_2-1_a-SG1 and for 1_a-Pr_2-1_a-SG1-0_a-Pr_2-1_a-SG1-P_0-SG1-P_0-Pr_2 were performed. The iterative synthesis was performed following the conditions described previously. After eight coupling, the resin was split in two parts, one part was purified and cleaved to obtain the octamer; the second part was involved in two additional iterative cycles to obtain the dodecamer. In

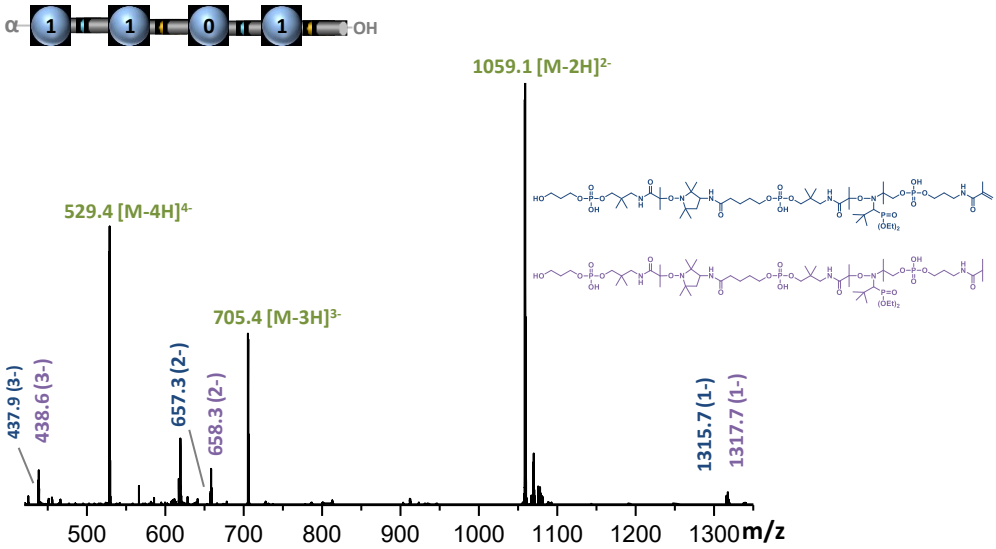


Figure IV.9 is depicted the negative ESI-HRMS obtained for the sequence-defined octamer.

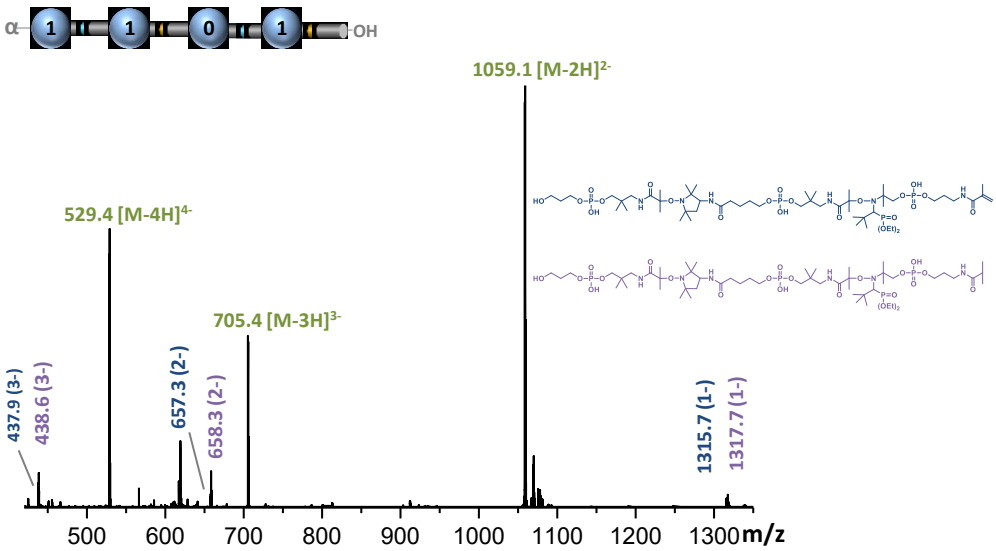


Figure IV.9: ESI-HRMS recorded for a sequence-defined polymer coding for 1_a-Pr_2-1_a-SG1-0_a-Pr_2-1_a-SG1. The peaks labeled in green correspond to the desired product. The blue and purple labeled peaks correspond to impurities.

The targeted product was detected at different charge states ranging from 2^- to 4^-. Moreover, impurities, probably generated form a disproportionation reactions during the second NRC of the PROXYL derivative were detected at low intensities (purple and blue labeled peaks). The proposed structures for the impurities found in ESI-HRMS are depicted in

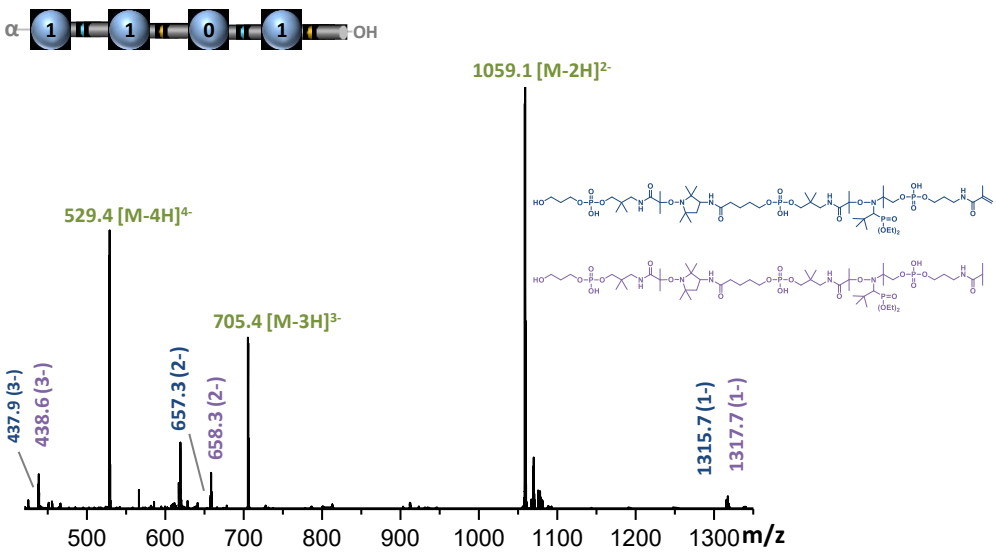


Figure IV.9. The presence of side-products in the sample is further confirmed from the size exclusion chromatogram registered in water/acetonitrile that indicates the presence of contaminations in the range of lower molecular weights. These impurities are probably originated from the disproportionation reactions during the NRC.

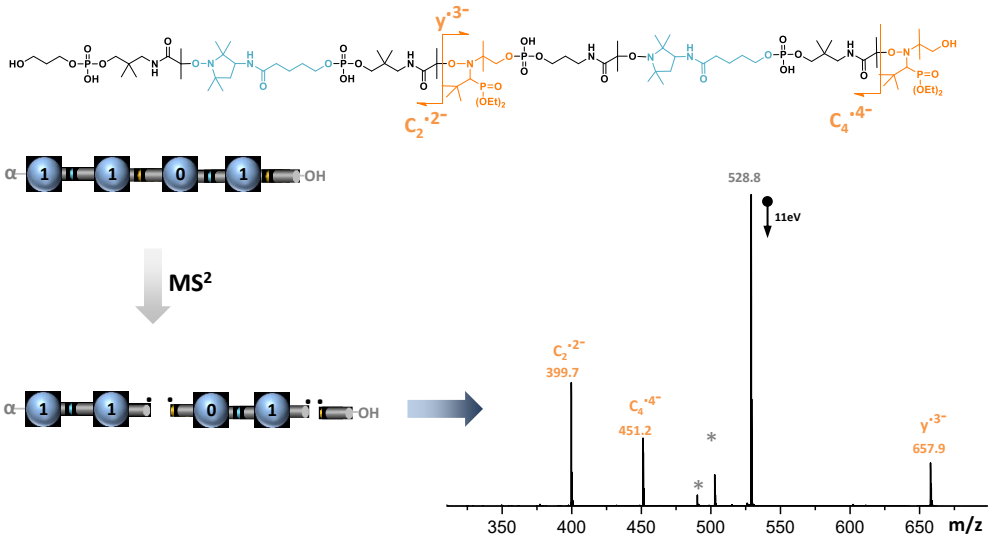


Figure IV.10: ESI-MS/MS of the precursor ion corresponding to 528. (grey labelled peak). The peaks labelled in orange correspond to the fragments originated from the SG1-based alkoxyamine bond cleavage. The peaks labelled with a star correspond to internal fragments.

After the MS analysis, the fully deprotonated species (grey labeled peaks in Figure IV.10) was selected as precursor ion for the selective cleavage of the SG1-based alkoxyamine bonds. At 11 eV, only the fragments originated from the cleavage of the SG1-based alkoxyamine bonds were detected in MS/MS (orange labeled peaks). This test further demonstrated that this combination of nitroxides, inserted into the polymer backbone, enables the stepwise selective alkoxyamine bond fragmentation. The stepwise selective alkoxyamine cleavage was

tested also for the dodecamer which was synthesized using the same solid-support. As shown in Figure IV. 11, the sequence-defined dodecamer was detected in negative HRMS as preponderant species at different charge states ranging from 3^- to 6^-.

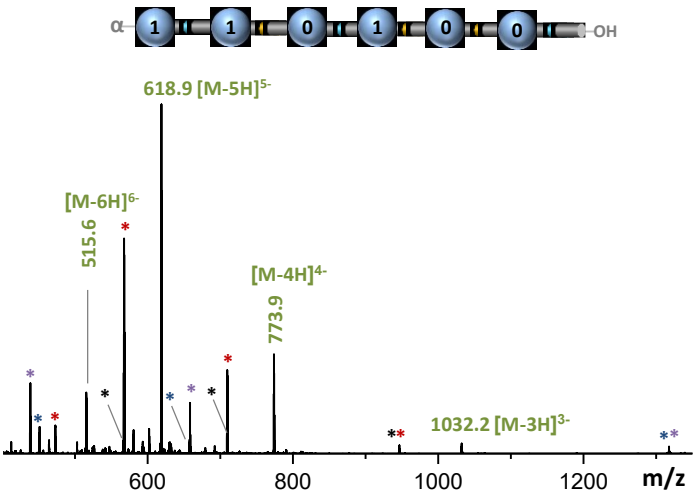


Figure IV.11: ESI-HRMS recorded for a sequence-defined polymer coding for1_a-Pr_2-1_a-SG1-0_a-Pr_2-1_a-SG1-P_0-SG1-P_0-Pr_2. The peaks labeled in green correspond to the desired product. The peaks labelled with a star correspond to impurities originated from disproportionation reactions.

However, minor impurities, generated form disproportionation reactions during NRC were detected also in this sample; these contaminations were found at different charge states (peaks labeled with a star in Figure IV.11). After ESI-HRMS analysis, the fully deprotonated species (grey labeled peak in Figure IV.12) was selected as precursor ion for the selective cleavage of the SG1-based alkoxyamine bonds in the backbone. As shown in Figure IV.12, the homolytic rupture was successfully achieved at 10 eV. Hence, in all the experiments the selective fragmentation of the SG1-based alkoxyamine bonds was achieved using ionization energies between 10 and 11eV. These promising results show that a stepwise selective cleavage is possible tuning the stability of the alkoxyamine bonds into the poly(alkoxyamine phosphodiester) backbone. However, the NRC reaction still needs to be optimized. For instance, the formation of side products during the coupling of Pr_2 with the growing chain may be avoided.

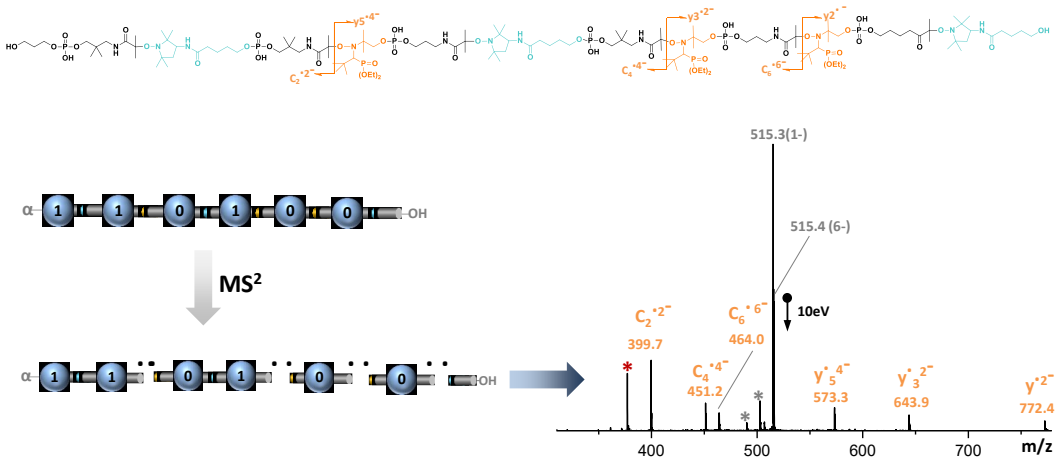


Figure IV.12: ESI-MS/MS of the precursor ion corresponding to 515.3. (grey labelled peak). The peaks labelled in orange correspond to the fragments originated from the SG1-based alkoxyamine bond cleavage. The peaks labelled with a star correspond to internal fragments.

3 Conclusions and perspectives

In this chapter it was demonstrated that the insertion of two alkoxyamine bonds having different stability in the poly(alkoxyamine phosphodiester) backbone allows to obtain a stepwise selective homolytic fragmentation. This first proof-of-concept, presented in this chapter, further demonstrates that the backbone modification of synthetic sequence-defined polymers could be exploited to obtain tailor-made features. Even taught the results presented herein seem very promising, the design of the building blocks and the molecular architecture could be further improved. In this context, a new molecular architecture in which the SG1-based alkoxyamine bonds are placed between each byte could be designed. Hence, the first alkoxyamine cleavage allows the production of fragments containing a single byte. Each fragment could be selected in MS and further fragmented using an higher ionization energy. This molecular architecture should allow the easy sequencing of long polymers chains. A similar concept was already proposed from our group,[184] however in this project, the second cleavage was performed for polyphosphates fragments which leads to complicated MS spectra compare to poly(alkoxyamine phosphodiester)s (see section 5.5 chapter 1). The molecular design proposed herein could facilitate the sequencing of the selected bytes. Moreover, hydroxy-functionalized nitroxides, containing coding unit could be inserted in the polymer backbone, enabling the formation of encoding dyads in MS^3 and ensuring a high storage capacity for the resulting digital polymers (see chapter 3). The general concept proposed herein is depicted in Figure IV.13.

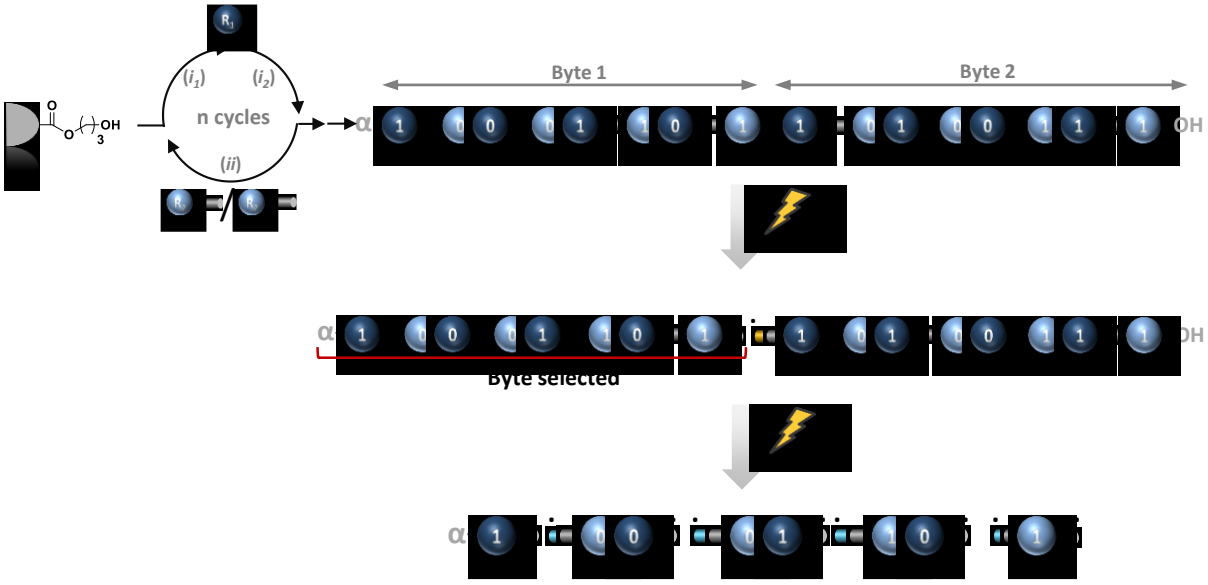


Figure IV.13: General concept proposed to obtain byte fragmentations and sequencing *via* stepwise selective alkoxyamine bond fragmentation. During the first fragmentation the sequence is fragmented in byte. The second fragmentation allows the byte fragmentation.

In the frame of the same concept, all the possible combination of dyads could be previously synthesized in solution and applied to an automated synthesizer in order to speed the iterative synthesis up and to obtain long sequences in a relatively short time.

Additionally, the stepwise fragmentation could be performed not only in MS conditions. Indeed, it was demonstrated that the alkoxyamine bond can be cleaved electrochemically [226] or *via* a thermal degradation.[227] Hence, the selective fragmentation of the polymer backbone may be performed using different kind of stimuli. Moreover, the backbone could be further modified adding more than two nitroxide-based monomers to obtain a multi-step selective fragmentation.

To conclude, the concept described in this chapter brings several interesting advantages to the field. Indeed, the implementation of this new molecular architecture could enable the synthesis of long polymers easy to sequence in MS/MS having a high storage capacity. The perspectives discussed in this section, highlights the high potential of this concept to further improve the features of sequence-defined poly(alkoxyamine phosphodiester)s as molecular data storage media.

General conclusion

In the frame of the PhD thesis, the new class of sequence-defined poly(alkoxyamine phosphodiester)s have been developed and extensively studied. The results presented in this thesis are placed in the sequence-defined polymers field, which is a relatively new area in polymer chemistry. The tools imported from other disciplines in the last years, enabled the synthesis of polymer having fully defined primary structures and transformed this subject into a broad interdisciplinary area. Digital polymers are sequence-defined polymers able to store information at the molecular level. These molecules are formed by a defined sequence of two different monomers intentionally defined as bit 0 and bit 1. Digital polymers must be monodisperse and perfectly defined to ensure a reliable molecular encryption. Moreover, their synthesis and sequencing should be fast and reliable. The synthesis of these molecules is usually achieved *via* a multi-step growth polymerization approach. In the frame of the PhD thesis, phosphoramidite coupling and nitroxide-radical coupling were implemented in the iterative protocol for the synthesis of sequence-defined poly(alkoxyamine phosphodiester)s. The monomer structures were carefully designed to allow the successive repetition of these reactions and to obtain a polymer backbone which can be easily modified by monomer design.

The studies presented in **Chapter II** show how the poly(alkoxyamine phosphodiester)s class was developed and following tested for different sequencing techniques. This class of digital polymers was synthesized with the aim to improve the readability of poly(phosphodiester)s by alkoxyamine bond insertion in the polymer backbone. The chemoselective protocol developed for poly(alkoxyamine phosphodiester)s allow to avoid deprotection steps and speeds the iterative protocol up. The synthesis of such sequence-defined polymers relies on phosphoramidite coupling and nitroxide-radical coupling. These reactions were intensively studied in the past years and were already employed for the synthesis of other digital polymers classes. In order to prove the reproducibility of the protocol, three different coding alphabets were synthesized. The monomer structures and the iterative synthesis conditions were carefully optimized to obtain uniform sequence-defined poly(alkoxyamine phosphodiester)s. Afterwards, a library of eight different uniform digital polymers was synthesized using three different alphabets and one hydroxy-functionalized nitroxide. All the digital polymers were recovered in good yields, highlighting the robustness of the iterative protocol as a new tool for the synthesis of digitally encoded sequence-defined polymers. In collaboration with the group of Prof. Laurence Charles the conventional MS/MS sequencing of poly(alkoyamine phosphodiester)s was successfully achieved. Indeed, it was found that the labile alkoxyamine bond inserted in the polymer backbone, is preferentially fragmented at low ionization energies originating an “easy to interpret” fragmentation pattern. This feature enables the fast and easy data extraction from this class of digital polymers. The sequencing by MS/MS was successfully achieved for all the digital polymers contained in the library, highlighting the efficiency and the reproducibility of MS/MS as a powerful tool for the digital polymers read-out.

The sequencing of digitally encoded poly(alkoxyamine phosphodiester)s was also tested using "*non*-conventional" techniques. Indeed, in a collaborative work with Prof. Hans Börner, this class of molecules was employed for fragmentation-free sequencing (F^2S). The concept relies on the generation of small percentages of specific truncated sequences and the full-length polymers during the iterative synthesis. In the frame of the project, a capping agent was designed and following employed during the phosphoramidite coupling to generate terminated-fragments. Moreover, the solid-state strategy was slightly modified. Indeed, a tetrapeptyde was added to the poly(alkoxyamine phosphodiester)s in order to ensure the analysis of the short fragments in a convenient m/z range. The resulting MS analysis allow to sequence the digital polymers avoiding MS/MS fragmentations however, some parameters have still to be optimized. The iterative protocol described in chapter II allows to obtain uniform sequence-defined poly(alkoxyamine phosphodiester)s in good yields however, these molecule show a low storage capacity compare to other digital polymers. For this reason the backbone of poly(alkoxyamine phosphodiester)s was modified to increase the data density.

The storage capacity is a crucial parameter in the frame of miniaturization of data storage devices. In this context, the studies presented **in Chapter III** describes how the modification of the polymer backbone, allow to obtain an enhanced storage capacity for poly(alkoxyamine phosphodiester)s maintaining the reading by MS/MS easy. To obtain an enhanced data density, four coding monomer (two phosphoramidites and two nitroxides) were employed in the iterative synthesis. The use of four coding units allows the information encryption in all the steps of the iterative cycle. Hence, the resulting digital polymers contain two bits of information for each repeating units. In the frame of the project, three different nitroxides-based alphabets were designed and tested. One of those was found to be suitable for the synthesis of poly(alkoxyamine phosphodiester)s encoding dyads. Indeed, the monomers pair forming this alphabet led to the synthesis of uniform digital polymers. Afterwards, a library of six digitally encoded poly(alkoxyamine phosphodiester)s encoding dyads was synthesized, using these new building blocks. The sequence-defined polymers were following sequenced via MS/MS. As for the first generation, the alkoxyamine bonds in the backbone are preferentially cleaved in MS/MS. This fragmentation generates encoding dyads that can be easily discriminated because of their different MS signature. Moreover, the data extraction from these digital polymers is even easier after MS charge state deconvolution. Indeed, two bits of information can be extracted from the single peak-to-peak distance in the ion series, making faster and easier the resulting sequencing. The sequencing of all the digital polymers encoding dyads was successfully achieved via MS/MS and MS charge state deconvolution. In addition, MS-DECODER software was implemented for the dyad extraction from poly(alkoxyamine phosphodiester)s. This software, using the MS/MS spectra as input enables the automatic digital polymers sequencing. MS-DECODER was tested for all the sequences present in the library and in all the cases allowed the data extraction form these molecules in a few milliseconds.

In **Chapter IV** the insertion of alkoxyamine bonds having different stability in the polymer backbone is described. The aim of this project is the synthesis of uniform poly(alkoxyamine phosphodiester)s which can be fragmented in specific parts of the backbone. In the frame of the doctoral studies, three hydroxy-functionalized nitroxides having different stability (T_3, SG1-OH and Pr_2) were synthesized. These building blocks were designed according to the studies performed in the past years on the nitroxide stability. Afterwards, different combinations of two nitroxides were employed for the synthesis of uniform sequence-defined polymers. The resulting molecules were following tested for the selective cleavage in MS/MS. It was found that the monomers pair SG1-OH/Pr_2 enables the selective alkoxyamine bond fragmentation. Indeed, the SG1-based alkoxyamine bonds present in the polymer backbone, are selectively fragmented at 10 eV. The selective rupture in MS/MS was obtained for a tetramer, an octamer and a dodecamer. The results presented in this chapter confirm the reproducibility of the method. However, the molecular architecture may be improved to increase the storage capacity or to perform a multi-step selective fragmentation.

The development of a new class of sequence-defined polymers for data storage applications was developed and presented in this thesis. The backbone modification of poly(alkoxyamine phosphodiester)s presented in **Chapter III** and **Chapter IV**, demonstrate that this new class of digital polymers is extremely versatile. Indeed, digital poly(alkoxyamine phosphodiester)s can be synthesized efficiently in a relatively short time moreover, the iterative protocol tolerates the insertion of different building blocks allowing the backbone modification by monomer design. The possibility to tune the backbone to obtain tailor-made features for the resulting digital polymers makes poly(alkoxyamine phosphodiester)s promising candidates for future molecular data storage applications.

Experimental part

1 Characterization techniques and instrumentation

1.1 NMR

^{1}H NMR (400 MHz), ^{13}C NMR (100.6 MHz), and ^{31}P NMR (161.92 MHz) spectra were recorded on a Bruker Avance 400 spectrometer equipped with Ultrashield magnet.

1.2 SEC

SEC analysis were performed in a mixture of 60% millipore water, 40% acetonitrile and 0.1 M $NaNO_3$(flow rate: 0.5 mL·min^{-1}) using apparatus DIONEX Ultimate 3000 series (degasser. pump, autosampler). The set-up was equipped with four columns Shodex OH-pak 30 cm (802.5HQ, 804HQ, 806HQ, 807HQ) and a pre-column with separation domain from 500 to 100000000 g·mol^{-1}. The polymers were detected with OPTILAB rEX (Wyatt Technology Corporation) refractometer and a multiangle light scattering detector DAWN HELEOS II (Wyatt Technology Corporation).

1.3 MS spectrometry

High resolution MS and MS/MS experiments were performed using a QqTOF mass spectrometer (QStar Elite, SCIEX, Concord, ON, Canada) with the ESI source operated in the negative mode (capillary voltage: -4200 V; cone voltage: -75 V). Ions were accurately mass measured in the orthogonal acceleration time-of-flight (oa-TOF) mass analyser, using the precursor ions as internal standards to calibrate MS/MS data. In this instrument, air was used as nebulizing gas (10 psi) while nitrogen was used as curtain gas (20 psi) and collision gas. Instrument control, data acquisition and data processing were achieved using Analyst software (QS 2.0) provided by Applied Biosystems. Charge state deconvolution of MS/MS data was performed with MagTran (Amgen Inc.), an implementation of the ZSCORE algorithm.[228] Polymer samples (a few mg) were first solubilized in methanol (SDS, Peypin, France) and then diluted (1/10 to 1/1000, v/v) in methanol supplemented with ammonium acetate (Sigma Aldrich, Saint Louis, MO) at a 3 mM concentration level, prior to injection in the ESI source at 10 µL/min using a syringe pump.

1.4 UV spectroscopy

The UV spectra were recorded on a Perkin Elmer lambda 25 UV-Vis spectrometer using Perkin Elmer UV WinLab.

1.5 Iterative synthesis

Oligomers syntheses were performed in an equipped argon/vacuum solid phase extraction glass tube (12 mL with frit 3 Å) and stirred with an IKA HS 260 Basic shaker.

2 Materials

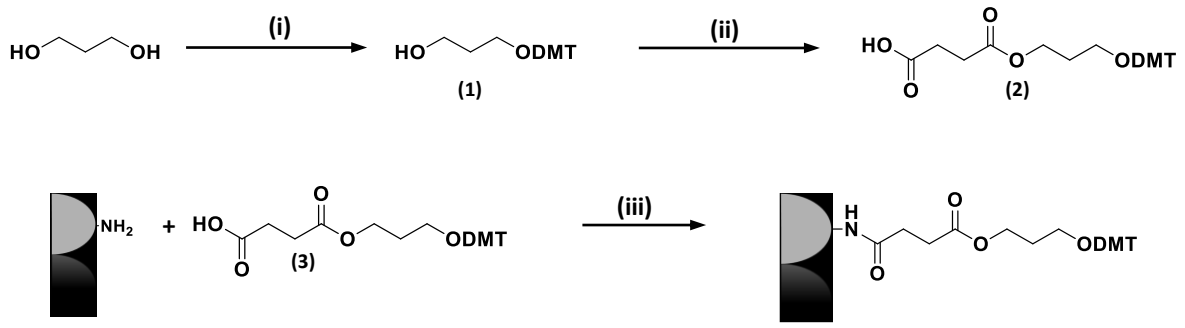
3-Amino-1-propanol (99%, Alfa Aesar), 3-amino-2,2-dimethyl-1-propanol (97%, TCI Europe), 4- amino-1-butanol (98%, TCI Europe), 5-amino-1-pentanol (>95%, TCI Europe), 4-amino-2-methyl-1-butanol (98%, TCI Europe), 4-amino-TEMPO (97%, TCI), Aminomethylated polystyrene HL (100-200 mesh, 1.14mmol/g, Merck/Novabiochem), 2-bromoisobutyryl bromide (97%, Alfa Aesar), 1,3-Propanediol, (99%, %, Alfa Aesar), Ɛ-caprolactone (97%, Sigma-Aldrich), succinic anhydride (>95%, TCI Europe), 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (97%, ABCR), 4,4'-dimethoxytriphenylmethyl chloride (DMT-Cl,≥97.0%, Sigma- Aldrich), copper(I) bromide (98%, Alfa Aesar), N,N-diisopropylethylamine (DIPEA, 99%, TCI), 4-Hydroxy-TEMPO (98%, TCI), 2,6-lutidine (98%, Alfa Aesar), 2-Propanol (99.8%, Sigma-Aldrich/Millipore), tris(2-dimethylaminoethyl)amine (99%, Alfa Aesar), 1-methylimidazole (NMI, 99%, Alfa Aesar), p-Toluenesulfonic acid monohydrate, (PTSA, 98.5%, Sigma-Aldrich), propan-2-ol (≥99.8%, Sigma-Aldrich), α-methyl-γ-butyrolactone, (>98 %, TCI), γ-hexalactone (98%, Sigma-Aldrich), diisopropylamine (purified by redistillation, 99.9%, Sigma-Aldrich), methyl iodide (99%, Alfa Aesar), γ-valerolactone (98%, Sigma-Aldrich), hexamethylphosphoramide (HMPA, 99% Sigma-Aldrich), triisopropylsilyl chloride (TIPSCl, 97%, Sigma-Aldrich), 1,5-pentanediol (97%, Sigma-Aldrich), 3-methyl-1,5-pentanediol (≥98.0%, Sigma-Aldrich), 1H-Benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate (PyBOP, 98%, Novabiochem/Merck), 2-bromoethanol (95%, Sigma-Aldrich), imidazole (99%, Alfa Aesar), iodobenzene diacetate (BAIB, >97%, TCI), TEMPO (>98%, Alfa Aesar), Sodium hydride (60% in mineral oil, Sigma-Aldrich), *n*-buthyllithium solution (2.5M in hexanes, Sigma-Aldrich), tetrabutylammonium fluoride solution (1.0 M in THF, Sigma-Aldich), Pyridinium dichromate, (PDC, 98%, Alfa Aesar), 3-Amino-PROXYL (99%, Acros Organics), Ethanol absolute (99.9%, Fisher-Chemical), acetic anhydride (99%, Alfa Aesar), (piperidine (99%, Alfa Aesar), methylamine (aqueous solution, 40%, Fluka), ammonium hydroxide solution (28-30% NH3, VWR), tetrazole (acetonitrile solution, 0.45M, Sigma-Aldrich), triethylamine (97%, Merck), trichloroacetic acid (>99%, Sigma- Aldrich), 1,4 dioxane (99.5%, Sigma-Aldrich), N,N-Dimethylformamide anhydrous, (DMF, 99.8%, Sigma-Aldrich) tetrahydrofuran (THF, > 99%, 250 ppm BHT, Sigma-Aldrich), anhydrous dichloromethane (Sigma-Aldrich), anhydrous acetonitrile (Sigma-Aldrich) and anhydrous dimethyl sulfoxide (DMSO) were used as received. Anhydrous tetrahydrofuran (THF) was obtained on a dry solvent station GTS100. The phosphoramidite compounds were kept in the freezer at -18°C.

3 Experimental section

3.1 Syntheses Chapter 2

3.1.1 Solid support modification

The solid support modification was performed following the procedure depicted in Scheme E.1.



Scheme E.1: General strategy employed for the solid-support modification.

3.1.1.1 Synthesis of intermediate 1

Propanediol (2 g, 26.2 mmol, 1 eq.) was dissolved in a mixture of 24 mL of anhydrous pyridine and 36 mL of anhydrous THF under argon atmosphere. DMT-Cl (809 g, 26.2 mmol, 1eq.) was added to the mixture in four parts every 45 minutes. 1 hour after the last addition, the reaction was quenched with MeOH. The solvent was dried, the residue taken in EtOAc (30mL) and washed with cold water (3x20mL). The organic layers were recovered, dried over $MgSO_4$ and the solvent was evaporated. Chromatographic column on silica gel (Cyclohexane/EtOAc/DIPEA) (80/20/2) gives the product in 65% yield. ^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 1.82-1.87 (m, 2H, HO- CH_2-C$\underline{H}_2$), 3.26-3.29 (t, 2H, DMT-O-C$\underline{H}_2$), 3.74-3.77 (m, 2H, HO- C$\underline{H}_2$-CH_2), 3.79 (s, 6H 2xC$\underline{H}_3$), 6.81-6.85 (m, 4H, Ar-$\underline{H}$), 7.19-7.23(m, 1H, Ar-$\underline{H}$), 7.26-7.34 (m, 4H AR-$\underline{H}$), 7.40-7.43 (m, 4H, Ar-$\underline{H}$). ^{13}C NMR (100 MHz, $CDCl_3$, δ, ppm): 32.59, 55.36, 62.12, 62.57, 86.63, 113.29, 126.92, 128.00, 128.24, 130.13, 136.32, 145.06, 158.62.

3.1.1.2 Synthesis of intermediate 2

In a round bottom flask intermediate 1 (1 g, 4.23 mmol, 1 eq.), DMAP (1.29 g, 10.6 mmol, 2.5 eq.) and succinic anhydride (1.27 g, 12.7 mmol, 3 eq.) were added under argon atmosphere. The mixture was dissolved in anhydrous pyridine (15mL) and stirred at 37°C. After 2.5 hours the reaction is quenched with MeOH, and the solvent was evaporated. The residue was taken in EtOAc and extracted with a solution 0.6M of KH_2PO_4 in water (20mL). The organic layer was dried over $MgSO_4$ and the solvent evaporated. The product was recovered as a brown-yellow oil in 97% yield. ^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 1.84-1.96 (m, 2H, DMT-O – CH_2-C$\underline{H}_2$), 2.54-2.65 (m, 4H, O=C-C$\underline{H}_2$-C$\underline{H}_2$-C=O), 3.14-3.16 (t, 2H, DMT-O-C$\underline{H}_2$), 3.79 (s. 6H 2x Ar-O-C$\underline{H}_3$),4.24-4.26 (t, 2H, O=C-O-C$\underline{H}_2$), 6.79-6.84 (m, 4H, Ar-$\underline{H}$), 7.25-7.34 (m, 8H Ar-$\underline{H}$), 7.40-7.43 (m, 1H, Ar-$\underline{H}$). ^{13}C NMR (100 MHz, $CDCl_3$, δ, ppm): 29.14, 31.08, 55.40, 59.24, 61.93, 81.60, 113.31, 127.22, 127.91, 129.27, 130.14, 139.59, 147.45, 157.72.

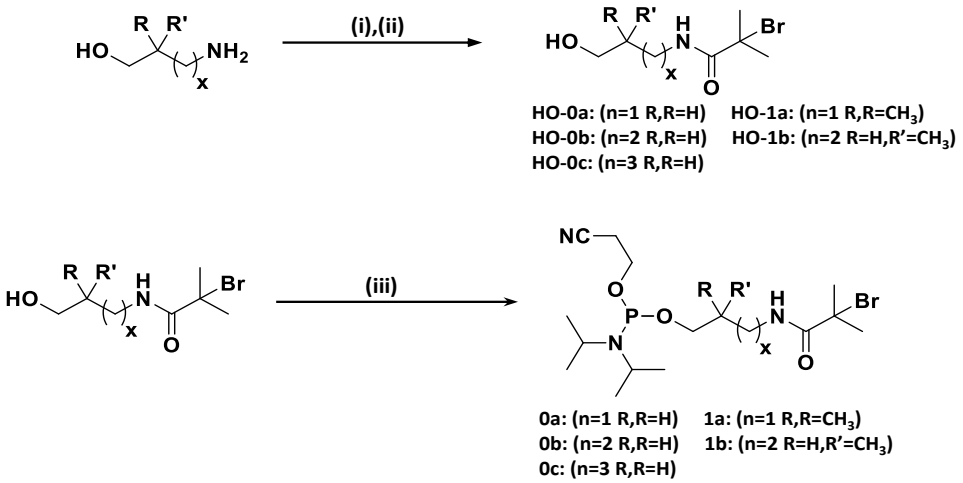
3.1.1.3 Synthesis of intermediate 3

The aminopolystirene resin (1.4 mmol/g) was placed into an argon/vacuum solid phase extraction glass tube and swollen in anhydrous dichloromethane (6mL) for 10 minutes. The solvent was removed. DMAP (85 mg, 0.7 mmol, 0.5 eq.), DCC (206.3 mg, 5.92 mmol, 4.2 eq.) and a solution of intermediate 2 (1g, 2.09mmol, 1.5 eq.) in 10 mL of anhydrous dichloromethane were added to the resin and the solution was shaken overnight. The day after the solvent was removed and the resin washed with MeOH (5X 10mL), acetonitrile (3x10mL), dichloromethane (3x10mL), and diethyl ether (3x10mL). The loading is measured on an aliquot of the modified resin. After DMT deprotection, the absorbance calculated at 504 nm allows to calculate the loading that in all the cases was between 0.69 and 0.92 mmol/g.

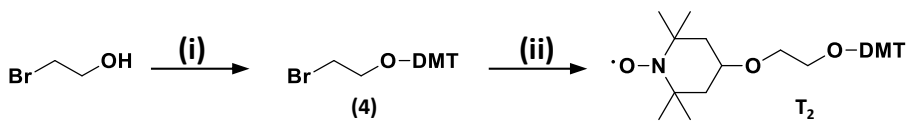
3.1.2 Monomers synthesis

The synthetic strategies employed for the syntheses of the coding monomers and the linkers are depicted in Scheme E.2.

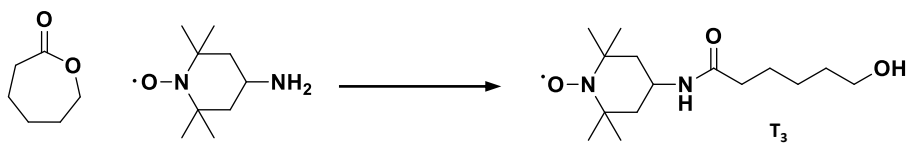
a) **Coding monomer synthesis**



b) **Linker T_2 synthesis**



c) **Linker T_3 synthesis**



Scheme E.2: General strategies for the syntheses of the monomers employed in this work. **a)** Coding monomers synthesis. **b)** Synthesis of linker T_2. **c)** Synthesis of linker T_3.

3.1.2.1 Synthesis of intermediate HO-0_a

The synthesis of intermediate HO-0_a was adapted according to a procedure from literature. [193] 3-Amino-1-propanol (5.0 g, 66.6 mmol, 1 eq.) was dissolved in 50 mL of anhydrous dichloromethane. Triethylamine (10.21 mL, 73.2 mmol, 1.1 eq.) was added and the solution was stirred at 0 °C for 15 min under argon atmosphere. 2-bromoisobutyryl bromide (8.23 mL, 66.6 mmol, 1 eq.) was added dropwise to the mixture. The reaction mixture was let to reach room temperature and stirred for 90 minutes. Afterward, the formed precipitate was removed and the filtrate was stirred with 20 mL of 5% KOH aqueous solution for 1h. The aqueous phase was separated and the organic layer was washed with NaOH 1 M (2 x 35 mL), HCl 1 M (2 x 35 mL), brine (1 x 40 mL), dried over Na_2SO_4 and evaporated. The intermediate HO-0_a was recovered in 80% of yield as pale yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 1.75 (m, 2H, -NH-CH_2-C$\underline{H}_2$-), 1.96 (s, 6H, 2x-C$\underline{H}_3$), 2.67 (s, 1H, -OH), 3.45 (m, 2H, -NH-C$\underline{H}_2$-CH_2-

), 3.67 (t, 2H, HO-C$\underline{H}_2$-), 7.08 (s, 1H, -NH). ^{13}C NMR (100 MHz, $CDCl_3$, δ, ppm): 31.65, 32.35, 37.49, 59.65, 62.18, 172.95.

3.1.2.2 Synthesis of intermediate HO-0b

The product was synthesized following the procedure described in section 3.1.2.1 using 4-amino 1-butanol as starting material. The product was recovered in 79% yield. ^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 1.77-1.75(m, 4H, -NH-CH_2-C$\underline{H}_2$-C$\underline{H}_2$-CH_2), 1.96 (s, 6H, 2x-C$\underline{H}_3$), 2.67 (s, 1H, -O$\underline{H}$), 3.45 (m, 2H, -NH-C$\underline{H}_2$-CH_2-), 3.67 (t, 2H, HO-C$\underline{H}_2$-), 7.08 (s, 1H, -NH). ^{13}C NMR (100 MHz, $CDCl_3$, δ, ppm): 25.95, 29.72, 30.79, 32.53, 40.22, 62.11, 62.86, 172.21.

3.1.2.3 Synthesis of intermediate HO-0_c

The product was synthesized following the procedure described in section 3.1.2.1 using 5-amino 1-pentanol as starting material. The product was recovered in 84% yield. ^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 1.42 (m, 2H, -NH-CH_2-CH_2-C$\underline{H}_2$-), 1.59 (m, 5H, -NH-CH_2-C$\underline{H}_2$-CH_2-CH_2-, -OH), 1.95 (s, 6H, 2x-C$\underline{H}_3$), 3.28 (m, 2H, -NH-C$\underline{H}_2$-CH_2-), 3.66 (t, 2H, HO-C$\underline{H}_2$-), 6.75 (s, 1H, -NH). ^{13}C NMR (100 MHz, $CDCl_3$, δ, ppm): 22.93, 29.01, 32.09, 32.53, 40.20, 62.44, 63.23, 171.99.

3.1.2.4 Synthesis of intermediate HO-1_a

The product was synthesized following the procedure described in section 3.1.2.1 and using 3-amino-2,2-dimethyl-1-propanol as starting material. The product was recovered in 74% yield after recrystallization in ethyl acetate/pentane. ^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 0.90 (s, 6H, -CH_2-C-(C$\underline{H}_3$)$_2$), 1.96 (s, 6H, 2x-C$\underline{H}_3$), 3.14 (d, 2H, -NH-C$\underline{H}_2$-), 3.19 (d, 2H, HO-C$\underline{H}_2$-), 3.43 (t, 1H, -O$\underline{H}$), 7.13 (s, 1H, -NH). ^{13}C NMR (100 MHz, $CDCl_3$, δ, ppm): 22.53, 32.52, 36.80, 47.30, 62.50, 68.62, 173.54.

3.1.2.5 Synthesis of intermediate HO-1_b

The product was synthesized following the procedure described in section 3.1.2.1using 4-amino-2-methyl-1-butanol as staring material. The product was recovered in 78% yield ^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 0.95 (d, 3H, -CH-C$\underline{H}_3$), 1.46 (m, 1H, -C$\underline{H}$-CH_3), 1.71 (m, 2H, -NH-CH_2-C$\underline{H}_2$-), 1.83 (s, 1H, -O$\underline{H}$), 1.94 (s, 6H, 2x-C$\underline{H}_3$), 3.24-3.43 (m, 2H, -NH-C$\underline{H}_2$-CH_2-), 3.44-3.59 (m, 2H, HO-C$\underline{H}_2$-), 6.92 (s, 1H, -NH). ^{13}C NMR (100 MHz, $CDCl_3$, δ, ppm): 16.75, 32.33, 32.95, 33.47, 38.40, 62.57, 67.42, 172.01.

3.1.2.6 Synthesis of coding monomer 0_a

Intermediate HO-0_a (1.0 g, 4.46 mmol, 1 eq.) was dissolved in anhydrous dichloromethane (8 mL) and DIPEA (3.88 mL, 22.27 mmol, 5 eq.). The solution was stirred at 0°C for 5 min, 1-methylimidazole (0.18 mL, 2.23 mmol, 0.5 eq.) under argon atmosphere and a solution of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (1.16 g, 4.91 mmol, 1.1 eq.) in anhydrous dichloromethane (2 mL) were added. The reaction mixture was stirred 30 min at 0 °C, then was let to reach room temperature and stirred for 90 minutes. The crude was extracted with a solution of saturated $NaHCO_3$ in water (12 mL), the organic phase was dried over Na_2SO_4, filtrated and the solvent evaporated. Chromatographic column on silica gel (n-pentane/ethyl acetate/triethylamine) (70/29/1) gives a colourless oil (75%). ^{1}H NMR (400 MHz, $CDCl_3$, δ,

ppm): 1.17 (d, 6H, -N-CH-(C$\underline{H}_3$)$_2$), 1.85 (m, 2H, -NH-CH$_2$-C$\underline{H}_2$-), 1.94 (s, 6H, 2x-C$\underline{H}_3$), 2.64 (t, 2H, NC-C$\underline{H}_2$-CH$_2$-), 3.33-3.45 (m, 2H, -NH-C$\underline{H}_2$-CH$_2$-), 3.55-3.92 (m, 6H, -N-C$\underline{H}$-(CH$_3$)$_2$, -P-O-CH$_2$-, NC-CH$_2$-CH$_2$-), 6.98 (s, 1H, -NH). ^{13}C NMR (100 MHz, CDCl$_3$, δ, ppm): 20.36, 20.44 , 24.59, 24.65, 30.27, 30.34, 32.49, 37.96, 43.01, 43.14, 58.30, 58.50, 61.39, 61.57, 62.80, 117.56, 171.92. ^{31}P NMR (161 MHz, CDCl$_3$, δ, ppm): 147.85.

3.1.2.7 Synthesis of coding monomer 0_b

The product was recovered as 80% yield. ^{1}H NMR (400 MHz, CDCl$_3$, δ, ppm): 1.15 (d, 6H, N-CH-(C$\underline{H}_3$)$_2$)), 1.17 (d, 6H, , -N-CH-(C$\underline{H}_3$)$_2$), 1.60-1.69 (m, 4H, C$\underline{H}_2$-C$\underline{H}_2$), 1.93 (s, 6H, 2x-C$\underline{H}_3$), 2.63 t, 2H, NC-C$\underline{H}_2$-CH$_2$-), 3.26-3.32 (m, 2H, -NH-C$\underline{H}_2$-CH$_2$-), 3.52-3.91 (m, 6H, -N-C$\underline{H}$-(CH$_3$)$_2$, -P-O-CH$_2$-, NC-CH$_2$-CH$_2$-), 6.77 (s, 1H, -NH). ^{13}C NMR (100 MHz, CDCl$_3$): 20.22, 20.34, 24.40, 24.53, 24.50, 24.67, 25.93, 28.32, 28.43, 32.42, 39.93, 42.8, 43.03, 58.04, 58.37, 62.91, 63.04, 63.12, 117.52, 171.82; ^{31}P NMR (127 MHz, CDCl$_3$):147.65.

3.1.2.8 Synthesis of coding monomer 0_c

The product was recovered as collarless oil in 83% yield. ^{1}H NMR (400 MHz, CDCl$_3$, δ, ppm): 1.17 (d, 6H, -N-CH-(C$\underline{H}_3$)$_2$), 1.19 (d, 6H, -N-CH-(C$\underline{H}_3$)$_2$), 1.42 (m, 2H, -NH-CH$_2$-CH$_2$-C$\underline{H}_2$-), 1.59 (m, 4H, -NH-CH$_2$-C$\underline{H}_2$-CH$_2$-C$\underline{H}_2$-), 1.95 (s, 6H, 2x-C$\underline{H}_3$), 2.64 (t, 2H, NC-C$\underline{H}_2$-CH$_2$-), 3.22-3.31 (m, 2H, -NH-C$\underline{H}_2$-CH$_2$-), 3.53-3.90 (m, 6H, -N-C$\underline{H}$-(CH$_3$)$_2$, -P-O-CH$_2$-, NC-CH$_2$-CH$_2$-), 6.73 (s,1H, -NH). ^{13}C NMR (100 MHz, CDCl$_3$, δ, ppm): 20.33, 20.39, 23.30 24.52, 24.58, 24.64, 28.97, 30.72, 30.79, 32.58, 40.29, 42.92, 43.05, 58.11, 58.32., 63.30, 63.46, 63.37, 117.56, 171.60. ^{31}P NMR (161 MHz, CDCl$_3$, δ, ppm): 147.37.

3.1.2.9 Synthesis of coding monomer 1_a

The product was recovered as colourless oil in 90% yield. ^{1}H NMR (400 MHz, CDCl$_3$, δ, ppm): 0.93 (s, 6H, -CH2-C-(C$\underline{H}_3$)$_2$), 1.19 (d, 12H, 2x-N-CH-(C$\underline{H}_3$)$_2$), 1.96 (s, 6H, 2x-C$\underline{H}_3$), 2.65 (t, 2H, NC-C$\underline{H}_2$-CH$_2$-), 3.16-3.26 (m, 2H, -NH-C$\underline{H}_2$-C-), 3.31-3.46 (m, 2H, -O-C$\underline{H}_2$-C-), 3.56-3.68 (m, 2H, -N-C$\underline{H}$-(CH$_3$)$_2$), 3.78-3.93 (m, 2H, NC-CH$_2$-C$\underline{H}_2$-), 7.10 (s, 1H, -NH). ^{13}C NMR (100 MHz, CDCl$_3$, δ, ppm): 20.31, 20.38, 22.33, 22.13, 24.49, 24.57, 24.13, 32.45, 32.50, 36.14, 36.2), 42.97, 43.09, 47.70, 58.20, 58.40, 63.01, 70.91, 71.06, 117.47, 171.86. ^{31}P NMR (161 MHz, CDCl$_3$, δ, ppm): 148.05.

3.1.2.10 Synthesis of coding monomer 1_b

The product was recovered as colourless oil in 92% yield. ^{1}H NMR (400 MHz, CDCl$_3$, δ, ppm): 0.96 (d, 3H, -CH-C$\underline{H}_3$), 1.17 (d, 6H, -N-CH-(C$\underline{H}_3$)$_2$), 1.19 (d, 6H, -N-CH-(C$\underline{H}_3$)$_2$), 1.42 (m, 1H, -C$\underline{H}$-CH$_3$), 1.74 (m, 2H, -NH-CH$_2$-C$\underline{H}_2$-), 1.94 (s, 6H, 2x-C$\underline{H}_3$), 2.64 (t, 2H, NC-C$\underline{H}_2$-CH$_2$-), 3.32 (m, 2H, -NH-C$\underline{H}_2$-CH$_2$-), 3.40-3.66 (m, 4H, -N-C$\underline{H}_2$-(CH3)2, -P-O-C$\underline{H}_2$-), 3.75-3.90 (m, 2H, NC-CH$_2$-C$\underline{H}_2$-), 6.75 (s, 1H, -NH). ^{13}C NMR (100 MHz, CDCl$_3$, δ, ppm): 16.76, 20.32, 20.39, 24.49, 24.51, 24.57, 24.58, 32.36, 32.38, 32.44, 32.95, 32.97, 38.26, 42.89, 43.01, 58.05, 58.24, 63.00, 68.32, 68.48, 117.59, 171.74. ^{31}P NMR (161 MHz, CDCl$_3$, δ, ppm): 147.73.

3.1.2.11 Synthesis of intermediate 4

2 bromoethanol (500 mg, 4 mmol, 1 eq.) was dissolved in a mixture of 6 mL of anhydrous pyridine and 10 mL of anhydrous THF under argon atmosphere. DMT-chloride (155 mg, 4 mmol, 1 eq.) was added to the solution and the mixture was stirred overnight. The day after, the reaction was quenched with MeOH and the solvent was evaporated. The crude was taken in EtOAc (20 mL), washed with water (2x15 mL) and dried over $MgSO_4$ anhydrous. Chromatographic column on silica gel EtOAc/Cyclohexane/DIPEA (20/80/2) gives the product in 82% yield. ^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 3.37-3.47 (m, 4H Br-C$\underline{H}_2$-C$\underline{H}_2$-O-), 3.79 (s. 6H, O-C$\underline{H}_3$), 6.83-6.85 (m, 4H Ar$\underline{H}$) 7.22-7.26 (m, 4H, Ar$\underline{H}$), 7.28-7.32 (m, 4H Ar$\underline{H}$), 7.33-7.37 (m, 1H, Ar$\underline{H}$). ^{13}C NMR (100 MHz, $CDCl_3$, δ, ppm): 31.41, 43.60, 55.37, 63.94, 846.45, 113.29, 126.97, 127.99, 128.25, 130.16, 136.10, 144.89.

3.1.2.12 Synthesis of linker T_2

Sodium hydride (94 mg, 60%wt in mineral oil, 2.35 mmol, 5 eq.) was dissolved in 5 mL of anhydrous THF and stirred for 10 minutes under argon atmosphere. To the solution 4 hydroxy TEMPO (96.7 mg, 0.56 mmol, 1.2 eq.) was added at 0°C, the solution was let to reach room temperature and stirred for 80 minutes. To the mixture, a solution of intermediate 4 (200 mg, 0.47mmol, 1 eq.) in THF (4mL) was dropped at 0°C and stirred overnight. The day after the reaction was quenched by carefully addition at 0°C of brine (60 mL). The THF was evaporated, the solution extracted with dichloromethane (3x50mL), dried over $MgSO_4$ anhydrous, filtrated and the solvent evaporated. Chromatography on silica gel EtOAc/Cyclohexane/DIPEA (20/80/2) gives the product in 37% yield. ESI-MS $[M+H]^+$, m/z expected for: $C_{32}H_{40}NO_5$ 519.2, found 520.8.

3.1.2.13 Synthesis of linker T_3

The synthesis of this molecule was adapted from the literature.[229] 4-Amino-TEMPO (800 mg, 4.6 mmol, 1eq.) was dissolved in Ɛ-caprolactone (1.57 g, 13.8 mmol, 3 eq.) and refluxed at 130°C. The solution was monitored by TLC and stopped after 120 minutes. The mixture was purified by chromatography on silica gel (dichloromethane/methanol) (95/5) yielding 900 mg of monomer as red oil (73%). ESI-MS $[M+H]^+$, m/z expected for $C_{15}H_{30}N_2O_3$ 286.2, found 287.2.

3.1.3 Solid-state iterative protocol

All the coupling reactions were performed under argon atmosphere using a peptide synthesis reactor equipped with a fritted glass plate and a stopcock.

3.1.3.1 Phosphoramidite coupling (step i_1 and i_2)

1 equivalent of modified polystyrene resin (0.7 mmol/g) was swollen for 10 minutes in dichloromethane. DMT deprotection was achieved treating the resin with a 3% trichloroacetic acid solution in dichloromethane and the solid support was washed with anhydrous acetonitrile. A solution of phosphoramidite monomer (0a, 1a 0b, 1b or 0c, 3.0 eq.) dissolved in anhydrous acetonitrile (2 mL) and tetrazole (4 eq., 0.45 M solution in acetonitrile) were added under argon atmosphere. The reactor was shaken at room

temperature and after 30 minutes, the solution was removed and the resin was washed with anhydrous acetonitrile. Step (i_2). The phosphite produced in step (i_1) was oxidized into a more stable phosphate treating the resin with 0.1 M iodine solution (I_2 in water/2,6-lutidine/THF, 2/20/80) for 15 min. The excess of iodine was then removed washing the solid-support with anhydrous dichloromethane.

3.1.3.2 Nitroxide-radical coupling (step ii)

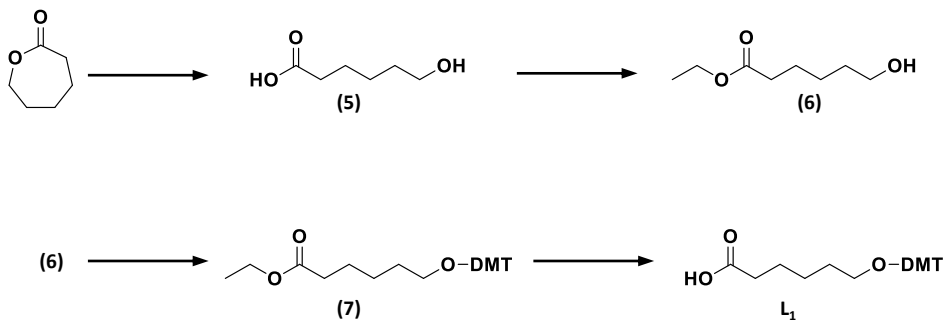
The resin was swollen in anhydrous dichloromethane and washed 2 times with anhydrous DMSO. A solution of hydroxy-functionalized nitroxides (T_1, T_2 or T_3, 5 eq.) and tris(2-dimethylaminoethyl)amine (3.3 eq.) in 2 mL of anhydrous DMSO was added under argon atmosphere. To the mixture, a solution of CuBr (3 eq.) in 2 mL of anhydrous DMSO was added; and the resin was shaken for 40 minutes. Afterwards, the solution was removed and the resin was washed with THF and shaken 2 time with a 2% solution of 4,4'-di-n-nonyl-2,2'-dipyridyl in dichloromethane for 5 minutes.

3.1.3.3 Cleavage and deprotection (step iii)

Traces of copper were removed washing several times the resin with a solution 2% sodium-EDTA in water/THF (2:1 v/v) for 5 minutes. Deprotection of the cyanoethyl group was achieved shaking the solid-support with a 10% piperidine solution in acetonitrile 2 times for 15 minutes. The polymer was cleaved from the support using a mixture of ammonia aqueous solution, methylamine aqueous solution and 1,4 dioxane (1:1:0.5 v/v/v) for 1h. The reaction mixture was filtered and degased for 30 min. Afterward, the solvent was removed, giving the resulting powders in overall yields between 40-60%. The products were analyzed by ESI-HRMS, SEC and NMR.

3.1.4 Monomer synthesis for F^2S sequencing

The synthesis of linker L_1 was done following the synthetic strategy depicted in Scheme E.3.



Scheme E.3: Synthetic strategy employed for the synthesis of linker L_1.

3.1.4.1 Synthesis of intermediate 5

NaOH (350 mg, 8.6 mmol, 2 eq.) was added to a solution of Ɛ Caprolactone (500 mg, 4.38 mmol 1 eq.) in 15 mL of water. The reaction mixture was stirred for 3 hours at room temperature. The mixture was then acidified until PH 3-4 *via* a carefully HCl addition. The mixture was extracted with EtOAc (3x 10mL) and the combined organic layers were dried over $MgSO_4$. The mixture was filtered and the solvent removed. The product was recovered in 82% yield as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$): 1.37-1.44 (m, 2H, HO-CH_2-CH_2-C$\underline{H}_2$); 1.54-1.69 (m, 4H, HO-CH_2-C$\underline{H}_2$-CH_2-C$\underline{H}_2$-CH_2-), 2.32-2.36 (m, 2H, O=C-C$\underline{H}_2$), 3.61-3.65 (t, 2H, HO-C$\underline{H}_2$). ^{13}C NMR (100 MHz, $CDCl_3$): 24.52, 25.34, 32.43, 34.04, 62.81, 179.18.

3.1.4.2 Synthesis of intermediate 6

PTSA (9.5 mg, 0.055mmol, 0.013eq.) was added to a solution of intermediate 5 (564 mg, 4.25 mmol, 1 eq.) in absolute ethanol (6 mL) and the mixture was stirred at 65°C for 4 hours. The solvent was evaporated, the residue was dissolved in EtOAc (20 mL) and washed with saturated $NaHCO_3$ (2x2 0mL) and brine (20 mL). The combined organic layers were dried over $MgSO_4$, filtered and the solvent was evaporated. The product was recovered as a yellow oil in 87% yield. ^{1}H NMR (400 MHz, $CDCl_3$): 1.22-1.26 (t, 3H, O-CH_2-C$\underline{H}_3$), 1.40-1.44 (m, 2 H, 0=C-CH_2-CH_2-C$\underline{H}_2$), 1.55–1.69 (m, 4H, 0=C-CH_2-C$\underline{H}_2$-CH_2-C$\underline{H}_2$), 2.30-2.32 (t, 2H, 0=C-C$\underline{H}_2$-CH_2-), 3.65 (t, 2H, HO-C$\underline{H}_2$), 4.10-4.13 (q, 2H, O=C-O-C$\underline{H}_2$). ^{13}C NMR (100 MHz, $CDCl_3$): 14.44, 24.86, 25.45, 32.41, 34.48, 60.47, 62.85, 173.99.

3.1.4.3 Synthesis of intermediate 7

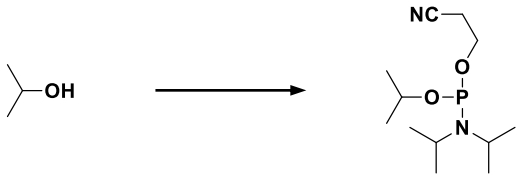
In a round bottom flask, DMAP (0.4 g, 3.27mmol, 0.5eq.), DMT-Cl (2.5 g, 7 mmol, 1.1 eq.) and intermediate 6 (1.1 g, 6.85 mmol, 1 eq.) were added under argon atmosphere. The mixture was dissolved in 13 mL of anhydrous pyridine and the mixture was stirred for 3 hours. The solvent was evaporated and the crude was taken in 50 mL of EtOAc and washed with water (2x50 mL). The combined aqueous layers were extracted with EtOAc (80 mL). The combined organic layers were dried over $MgSO_4$, filtered and the solvent was evaporated. Chromatographic column on silica gel (EtOAc/cyclohexane/TEA) (7/91/2) gives the product as a white oil in 95% yield. ^{1}H NMR (400 MHz, $CDCl_3$): 1.25 (t, 3H, O=C-CH_2C$\underline{H}_3$), 1.40-1.44 (m, 2 H, 0=C-CH_2-CH_2-C$\underline{H}_2$), 1.55–1.69 (m, 4H, 0=C-CH_2-C$\underline{H}_2$-CH_2-C$\underline{H}_2$), 2.30-2.32 (t, 2H, 0=C-C$\underline{H}_2$-CH_2-), 3.65 (t, 2H, DMT-O-C$\underline{H}_2$), 3.79 (s. 6H 2x Ar-O-C$\underline{H}_3$), 4.10-4.13 (q, 2H, O=C-O-C$\underline{H}_2$), 6.79-6.84 (m, 4H, Ar-$\underline{H}$), 7.25-7.34 (m, 8H Ar-$\underline{H}$), 7.40-7.43 (m, 1H, Ar-$\underline{H}$). ^{13}C NMR (100 MHz, $CDCl_3$): 14.42, 25.13, 26.15, 29.98, 34.54, 55.37, 60.35, 63.38, 85.88, 113.12, 126.74, 127.83, 128.38, 130.27, 136.81, 145.50, 158.35, 173.94.

3.1.4.4 Synthesis of linker L_1

An aqueous solution of NaOH (6M 10 mL) was added to a solution of intermediate 8 (2.36 g, 5.1 mmol, 1 eq.) in absolute ethanol (5 mL). The reaction was stirred for 6 hours the saturated NH_4Cl was added to the mixture until a white precipitate was formed. The mixture was filtered and the residue was acidified with a solution of citric acid in water (1%m/m, 100mL) and extracted with dichloromethane (3x60 mL). The organic layers were washed

with water (120 mL), dried over $MgSO_4$, filtered and the solvent was evaporated. The product was recovered as a foam in 85%yield. ^{1}H NMR (400 MHz, $CDCl_3$): 1.41-1.66 (m, 2H, DMT-O-CH_2-CH_2-C$\underline{H}_2$), 1.58-1.66 (m, 4H, DMT-O-CH_2-C$\underline{H}_2$-CH_2-C$\underline{H}_2$-CH_2-), 2.18 (s, 6H, Ar-O-C$\underline{H}_3$) 2.33-2.36 (t, 2H, DMT-O-C$\underline{H}_2$), 3.06-3.08 (t, 2H, O=C-CH_2), 6.81-6.85 (m, 4H, Ar-$\underline{H}$), 7.19-7.23(m, 1H, Ar-$\underline{H}$), 7.26-7.34 (m, 4H AR-$\underline{H}$), 7.40-7.43 (m, 4H, Ar-$\underline{H}$). ^{13}C NMR (100 MHz, acetone-d6): 25.64; 26.73, 30.58, 34.20, 55.51, 63.87, 86.58, 113.80, 127.49, 128.52, 129.05, 130.97, 137.59, 146.71, 159.54, 174.66.

3.1.4.5 Capping agent synthesis.



Scheme E.4: Synthetic strategy employed for the synthesis of the capping agent.

A solution of HPLC-grade propan-2-ol (246.8 mg, 4.1 mmol, 1.0 eq.) and DIPEA (3.12 g, 24.1 mmol, 6 eq.) was dissolved in 10 mL anhydrous dichloromethane. The mixture was stirred for 1 h over activated molecular sieves (3 Å) then 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite was added dropwise at 0 °C and 10 min later the reaction mixture was allowed to reach room temperature. After an additional 40 min the reaction mixture was filtered and washed with EtOAc (50 mL). The solvent was removed and the crude was purified by chromatographic column on silica gel (ethyl acetate/cyclohexane/ DIPEA) (40/60/2) to obtain the product as a colourless oil in 93 % yield. ^{1}H NMR (400 MHz, $CDCl_3$, δ): 1.18 (m, 12H, N(CH(C$\underline{H}_3$)$_2$)$_2$), 1.23-1.26 (m, 6H, OCH(C$\underline{H}_3$)2), 2.62-2.64 (m, 2H, $OCH_2C\underline{H}_2CN$), 3.58-3.61 (m, 2H, N(C$\underline{H}$(CH_3)2)2), 3.79-3.82 (m, 2H, $OC\underline{H}_2CH_2CN$), 4.07-4.09 (m, 1H, OC$\underline{H}$(CH_3)2). ^{13}C NMR (100.6 MHz, $CDCl_3$, δ, ppm): 20.4, 24.5, 24.8, 43.1, 58.3, 67.1, 117.8. ^{31}P NMR (161.9 MHz, $CDCl_3$, δ): 145.7.

3.1.5 Solid-state protocol for F²S

The general iterative protocol strategy assessed for F²S is depicted in Figure E.1.

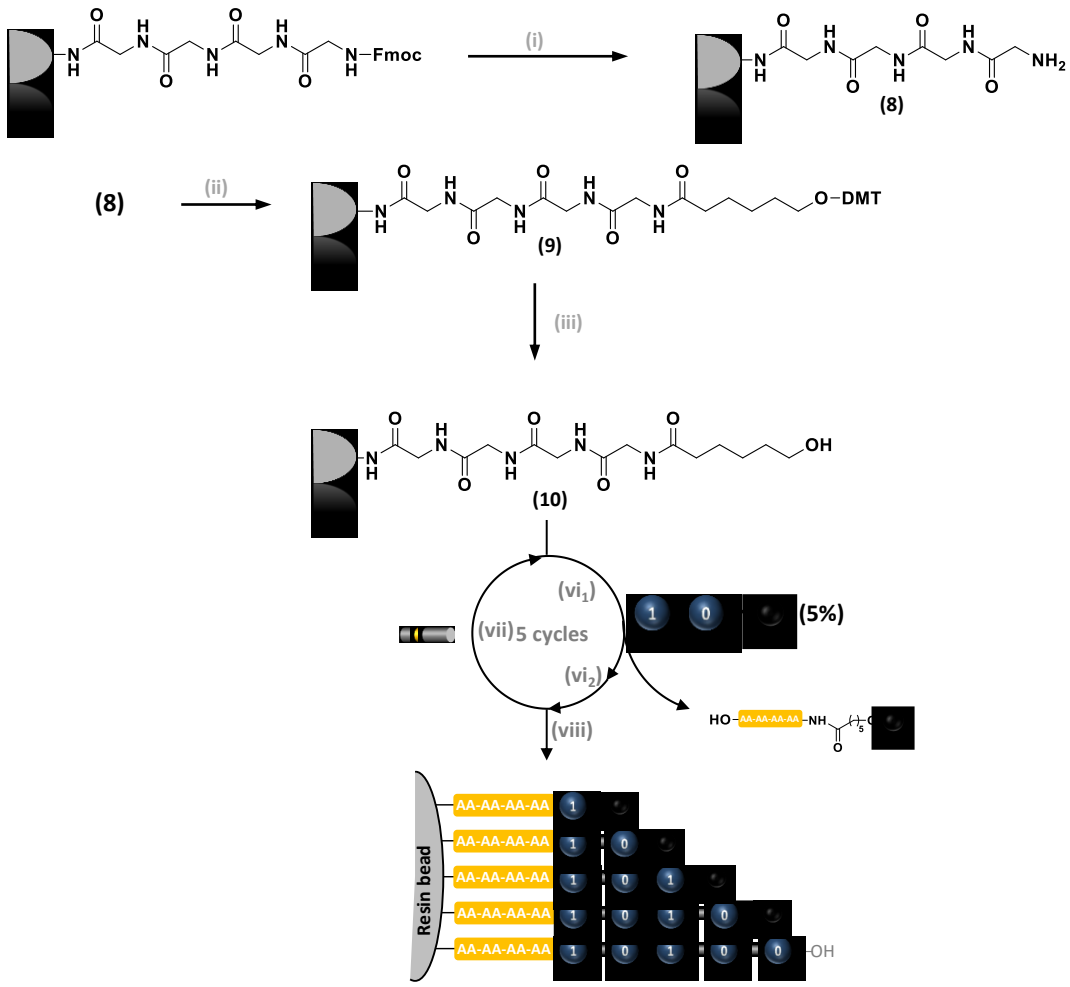


Figure E.1: General strategy employed for the development of poly(alkoxyamine phosphodiester)s for fragmentation-free sequencing.

3.1.5.1 FMOC deprotection (8)

The aminopolystirene resin containing the tetrapeptide (L=0.49 mmol/g) (101 g, 0.25 mmol, 1 eq.) was swollen in dichloromethane for 10 minutes. The support was treated with a solution of 25% piperidine in DMF for 20 minutes. Afterwards, the solvent was removed and the resin washed several times with DMF and dichloromethane.

3.1.5.2 Linker L_1 coupling (9)

The resin containing the unprotected polypeptidic sequence was swollen in dichloromethane. Afterwards, DMAP (45 mg, 0.37 mmol, 1.5 eq.), DCC (15 4mg, 0.75 mmol, 3 eq.) and a solution of pentanoic acid (271 mg, 0.55 mmol, 2.5 eq.) in 6 mL of anhydrous dichloromethane were added and the resin was shaken overnight. The day after, the solvent was removed and the solid support was washed with DMF (3x10 mL) and dichloromethane (3x5 mL). The resin was capped for 1 hour with a solution of acetic anhydride in pyridine in

1/5. Afterwards, the solvent was removed and the resin washed with dichloromethane (5x3 mL) and ether (5x3 mL).

3.1.5.3 Poly(alkoxyamine phosphodiester)s synthesis on solid support.
After DMT deprotection with TCA 3% in dichloromethane the iterative protocol was performed as described in section 3.1.3 using a mixture of the corresponding monomer (0_a or 1_a) and 5% of capping agent for the phosphoramidite coupling (step iv). The sequence was cleaved from the resin using an aqueous solution of ammonia overnight. The mixture was filtered and bubbled under argon for 30 min in order to remove the excess of ammonia and the solvent was removed.

3.2 Additional figures chapter2

3.2.1 MS analysis and MS/MS sequencing.

All the synthesized sequence defined polymers were analyzed by the group of Prof. Laurence Charles at AIX Marseille. The samples were dissolved in methanol and then diluted with a methanolic solution of ammonium acetate (3 mM) and introduced in the ESI source at a 5 µL.min^{-1} flow rate using a syringe pump.

Table E.1: ESI-HRMS characterization of the digital polymers synthesized and studied in this work.

	sequence	*mass*	*m/z$_{th}$*	*m/z$_{exp}$*
1	α-$0_aT_20_aT_20_aT_2$	1339.7	672.3598[a]	1339.6910
2	α-$0_aT_30_aT_30_aT_3$	1549.9	515.6196[a]	515.6187
3	α-$0_aT_30_aT_30_aT_30_aT_3$	2041.2	509.2819[b]	509.2814
4	α-$1_aT_30_aT_30_aT_31_aT_3$	2097.2	523.2975[b]	523.2977
5	α-$(1_aT_3)_2(0_aT_3)_31_aT_31_aT_3$	4118.4	514.0417[c]	514.0429
6	α-$0_bT_30_bT_30_bT_31_bT_3$	2097.2	523.2975[b]	523.2970
7	α-$0_bT_30_bT_31_bT_31_bT_3$	2125.3	530.3035[b]	530.3060
8	α-$0_bT_31_bT_3$	4174.4	521.0496[c]	521.0495
9	α-$0_cT_30_cT_30_bT_30_cT_3$	2153.3	537.3132[b]	537.3123

a $[M - 3H]^{3-}$, b to $[M - 4H]^{4-}$; c $[M - 8H]^{8-}$ .measured at isotopic maximum

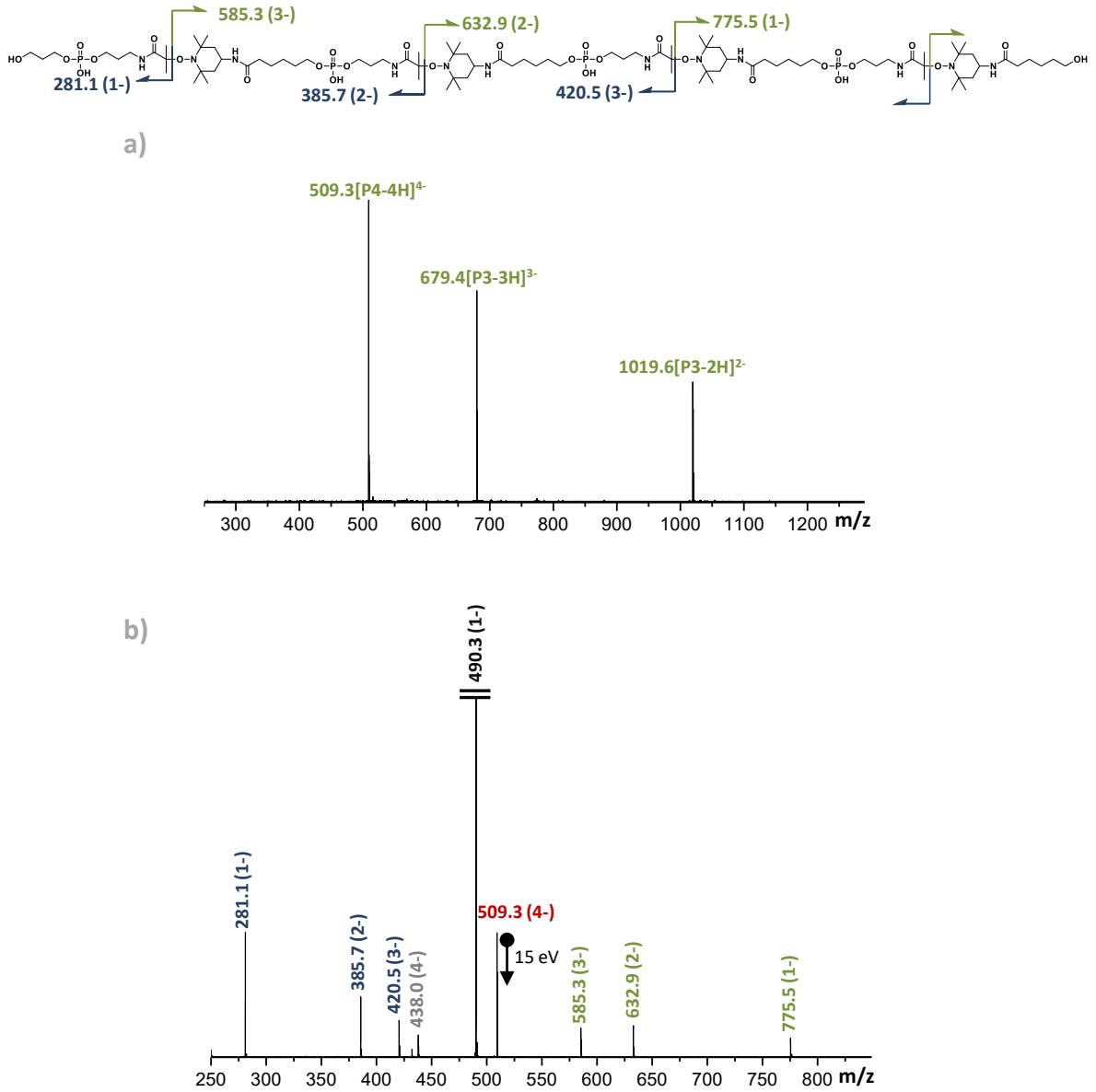


Figure E.2: Example of Sequencing of poly(alkoxyamine phosphodiester)s **a)** Negative ion mode ESI mass spectrum of a poylmer coding for, α-$0_aT_30_aT_30_aT_30_aT_3$ (entry 6 Table E.1). **b)** MS/MS sequencing of $[P - 4H]^{4-}$ at m/z 509.3 (red labelled) (collision energy : 15 eV) and corresponding dissociation scheme.

3.2.2 SEC.

The hexadecamer (entry 8 Table E.1) was analyzed by SEC in a mixture of 60% millipore water, 40% acetonitrile and 0.1 M $NaNO_3$(flow rate: 0.5 mL•min-1). The resulting sequence defined polymer showed a dispersity index (Đ) of 1.019.

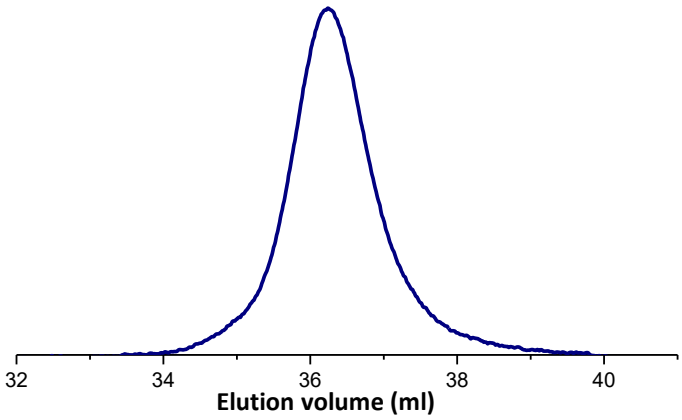


Figure E.3: Size exclusion chromatogram recorded in water/acetonitrile for oligomer 9 in Table E.1

3.2.3 NMR.

3.2.3.1 Coding monomer NMR

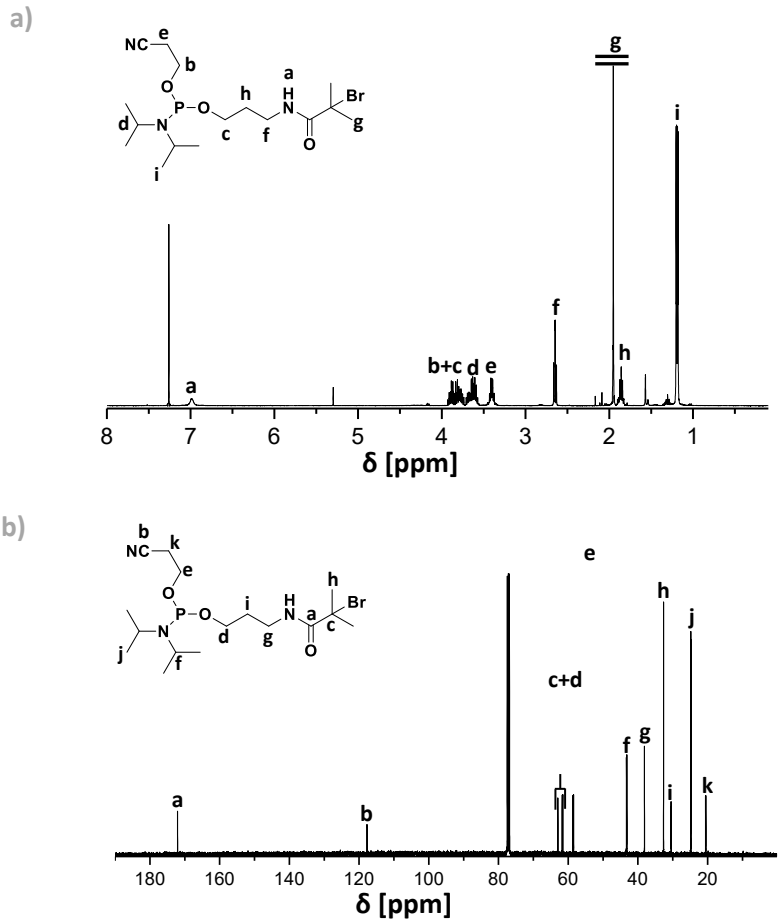


Figure E.4: NMR recorder in $CDCl_3$ for the coding monomer 0_a. **a)** ^{1}H-NMR **b)** ^{13}C-NMR.

3.2.3.2 Poly(alkoxyamine phosphodiester)s NMR

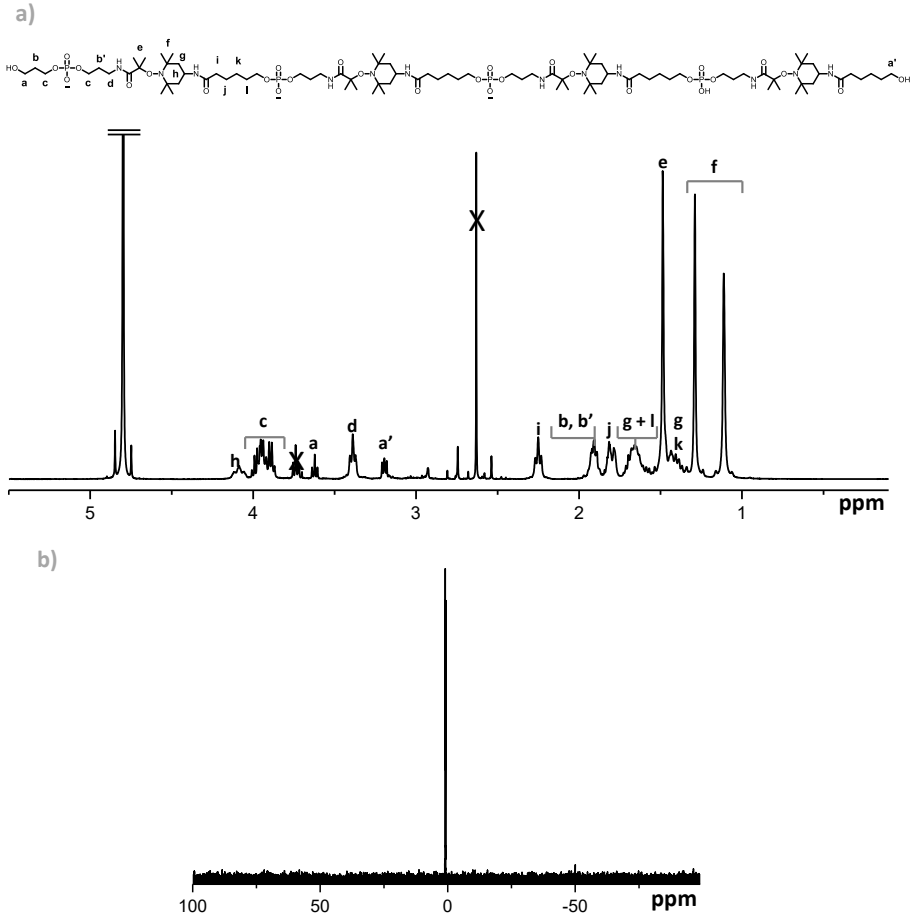


Figure E.5: NMR spectra recorded in D_2O for entry 3 in Table E.1 corresponding to the sequence α-0a$T_3$0a$T_3$0a$T_3$0aT_3. **a)** ^{1}H NMR, **b)** ^{31}P-NMR.

3.2.3.2 Capping agent NMR

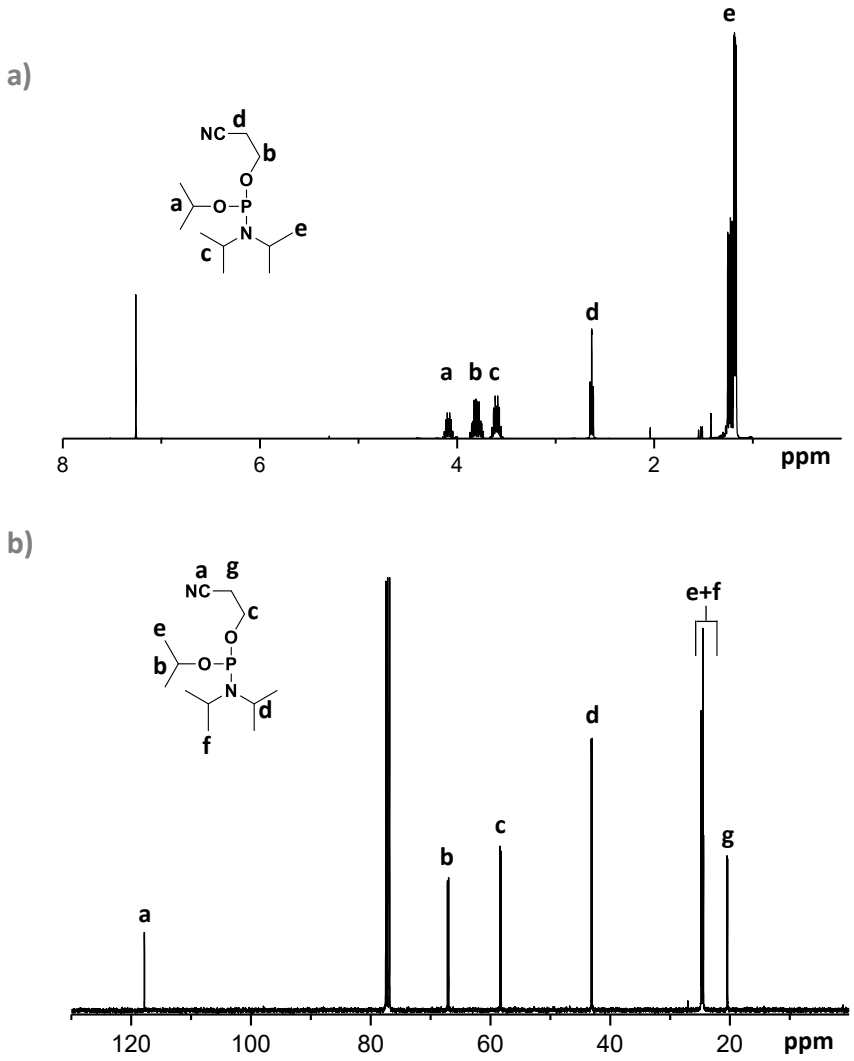
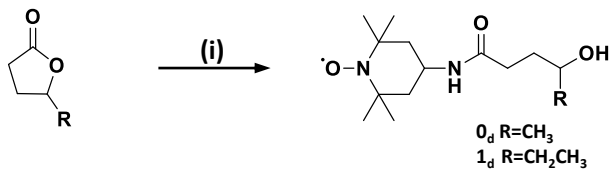


Figure E. 6: NMR of the capping agent recorded in $CDCl_3$: **a)** ^{1}H-NMR; **b)** ^{31}P-NMR.

3.3 Syntheses Chapter 3

3.3.1 Synthesis of alphabet 1

The synthesis of the monomers forming alphabet 1 was done following the synthetic strategy depicted in Scheme E.5.



Scheme E.5: Synthetic strategy employed for the synthesis of alphabet 1.

3.3.1.1 Synthesis of coding nitroxide 0_d

4 amino TEMPO (800 mg, 4.6 mmol, 1 eq.) was dissolved in γ-valerolactone (1.38 g, 13.8 mmol, 3 eq.). The mixture was refluxed at 130°C for 2.5 hours and the crude was purified by

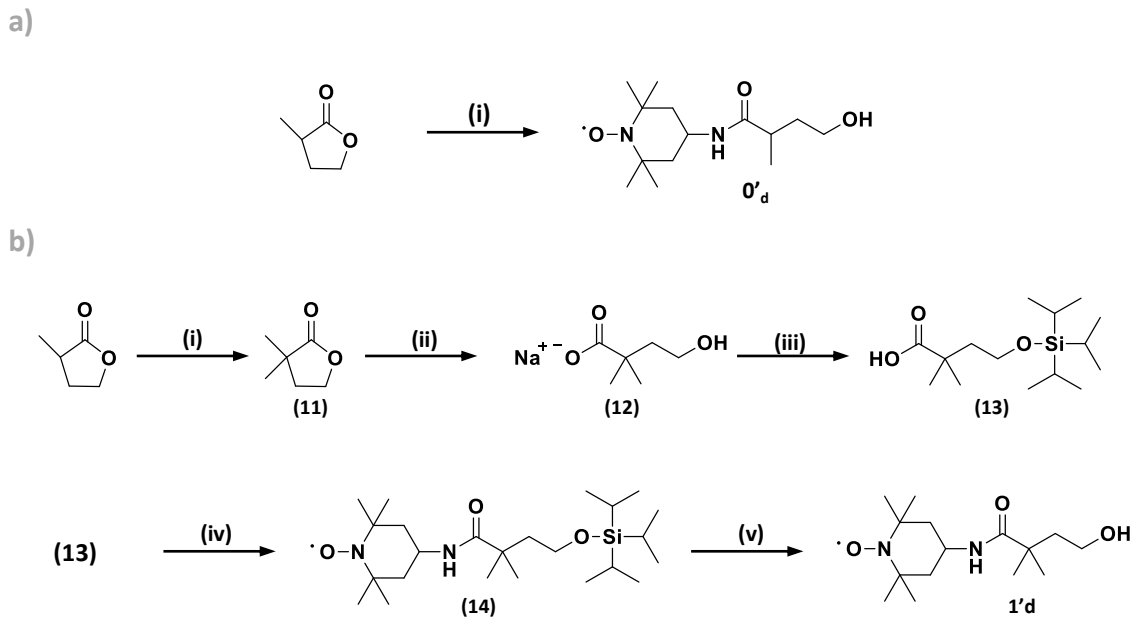
chromatographic column on silica gel (dichloromethane/methanol) (95/5). The product was recovered as a red oil in 82% yield. ESI-MS $[M+H]^+$ m/z expected for $C_{14}H_{28}N_2O_3^{+\bullet}$ 272.2094, found 272.2092.

3.3.1.2 Synthesis of coding nitroxide 1_d

4 amino TEMPO (800 mg, 4.6 mmol, 1 eq.) was dissolved in γ-hexanolactone (1.62 g, 13.8mmol, 3 eq.). The mixture was refluxed at 130°C for 2.5 hours and the crude was purified by chromatographic column on silica gel (dichloromethane/methanol) (95/5). The product was recovered in 89% yield. ESI-MS $[M+H]^+$ m/z expected for $C_{15}H_{30}N_2O_3^{+\bullet}$ 286.2251, found 286.2249.

3.3.2 Synthesis of alphabet 2

The synthesis of the monomers forming alphabet 2 was done following the synthetic strategies depicted in Scheme E.6.



Scheme E.6: Synthetic strategies employed for the synthesis of alphabet 2. **a)** Synthesis of monomer $0'_d$; **b)** Synthesis of monomer $1'_d$.

3.3.2.1 Synthesis of coding nitroxide $0'_d$

4 amino TEMPO (800 mg, 4.6 mmol, 1 eq.) was dissolved in α-methyl-γ-butyrolactone (1.42 g, 13.8 mmol, 3 eq.) and stirred at 130°C for 2.5h. The mixture was purified by chromatography on silica gel (dichloromethane/methanol) (95/5). The product was recovered oil in 85% yield. ESI-MS $[M+H]^+$ m/z expected for $C_{14}H_{28}N_2O_3^{+\bullet}$ 272.2094, found 272.2096.

3.3.2.2 Synthesis of intermediate 11

The synthesis of intermediate 11 was adapted according to a procedure from the litearature[230] In a two-necked round bottom flask, diisopropylamine (7.2 mL, 40.97 mmol, 1.2 eq.) was dissolved in 100 mL of anhydrous THF, and the solution was cooled down to 0°C

under argon atmosphere. A solution of n-BuLi in THF 2.5 M (16 mL, 40.97 mmol, 1.2 eq.) was added dropwise to the solution and stirred for 30 minutes. The mixture was cooled to -78°C and a solution of α-methyl-γ- butyrolactone (4 g, 40.97 mmol, 1 eq.) in 10 mL of anhydrous THF was dropped. After 30 min a solution of methyl iodide (3.22 mL, 53.26 mmol, 1.3 eq.) and hexamethylphosphoramide (3.46 mL, 20.48 mmol, 0.5 eq.) was dropped and stirred at -78°C for 4 h and then overnight at r.t. The day after, the reaction was quenched by carefully addition of a 10% HCl solution in water until PH1. The THF was then removed and the aqueous layer was extracted with dichloromethane (3x50 mL). The organic layer was washed with water (1x 50 mL), 5% $NaHCO_3$ solution (2x 50mL), brine (1x 50mL), dried over $MgSO_4$, filtered and the solvent evaporated. Chromatographic column on silica gel (dichloromethane/methanol) (95/5) gives a red brown oil which is distilled under vacuum (bp 64-66°C, 3.2 mmHg), giving the intermediate 11 in 65% yield. ^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 1.27 (s, 6H, -CH_2-C-($C\underline{H}_3$)$_2$), 2.11 (t, 2H, -O-$C\underline{H}_2$-C-(CH_3)$_2$. 4.26 (t, 2H, -$C\underline{H}_2$-O-). ^{13}CNMR (100 MHz, $CDCl_3$, δ, ppm): 24.33, 37.24, 38.73, 64.83, 182.41.

3.3.2.3 Synthesis of intermediate 12

The synthesis of intermediate 12 was adapted according to an article from the literature.[231] Intermediate 11 (250 mg, 2.19 mmol, 1 eq.) was dissolved in a mixture of water (1.4 mL) and 1-4 dioxane (1.4 mL). Then, sodium hydroxide (87.6 mg, 2.19 mmol, 1 eq.) was added to the mixture. The solution was stirred for 45 minutes, the solvent was dried and the product was recovered in quantitative yield as sodium salt. ^{1}H NMR (400 MHz, DMSO-d6, δ, ppm): 0.99 (s, 6H, -CH_2-C-($C\underline{H}_3$)$_2$), 1.51 (t, 2H, -$C\underline{H}_2$-C-(CH_3)$_2$), 3.44 (t, 2H, -$C\underline{H}_2$-OH). ^{13}C-NMR (100 MHz, DMSO-d6, δ, ppm): 27.55, 41.73, 42.89, 58.43, 181.82.

3.3.2.4 Synthesis of intermediate 13

The synthesis of intermediate 13 was adapted from the literature.[232] The intermediate 12 (313 mg, 2.03 mmol, 1 eq.) was dissolved in 6.5 ml of anhydrous DMF. To the solution, triisopropylsilyl chloride (0.43 ml, 2.03 mmol, 1 eq.), dissolved in 23 mL of anhydrous DMF was dropped under argon atmosphere and stirred overnight. The reaction mixture was diluted with Et_2O (30 mL) and the formed precipitate was filtered out; the liquid phase was then washed with water (70 mL). The diethyl ether was extracted with a saturated $NaHCO_3$ solution (30mL). Afterwards, the combined aqueous layers were acidified to PH 3.5-4 with a 10% HCl solution and extracted with Et_2O (3x 70mL). The combined organic layers were dried over $MgSO_4$ and the solvent removed, giving the intermediate 13 as colorless oil, in 62% yield, ^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 1.05 (s. 21H, -O-Si-(iPr)3), 1.25 (s, 6H, -CH2-C-($C\underline{H}_3$)$_2$), 1.87 (t, 2H, -$C\underline{H}_2$-C-(CH_3)$_2$), 3.80 (t, 2H, -$C\underline{H}_2$-O-TIPS). ^{13}C-NMR (100 MHz, CDCl3, δ, ppm): 12.41, 25.16, 40.82, 42.25, 60.31, 181.82.

3.3.2.5 Synthesis of intermediate 14

4-Amino-TEMPO (277 mg 1.62 mmol, 1eq.), intermediate 13 (467 mg, 1.62 mmol, 1eq.) and DIPEA (0.85 mL 4.86 mmol, 3eq.) were dissolved in 23 mL of anhydrous dichloromethane and cooled down to 0°C. To the solution, PyBOP (1.3 g, 2.43 mmol, 1.5 eq.) was added and

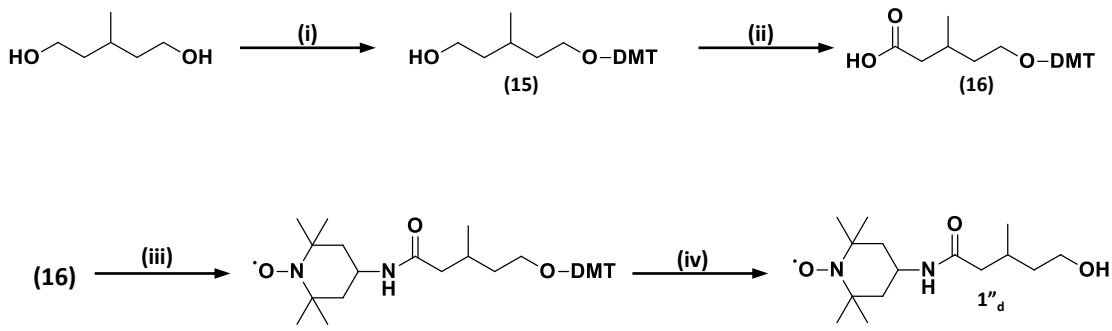
after 10 minutes the reaction mixture was let to reach room temperature and stirred overnight under argon atmosphere. The day after, the reaction was quenched with ammonium chloride saturated solution and the solvent was evaporated. The aqueous layer was extracted with ethyl acetate (3x 40mL). The organic layer was then washed with brine (60 mL), dried over $MgSO_4$ and the solvent was evaporated. Chromatographic column on silica gel (EtOAc/Cyclohexane) (35/65) gives the product in 75% of yield. ESI-MS $[M+H]^+$, m/z expected for $C_{23}H_{46}N_2O_3Si$ 441.35, found 442.35.

3.3.2.6 Synthesis of nitroxide $1'_d$

The intermediate 14 (353 mg, 0.8 mmol, 1eq.) was dissolved in10 mL of THF. To the solution, 2.8 mL of TBAF (1M in THF, 2.8 mmol, 3.5 eq.) was added. The reaction was monitored by thin layer chromatography and was stopped after 2.5 hours. The solvent was evaporated and purified by chromatographic column on silica gel (dichloromethane/methanol) (95/5). After purification, the monomer $1'_d$ was recovered in 98% yield as a red yellow powder. ESI-MS $[M+H]^+$, m/z expected for $C_{15}H_{30}N_2O_3^{+\bullet}$ 286.2251, found 286.2251.

3.3.3 Synthesis of alphabet 3

The first synthetic strategy employed to obtain alphabet 3 was done following the synthetic pathway described in Scheme E.7. However, the alphabet was obtained following the synthetic pathway described in Scheme E.8.



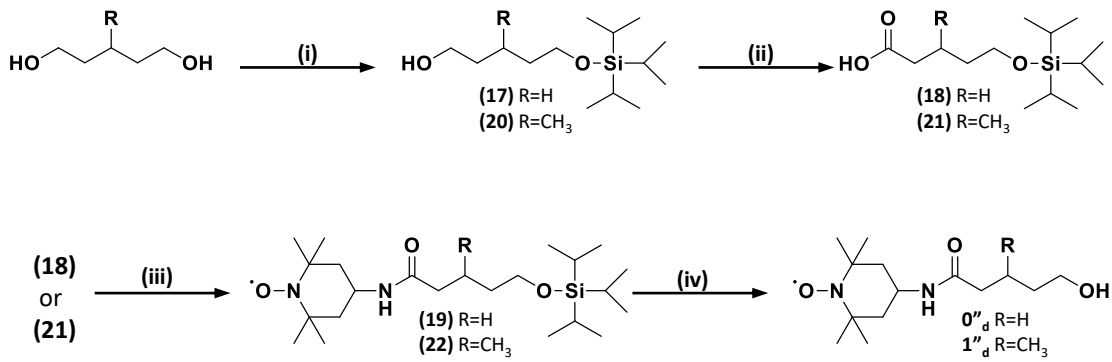
Scheme E.7: First synthetic strategy employed for the synthesis of alphabet 3.

3.3.3.1 Synthesis of intermediate 15

3-methyl 1-5 pentanediol (1.5 g, 12.7 mmol, 1eq.) was dissolved in a mixture of 12 mL of anhydrous pyridine and 18 mL of anhydrous THF. To the mixture, DMT-Cl (4.3g, 12.7 mmol, 1 eq.) was added under argon atmosphere in 4 aliquots; one each 30 minutes. After the last addition, the reaction mixture was stirred overnight under argon atmosphere. The day after, the solvent was evaporated ant the residue taken in EtOAc (30 mL). The solution was washed with cold water (2x 20 mL) and dried over $MgSO_4$. Chromatographic column on silica gel (cyclohexane/EtOAc/DIPEA) (80/20/2) gives the product in 64% yield. ^1H NMR (400 MHz, $CDCl_3$, δ, ppm): 0.82-0.84 (d, 2H, $CH\text{-}C\underline{H}_3$), 1.62-1.65 (m, 1 H, $C\underline{H}\text{-}CH_3$), 1.64–1.79 (m, 4H, O-$CH_2$-$C\underline{H}_2$-CH-$C\underline{H}_2$), 3.04-3.15 (m, 2H, $C\underline{H}_2$-O-DMT), 3.61-3.68 (m, 2H, $C\underline{H}_2$-OH), 3.79 (s. 6H 2x

Ar-O-C<u>H</u>$_3$), 6.79-6.84 (m, 4H, Ar-<u>H</u>), 7.25-7.34 (m, 8H Ar-<u>H</u>), 7.40-7.43 (m, 1H, Ar-<u>H</u>). ^{13}C-NMR (100 MHz, CDCl3, δ, ppm): 20.17, 26.25, 40.07, 40.09, 55.75, 64.74, 61.70, 113.12, 126.74, 127.83, 128.38, 130.27, 136.81, 145.50, 158.35, 173.94.

The synthesis of the monomers forming alphabet 3 was done following the synthetic strategy depicted in Scheme E.8



Scheme E.8: Synthetic strategy employed for the synthesis of alphabet 3.

3.3.3.2 Synthesis of intermediate 17

Synthesis of intermediate 17 was adapted from the literature. [233]In a 2 necked round bottom flask, 1,5-pentanediol (5 g, 48 mmol, 1eq.) and imidazole (7.2g, 105 mmol, 2.2 eq.) were dissolved in 70 mL of anhydrous DMF and the solution was cooled down at 0ºC. Into an addition funnel, TIPSCl (10.17 g, 52 mmol, 1.1 eq.) was dissolved in 30 mL of anhydrous DMF and the resulting solution was dropped into the reaction mixture at 0ºC. After 5 hours, Et_2O (60 mL) and Brine (60 mL) were added to the mixture. The organic layer was separated and the aqueous layer was further extracted with Et_2O (3x 60mL). The solvent was removed and the organic layers were washed with water and dried over Na_2SO_4. The solvent was evaporated and the crude was purified by chromatographic column on silica gel (EtOAc/Cyclohexane) (35/65) giving the intermediate 15 in 74% yield. $_1$H NMR (400 MHz, $CDCl_3$, δ, ppm): 1.04-1.08 (m 21H (I<u>Pr</u>)$_3$), 1.40-1.47 (m 2H, -C<u>H</u>$_2$-CH$_2$-CH$_2$-OH), 1.54-1.63 (m 4H, C<u>H</u>$_2$-CH$_2$-OH and C<u>H</u>$_2$-CH$_2$-OTIPS), 3.62-3.66 (t 2H, -C<u>H</u>$_2$-OTIPS), 3.67-3.70 (t 2H, -C<u>H</u>$_2$-OH). ^{13}C NMR (100 MHz, CDCl3, δ, ppm): 12.17, 18.10, 22.19 32.71, 32.83, 63.44, 63.17.

3.3.3.3Synthesis of intermediate 18

Synthesis of intermediate 18 was done according to a procedure from the literature.[234] In a round bottom flask intermediate 17 (3.6 g, 13.8 mmol, 1 eq.), TEMPO (647 mg, 4.6 mmol, 0.3 eq.) and (diacetoxyiodo)benzene (13.7 g, 41.4 mmol, 3 eq.) were dissolved in a mixture of water (12 mL) and dichloromethane (14 mL).The reaction was stirred at 0ºC, monitored by TLC and stopped after 3 hours. The crude was diluted with dichloromethane (20 mL) and washed with saturated $Na_2S_2O_3$ solution (3x 30mL). The organic layer was dried over Na_2SO_4 and the solvent was evaporated. Chromatographic column on silica gel (cyclohexane/EtOAc) (90/10) gives product as pale yellow oil in 62% yield. ^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 1.04-

1.08 (m 21H $(IPr)_3$), 1.56-1.62 (m 2H, HO-C=O-CH_2-C$\underline{H}_2$), 1.70-1.77 (m 2H, C$\underline{H}_2$-CH_2-OTIPS), 2.38-2.42 (m, 2H, -C$\underline{H}_2$-OTIPS-), 3.69-3.72(m, 2H. HO-C=O-C$\underline{H}_2$), 10.92 (s, 1H, -C=O-O$\underline{H}$). ^{13}C NMR (100 MHz, $CDCl_3$, δ, ppm): 12.12, 18.13, 21.31, 32.37, 34.05, 62.98, 180.43.

3.3.3.4 Synthesis of intermediate 19

Synthesis of intermediate 19 was adapted according to a procedure from the literature.[235] 4-Amino-TEMPO (445 mg, 2.6 mmol, 1.1eq.), intermediate 18 (683 mg 2.36 mmol, 1eq.) and DIPEA (1.2 mL, 7.08 mmol, 3eq.) were dissolved in 40 mL of anhydrous dichloromethane under argon atmosphere. The mixture was cooled down to 0°C then PyBOP (1.84 g, 3.54 mmol, 1.5eq.) was added and after 10 minutes, the reaction mixture was allowed to reach room temperature and stirred overnight The day after, the reaction was quenched with ammonium chloride saturated solution and the solvent was evaporated. The aqueous layer was extracted with EtOAc (3x40 mL), washed with brine (30 mL) and dried over $MgSO_4$. Chromatographic column on silica gel (EtOAc/Cyclohexane) (35/65) gives product in 77% yield. ESI-MS$[M+H]^+$, m/z expected for $C_{23}H_{48}N_2O_3Si^{+\bullet}$ 427.34, found 428.34.

3.3.3.5 Synthesis of coding nitroxide 0”$_d$

The intermediate 19 (879 mg, 1.99mmol, 1eq.) was dissolved in 15 mL of THF. To the solution, 6.97 mL of TBAF (1M in THF, 2.8mmol, 3.5 eq.) were added. The reaction was stopped after 2.5 hours. Afterwards, the solvent was dried and chromatographic column on silica gel (dichloromethane/methanol) (95/5) gives the monomer 0”d in 98% yield as a red oil. ESI-MS $[M+H]^+$, m/z expected for $C_{14}H_{28}N_2O_3^{+\bullet}$ 271.20, found 272.21.

3.3.3.6 Synthesis of intermediate 20

The protocol employed for the synthesis of intermediate 20 was the same as the one described previously for intermediate 17 but using 3-methyl-1,5-pentanediol as starting material. The product was recovered as colorless oil in 70% yield. ^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 0.94-0.92 (d, 3H, CH_2-CH(C$\underline{H}_3$)-CH_2-), 1.08-1.04 (m 21H $(\underline{IPr})_3$), 1.67-1.46 (m,2H, -C$\underline{H}_2$-CH_2-OTIPS), 2.25-2.05 (m, 2H, -C$\underline{H}_2$-CH_2-OH),. 2.46-2.41 (m, 1H, -CH_2-C$\underline{H}$(CH_3)-CH), 3.79-3.72 (m, 4H, -C$\underline{H}_2$-OH- and -C$\underline{H}_2$-OTIPS-). ^{13}C NMR (100 MHz, $CDCl_3$, δ, ppm): 12.16, 18.18, 20.17, 26.25, 40.07, 40.09, 61.34, 61.70.

3.3.3.7 Synthesis of intermediate 21

The protocol employed for the synthesis of intermediate 21 was the same as the one described previously for intermediate 18 but using intermediate 20 as starting material. The product was recovered as a pale-yellow oil in 65% yield.^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 1.04-1.05 (d, 3H, CH_2-CH(C$\underline{H}_3$)-CH_2-), 1.06-1.08 (m 21H $(\underline{IPr})_3$), 1.37-1.45 (d, 3H, -CH_2-CH(C$\underline{H}_3$)-CH_2-), 1.53-1.62 (m 2H, C$\underline{H}_2$-CH_2-OTIPS), 2.14-2.25 (m, 2H,. HO-C=O-C$\underline{H}_2$), 2.41-2.46 (m, 1H, CH_2-C$\underline{H}$(CH_3)-CH_2-), 3.70-3.80 (m, 2H, -C$\underline{H}_2$-OTIPS-), 10.77 (s, 1H, -C=O-O$\underline{H}$), ^{13}C NMR (100 MHz, CDCl3, δ, ppm): 12.13, 18.11, 19.95, 27.44, 39.54, 41.77, 61.46, 179.61.

3.3.3.8 Synthesis of intermediate 22

The protocol employed for the synthesis of intermediate 22 was the same as the one employed for intermediate 19, but using intermediate 21 as starting material. The product was recovered as a colorless oil in 73% yield. ESI-MS [M+H]+, m/z expected for $C_{24}H_{50}N_2O_3S^{i+\bullet}$ 441.36, found 442.36.

3.3.3.9 Synthesis of coding nitroxide $1''_d$

The protocol employed for the synthesis of intermediate $1''_d$ was the same as the one employed for $0''_d$, but using 22 as starting material. The product $1''_d$ was recovered as a red oil in 98% yield. ESI-MS $[M+H]^+$, m/z expected for $C_{15}H_{30}N_2O_{3+\bullet}$ 285.22, found 286.22.

3.3.4 Iterative solid-state protocol

The iterative protocol for the sequence-define polymers synthesis was performed in the same condition described in section 3.1.3. All the products were recovered in overall yields between 73 and 90% as shown in Table III.2. The resulting products were analyzed by ESI-HRMS, SEC and NMR and were sequenced by ESI-MS/MS.

3.4 Additional figures chapter 3

3.4.1 MS analysis and sequencing

All the sequencing synthesized using alphabet 3 were recovered in good yield moreover the molecular weight found were in line with the theoretical ones. The ESI-HRMS characterization of these sequence-defined polymers is summarized in Table E.2.

Table E.2: Negative ESI-HRMS characterization of the sequence-coded polymers synthesized using alphabet 3.

	sequence	*yield*	*m/z$_{th}$*	*m/z$_{exp}$*
1	α-1_a-$1''_d$-1_a-$1''_d$	90%	556.3263	556.3261[a]
2	α-0_a-$1''_d$-0_a-$1''_d$-1_a-$0''_d$	73%	780.9408	780.9429[a]
3	α-1_a-$0''_d$-0_a-$1''_d$-1_a-$0''_d$	75%	787.9486	787.9510[a]
4	α-1_a-$0''_d$-1_a-$1''_d$-0_a-$0''_d$	78%	787.9486	787.9482[a]
5	α-0_a-$1''_d$-0_a-$0''_d$-1_a-$0''_d$-0_a-$0''_d$	93%	1012.5632	1012.5669[a]
6	α-1_a-$0''_d$-0_a-$0''_d$-1_a-$1''_d$-0_a-$1''_d$	85%	1033.5867	1033.5895[a]

[a] $[M-2H]^{2-}$ measured at isotopic maximum.

One example of dyad extraction of Poly(alkoxyamine phosphodiester)s by MS/MS is depicted in Figure E.7.

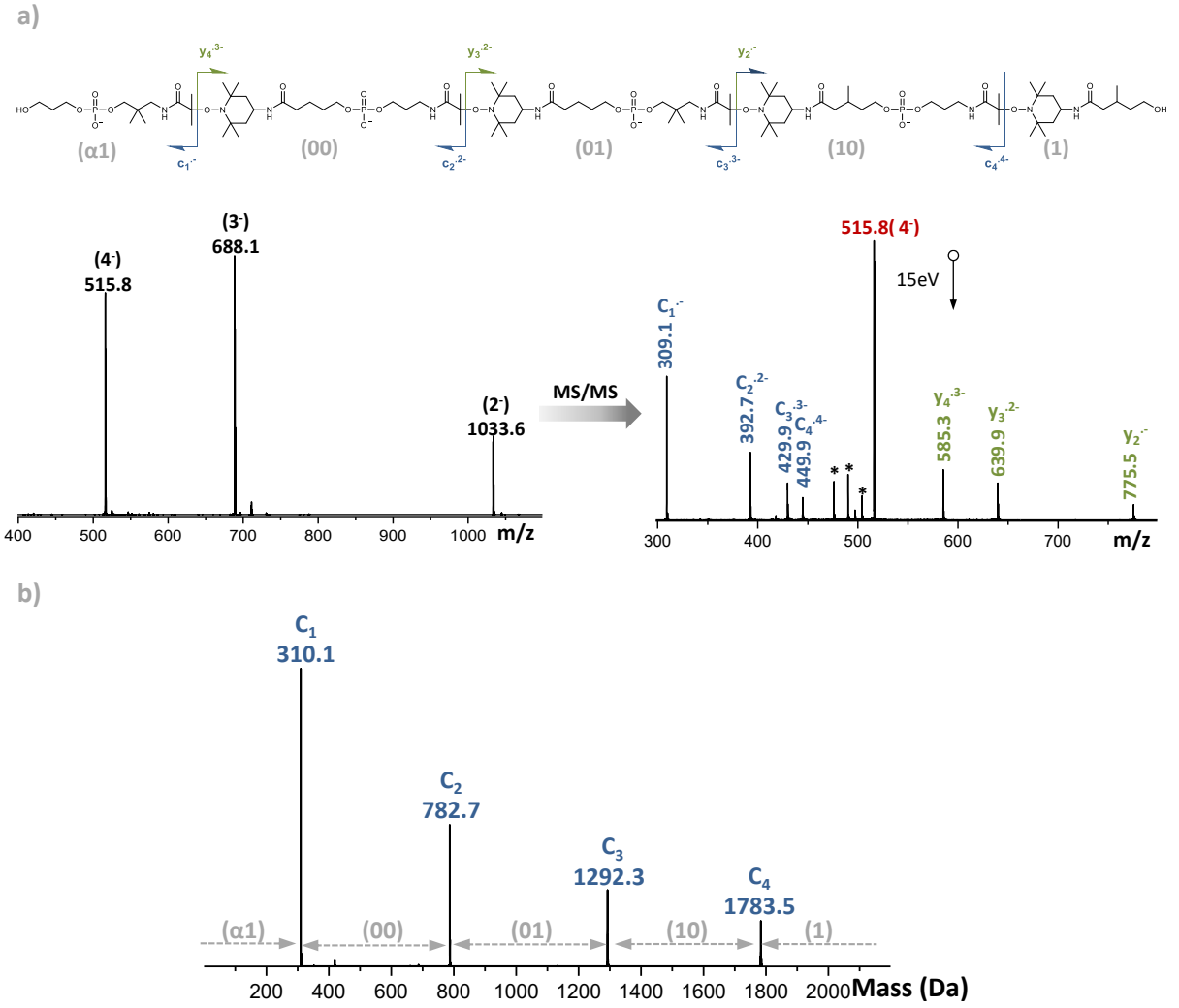


Figure E.7: Mass spectrometry analysis of the sequence-coded oligomer entry 6 tab.4: **a)** Negative ion mode ESI mass spectrum (left); molecular structure of the fully deprotonated oligomer with fragments produced in tandem MS; ESI-MS/MS obtained for this oligomer (right). The peaks not labeled in tandem MS correspond to internal fragments generated during secondary dissociation reactions .**b)** Charge state deconvolution (right) of the $c_i^{\bullet z-}$ ion series (blue labeled in MS/MS).

3.4.2 NMR

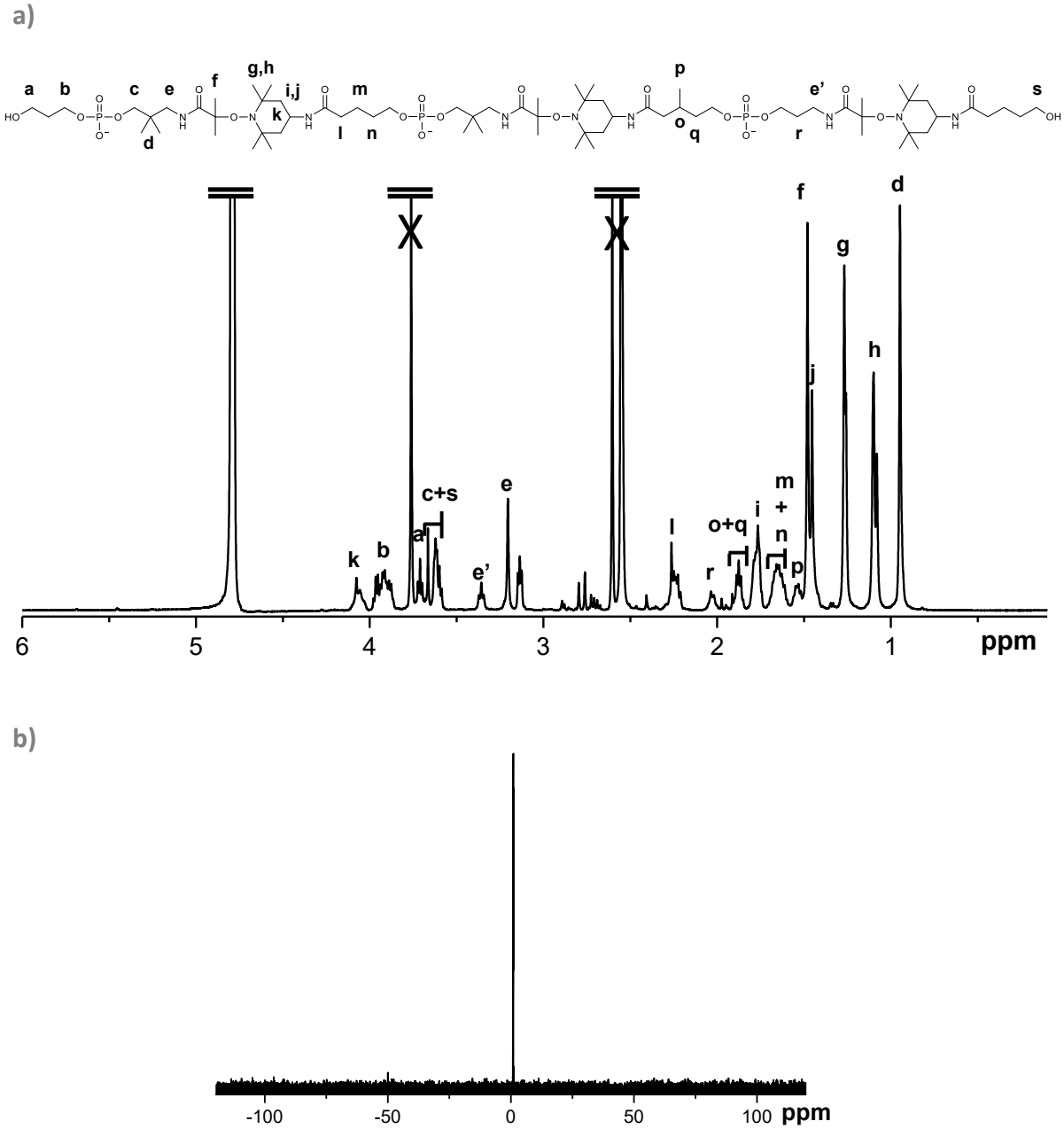


Figure E.8: Examples of NMR spectra obtained for oligomer 3 in Table E.2 recorded in D_2O: **a)** ^{1}H-NM, **b)** ^{13}C-NMR.

3.4.3 SEC

The Poly(alkoxyamine phosphodiester)s encoding dyads were analyzed by Size exclusion chromatography in a mixture of 60% millipore water, 40% acetonitrile and 0.1 M $NaNO_3$(flow rate: 0.5 mL•min-1).

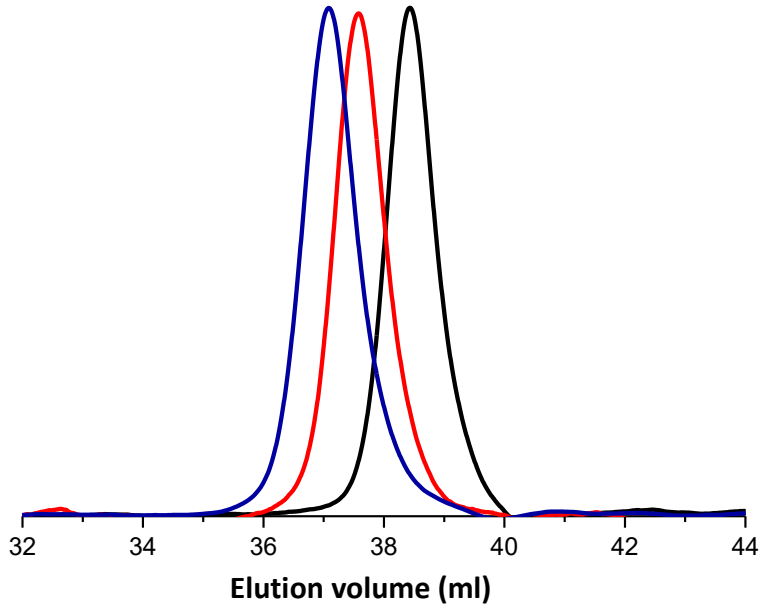


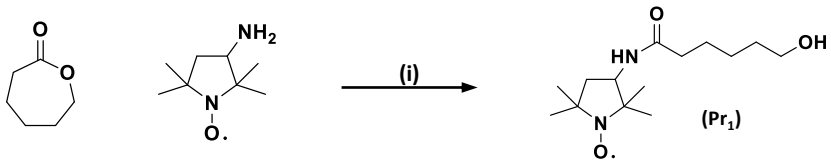
Figure E.9: Size exclusion chromatograms recorded in water/acetonitrile for oligomers P6 (blue), P4 (red) and P2 (black) in Table E.2.

3.5 Syntheses Chapter 4

As already discussed in Chapter 4, the hydroxy-functionalized nitroxide T_3 was synthesized following the synthetic strategy depicted in section 3.1.2.13. The monomer SG1-OH was supplied by Jean-Louis Clement from Marseille University.

3.5.1 Synthesis of monomer Pr_1

The synthetic strategy employed for the synthesis of monomer Pr_1 is depicted in Scheme E.9.



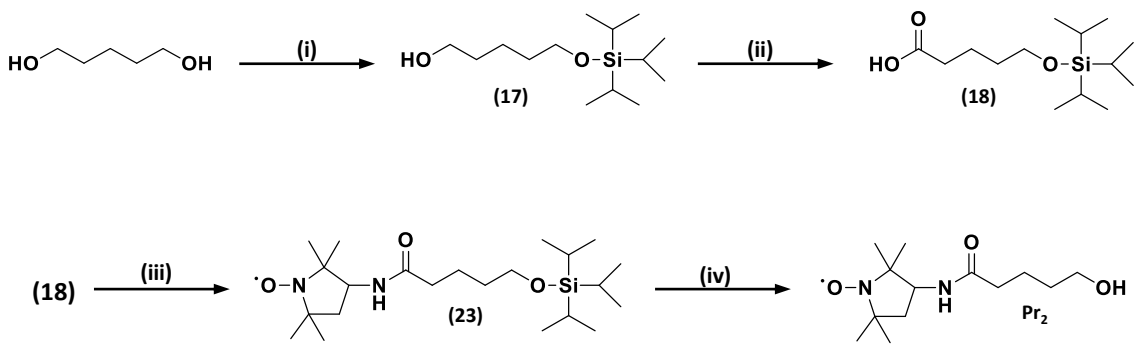
Scheme E.9: Synthetic strategy employed for the synthesis of Pr_1.

The synthesis of this molecule was done modifiying the procedure described in section 3.1.2.13. 3-Amino-PROXYL (250 mg, 1.59 mmol, 1eq.) was dissolved in Ɛ-caprolactone (544 mg, 4.77 mmol, 3 eq.) and refluxed at 140°C. The solution was monitored by TLC and stopped after 190 minutes. The mixture was purified by chromatography on silica gel (dichloromethane/methanol) (95/5) yielding 332 mg of product which was analyzed by

positive ESI-HRMS however, as already discussed in chapter 4, the analysis evidenced that the recovered product was not pure.

3.5.2 Synthesis of monomer Pr_2

The synthetic strategy employed for the synthesis of monomer Pr_1 is depicted in Scheme E.10.



Scheme E.10: Synthetic strategy employed for the synthesis of Pr_2.

3.5.2.1 Synthesis of intermediate 17

The synthesis of intermediate 17 was done following the same synthetic strategy described in section 3.3.3.2.

3.5.2.2 Synthesis of intermediate 18

The synthesis of intermediate 18 was done following the same synthetic strategy described in section 3.3.3.3.

3.5.2.3 Synthesis of intermediate 23

Intermediate 22 (190 mg, 0.66 mmol, 1 eq.), DIPEA (255mg, 1.98 mmol, 3 eq.) and 3-Amino-PROXYL (103 mg, 0.66mmol, 1 eq.) were dissolved in 20 mL of anhydrous DCM and the solution was cooled to 0°C. To the reaction mixture, PyBOP (515 mg, 0.99 mmol, 1.5 eq.) was added and the solution was allowed to reach room temperature and stirred overnight. The day after, the reaction was quenched with ammonium chloride saturated solution and the solvent was evaporated. The aqueous layer was extracted with EtOAc (3x20 mL), washed with brine (20 mL) and dried over MgSO4. Chromatographic column on silica gel (EtOAc/Cyclohexane) (30/70) gives product in 67% yield. ESI-MS $[M+H]^+$, m/z expected for $C_{23}H_{47}N_2O_3Si^{\bullet}$ 413.32, found 414.33.

3.5.2.4 Synthesis of Monomer Pr_2

The synthesis of Pr_2 was done following the same synthetic strategy described in section 3.3.2.6, using intermediate 23 as starting material. ESI-MS $[M+H]^+$, m/z expected for $C_{13}H_{25}N_2O_3^{\bullet}$, 257.18 found 258.19.

3.5.3 Solid state Iterative protocol

3.5.3.1 Phosphoramidite coupling (step i_1 and i_2)

The step (i) of the iterative protocol was performed using the same conditions described in section 3.1.3.1 using 3eq. of the corresponding monomer (0a, or 1a).

3.5.3.2 Nitroxide-radical coupling (step ii)

The hydroxy-functionalized nitroxides coupling (SG1-OH, T_3 and Pr_2) were performed using the same conditions described in section 3.1.3.2. However the reaction time was increased to 1 hour for nitroxide SG1-OH and nitroxide Pr_2.

3.5.3.3 Cleavage and deprotection (step iii)

The final purification and cleavage were done following the same condition described in section 3.1.3.3. All the products were recovered in overall yields between 70 and 85%. The resulting products were analyzed by ESI-HRMS, SEC and NMR.

3.6 Additional figures chapter 4

3.6.1 MS analysis and selective cleavage

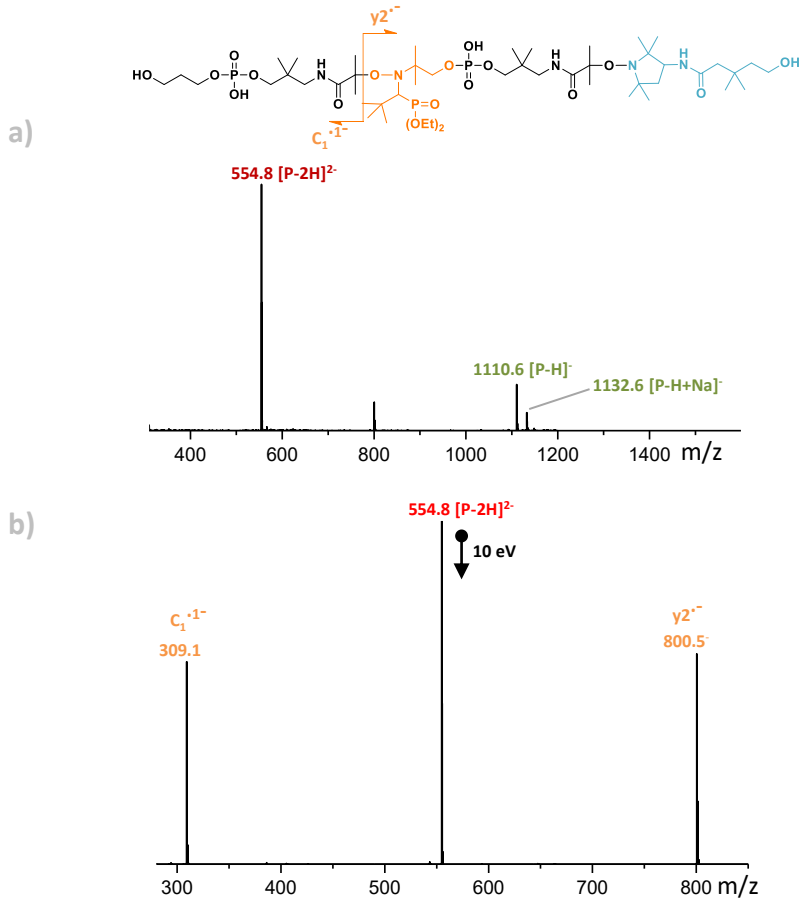


Figure E.10: Selective fragmentation of the SG1-based alkoxyamine bonds obtained at 10 eV for a polymer coding for1_a-SG1-1_a-Pr_2.

3.6.2 SEC

Size exclusion chromatography registered in water/acetonitrile for the polymer Octamer coding for 1_a-Pr_2-1_a-SG1-0_a-Pr_2-1_a-SG1.

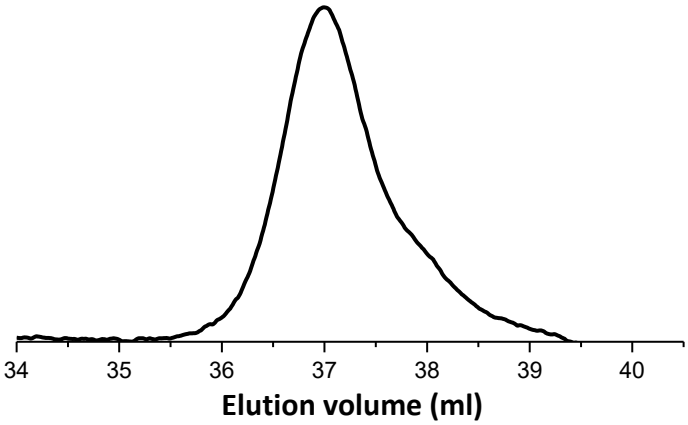


Figure E.11: Size exclusion chromatograms recorded in water/acetonitrile for oligomers coding for 1_a-Pr_2-1_a-SG1-0_a-Pr_2-1_a-SG1.

3.6.3 NMR

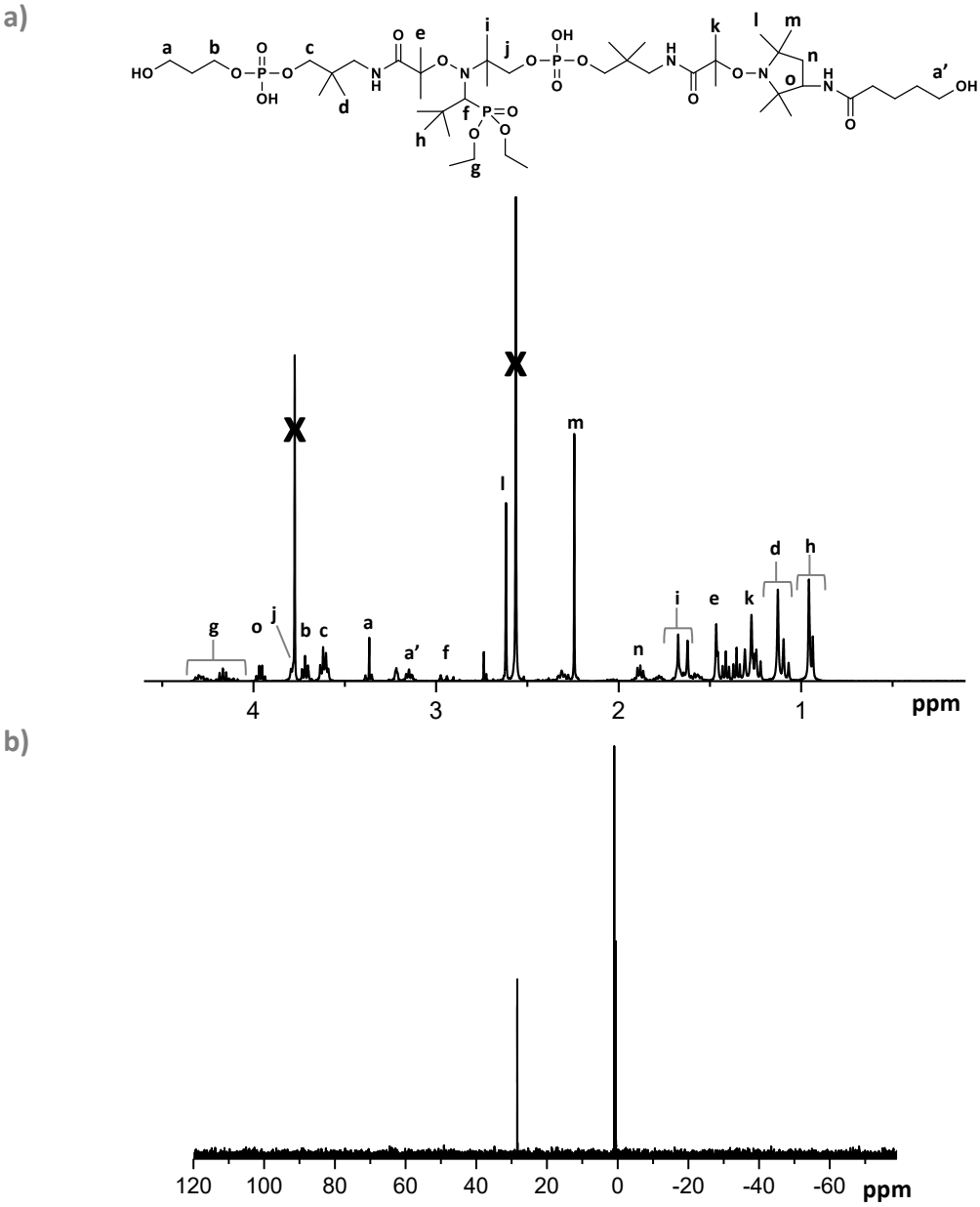


Figure E.12: NMR spectra recorded in D_2O for an oligomer coding for 1_a-SG1-1_a-Pr_2: **a)** ^{1}H-NM, **b)** ^{31}P-NMR.

References

[1] D. R. John Gantz, *external Publication of IDC, Information and Data* **2010**, 1-6.

[2] M. K. G. D. Limbachiya, *arXiv* **2015**.

[3] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, E. Birney, *Nature* **2013**, *494*, 77.

[4] S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, *Nature* **2017**, *547*, 345.

[5] H. Colquhoun, J.-F. Lutz, *Nat. Chem.* **2014**, *6*, 455.

[6] J.-F. Lutz, *Macromolecules* **2015**, *48*, 4759-4767.

[7] J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science* **2013**, *341*, 1238149.

[8] J.-F. Lutz, J.-M. Lehn, E. W. Meijer, K. Matyjaszewski, *Nat. Rev. Mat.* **2016**, *1*, 16024.

[9] H. Mutlu, J. F. Lutz, *Angew. Chem. Int.Ed.* **2014**, *53*, 13010-13019.

[10] R. B. Merrifield, *Angew. Chem. Int.Ed. in English* **1985**, *24*, 799-810.

[11] R. K. Roy, A. Meszynska, C. Laure, L. Charles, C. Verchin, J.-F. Lutz, *Nat. Commun.* **2015**, *6*, 7237.

[12] A. Al Ouahabi, M. Kotera, L. Charles, J.-F. Lutz, *ACS Macro Lett.* **2015**, *4*, 1077-1080.

[13] Jean-Arthur. Amalian, Abdelaziz Al Ouahabi, Gianni Cavallo, Niklas. Felix. König, Salomé. Poyer, Jean-François Lutz, Laurence Charles, *J. Mass Spectrom.* **2017**, *52*, 788-798.

[14] L. Charles, C. Laure, J.-F. Lutz, R. K. Roy, *Macromolecules* **2015**, *48*, 4319-4328.

[15] P. A. Wender, V. A. Verma, T. J. Paxton, T. H. Pillow, *Acc. Chem. Res.***2008**, *41*, 40-49.

[16] R. S. Youngquist, G. R. Fuentes, M. P. Lacey, T. Keough, *J. Am. Chem. Soc.* **1995**, *117*, 3900-3906.

[17] PlasticsEurope, Annual report, **2017**.

[18] P. C. Painter, M. M. Coleman, *Fundamentals of polymer science : an introductory text*, 2nd ed. ed., Lancaster, **1997**.

[19] J.-F. Lutz, M. Ouchi, M. Sawamoto, T. Y. Meyer, in *ACS Symp. Ser., Vol. 1170*, **2014**, p. 313-325.

[20] J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science* **2013**, *341*.

[21] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737.

[22] F. Crick, *Nature* **1970**, *227*, 561.

[23] S. Hahn, *Nat. Struct. Mol. Biol.* **2004**, *11*, 394.

[24] D. L. Nelson, M. M. Cox, *Principles Of Biochemistry*, 5th ed., Lehninger, New York, **2008**.

[25] I. R. Lehman, M. J. Bessman, E. S. Simms, A. Kornberg, *J. Biol. Chem.* **1958**, *233*, 163-170.

[26] M. Meselson, F. W. Stahl, *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 671-682.

[27] M. B. Hoagland, P. C. Zamecnik, M. L. Stephenson, *Biochim. Biophys. Acta* **1957**, *24*, 215-216.

[28] S. Brenner, S. Benzer, L. Barnett, *Nature* **1958**, *182*, 983.

[29] C. B. Anfinsen, *Science* **1973**, *181*, 223-230.

[30] L. Pauling, R. B. Corey, H. R. Branson, *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 205-211.

[31] R. B. Merrifield, *Angew. Chem. Int. Edit.* **1985**, *24*, 799-810.

[32] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149-2154.

[33] M. Szwarc, *Nature* **1956**, *178*, 1168.

[34] M. Ouchi, T. Terashima, M. Sawamoto, *Chem. Rev* **2009**, *109*, 4963-5050.

[35] K. Matyjaszewski, J. Xia, *Chem. Rev* **2001**, *101*, 2921-2990.

[36] G. Moad, E. Rizzardo, S. H. Thang, *Polymer* **2008**, *49*, 1079-1131.

[37] J. K. Stille, *J. Chem. Educ.* **1981**, *58*, 862.

[38] W. H. Carothers, *Chem. Rev* **1931**, *8*, 353-426.

[39] P. J. Flory, *Chem. Rev* **1946**, *39*, 137-197.

[40] M. D. Schulz, K. B. Wagener,Macromol. Chem. Phys. **2014**, *215*, 1936-1945.

[41] K. B. Wagener, J. M. Boncella, J. G. Nel, *Macromolecules* **1991**, *24*, 2649-2657.

[42] M.-A. Berthet, Z. Zarafshani, S. Pfeifer, J.-F. Lutz, *Macromolecules* **2010**, *43*, 44-50.

[43] K. Satoh, S. Ozawa, M. Mizutani, K. Nagai, M. Kamigaito, *Nat. Commun.* **2010**, *1*, 6.
[44] B. S. Aitken, C. F. Buitrago, J. D. Heffley, M. Lee, H. W. Gibson, K. I. Winey, K. B. Wagener, *Macromolecules* **2012**, *45*, 681-687.
[45] A. M. Kushner, Z. Guan, *Angew. Chem. Int.Ed.* **2011**, *50*, 9026-9057.
[46] H. Mutlu, L. M. de Espinosa, M. A. R. Meier, *Chem.Soc.Rev.* **2011**, *40*, 1404-1445.
[47] L. Billiet, D. Fournier, F. Du Prez, *Polymer* **2009**, *50*, 3877-3886.
[48] C. Zhang, Q. Wang, *Macromol. Rapid Commun.* **2011**, *32*, 1180-1184.
[49] C. Zhang, J. Ling, Q. Wang, *Macromolecules* **2011**, *44*, 8739-8743.
[50] T. Yokozawa, Y. Ohta, *Chem. Rev* **2016**, *116*, 1950-1968.
[51] J. Nicolas, Y. Guillaneuf, C. Lefay, D. Bertin, D. Gigmes, B. Charleux, *Prog. Polym. Sci.* **2013**, *38*, 63-235.
[52] K. Matyjaszewski, Ziegler, M. J., Arehart, S. V., Greszta, D. and Pakula, T., *J. Phys. Org. Chem.* **2000**, *13*, 775–786.
[53] C. J. Hawker, A. W. Bosman, E. Harth, *Chem. Rev* **2001**, *101*, 3661-3688.
[54] J.-F. Lutz, *Polym. Chem.* **2010**, *1*, 55-62.
[55] G. Gody, T. Maschmeyer, P. B. Zetterlund, S. Perrier, *Nat. Commun.* **2013**, *4*, 2505.
[56] A. Anastasaki, V. Nikolaou, G. S. Pappas, Q. Zhang, C. Wan, P. Wilson, T. P. Davis, M. R. Whittaker, D. M. Haddleton, *Chemical Science* **2014**, *5*, 3536-3542.
[57] F. S. Bates, M. A. Hillmyer, T. P. Lodge, C. M. Bates, K. T. Delaney, G. H. Fredrickson, *Science* **2012**, *336*, 434-440.
[58] G. Gody, P. B. Zetterlund, S. Perrier, S. Harrisson, *Nat. Commun.* **2016**, *7*, 10514.
[59] S. Ida, M. Ouchi, M. Sawamoto, *J. Am. Chem. Soc.* **2010**, *132*, 14748-14750.
[60] D. Oh, M. Ouchi, T. Nakanishi, H. Ono, M. Sawamoto, *ACS Macro Lett.* **2016**, *5*, 745-749.
[61] D. Benoit, C. J. Hawker, E. E. Huang, Z. Lin, T. P. Russell, *Macromolecules* **2000**, *33*, 1505-1507.
[62] S. Pfeifer, J.-F. Lutz, *J. Am. Chem. Soc.* **2007**, *129*, 9542-9543.
[63] N. Baradel, S. Fort, S. Halila, N. Badi, J.-F. Lutz, *Angew. Chem. Int.Ed.* **2013**, *52*, 2335-2339.
[64] J.-F. Lutz, *Macromol. Rapid Commun.* **2017**, *38*, 1700582-n/a.
[65] J. C. Barnes, D. J. C. Ehrlich, A. X. Gao, F. A. Leibfarth, Y. Jiang, E. Zhou, T. F. Jamison, J. A. Johnson, *Nat. Chem.* **2015**, *7*, 810.
[66] S. Binauld, D. Damiron, L. A. Connal, C. J. Hawker, E. Drockenmuller, *Macromol. Rapid Commun.* **2011**, *32*, 147-168.
[67] P. H. Toy, K. D. Janda, *Acc. Chem. Res.* **2000**, *33*, 546-554.
[68] R. B. Merrifield, *Science* **1965**, *150*, 178-185.
[69] A. R. Mitchell, *J. Pept. Sci.* **2008**, *90*, 175-184.
[70] C.-D. Chang, J. Meienhofer, *Int. J. Pept. Protein Res.* **1978**, *11*, 246-249.
[71] R. Gedye, F. Smith, K. Westaway, H. Ali, L. Baldisera, L. Laberge, J. Rousell, *Tetrahedron Lett.* **1986**, *27*, 279-282.
[72] R. J. Giguere, T. L. Bray, S. M. Duncan, G. Majetich, *Tetrahedron Lett.* **1986**, *27*, 4945-4948.
[73] H. M. Yu, S. T. Chen, K. T. Wang, *J. Org. Chem.* **1992**, *57*, 4781-4784.
[74] F. Rizzolo, C. Testa, D. Lambardi, M. Chorev, M. Chelli, P. Rovero, A. M. Papini, *J. Pept. Sci.* **2011**, *17*, 708-714.
[75] F. Guillier, D. Orain, M. Bradley, *Chem. Rev* **2000**, *100*, 2091-2158.
[76] N. Badi, J.-F. Lutz, *Chem. Soc. Rev.* **2009**, *38*, 3383-3390.
[77] S. C. Solleder, R. V. Schneider, K. S. Wetzel, A. C. Boukis, M. A. R. Meier, *Macromol. Rapid Commun.* **2017**, *38*, 1600711-n/a.
[78] R. L. Letsinger, V. Mahadevan, *J. Am. Chem. Soc.* **1965**, *87*, 3526-3527.
[79] M. D. Matteucci, M. H. Caruthers, *J. Am. Chem. Soc.* **1981**, *103*, 3185-3191.
[80] S. L. Beaucage, M. H. Caruthers, *Tetrahedron Lett.* **1981**, *22*, 1859-1862.
[81] S. L. Beaucage, R. P. Iyer, *Tetrahedron* **1992**, *48*, 2223-2311.
[82] J.-F. Lutz, H. G. Börner, *Prog. Polym. Sci.* **2008**, *33*, 1-39.
[83] H. G. Börner, *Prog. Polym. Sci.* **2009**, *34*, 811-851.

[84] S. Pfeifer, Z. Zarafshani, N. Badi, J.-F. Lutz, *J. Am. Chem. Soc.* **2009**, *131*, 9195-9197.
[85] L. Oswald, A. Al Ouahabi, L. Charles, J.-F. Lutz, *Chem. Eur. J.* **2016**, *22*, 3462-3469.
[86] K. E. Geckeler, Progress in *Polymer Synthesis/Polymer Engineering*, Springer Berlin Heidelberg, Berlin, Heidelberg, **1995**, pp. 31-79.
[87] L. Andersson, L. Blomberg, M. Flegel, L. Lepsa, B. Nilsson, M. Verlander, *J. Pept. Sci.* **2000**, *55*, 227-250.
[88] E. Bayer, M. Mutter, *Nature* **1972**, *237*, 512.
[89] H. Hayatsu, H. G. Khorana, *J. Am. Chem. Soc.* **1967**, *89*, 3880-3887.
[90] A. Meszynska, N. Badi, H. G. Borner, J.-F. Lutz, *Chem. Comm.* **2012**, *48*, 3887-3889.
[91] S. P. Douglas, D. M. Whitfield, J. J. Krepinsky, *J. Am. Chem. Soc.* **1991**, *113*, 5095-5097.
[92] T. T. Trinh, C. Laure, J.-F. Lutz, *Macromol. Chem. Phys.* **2015**, *216*, 1498-1506.
[93] Y. Hao, A. D. T. L., A. Ulrich, H. Robert, *Angew. Chem. Int.Ed.* **2017**, *56*, 5040-5044.
[94] T. G. W. Edwardson, K. M. M. Carneiro, C. J. Serpell, H. F. Sleiman, *Angew. Chem. Int.Ed.* **2014**, *53*, 4567-4571.
[95] A. Al Ouahabi, L. Charles, J.-F. Lutz, *J. Am. Chem. Soc.* **2015**, *137*, 5629-5635.
[96] C. Cho, E. Moran, Cherry, J. Stephans, S. Fodor, C. Adams, A. Sundaram, J. Jacobs, P. Schultz, *Science* **1993**, *261*, 1303-1305.
[97] R. Warrass, P. Walden, K.-H. Wiesmüller, G. Jung, *Letters in J. Pept. Sci.* **1998**, *5*, 125-128.
[98] A. Boeijen, R. M. J. Liskamp, *Eur. J. Org. Chem.* **1999**, *1999*, 2127-2135.
[99] D. Jönsson, A. Undén, *Tetrahedron Lett.* **2002**, *43*, 3125-3128.
[100] N. G. Angelo, P. S. Arora, *J. Am. Chem. Soc.* **2005**, *127*, 17134-17135.
[101] N. Zydziak, W. Konrad, F. Feist, S. Afonin, S. Weidner, C. Barner-Kowollik, *Nat. Commun.* **2016**, *7*, 13672.
[102] P. R. L. Malenfant, J. M. J. Frechet, *Chem. Comm.* **1998**, 2657-2658.
[103] U. Mehmood, A. Al-Ahmed, I. A. Hussein, *Renew. Sust. Energ. Rev.* **2016**, *57*, 550-561.
[104] O. Renaudet, J.-L. Reymond, *Org. Lett.* **2004**, *6*, 397-400.
[105] D. Karamessini, B. E. Petit, M. Bouquey, L. Charles, J.-F. Lutz, *Adv. Funct. Mater.* **2017**, *27*, 1604595-n/a.
[106] D. Karamessini, S. Poyer, L. Charles, J.-F. Lutz, *Macromol. Rapid Commun.* **2017**, *38*, 1700426-n/a.
[107] K. Rose, J. Vizzavona, *J. Am. Chem. Soc.* **1999**, *121*, 7034-7038.
[108] L. Hartmann, E. Krause, M. Antonietti, H. G. Börner, *Biomacromolecules* **2006**, *7*, 1239-1244.
[109] J. W. Grate, K.-F. Mo, M. D. Daily, *Angew. Chem. Int.Ed.* **2016**, *55*, 3925-3930.
[110] B. M. Trost, *Science* **1983**, *219*, 245-250.
[111] C.-H. Wong, S. C. Zimmerman, *Chem. Comm.* **2013**, *49*, 1679-1695.
[112] M. Schelhaas, H. Waldmann, *Angew. Chem. Int.Ed.* **1996**, *35*, 2056-2083.
[113] R. N. Zuckermann, J. M. Kerr, S. B. H. Kent, W. H. Moos, *J. Am. Chem. Soc.* **1992**, *114*, 10646-10647.
[114] J. Sun, R. N. Zuckermann, *ACS Nano* **2013**, *7*, 4715-4732.
[115] R. Luxenhofer, C. Fetsch, A. Grossmann, *J. Polym. Sci. Pol. Chem.* **2013**, *51*, 2731-2752.
[116] T. D. Michelle, K. Rinki, E. B. Annelise, *Curr. Pharm. Des.* **2011**, *17*, 2732-2747.
[117] R. J. Simon, R. S. Kania, R. N. Zuckermann, V. D. Huebner, D. A. Jewell, S. Banville, S. Ng, L. Wang, S. Rosenberg, C. K. Marlowe, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 9367-9371.
[118] Á. Furka, F. SebestyÉN, M. Asgedom, G. DibÓ, *Int. J. Pept. Protein Res.* **1991**, *37*, 487-493.
[119] K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski, R. J. Knapp, *Nature* **1991**, *354*, 82.
[120] M. C. Desai, R. N. Zuckermann, W. H. Moos, *Drug Dev. Res.* **1994**, *33*, 174-188.
[121] 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.
[122] S. Martens, J. Van den Begin, A. Madder, F. E. Du Prez, P. Espeel, *J. Am. Chem. Soc.* **2016**, *138*, 14182-14185.
[123] C. J. Leumann, *Bioorg. Med. Chem.* **2002**, *10*, 841-854.

[124] G. Fiers, D. Chouikhi, L. Oswald, A. Al Ouahabi, D. Chan-Seng, L. Charles, J.-F. Lutz, *Chem. Eur. J.* **2016**, *22*, 17945-17948.
[125] T. T. Trinh, L. Oswald, D. Chan-Seng, J.-F. Lutz, *Macromol. Rapid Commun.* **2014**, *35*, 141-145.
[126] M. Porel, C. A. Alabi, *J. Am. Chem. Soc.* **2014**, *136*, 13162-13165.
[127] M. Matsugi, D. P. Curran, *Organic Lett.* **2004**, *6*, 2717-2720.
[128] M. Porel, D. N. Thornlow, N. N. Phan, C. A. Alabi, *Nat. Chem.* **2016**, *8*, 590.
[129] R. Kakuchi, *Angew. Chem. Int.Ed.* **2014**, *53*, 46-48.
[130] Passerini M. , S. L., *Gazz. Chim. Ital.* **1921**, *51*, 126–129.
[131] S. C. Solleder, M. A. R. Meier, *Angew. Chem. Int.Ed.* **2014**, *53*, 711-714.
[132] S. C. Solleder, S. Martens, P. Espeel, F. Du Prez, M. A. R. Meier, *Chem. Eur. J.* **2017**, *23*, 13906-13909.
[133] I. Ugi, C. Steinbrückner, *Angew. Chem.* **1960**, *72*, 267-268.
[134] A. Dömling, I. Ugi, *Angew. Chem. Int.Ed.* **2000**, *39*, 3168-3210.
[135] O. E. Vercillo, C. K. Z. Andrade, L. A. Wessjohann, *Org. Lett.* **2008**, *10*, 205-208.
[136] A. M. Michelson, A. R. Todd, *J. Chem. Soc.* **1955**, 2632-2638.
[137] P. T. Gilham, H. G. Khorana, *J. Am. Chem. Soc.* **1958**, *80*, 6212-6222.
[138] Y. Lapidot, H. G. Khorana, *J. Am. Chem. Soc.* **1963**, *85*, 3852-3857.
[139] Y. Lapidot, H. G. Khorana, *J. Am. Chem. Soc.* **1963**, *85*, 3857-3862.
[140] G. M. Tener, *J. Am. Chem. Soc.* **1961**, *83*, 159-168.
[141] R. L. Letsinger, V. Mahadevan, *J. Am. Chem. Soc.* **1966**, *88*, 5319-5324.
[142] R. C. Pless, R. L. Letsinger, *Nucleic Acids Res.* **1975**, *2*, 773-786.
[143] R. L. Letsinger, W. B. Lunsford, *J. Am. Chem. Soc.* **1976**, *98*, 3655-3661.
[144] R. L. Letsinger, J. L. Finnan, G. A. Heavner, W. B. Lunsford, *J. Am. Chem. Soc.* **1975**, *97*, 3278-3279.
[145] M. D. Matteucci, M. H. Caruthers, *Tetrahedron Lett.* **1980**, *21*, 719-722.
[146] V. Amarnath, A. D. Broom, *Chem. Rev* **1977**, *77*, 183-217.
[147] H. Köster, A. Stumpe, A. Wolter, *Tetrahedron Lett.* **1983**, *24*, 747-750.
[148] S. P. Adams, K. S. Kavka, E. J. Wykes, S. B. Holder, G. R. Galluppi, *J. Am. Chem. Soc.* **1983**, *105*, 661-663.
[149] M. H. Caruthers, *Journal of Biol. Chem.* **2013**, *288*, 1420-1427.
[150] Z. Jia, C. A. Bell, M. J. Monteiro, *Macromolecules* **2011**, *44*, 1747-1751.
[151] Z. Jia, C. A. Bell, M. J. Monteiro, *Chem. Comm.* **2011**, *47*, 4165-4167.
[152] H. Otsuka, K. Aotani, Y. Higaki, A. Takahara, *J. Am. Chem. Soc.* **2003**, *125*, 4064-4065.
[153] D. Yang, C. Feng, J. Hu, *Polym. Chem.* **2013**, *4*, 2384-2394.
[154] S. D. H., *J. Polym. Sci. Pol. Chem.* **2005**, *43*, 5748-5764.
[155] K. Matyjaszewski, B. E. Woodworth, X. Zhang, S. G. Gaynor, Z. Metzner, *Macromolecules* **1998**, *31*, 5955-5957.
[156] V. Percec, T. Guliashvili, J. S. Ladislaw, A. Wistrand, A. Stjerndahl, M. J. Sienkowska, M. J. Monteiro, S. Sahoo, *J. Am. Chem. Soc.* **2006**, *128*, 14156-14165.
[157] B. M. Rosen, V. Percec, *Chem. Rev* **2009**, *109*, 5069-5119.
[158] Q. Fu, W. Lin, J. Huang, *Macromolecules* **2008**, *41*, 2381-2387.
[159] Q. Fu, Z. Zhang, W. Lin, J. Huang, *Macromolecules* **2009**, *42*, 4381-4383.
[160] R. Jing, W. Lin, G. Wang, J. Huang, *J. Polym. Sci. Pol. Chem.y* **2011**, *49*, 2594-2600.
[161] W. Lin, Q. Fu, Y. Zhang, J. Huang, *Macromolecules* **2008**, *41*, 4127-4135.
[162] E. H. H. Wong, C. Boyer, M. H. Stenzel, C. Barner-Kowollik, T. Junkers, *Chem. Comm.* **2010**, *46*, 1959-1961.
[163] Andrew McAfee, E. Brynjolfsson, *Harv. Bus. Rev.* **2012**, *90*, 60-68.
[164] A. Extance, *Nature* **2016**, *537*.
[165] G. M. Church, Y. Gao, S. Kosuri, *Science* **2012**.
[166] Y. Erlich, D. Zielinski, *Science* **2017**, *355*, 950-954.
[167] H. Mutlu, J.-F. Lutz, *Angew. Chem. Int.Ed.* **2014**, *53*, 13010-13019.

[168] J. A. Amalian, A. A. Ouahabi, G. Cavallo, N. F. König, S. Poyer, J. F. Lutz, L. Charles, *J. Mass. Spectrom* **2017**, *52*, ii-ii.
[169] M. Boukhet, N. F. König, A. A. Ouahabi, G. Baaken, J.-F. Lutz, J. C. Behrends, *Macromol. Rapid Commun.* **2017**, *38*, 1700680-n/a.
[170] L. Charles, G. Cavallo, V. Monnier, L. Oswald, R. Szweda, J.-F. Lutz, *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1149-1159.
[171] T. T. Trinh, L. Oswald, D. Chan-Seng, L. Charles, J.-F. Lutz, *Chem. Eur. J.***2015**, *21*, 11961-11965.
[172] R. K. Roy, C. Laure, D. Fischer-Krauser, L. Charles, J.-F. Lutz, *Chem. Comm.* **2015**, *51*, 15677-15680.
[173] T. Maeda, H. Otsuka, A. Takahara, *Prog. Polym. Sci.* **2009**, *34*, 581-604.
[174] J.-M. Lehn, *Prog. Polym. Sci.* **2005**, *30*, 814-831.
[175] O. Bayer, *Angew. Chem.* **1947**, *59*, 257-272.
[176] D. Randall, S. Lee, *The Polyurethanes Book* **2003**.
[177] P. A. Wender, J. B. Rothbard, T. C. Jessop, E. L. Kreider, B. L. Wylie, *J. Am. Chem. Soc.* **2002**, *124*, 13382-13383.
[178] Ufuk S. Gunay, Benoît E. Petit, D. Karamessini, A. Al Ouahabi, J.-A. Amalian, C. Chendo, M. Bouquey, D. Gigmes, L. Charles, J.-F. Lutz, *Chem* **2016**, *1*, 114-126.
[179] J.-A. Amalian, S. Poyer, B. E. Petit, S. Telitel, V. Monnier, D. Karamessini, D. Gigmes, J.-F. Lutz, L. Charles, *International J. Mass. Spectrom* **2017**, *421*, 271-278.
[180] R. J. Zdrahala, I. J. Zdrahala, *J. Biomater. Appl.* **1999**, *14*, 67-90.
[181] K. Denise, S.-Y. Teresa, P. Salomé, K. Evgeniia, C. Laurence, L. Didier, L. Jean-Francois, *Angew. Chem. Int.Ed.*, *0*.
[182] M. Caruthers, *Science* **1985**, *230*, 281-285.
[183] J. A. Amalian, A. Al Ouahabi, G. Cavallo, N. F. König, S. Poyer, J. F. Lutz, L. Charles, *J. Mass. Spectrom* **2017**, *52*, 788-798.
[184] A. Al Ouahabi, J.-A. Amalian, L. Charles, J.-F. Lutz, *Nat. Commun.* **2017**, *8*, 967.
[185] D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. S. Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Pershin, J. M. Ramsey, R. Riehn, G. V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin, J. A. Schloss, *Nat. Biotechnol.* **2008**, *26*, 1146.
[186] N. F. König, A. Al Ouahabi, S. Poyer, L. Charles, J.-F. Lutz, *Angew. Chem. Int.Ed.* **2017**, *56*, 7297-7301.
[187] K. N. Felix, T. Sofia, P. Salomé, C. Laurence, L. Jean-François, *Macromol. Rapid Commun.* **2017**, *38*, 1700651.
[188] L. Jean-François, *Macromol. Rapid Commun.* **2017**, *38*, 1700582.
[189] S. A. Mcluckey, G. J. Van Berkel, G. L. Glish, *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 60-70.
[190] I. S. Young, P. S. Baran, *Nat. Chem.* **2009**, *1*, 193.
[191] J. Kulis, C. A. Bell, A. S. Micallef, Z. Jia, M. J. Monteiro, *Macromolecules* **2009**, *42*, 8218-8227.
[192] E. Rizzardo, D. H. Solomon, *Polym. Bull.* **1979**, *1*, 529-534.
[193] S. E. Averick, S. K. Dey, D. Grahacharya, K. Matyjaszewski, S. R. Das, *Angew. Chem. Int.Ed.* **2014**, *53*, 2739-2744.
[194] O. I. Kolodiazhnyi, A. Kolodiazhna, *Tetrahedron: Asymmetry* **2017**, *28*, 1651-1674.
[195] D. F. Swinehart, *J. Chem. Educ.* **1962**, *39*, 333.
[196] M. P. Reddy, P. J. Voelker, *Int. J. Pept. Protein Res.* **1988**, *31*, 345-348.
[197] B. H. Dahl, J. Nielsen, O. Dahl, *Nucleic Acids Res.* **1987**, *15*, 1729-1743.
[198] S. Kitagawa, M. Munakata, N. Miyaji, *Inorg. Chem.* **1982**, *21*, 3842-3843.
[199] R. G. Bhirud, T. S. Srivastava, *Eur. J. Inorg. Chem.* **1990**, *40*, 331-338.
[200] L. Wencheng, H. Bing, F. Qiang, W. Guowei, H. Junlian, *J. Polym. Sci. Pol. Chem.* **2010**, *48*, 2991-2999.
[201] G. Gryn'ova, C. Y. Lin, M. L. Coote, *Polym. Chem.* **2013**, *4*, 3744-3754.
[202] J. Gierlich, G. A. Burley, P. M. E. Gramlich, D. M. Hammond, T. Carell, *Org. Lett.* **2006**, *8*, 3639-3642.

[203] S. Y. R., F. G. R., L. M. P., K. Thomas, B. T. A., *Rapid Commun. Mass Spectrom.* **1994**, *8*, 77-81.
[204] G. M. Church, Y. Gao, S. Kosuri, *Science* **2012**, 1226355.
[205] F. W. McLafferty, *Anal. Chem.* **1959**, *31*, 82-87.
[206] W. F. Meek, J. B. Entrikin, *J. Chem. Educ.* **1964**, *41*, 420.
[207] L. M. Sayre, *J. Am. Chem. Soc.* **1986**, *108*, 1632-1635.
[208] E. J. Corey, G. Schmidt, *Tetrahedron Lett.* **1979**, *20*, 399-402.
[209] G. Tojo, M. Fernández, in *Oxidation of Primary Alcohols to Carboxylic Acids*, Springer, **2007**, pp. 33-41.
[210] G. Tojo, M. Fernández, in *Oxidation of Primary Alcohols to Carboxylic Acids*, Springer, **2007**, pp. 79-103.
[211] I. Paterson, M. Tudge, *Angew. Chem.* **2003**, *115*, 357-361.
[212] A. Burel, C. Carapito, J.-F. Lutz, L. Charles, *Macromolecules* **2017**, *50*, 8290-8296.
[213] K. Jeong, S. Kim, P. A. Pevzner, *Bioinformatics* **2013**, *29*, 1953-1962.
[214] J. Shendure, H. Ji, *Nat. biotechnol.* **2008**, *26*, 1135.
[215] M. Boukhet, N. F. König, A. Al Ouahabi, G. Baaken, J.-F. Lutz, J. C. Behrends, *Biophys. J.* **2018**, *114*, 182a.
[216] R. J. Wojtecki, M. A. Meador, S. J. Rowan, *Nat. Mater.* **2010**, *10*, 14.
[217] M. Herder, J.-M. Lehn, *J. Am. Chem. Soc.* **2018**, *140*, 7647-7657.
[218] C. e. Yuan, M. Z. Rong, M. Q. Zhang, Z. P. Zhang, Y. C. Yuan, *Chem. Mater.* **2011**, *23*, 5076-5081.
[219] D. Bertin, D. Gigmes, S. R. A. Marque, P. Tordo, *Macromolecules* **2005**, *38*, 2638-2650.
[220] G. Moad, E. Rizzardo, *Macromolecules* **1995**, *28*, 8722-8728.
[221] S. Marque, C. Le Mercier, P. Tordo, H. Fischer, *Macromolecules* **2000**, *33*, 4403-4410.
[222] D. Bertin, D. Gigmes, S. R. Marque, P. Tordo, *Macromolecules* **2005**, *38*, 2638-2650.
[223] D. Benoit, S. Grimaldi, S. Robin, J.-P. Finet, P. Tordo, Y. Gnanou, *J. Am. Chem. Soc.* **2000**, *122*, 5929-5939.
[224] G. Audran, P. Brémond, S. R. A. Marque, *Chem. Commun.***2014**, *50*, 7921-7928.
[225] M. Sindt, B. Stephan, M. Schneider, J. Mieloszynski, *Phosphorus, Sulfur, and Silicon and the Related Elements* **2001**, *174*, 163-175.
[226] L. Zhang, E. Laborda, N. Darwish, B. B. Noble, J. H. Tyrell, S. Pluczyk, A. P. Le Brun, G. G. Wallace, J. Gonzalez, M. L. Coote, S. Ciampi, *J. Am. Chem. Soc.* **2018**, *140*, 766-774.
[227] S. Marque, C. Le Mercier, P. Tordo, H. Fischer, *Macromolecules* **2000**, *33*, 4403-4410.
[228] Z. Zhang, A. G. Marshall, *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 225-233.
[229] W. R. Couet, R. C. Brasch, C. Sosnovsky, J. Lukszo, I. Prakash, C. T. Gnewech, T. N. Tozer, *Tetrahedron* **1985**, *41*, 1165-1172.
[230] D. J. Canney, K. D. Holland, J. A. Levine, A. C. McKeon, J. A. Ferrendelli, D. F. Covey, *J. Med. Chem.* **1991**, *34*, 1460-1467.
[231] K. P. Jayasundera, S. J. Brodie, C. M. Taylor, *Tetrahedron* **2007**, *63*, 10077-10082.
[232] P. Renton, L. Shen, J. Eckert, G. M. Lee, D. Gala, G. Chen, B. Pramanik, D. Schumacher, *Org. Process Res. Dev.* **2002**, *6*, 36-41.
[233] M. Rodríguez-Rodríguez, E. Gras, M. A. Pericàs, M. Gómez, *Chem. Eur. J.* **2015**, *21*, 18706-18710.
[234] G. V. M. Sharma, P. S. Reddy, *Eur. J. Org. Chem.* **2012**, *2012*, 2414-2421.
[235] G. Venkateswar Reddy, R. Satish Chandra Kumar, G. Shankaraiah, K. Suresh Babu, J. Madhusudana Rao, *Helv. Chim. Acta* **2013**, *96*, 1590-1600.

List of publications

- **Orthogonal Synthesis of "Easy-to-Read" Information-Containing Polymers Using Phosphoramidite and Radical Coupling Steps**
  Gianni Cavallo, Abdelaziz Al Ouahabi, Laurence Oswald, Laurence Charles, Jean-François Lutz. , J. *Am. Chem. Soc*., 2016, 138 (30), pp 9417–9420.

- **MS/MS-Assisted Design of Sequence-Controlled Synthetic Polymers for Improved Reading of Encoded Information**
  Laurence Charles, Gianni Cavallo, Valérie Monnier, Laurence Oswald, Roza Szweda, Jean-François Lutz. , *J. Am. Soc. Mass Spectrom*.; 2017, 28 (6), pp 1149-1159.

- **Controlling the structure of sequence-defined poly(phosphodiester)s for optimal MS/MS reading of digital information**
  Jean-Arthur. Amalian, Abdelaziz Al Ouahabi, Gianni Cavallo, Niklas. Felix. König, Salomé. Poyer, Jean-François Lutz, Laurence Charles. *J. Mass Spectrom.* 2017, 52 (11), pp 788-798.

- **Cleavable binary dyads: simplifying data extraction and increasing storage density in digital polymers**
  Gianni Cavallo, Salomé Poyer, Jean-Arthur Amalian, Florent Dufour, Alexandre Burel, Christine Carapito, Laurence Charles, Jean-François Lutz*. Angew. Chem. Int. Ed.*; 2018, 57,pp 6266-6269

Résumé en anglais

Uniform sequence-defined polymers were synthesized using a new iterative (AB+CD) strategy involving two orthogonal reactions. This approach allowed the protecting-group free synthesis of uniform poly(alkoxyamine phosphodiester)s. These molecules, having a defined sequence of comonomers defined as bits 0 and 1, enable the data storage of binary information at the molecular level. Interestingly, poly(alkoxyamine phosphodiester)s were found to be extremely easy to sequence. Indeed, the cleavage of the labile alkoxyamine bond in MS/MS generates “easy-to-read” fragmentation patterns. The sequencing was also tested using non-conventional techniques as fragmentation-free sequencing. Furthermore, the poly(alkoxyamine phosphodiester) backbone was modified using an extended alphabet. This optimization increases the storage capacity maintaining the read-out by MS/MS easy. Finally, two alkoxyamine bonds, having different stabilities, were inserted in the poly(alkoxyamine phosphodiester) backbone to obtain sequence-defined polymers which can be fragmented in defined poistions of the chain using different *stimuli*.

Key words: Sequence-defined polymers, digital polymers, iterative synthesis, MS/MS sequencing