UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE 222

Institut de Science et d'Ingénierie Supramoléculaires (ISIS)

THÈSE présentée par :

Thibault SCHUTZ

soutenue le : **13 décembre 2023**

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

Discipline/ Spécialité : Chimie des polymères

Design et synthèse de poly(phosphodiesters) numériques pour un séquençage facilité

THÈSE dirigée par :

M. LUTZ Jean-François	Directeur de recherche au CNRS, ISIS, Strasbourg

RAPPORTEURS :

M. WOISEL Patrice	Professeur des universités, UMET, Centrale Lille
Mme GUILLAUME Sophie	Directrice de recherche au CNRS, ISCR, Rennes

---

AUTRES MEMBRES DU JURY :

M. NIERENGARTEN Jean-François	Directeur de recherche au CNRS, LIMA, Strasbourg
Mme CHARLES Laurence	Professeure des universités, ICR, Marseille

Remerciements

Qui aurait cru qu'un élève ordinaire comme moi, candidaterait un jour pour devenir docteur en chimie ? Certainement peu de personnes. Et pourtant, après toutes ces années à Strasbourg, j'écris ces lignes afin d'espérer obtenir un doctorat. Mes études s'achèvent (ou presque) après ces trois années de thèse bien riches en émotion, autant sur le plan professionnel que personnel, et pour cela, j'aimerais remercier un bon nombre de personnes.

Tout d'abord, merci à vous, rapporteurs, Mme **Guillaume** et M. **Woisel**, et membres du Jury, Mme **Charles** et M. **Nierengarten** d'avoir accepté d'évaluer mes travaux de thèse.

Il me paraît évident de commencer par remercier aussi mon directeur de thèse, **Jean-François**, sans qui tout cela ne serait arrivé. Merci d'avoir permis mon épanouissement au sein de ton groupe. Je suis heureux d'avoir effectué ma thèse avec toi, dans ce domaine si original et captivant. Notre déménagement à l'ISIS nous a permis de mieux te découvrir en passant plus de moments conviviaux, ce qui a été un gros plus pour ma dernière ligne droite. Merci de m'avoir permis de voyager et d'aller à divers conférences (Bordeaux, Göteborg, Porquerolles) où j'ai pu faire de belles rencontres. J'ai apprécié les meetings du vendredi matin qui souvent commençaient par : « Alors le racing ? ». Enfin, je ne peux que t'être reconnaissant d'avoir rendu possible, au cours de cette thèse, ce qui restera sans doute l'un des plus beaux jours de ma vie : mon mariage.

Les collaborateurs à Marseille ont bien sûr été très importants pour l'accomplissement de cette thèse. **Laurence** qui effectue un travail d'analyse incroyable avec des rapports magnifiques, **Isaure** qui m'a simplifié la vie avec son code python pour la recherche des Tags et pour ses explications très claires sur les paramètres MS utilisés, **Jean-Louis** et **Didier** pour les discussions autour des nitroxides, puis **Kévin** dont j'ai également repris le flambeau qui avait fait du bon travail avec le RISC2.

Après le chef et les collaborateurs, viennent les collègues. Pendant trois ans, j'ai pu rencontrer un grand nombre de personne et bien-entendu au sein du même groupe. La première personne que je tiens à remercier est bien-sûr **Svetlana**, puisque c'est avec elle seule que j'ai passé l'intégrité de mon expérience. Merci beaucoup Svetlana de m'avoir supporté pendant ces années. Un jour peut-être, tes plaintes incessantes sur ton quotidien, la vie et les injustices me manqueront. Mais pour l'heure, tu as la main sur le cœur et j'étais heureux de pouvoir partager du temps avec une personne aussi adorable que toi. Parmi les autres doctorants qui me laisseront un souvenir nostalgique, j'aimerais remercier **Benoit** et **Hugo**, les PEPR *guys* qui sont également pleins de générosité. Même si vous n'étiez là qu'une année et demie avec moi, j'ai l'impression d'avoir fait beaucoup plus avec vous. La musique dans le labo, les discussions scientifiques, sport, vélo (et bien d'autres), les bières en *after work,* mais aussi Agricas et Abioco : partager avec vous ces paradis du manger me manqueront c'est certain. **Georgette**, je voulais également te remercier. Grâce à toi et ton efficacité rare pour une stagiaire, j'ai pu avancer rapidement dans une période où le temps était compté. Tu es intelligente et de bonne volonté, j'ai été très heureux de pouvoir t'encadrer. J'espère avoir pu contribuer à ton épanouissement au

sein du labo et à ton enrichissement. **Aziz**, même si malheureusement je n'ai pas eu l'occasion de travailler beaucoup avec toi, je me souviendrai longtemps de ton intelligence et tes solutions à tous problèmes. Tu m'as été d'une très grande aide. **Laurence**, merci pour ton aide, ton expertise, et la gestion du labo. Tu m'as été d'une grande aide pour la synthèse des monomères et grâce à toi, j'ai pu économiser beaucoup de temps pour mon projet. Les autres membres du groupe, d'abord à l'ICS puis à l'ISIS, ont également participé au bon déroulement de ma thèse : **Éline** qui m'a passé le flambeau, **Capucine** qui s'occupait de tout au labo, **Ian** et sa cool attitude, **Maria** avec qui j'ai pu apprendre pas mal de choses sur la synthèse d'ARN, **Paul** qui sans doute est le chimiste le plus fou et intelligent avec qui j'ai travaillé, **Sven** qui nous a rendu une petite visite du KIT, **Alex** PEPR *guy*, puis les autres : **Itab**, **Jean-Marc** à l'accueil et les débrief Racing, les membres du CMP, les membres de l'ICS avec qui j'ai pu discuter ou avoir de l'aide… Et la nouvelle génération d'étudiants du labo avec **Zhaozheng**, **Ala** et **Ranajit**, puis notre secrétaire multifonction **Corinne**, qui nous a beaucoup aidé dans notre installation à l'ISIS.

Merci à ceux avec qui je mangeais le midi, d'abord au RU à Cronembourg, on avait de la chance puisque c'était l'un des meilleurs de Strasbourg ! **Quentin**, **Alexandre**, **Souhail, Franck …** Et **Duncan** : depuis qu'on s'est connu au lycée, j'ai pas mal suivi tes traces, jusqu'au même institut ! C'était une réelle chance de t'avoir eu deux étages en dessous du miens pendant 2 ans, merci pour ton aide, tes conseils, les discussions, les cafés … J'espère te garder comme bon pote ! D'autres amis ont participé à cette thèse de prêt ou de loin, comme mes anciens camarades de licence, **Momo**, **Saifou**, **Baptiste** et **Dom, Alex** de mon Master. Merci **Stefan**, **Patrick** et **Stéphane** pour votre aide durant mon master chimie verte et mes stages, notamment à Montréal, où j'ai pu rencontrer **Léa** et **Harry**. Les potes que j'ai pu découvrir en conférence au JEPO, Bordeaux et Göteborg, particulièrement **Dijwar** et **Clémence**.

Il y a également à remercier tous ceux que j'ai pu rencontrer dans notre nouvelle aventure à l'ISIS, et que j'avais déjà pu connaître de mon Master 2 à l'ISIS 2. **Amir**, merci de m'avoir recommandé et de m'avoir permis d'autant apprendre dans ton labo. Ceci était surtout grâce à **Paulo** et **Stella**, qui m'ont beaucoup aidé lors de ce stage M2 : travailleurs, intelligents et également très fort quand il s'agit d'aller au boire un coup ! **Claudio**, **Joris**, **Emilie**, **Cyprien**, **Michel**, tous les potes du Moran's lab et aussi les potes et permanents de l'ISIS : **Jean-Louis**, **Cyril** et **Wahnyalo** et **Hava** pour les bons conseils d'analyse MS et votre gentillesse, **Arthur** et **Philippe** les MacGyvers de l'institut toujours là quand il faut.

Finalement, j'aimerais aussi remercier mes amis qui n'ont rien à voir avec cette science mais avec qui j'ai pu avoir des moments de répits et exercer mes talents de vulgarisateurs, mais aussi ma famille, **Maman**, **Papa**, mon **frère**, **Pri**, **Lilia** et **Eléna**, **Mami**, **Mémé**, (**Tati**, **Papi**, **Pépé**) qui même avec le concours de ma thèse en 180 secondes, avaient du mal à comprendre ce que je faisais. Merci pour votre soutien et d'avoir essayé de jouer le jeu avec un domaine qui est complétement nouveau pour vous. Vous m'avez permis de me sentir bien et heureux pendant cette thèse. La belle-famille aussi, **Patricia**, **Fabrice**, **Maëwen** et **Hugo**, merci à vous d'avoir été là.

Et enfin, j'aimerais remercier ma moitié, ma meilleure amie, ma pitchoun, ma femme, **Maëlys**. Merci d'avoir été là pendant cette thèse, j'ai pu sentir en toi, la pression que je ne sentais pas, ce qui pouvait me donner de temps en temps des petits coups de fouets utiles. Tu as été mon réconfort quand ça n'allait pas, tu m'as laissé travailler comme je voulais et malgré parfois, les heures tardives, tu m'as donné l'opportunité de souffler le temps de vacances ou de week-ends. Grâce à toi, je ne rentrais pas à la maison seul le soir, je savais que ma pitchoun et **Milky** m'attendaient.

J'espère sincèrement que ce manuscrit vous en apprendra davantage sur le stockage d'information à l'échelle moléculaire et vous souhaite une agréable lecture.

Comme on dit chez nous, Jetzt geht's los !

Tables des Matières

Remerciements....3
Tables des Matières....6
Liste des Abréviations....9
Liste des Figures....12
Liste des Schémas....15
Liste des Tableaux....16
Introduction Générale....17
Chapitre I.... 25
1. Contexte et nécessité d'un stockage d'information à l'échelle moléculaire....26
1.1. Vers un déferlement de données numériques dans le monde ?....26
1.2. Existe-t-il des solutions ?....27
1.2.1. Alternatives électroniques et optiques....27
1.2.2. L'ADN, un moyen de stockage dense et robuste....28
1.2.3. Stockage de données dans l'ADN, différentes preuves de concept....30
2. Stockage d'information dans des macromolécules biologiques naturelles....31
2.1. Stockage et synthèse d'information dans des systèmes biologiques....31
2.1.1. L'ADN : structure et biosynthèse....31
2.1.2. L'ARN : structure et biosynthèse....35
2.1.3. Les protéines : structures et biosynthèse....36
2.2. Synthèse chimique de macromolécules biologiques....37
2.2.1. Synthèse de polypeptides (protéines)....37
2.2.2. Synthèse d'oligonucléotides naturels....39
2.2.3. Principes de la méthode phosphoramidite....45
2.2.4. Synthèse sur micro-puces utilisant la chimie de phosphoramidite....52
2.3. Méthodes de séquençage....54
2.3.1. 1ère génération : W. Gilbert et F. Sanger....54
2.3.2. 2e génération : séquençage de nouvelle génération (*Next Generation Sequencing*, NGS)....55

2.3.3. 3e génération : séquençage en temps réel de molécules singulières 56
2.4. Limites des systèmes biologiques dans le stockage d'information 59
3. Stockage d'information dans des macromolécules synthétiques abiotiques 60
3.1. Comment stocker l'information dans des macromolécules synthétiques ? Principes et avantages 61
3.2. Les polymères à séquences contrôlée pour le stockage d'information 62
3.2.1. Approches itérative « AB+AB » : la chimie de phosphoramidite 65
3.2.2. Approches itératives orthogonales 68
3.2.3. Autres approches de stockage moléculaire 77
3.3. Méthodes de séquençage 78
3.3.1. Principe et optimisation du séquençage par spectrométrie de masse 78
3.3.2. Autres méthodes pour décoder l'information dans les polymères numériques 81
3.4. La chimie phosphoramidite comme moyen de stockage synthétique : les dernières avancées 83
3.4.1. Optimisation de séquençage à l'aide d'espaceurs 83
3.4.2. Alphabets augmentés 86
3.4.3. Limites et objectifs de la thèse 88
4. Conclusion et perspectives 90
Chapitre II 93
1. Introduction 94
2. Nécessité d'un nouveau design à partir du *RISC2* 94
2.1. Synthèse du monomère *RISC2* 94
2.2. Règles imposées par le séquençage pseudo-MS^3 pour le choix d'un marqueur d'octet. 95
2.3. Idées de conception pour un marqueur de masse clivable 96
2.4. Vérification du respect des règles de séquençage avec le nouveau design amino-TEMPO 99
2.5. Le design *TISP* : optimisation et synthèse 102
3. Incorporation du *TISP* dans une chaîne polymère contrôlée 103
3.1. Méthodes pour la synthèse de polymères numériques 103
3.2. Comment synthétiser et purifier les polymères ? 104

3.3. Incorporation du *TISP* dans la séquence 106
3.4. Séquençage du polymère : résultat et discussion 109
4. *TISP2* – synthèse et incorporation dans la chaîne polymère 112
4.1. Réduction de l'alcyne 112
4.2. Stratégie de protection/déprotection 117
4.3. *TISP2* est-il un bon choix ? Résultats et discussions 121
5. Conclusion et perspectives 123
Chapitre III. 127
1. Introduction 128
2. Synthèse du *TISP*3 128
2.1. Schéma général 128
2.2. Quels sont les marqueurs problématiques ? Discussion des rendements obtenus 130
3. Incorporation du *TISP*3 dans une chaîne polymère contrôlée 133
3.1. Essais préliminaires sur support solide d'ADN 133
3.2. Utilisation d'un support solide universel 137
4. Contrôle de l'efficacité de couplage des monomères *TISP*3 152
4.1. Méthodes et principes : quantification du cation DMT 152
4.2. Résultats et discussion 154
4.3. Tests finaux pour l'incorporation des marqueurs clivables *TISP3* en chaîne polymère 159
5. Synthèse du *TISP*4 : Recherches d'optimisation d'efficacité de couplage 161
5.1. Nouveau design : *TISP* 4 161
5.2. Efficacité de couplage du *TISP*4 164
5.3. Design *TISP*4 ou *TISP*3 ? Gain de masse et économie d'atome 165
6. Conclusion et perspectives 166
Conclusion Générale & Perspectives 169
Partie Expérimentale 177

Liste des Abréviations

A

A : Adénine
ABI : *Applied Biosystems* (nom d'un robot synthétiseur d'ADN)
CAN : Acétonitrile
ADN : Acide désoxyribonucléique
AMA : Ammonium/MethylAmine (NH_3 (aq)/$MeNH_2$ (aq))
ARN : Acide ribonucléique
ARNt : ARN de transfert
Aq : Aqueux
ATP : Adénosine triphosphate

C

C : Cytosine
CCS : *Collision Cross Section* (section efficace de collision)
CID : *Collision Induced Dissociation* (dissociation induite par collision)
Covid-19 : *Coronavirus disease 2019* (maladie du coronavirus 2019)
CPG *: Controlled Pore Glasses* (pore vitreux contrôlés)
CuAAC : *Cupper (I) Alkyne Azide Cycloaddition* (cycloaddition 1,3-dipolaire de Huisgen catalysée au Cuivre(I))

D

DBMT : Mesitylènesulfonyle-4,6-dinitrobenzyltriazole
DBU : 1,8-diazabicyclo(5.4.0)undec-7-ène
DCC : Dicyclohexylcarbodiimide
ddNTP : *Didesoxynucleotide Phosphate* (phosphate de didésoxynucléotide)
DIC : *N,N'*-diisopropylcarbodiimide
DIPEA : *N*-ethyl-*N,N*-diisopropylamine, (base de Hunig)
DMAP : *N*-methylimidazole
DMT : Di-*p*-methoxytrityl
dNTP : *Desoxynucleotide Phosphate* (phosphate de désoxynucléotide)
DOSY : *Diffusion Ordered SpectroscopY* (spectroscopie par diffusion ordonnée)
DSC : *N,N'*-disuccinimidyle
DTT : Dithiothreitol

E

ESI : *Electrospray Ionisation* (ionisation par electrospray)

G

G : Guanine
Gb : *Gigabytes* (gigaoctets)

H

HDD : *Hard Drive Disk*
HPLC : *High Performance Liquid Chromatography* (chromatographie en phase liquide à haute performance)
HRMS : *High Resolution Mass Spectrometry* (spectrométrie de masse à haute résolution)

L

L : Connecteur
LBL : *Layer By Layer* (couche par couche)
LCAA : *Long Chain Alkyl Amine* (longue chaîne alkyle amine)

M

MAC : Methoxyacétyl
MALDI : *Matrix Assisted Laser Desorption Ionisation* (désorption-ionisation laser assistée par matrice)
Me6TREN : Tris[2-(diméthylamino)ethyl]amine
MMT : Monomethoxytrityl
Mo : *Megabytes* (Mégaoctets)
MS : Spectrométrie de Masse
MsCl : Chlorure de mesitylènesuflonyle
MSNT : Mesitylènesulfonyle-3-nitrotriazole
MS/MS : Spectrométrie de masse en tandem

N

NGS : *Next Generation Sequencing* (séquençage de nouvelle génération)
NPPOC : 3'-Nitrophénylpropyloxycarbonyl
Nt : Nucléotides
NTP : Nucléoside 5’-Triphosphate

O

ONT : *Oxford Nanopore Technology*

P

PAC : Phénoxyacétyl
PCA *: Polymerase Cycling Assembly* (assemblage par cycle polymérase)
PcL : poly(acide phényllactique-co-lactique)
PCR : *Polymer Chain Reaction* (réaction de polymérisation en chaîne)
PG : *Protecting Group* (groupe protecteur)
PMDETA : N″,N″-pentamethyldiethylenetriamine
PLA : Poly-Lactide Acid (acide poly-lactique)
POS : *Phosphoramidite Oligonucleotide Synthesis*
PS : Polystyrène

R

RCA : *Rolling Circle Amplification*
RAM : *Random Access Memory*
RPE *:* Résonance Paramagnétique Électronique

S

SBS *: Sequencing By Synthesis* (séquençage par synthèse)
SEC : *Size-Exclusion Chomatography* (chromatographie d'exclusion stérique)
SMRT : *Single-Molecule Real-Time* (séquençage d'une molécule en temps réel)
SOLiD : *Sequencing by Oligonucleotide Ligation and* Detection (séquençage par ligation et detection)
SSD : Solid-State Drive (disque statique semi-conducteur)
SPS : *Solid Phase Synthesis* (synthèse sur support solide)
SPPS : *Solide Phase Peptide Synthesis* (synthèse peptidique sur support solide)

T

T : Thymine
t.a. : Température ambiante
TBAF : Fuorure de tétrabutylammonium
TBDMS : Tertiobutyldimethylsylile
TCA : Acide trichloroacétique
TEA : Triethylamine
TEMPO : 2,2,6,6-tétraméthylpipéridine-1-oxyl
TdT : *Terminal Deoxynucleotidyl Transferase*
THF : Tetrahydrofuran
TiETOS : *Template-independent Enzymatic Oligonucleotide Synthesis*
TMS : Triméthylsylile
TOF : *Time Of Flight* (temps de vol)
Trityl : Triphénylméthyl

U

UV : Ultra violet

Liste des Figures

Figure IG. 1: Principe du stockage d'information numérique à l'échelle moléculaire. 18
Figure IG. 2 : Alphabets augmentés pour une densité d'information accrue dans la chaîne polymère. 20
Figure IG. 3 : Principe de séquençage pseudo-MS^3 d'un polymère numérique. 21
Figure IG. 4 : Long polymère à haute capacité de stockage codant le portrait d'Antoine de Lavoisier 21
Figure IG. 5 : Plan général de la stratégie approchée pour les travaux de cette thèse. . 22

Figure I. 1 : Principe du stockage de données dans l'ADN. 28
Figure I. 2 : Structure de l'ADN. 31
Figure I. 3 : Principe de conservation d'information lors de la réplication de l'ADN. 32
Figure I. 4 : Biosynthèse de l'ADN par l'ADN polymérase. 33
Figure I. 5 : Principe de synthèse enzymatique de l'ADN par TdT. 34
Figure I. 6 : Principes de synthèse par assemblage de Gibson et PCA. 35
Figure I. 7 : Biosynthèse de l'ARN par l'ARN polymérase suivie de la biosynthèse de protéine grâce à la lecture de l'ARNm par le ribosome. 36
Figure I. 8 : Principe de la SPPS. 38
Figure I. 9 : Machine supramoléculaire pour la construction de polypeptides. 39
Figure I. 10 : Délocalisation des électrons du cation DMT impliquant une signature UV caractéristique. 40
Figure I. 11 : Développement des supports solides. 42
Figure I. 12 : Principe de synthèse sur micro-puces. 53
Figure I. 13 : Principe du séquençage d'ADN développé par Sanger et Guilbert. 55
Figure I. 14 : Méthode de séquençage par « amplification par pont ». 56
Figure I. 15 : Principe de séquençage par SMRT. 57
Figure I. 16 : Principe du séquençage par nanopores. 58
Figure I. 17 : Principe de stockage d'information numérique dans un polymère synthétique. 61
Figure I. 18 : Les trois différentes catégories regroupant les polymères à séquence contrôlée non naturels. 64
Figure I. 19 : **A.** Stratégies « AB + CD » ou « AA + BB » et **B.** « AB + AB ». 65
Figure I. 20 : Unités codantes utilisées dans les polyphosphodiesters numériques. 66
Figure I. 21 : Polymère à séquence définie non naturel encodant le mot « Macromolecule » à l'aide du code ASCII. 67
Figure I. 22 : Principe d'encodage en utilisant la chimie CuAAC.[16] 69
Figure I. 23 : Principe d'encodage en utilisant des poly(alcoxyamines amides).[19] 70
Figure I. 24 : Principe de synthèse d'oligomères à séquence définie combinant chimie phosphoramidite et couplages radicalaires.[305] 71
Figure I. 25 : Synthèse d'oligocarbamates à séquence définie.[308] 72
Figure I. 26 : Utilisation des thiolactones pour la synthèse de polymères à séquence définie. 74

Figure I. 27 : Réaction de Passerini couplée à la thiol-ene pour la formation d'oligomères à séquence définie. 75
Figure I. 28 : Synthèse de polyester par méthode convergente pour du stockage d'information. 76
Figure I. 29 : Représentation du séquençage MS/MS d'un tétra(alcoxyamine amide).[19] 79
Figure I. 30 : Principe d'analyse par spectroscopie de masse à mobilité ionique. 81
Figure I. 31 : Structure d'un octet fonctionnalisé d'un espaceur clivable et d'un marqueur de masse. 83
Figure I. 32 : Illustration simplifiée du principe de séquençage pseudo-MS^3. 84
Figure I. 33 : Différents modes de dissociation du phosphate lors de l'analyse pseudo-MS^3. 84
Figure I. 34 : Alphabets augmentés à 4 et 8 monomères. 87
Figure I. 35 : Nouveau design pour un marqueur de masse clivable. 89

Figure II. 1 : Tableaux expliquant la vérification des règles à respecter dans le choix d'un marqueur. 100
Figure II. 2 : Illustration des contraintes liées aux règles à respecter pour l'analyse pseudo-MS^3. 101
Figure II. 3 : Design utilisé pour l'étude d'incorporation des monomères **TISP** en chaîne polymère contrôlée. 104
Figure II. 4 : Robot synthétiseur ABI3900 utilisé dans ces travaux. *Tétrazole, Anhydride acétique et I_2. 105
Figure II. 5 : Montage de purification des polymères synthétisés. 106
Figure II. 6 : Spectres ESI-MS en mode négatif des échantillons **P1** et **P2**. 107
Figure II. 7 : Défaut de masse lorsqu'une unité **0** ou **1** est manquante. 108
Figure II. 8 Spectre ESI-MS en mode négatif de l'échantillon **P3**. 109
Figure II. 9 : Spectre MS/MS et MS^3 de **P3**. 110
Figure II. 10 : Différents états de saturation du dérivé alcyne confirmé par RMN. 113
Figure II. 11 : Spectre RMN du produit secondaire observé (haut) correspondant à une déshydratation du produit désiré (bas). 114
Figure II. 12 : Contrôle de la formation du produit secondaire par RMN à différents temps de réactions. 115
Figure II. 13 : Spectre de masse validant la nature du **TISP2**. 119
Figure II. 14 : Interprétation de l'impureté majoritaire dans l'échantillon **TISP2**. 120
Figure II. 15 : Analyses ESI-MS et MS/MS du polymère **P4** impliquant le **TISP2**. 122
Figure II. 16 : Récapitulatif des stratégies **TISP** et **TISP2** avec une perspective sur **TISP3**. *Rendement globale pouvant être amélioré comme discuté précédemment partie 4.2. **Rendement théorique en se basant sur la synthèse du **RISC2** déjà rapporté. 123

Figure III. 1 : Analyse ESI-MS et MS/MS de **P5**. 134
Figure III. 2 : Analyse MS/MS et pseudo-MS^3 de l'ion m/z = 1825.2 Da de **P5**. 135

Figure III. 3 : Analyse HPLC de **P5**. 135
Figure III. 4 : Spectre ESI(-)-MS de **P6** montrant des agglomérats de TCA (#)..... 136
Figure III. 5 : Exemple de polymère d'ADN attaché sur un support solide CPG contenant une base nucléique Thymidine, avant et après clivage/déprotection. 137
Figure III. 6 : Différents types de supports universels. 138
Figure III. 7 : Spectre ESI-MS en mode négatif de **P7**. 140
Figure III. 8 : Spectres ESI-MS en mode négatif **A. P12** et **B. P13**. 143
Figure III. 9 : Spectre ESI-MS en mode négatif de **P14**, et comparaison semi-quantitative des intensités du polymère ciblé **P14**–**P17** et de l'impureté +293 Da. 144
Figure III. 10 : Comparaison semi-quantitative des intensités du polymère ciblé **P18**–**P20** et de l'impureté +293 Da..... 145
Figure III. 11 : Spectre ESI-MS en mode négatif de **P22**, et comparaison semi-quantitative des intensités du polymère ciblé **P21**-**P22** et de l'impureté +293 Da. 146
Figure III. 12 : Spectres ESI-MS en mode négatif de l'échantillon **P27**. 148
Figure III. 13 : Spectres ESI-MS en mode négatif de **P23**. 148
Figure III. 14 : Spectres ESI-MS en mode négatif des échantillons **P28** et **P29**..... 149
Figure III. 15 : Spectres ESI-MS en mode négatif des échantillons **P30** et **P31**..... 151
Figure III. 16 : **A.** Spectres ESI-MS en mode négatif et **B.** MS/MS de l'ion $[P1 - 6H]^{6-}$ à m/z 742.8 de **P32**. 152
Figure III. 17 : Principe de quantification du sel de DMT par analyse UV permettant d'estimer l'efficacité de couplage. 153
Figure III. 18 : Comparaison d'estimation de couplage entre les positions A/T/C (**EC1**) et 5/6/7 (**EC2**). 155
Figure III. 19 : Diagramme fluidique du robot Expedite 8909.[382] 156
Figure III. 20 : Diagramme affiché sur le robot Expedite montrant les quantités de trityles estimées. 157
Figure III. 21 : Nomenclature des espaceurs testés pour leur efficacité de couplage. . 158
Figure III. 22 : Diagrammes trityles **EC1** (gauche) et **EC2** (droite). 159
Figure III. 23 : Propositions de structures identifiées par analyses MS et RMN dans les échantillons **TISP4-E1** et **TISP4-E2**. 163
Figure III. 24 : Étude de l'efficacité de couplage de l'espaceur **TISP4**. 164
Figure III. 25 : Comparaison des stratégies **RISC2**, **TISP3** et **TISP4** pour l'économie moléculaire massique. 166

Figure C&P. 1 : Principaux faits marquants de cette thèse. 171
Figure C&P. 2 : Concept général étudié pour la synthèse de poly(phosphodiesters) biohybrides utilisés pour l'auto-assemblage guidé par l'ADN. 174

Liste des Schémas

Schéma I. 1 : Couplage phosphodiester catalysé par DCC. 40
Schéma I. 2 : Approche phosphotriester. 42
Schéma I. 3 : Approche phosphite triester en utilisant la matrice inorganique à base de silice de Caruthers.[166] 44
Schéma I. 4 : Synthèse du monomère phosphoramidite. 45
Schéma I. 5 : Cycle itératif pour la synthèse d'oligonucléotides par approche phosphoramidite. 49
Schéma I. 6 : Exemple de dépurination avec la guanosine.[218] 50
Schéma I. 7 : Cycle itératif pour la synthèse de polyphosphodiesters synthétique par approche phosphoramidite. 67
Schéma I. 8 : Principe de synthèse de peptoïdes par voie supportée (Zuckerman).[300] 68
Schéma I. 9 : Différents espaceurs clivables pour optimiser l'analyse pseudo-MS^3 85

Schéma II. 1 : Synthèse du RISC2 adapté de la littérature.[355] 95
Schéma II. 2 : Proposition de synthèse d'un nouveau marqueur clivable à l'aide d'un couplage de Miyaura Suzuki. 96
Schéma II. 3 : Idée de conception de marqueurs d'octets en variant le nitroxide TEMPO. 97
Schéma II. 4 : Deux approches pour la synthèse de l'aminoalcool-TEMPO fonctionnalisé **T2**. 98
Schéma II. 5 : Exemple de série de marqueur utilisable dans une même chaîne polymère. 101
Schéma II. 6 : Approche **TISP**/**TISP2** – Sonogashira. 103
Schéma II. 7 : Mécanisme conduisant à l'élimination du TEMPO de la partie nitroxide des alcoxyamines **TISP**. 111
Schéma II. 8 : Comparaison des délocalisations électroniques entre le **RISC2** et le **TISP**. 112
Schéma II. 9 : Approche **TISP2** – stratégie protection/déprotection avec le TBDMS. .. 118

Schéma III. 1 : Synthèse du marqueur clivable **TISP3**. 129
Schéma III. 2 : Impuretés identifiées en MS correspondant à une mauvaise déphosphorylation. 141
Schéma III. 3 : Design et synthèse du **TISP4** comportant une chaîne alkyle rallongée. 162

Liste des Tableaux

Tableau I. 1 : Les principaux groupes protecteurs utilisés par approche phosphoramidite. 52

Tableau II. 1 : Différentes conditions d'hydrogénation testées pour la réduction de l'alcyne. 113
Tableau II. 2 : Comparaison de différentes conditions utilisées pour l'hydrogénation de l'alcyne **c1**..... 116
Tableau II. 3 : Différents tests d'hydrogénation réalisés sur l'alcyne protégé au DMT. 117

Tableau III. 1 : Rendements obtenus pour la préparation des différents marqueurs **TISP3**. 131
Tableau III. 2 : Conditions de clivage pour l'unysupport PS. 139
Tableau III. 3 : Tests effectués vérifiant l'implication d'un monomère trop encombré dans la mauvaise déphosphorylation du support. 142
Tableau III. 4 : Polymères à tribloc (**P23**-**P26**) et dibloc (**P27**) incorporant les différents marqueurs clivables **TISP3** (Schéma III. 1) préparés dans cette thèse. 147
Tableau III. 5 : Nombre d'injections nécessaires pour atteindre la colonne en fonction des positions. 156
Tableau III. 6 : Efficacité de couplage des différents **TISP3** (Figure III. 21). 158
Tableau III. 7 : Séquences polymères incorporant les dérivés **TISP3**. 160

Introduction Générale

L'Homme est capable de stocker et transmettre de l'information depuis des millénaires. D'abord gravée dans la roche, ensuite inscrite sur des papyrus ou du papier, et de nos jours principalement enregistrée sur des bandes magnétiques, la conservation et la transmission de l'information jouent un rôle crucial dans notre évolution. Aujourd'hui, l'ère moderne dans laquelle nous vivons nécessite plus que jamais de nouvelles solutions pour le stockage d'information. En effet, le développement des nouvelles technologies numériques génère d'énormes quantités de données (5 fois plus de données sont estimées pour 2025 par rapport à 2018).[1] Ceci s'accompagne du besoin d'un grand volume de matériaux pour fabriquer les appareils, et d'une consommation d'énergie en conséquence. Par exemple, la demande en Sillicium, élément essentiel pour construire les moyens de stockages conventionnels d'aujourd'hui, risquent de dépasser l'offre d'ici 2030.[2] La création exponentielle d'information numérique conduit également à la fabrication de nouveaux centres de données énergivores (1% de la consommation mondiale d'électricité)[3] causant des impacts environnementaux à long et court termes. Notamment, une production de gaz à effet de serre accrue, qui est une des principales causes du dérèglement climatique observé sur notre planète.[4] La recherche d'alternatives aux moyens de stockage traditionnels est par conséquent devenue un défi de taille pour les années à venir. Il existe de nombreuses possibilités en s'appuyant sur les technologies d'aujourd'hui. Particulièrement, en optimisant la densité d'information permettant une réduction du coût énergétique et matériel.[5] Toutefois, la recherche de nouvelles méthodes de conservation demeure particulièrement attractive pour satisfaire les besoins à tenir.

Dans ce contexte, en prenant exemple sur la nature et particulièrement le stockage de l'information génétique dans l'ADN, il a été démontré que l'information numérique peut être stockée dans des polymères biologiques.[2,6,7] En utilisant les différentes bases nucléiques A, T, G et C, un code binaire peut être encodé en contrôlant parfaitement la séquence de la macromolécule naturelle (Figure IG. 1). Dans ce cas, chaque monomère peut stocker 2 bits par unité (00, 01, 10, et 11).

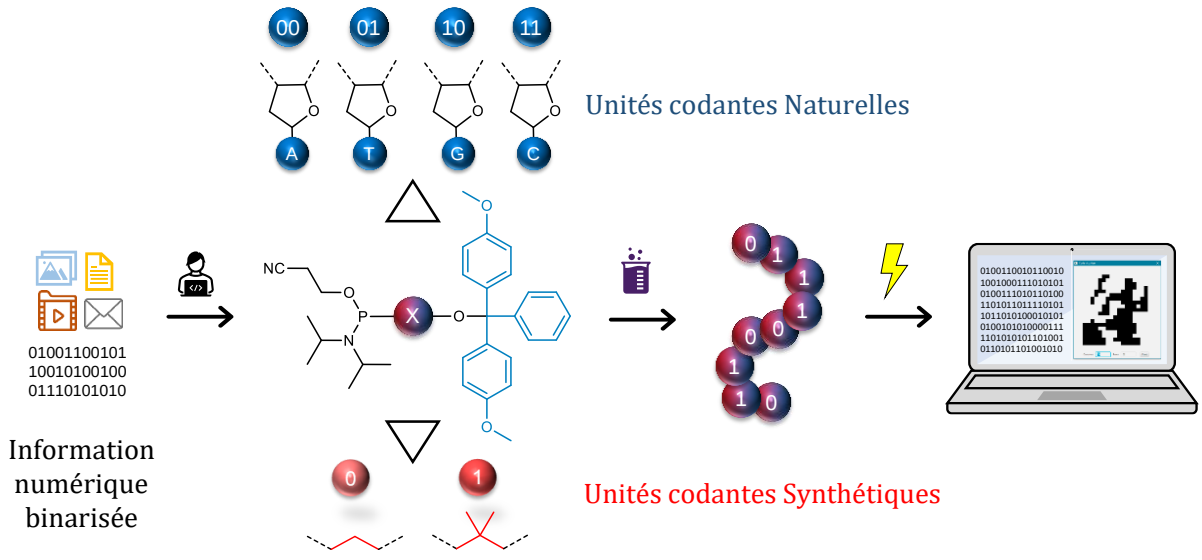


Figure IG. 1: Principe du stockage d'information numérique à l'échelle moléculaire.

La densité de stockage atteinte dans ce type de polymère est un atout considérable pour la conservation de données. En outre, la capacité de conservation théorique d'une telle macromolécule est de 5 x 10^{11} Gb par gramme,[8] ce qui permettrait, théoriquement, de stocker toutes les données de la planète dans environ 1 kg d'ADN. Par ailleurs, leur durée de vie est également un moyen de conservation attractif. Souligné par la reconstruction complète du génome d'un homme Neandertal âgé de 5300 ans,[9] l'information peut être conservée à travers des millénaires en étant stockée dans des conditions adaptées. De ce fait, des preuves de concept ont largement été démontrées pour un stockage d'information à l'échelle moléculaire.[10,11] Particulièrement en 2012, l'équipe dirigée par Church a réalisé une avancée remarquable en stockant pour la première fois un livre contenant à la fois du texte et des images sous forme de séquences d'ADN. Ces données ont pu ensuite être séquencées et récupérées avec un taux d'erreur minimal (10 bits d'erreur sur 5,27 millions).[8] Par ailleurs, les avancées récentes dans le domaine du séquençage de ces polymères, notamment à l'échelle miniaturisée, témoignent du potentiel prometteur de cette technologie de stockage. Particulièrement, les dispositifs MinION de la taille d'une clef USB qui utilisent les nanopores de protéines pour retrouver l'enchainement précis des bases nucléiques.[12]

Malgré ces différentes preuves de concept innovantes, l'utilisation de polymères biologiques pour un stockage d'information numérique peut présenter quelques contraintes. Étant limité à un alphabet à quatre monomères (ATGC), l'enchainement de répétition de bases peut être problématique,[13] et l'expansion de densité d'information limitée. Par ailleurs, la sensibilité de certaines bases à l'humidité et/ou aux irradiations UV a été démontrée, pouvant causer des erreurs d'écriture ou de lecture.[14] Par conséquent, le manque de flexibilité dans la construction de tels polymères naturels peut être compromettant pour un stockage numérique durable. Ainsi, les scientifiques se sont intéressés à d'autres polymères pour conserver les données numériques de manière pérenne, et particulièrement, les polymères synthétiques.[15]

Depuis la première preuve de concept de stockage numérique en 2014, utilisant la chimie des triazoles amides pour la construction de polymères non naturels à séquence définie,[16] de nombreux groupes de recherches ont cherché à optimiser la chimie synthétique pour fabriquer de telles macromolécules. Diverses chimies peuvent être employées à cette fin, telles que celle des thiolactones,[17] des CuAAC,[18] des nitroxides,[19] des carbamates,[20] etc. Ces différents types de stratégies emploient dans la plupart des cas des support solides, développés par Merrifield dans les années 1960.[21] D'abord très utiles au développement de la synthèse d'ADN en utilisant la chimie de phosphoramidite,[22] ces résines solides, au même titre que les supports solubles, ont été grandement utilisés pour la préparation de polymères synthétiques à séquence définie.[23] L'emploi de telles matrices inorganiques permet l'ajout consécutif de monomères, évitant des étapes de purifications fastidieuses et couteuses, augmentant ainsi les rendements. Ce processus a par la suite pu être automatisé,[24] ouvrant un grand nombre de possibilité pour la construction de polymères numériques.

En 2015, notre groupe a profité des années d'optimisation de la chimie de phosphoramidite automatisée pour construire une macromolécule synthétique (DP = 104)

en contrôlant parfaitement la séquence.[25] Cette première preuve de concept a utilisé deux monomères phosphoramidites synthétiques de masse différentes, permettant un stockage à l'échelle moléculaire sous formes de bits 0 et 1. Par la suite, la mise en œuvre d'alphabets augmentés a permis de faire croître la densité d'information dans ces macromolécules synthétiques.[26] Effectivement, avec 4 ou 8 monomères codant pour 2 ou 3 bits par unité, il est théoriquement possible d'encoder 2 à 3 fois plus d'information respectivement (Figure IG. 2). Dans ce dernier cas, il est notamment possible d'obtenir une densité supérieure à celle de l'ADN.

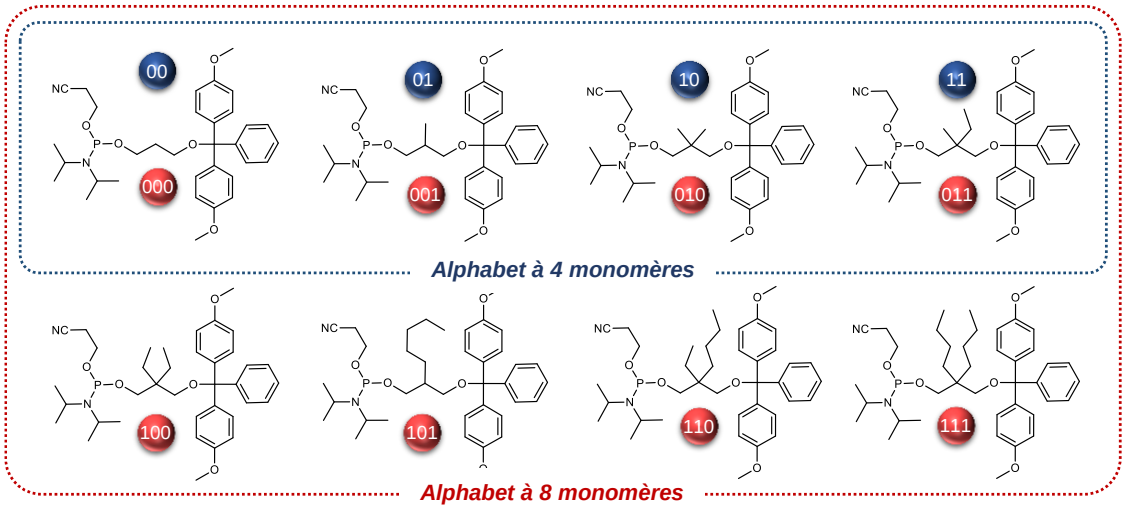


Figure IG. 2 : Alphabets augmentés pour une densité d'information accrue dans la chaîne polymère.

Toutefois, la quantité d'information stockée à l'échelle moléculaire n'est pas le seul paramètre à améliorer. Une lecture facile, rapide et efficace est essentielle pour une application durable d'un tel outil de conservation de données. En outre, dans le cas des polymères synthétiques, la méthode d'analyse la plus utilisée est la spectrométrie de masse (MS) : grâce aux masses molaires distinctes de chaque monomère, cette technique d'analyse permet le décodage de la macromolécule numérique. Cependant, l'allongement de la séquence rend difficile sa lecture, particulièrement à cause d'une augmentation des signaux.[25] En 2017, afin de faciliter leur analyse, notre groupe a utilisé des espaceurs alcoxyamines clivables (*SPACER1*, Figure IG. 3) et des marqueurs de masse différents inclus entre les sous-segments d'informations, permettant un séquençage optimisé par spectrométrie de masse en tandem (MS/MS).[27]

Grâce à leur énergie de liaison fragile, les alcoxyamines NO-C peuvent dans un premier temps être sélectivement clivés (MS^2), avant d'appliquer une énergie plus forte sur l'ion précurseur isolé (pseudo-MS^3), favorisant la dissociation des liaisons phosphates. Les derniers fragments ainsi formés permettent de retrouver l'enchainement précis des unités codantes du polymère numérique (Figure IG. 3).

Cependant, les alcoxyamines utilisés dans cette étude n'étaient pas adaptées à une lecture automatisée du spectre MS par un logiciel développé à cette fin (MS DECODER).[28] En cause, des réactions radicalaires secondaires des fragments radicaux nitroxides formés lors des analyses MS/MS. Afin de pallier ce défaut d'analyse, la

structure de l'espaceur alcoxyamine a été révisée en insérant un groupe méthyle en α et un cycle aromatique rigide en β de la liaison NO-C stabilisant le radical induit en expérience MS, et compromettant la formation de sous-produit non désiré (***RISC2***, Figure IG. 3).

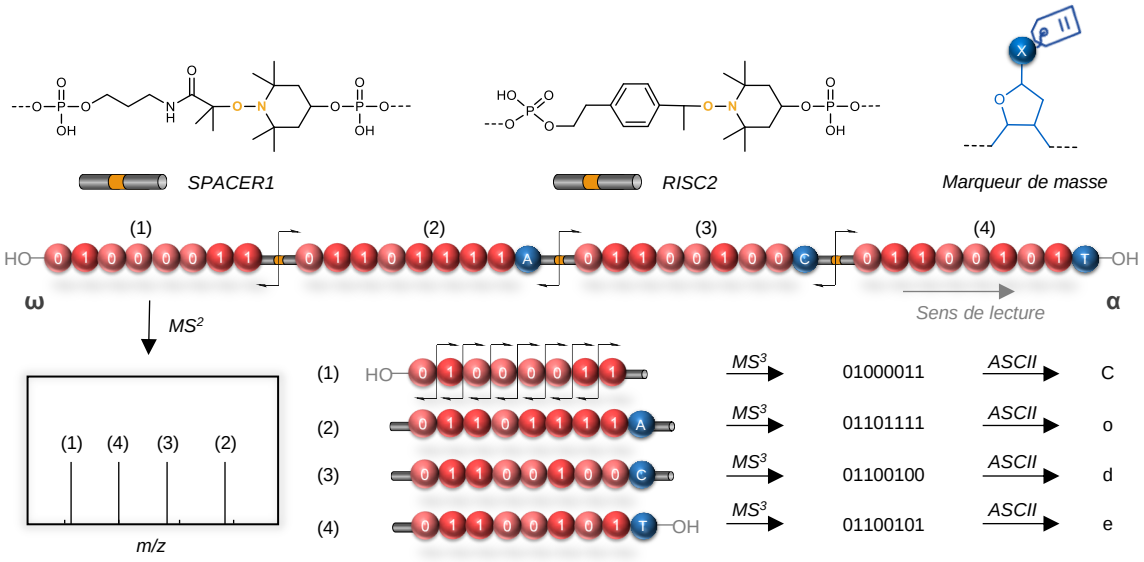


Figure IG. 3 : Principe de séquençage pseudo-MS³ d'un polymère numérique.

Afin de démontrer l'efficacité de ce nouveau design, le nouvel alcoxyamine a été employé dans la synthèse d'un long polymère comportant 11 blocs (11 octets). Chacun d'eux est séparé d'espaceurs clivable, et fonctionnalisés de marqueurs de masse permettant l'identification des fragments dans le spectre MS sans contrainte isobarique entre les segments d'information. En utilisant cette stratégie avec un alphabet à 8 monomères, le portrait d'Antoine de Lavoisier a pu être encodé et décodé automatiquement dans une seule et unique chaîne polymère avec un record de 440bits/chaîne (Figure IG. 4).[29]

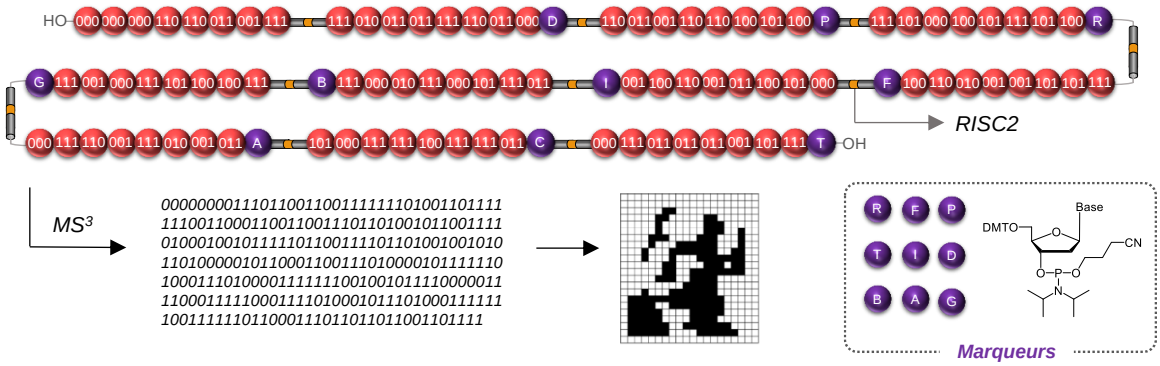


Figure IG. 4 : Long polymère à haute capacité de stockage codant le portrait d'Antoine de Lavoisier

Cependant, la construction d'un tel polymère a nécessité de nombreux éléments constitutifs et, en particulier, l'utilisation à la fois d'un espaceur alcoxyamine permettant un clivage contrôlé de la chaîne et de marqueurs de masse permettant l'identification des fragments (Figure IG. 5A et B). En outre, ces deux types de monomères sont dit « non informationnels » puisqu'ils facilitent le séquençage de la macromolécule numérique par

pseudo-MS^3, et sont inutiles pour un stockage d'informations artificielles. Par ailleurs, un des défis dans la synthèse de polymères à séquence définie est la longueur de chaîne maximale limitée par la synthèse chimique : par exemple, même avec des efficacités de couplages >99%, le rendement d'un polymère comportant 250 unités serait seulement de 6%. Afin de réduire le nombre de blocs de constructions non informationnels utilisé dans de telles synthèses, et ainsi augmenter la densité de stockage dans la chaîne polymère, un des objectifs de cette thèse a été d'élaborer et de synthétiser de nouveaux macromonomères contenant des alcoxyamines jouant à la fois le rôle d'espaceurs clivable et de marqueurs de masse (Figure IG. 5C).

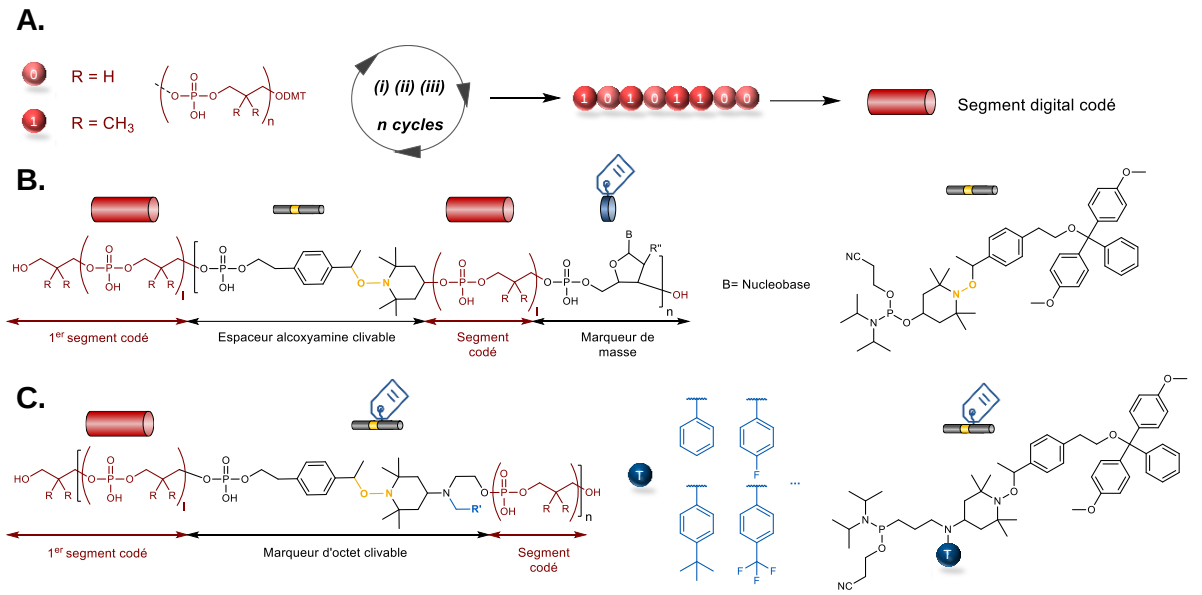


Figure IG. 5 : Plan général de la stratégie approchée pour les travaux de cette thèse.
***A.** Principe de construction de polymère numérique par chimie de phosphoramidite automatisée. i) déprotection du DMT, ii) couplage phosphoramidite et iii) oxydation. **B.** Design précédemment étudié par notre groupe. **C.** Nouveau design étudié dans cette thèse.*

Afin de développer les résultats obtenus au cours de cette thèse, ce manuscrit est divisé en 4 chapitres.

Dans le **Chapitre I**, le contexte dans lequel s'inscrit cette thèse ainsi qu'un état de l'art du stockage numérique à l'échelle moléculaire sont détaillés. Notamment, l'historique des développements réalisés dans la synthèse et le séquençage de polymères naturels, précisément l'ADN et la chimie de phosphoramidite. Après avoir démontré différentes preuves de concept dans la conservation de données de telles macromolécules, leurs limites seront exposées. Ceci permet d'identifier l'avantage d'utiliser des polymères synthétiques à séquence définie, pour une même application. Divers exemples sont donnés, employant différents types de chimie. Finalement, les développements effectués en majorité par notre groupe autour de la chimie de phosphoramidite synthétique sont montrés, introduisant les voies d'optimisation possibles, et les objectifs de cette thèse : la construction d'un marqueur alcoxyamine clivable pour un séquençage MS/MS facilité.

Le **Chapitre II** détaille les méthodes et les règles importantes à suivre pour une analyse MS optimale. Ensuite, les premiers essais réalisés pour la construction d'un espaceur

clivable sont démontrés. Une nouvelle stratégie utilisant un alcyne et un amino-TEMPO modifié est introduite (***TISP*** – ***T****ag* ***I****nside the* ***SP****acer*). Puis, l'incorporation du nouvel espaceur obtenu en chaîne polymère est présentée. Alors que l'alcyne a présenté des comportements radicalaires non désirés lors d'analyses MS/MS des macromolécules numériques, la réduction de l'alcyne a été étudiée (***TISP2***). Ces expériences peu concluantes ont conduit à tester d'autres alternatives, également discutées dans cette partie.

Le **Chapitre III** aborde une nouvelle approche pour la synthèse d'un marqueur clivable. Inspiré des travaux publiés par notre groupe (***RISC2***) et des développements effectués dans le Chapitre II, la nouvelle stratégie a été mise en œuvre efficacement (***TISP3***). Les résultats probants obtenus permettent de décrire la synthèse d'une série de douze marqueurs clivables. Parallèlement, l'utilisation de nouveaux supports solides universels pour la construction de polymères est montrée. De nombreuses optimisations ont dû être effectuées, et des résultats encourageants sont détaillés. Puis, l'incorporation des nouveaux marqueurs clivables ***TISP3*** en chaîne polymère utilisant ces nouveaux supports est exposée. Alors que certains espaceurs ne semblent pas avoir correctement été insérés dans la macromolécule numérique, l'étude de leur efficacité de couplage en chimie de phosphoramidite automatisée sera discutée. Enfin, les résultats obtenus qui ont permis d'identifier les marqueurs de masses favorables pour synthétiser des polymères numériques, et les séquencer facilement par MS/MS, sont décrits.

Cette thèse a d'abord été effectuée à l'Institut Charles Sadron de l'Université de Strasbourg au sein de l'équipe Chimie Macromoléculaire de Précision sous la direction du Dr. Jean-François Lutz (Octobre 2020 – Janvier 2023), puis à l'Institut de Science et d'Ingénierie Supramoléculaire (ISIS - Janvier 2023 – Décembre 2023) dans la même Université et sous la même tutelle. Ces travaux s'inscrivent dans le projet ANR « Shapenread » en étroite collaboration avec l'équipe du Prof. Laurence Charles à l'Institut de Chimie Radicalaire de l'Université d'Aix-Marseille. Toutes les données MS et MS/MS montrées dans ce manuscrit sont le fruit de cette collaboration. Le projet ANR a pour objectif d'optimiser le séquençage de polymère numérique par MS, de manière rapide et efficace. Pour ce faire, la technologie IMS (spectrométrie de mobilité ionique) a été identifiée comme étant prometteuse. Alors que le design antérieur nécessite une expérience MS/MS pour lire l'information contenue dans chaque segment d'information, l'IMS permet de séparer les différents octets et de les fragmenter séparément en une seule et unique expérience. Les espaceurs conçus et démontrés dans cette thèse ont pour but de réduire le nombre de monomère non informationnels dans la chaîne polymère, mais aussi de faciliter la séparation des segments d'information par IMS. Les tests effectués pour de telles analyses ont été réalisé par Isaure Carvin dans le cadre de sa thèse à Marseille. Les résultats trop préliminaires acquis jusqu'aujourd'hui ne permettent pas une discussion détaillée, et ne sont donc pas montrés dans ce manuscrit.

Chapitre I.

État de l’art : stockage de l’information à l’échelle moléculaire

1. Contexte et nécessité d'un stockage d'information à l'échelle moléculaire

1.1. Vers un déferlement de données numériques dans le monde ?

L'ensemble des données numériques mondiales (toutes les données numériques créées, capturées ou encore répliquées) connaît une croissance exponentielle selon *International Data Corporation* (IDC) et devrait passer de 33 ZB en 2018, à 175 ZB en 2025 (1 ZB = 10^{12} GB).[1] Malgré le fait que ces données soient en grande partie créées pour à posteriori être consommées et supprimées en l'espace de quelques millisecondes, il existe une demande de capacité de stockage croissante importante.

Depuis des millénaires, l'Homme n'a cessé de vouloir stocker et transmettre de l'information. Les premières traces de communication et de support de stockage remontent à la période du Néolithique il y a plus de 6000 ans, où la communication se faisait à travers des outils et des dessins. Depuis les premières traces d'écritures apparues pendant la Haute Antiquité il y a plus de 3000 ans, les techniques que connaissaient autre fois les Mésopotamiens ont considérablement évoluées.[30] En passant des dessins de caverne, au papyrus puis aux premières bandes magnétiques comme stockage d'information, les hommes ont perpétuellement essayé d'innover et de trouver de meilleures technologies. A la fin du 20e siècle, les appareils de stockages les plus répandus étaient les bandes magnétiques, et les appareils optiques (CD et DVD). Aujourd'hui, les plus utilisés sont les semi-conducteurs (RAM et SSD). Ces derniers sont composés en majorité de Silicium et, malgré son abondance à la surface de notre planète (2e élément le plus abondant), un épuisement des ressources pourrait se faire ressentir d'ici quelques dizaines d'années. En effet, d'après des estimations réalisées par W. L. Hughes et ses collègues,[31] les besoins en Silicium seraient de l'ordre de 2.4 x 10^{9} kg en 2040, alors que l'estimation d'offre de ce même élément avoisinerait 2.4 x 10^{7} kg pour la même année. La demande devrait donc selon ces évaluations dépasser l'offre d'ici moins de vingt ans. Une autre problématique de ces appareils de stockage est leur composition en éléments rares. Bien que des efforts soient mis en place pour pallier au manque de ressources (e.g. recyclage),[32–34] répondre à la demande émergente de certains métaux reste un enjeu.[35,36] Ces éléments de terre rare se trouvent en petite concentration sur notre planète et entrainent des procédés d'extraction dangereux pour l'Homme et l'environnement.[37,38] Ils peuvent également générer des conflits géopolitiques importants du fait de leur répartition inégale dans le monde (ex. la Chine est le premier producteur mondial en fournissant plus de 90% des éléments se trouvant sur Terre d'après United States Geological Survey).

Cette croissance de données numériques conduit également à la construction de nombreux centres de données dans le monde (*Datacenters*). Ces bâtiments érigés exclusivement pour stocker de l'information numérique à grande échelle posent des problèmes environnementaux majeurs.[39] En complément de l'espace et des matières premières nécessaires à leur construction et leur fonctionnement, ces centres posent un problème de consommation d'électricité. À eux seuls, ils consomment 200 TWh par an,

soit 1% de la demande d'électricité mondiale. De plus, afin de refroidir les appareils, une immense quantité d'eau est indispensable : il a été estimé que les centres de données américains ont utilisés environ 100 milliards de litres d'eau en 2014.[3] D'après *Cloudscene*, il y aurait plus de 8500 centres de données répartis dans le monde.[40]

1.2. Existe-t-il des solutions ?

Après avoir abordé les problématiques du stockage de données numériques dans les années à venir, il paraît évident que la recherche d'alternatives aux méthodes de stockage conventionnelles est devenue une priorité. Deux principaux types de solution existent pour y remédier : des moyens plus classiques de stockage en se basant sur les technologies déjà existantes telles que le codage par différence de champ électrique ou diffraction de lumière (1.2.1), et une nouvelle technologie émergente qui n'était pas exploitée il y a quelques dizaines d'années auparavant, le stockage à l'échelle moléculaire : l'acide désoxyribonucléique (ADN) (1.2.2).

1.2.1. Alternatives électroniques et optiques

Dans les disques durs conventionnels (HDD ou *Hard Disc Drive*), un plateau en rotation recouvert d'une fine couche magnétique sert à stocker l'information sous forme de code binaire (0 et 1). Une "tête" se déplace sur le plateau inscrivant le code sous forme de minuscules zones de Nord ou de Sud magnétique sur ce dernier. Pour relire les données, la tête se rend au même endroit et interprète les 0 et 1 initialement stockés.[41] Pour les semi-conducteurs (RAM ou *Random Access Memory*), l'information doit être accessible rapidement et est donc stockée de manière temporaire en se basant sur la charge électrique de l'électron. Un voltage faible codera pour 0 tandis qu'un voltage plus élevé codera pour 1.[42] Cependant, plus la demande de stockage est élevée, plus ces dispositifs nécessitent une large quantité de micro-constituants (micro-processeurs, puces, transistors, etc.). Cela freine ainsi la capacité de ces appareils à évoluer dans leur rapport taille/efficacité.

Pour pallier ces limites, la spintronique s'est avérée être une nouvelle classe de moyen de stockage intéressante. Alors que les semiconducteurs conventionnels n'utilisent que la charge électrique de l'électron pour fonctionner, la spintronique repose sur l'utilisation du spin de l'électron et de sa charge négative pour transporter et stocker des informations sous forme de bit 0 (spin up) et de bit 1 (spin down). Elle combine les propriétés des semi-conducteurs basées sur la charge (circuits intégrés) et les propriétés magnétiques basées sur le spin (disque dur de stockage) dans une seule puce afin d'effectuer tous les processus dans un seul dispositif.[42]

Afin d'améliorer les propriétés et l'efficacité des technologies bien connues, plusieurs groupes de recherches se sont également intéressés aux bénéfices de remplacer l'air par de l'hélium dans les disques durs. Dans ce cas, la trainée de la « tête » sur les plateaux en rotation est réduite, ce qui permet également de réduire la consommation d'énergie, d'améliorer la circulation de l'air et de faire tourner les plateaux plus rapidement.[43]

D'autres innovations dans ce domaine ont émergées ces dernières années, telles que l'utilisation de nanomatériaux (nanoparticules de métaux, oxide de graphène, etc.),[5] ou

encore le stockage optique en 3[44,45] et 5[46] dimensions permettant de repousser les limites de densité de stockage.

Beaucoup d'efforts sont investis dans la recherche d'alternatives pour un stockage de données durable et efficace mais, dans la plupart des cas, la technologie n'est encore qu'au stade de développement et n'est pas encore sur le point d'être industrialisée. Bien que ce soit une énième preuve de concept, le stockage de donnée à l'échelle moléculaire, en particulier dans l'ADN, est également une piste à considérer. Cette voie semble être plus accessible que d'autres stratégies du fait des propriétés attractives et du grand nombre d'études concernant la synthèse et la caractérisation de cette macromolécule depuis la découverte de sa structure en 1953.[47]

1.2.2. L'ADN, un moyen de stockage dense et robuste

L'acide désoxyribonucléique stocke l'information génétique grâce à sa structure primaire en utilisant un code à 4 unités, les 4 bases nucléiques de l'ADN : Adénine, Thymine, Guanine et Cytosine (A, T, G et C respectivement). Ce principe peut également être utilisé pour stocker de l'information numérique. En effet, en exploitant ces bases naturelles comme étant un alphabet à 4 variables, un code peut être défini sur le même principe que le langage binaire utilisé en informatique (e.g. A=00, T=01, C=10 et G=11, Figure I. 1).

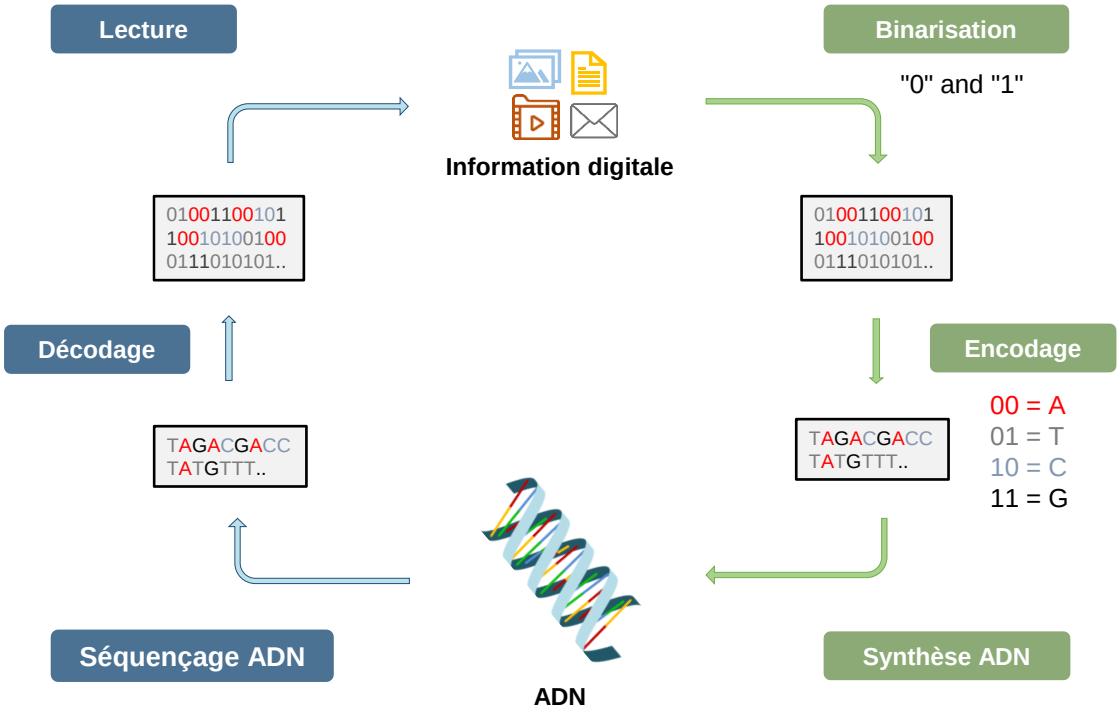


Figure I. 1 : Principe du stockage de données dans l'ADN.

Un code ASCII (American Standard Code for Information Interchange) est une norme de codage de caractères qui est généralement utilisée pour encoder de l'information en informatique. Une suite de 8 bits (= un octet) permet de traduire un caractère parmi les 256 présents dans cet alphabet en code binaire. Ceux-ci comprennent non seulement l'alphabet, mais aussi les chiffres, les ponctuations et les caractères spéciaux.

Cette alternative aux moyens de stockage conventionnel présente une multitude d'avantages dont le plus significatif est le gain d'espace. En effet, l'information encodée dans les bases nucléiques de l'ADN possède une densité de stockage volumétrique dépassant de loin les mémoires électroniques actuelles. Par exemple, le génome humain peut stocker l'équivalent d'1 Gb d'information par cellule. Par ailleurs, en stockant 2 bits par base nucléique et en se basant sur un seul brin, l'ADN bénéficie quant à lui d'un stockage théorique maximal de 5×10^{11} Gb par gramme.[8] Si une telle densité pouvait être atteinte dans des systèmes de stockage d'information, tous les besoins en stockage estimés pour 2040 pourraient tenir dans seulement 1 kg d'ADN, ce qui répondrait aux problèmes d'offres de métaux indispensables pour les technologies actuelles, telles que le Silicium, évoqué dans la sous partie précédente.[31]

L'avantage de stocker l'information dans des polymères d'ADN n'est pas seulement le gain de densité et le remplacement de certains métaux (Silicium). Avec l'augmentation de la production de donnée, nous avons pu voir précédemment (1.1) que la consommation électrique mondiale globale était affectée. Le stockage dans des macromolécules biologiques permettrait de réduire considérablement les coûts énergétiques. En effet, contrairement aux *datacenters* qui ont besoin d'être refroidis pour ne pas surchauffer, les polymères biologiques peuvent être stockés dans des conditions particulières sans demande énergétique supplémentaire. Pour ce faire, plusieurs façons de stocker l'information de manière efficace existe.[48] Pour un stockage à long terme avec un accès limité (1 fois tous les dix ans), l'encapsulation dans des particules de Silice est à ce jour la meilleure option. En théorie, grâce aux travaux de Grass et ses collègues, il a pu être estimé qu'en utilisant l'encapsulation, il serait possible de maintenir l'information intacte pendant 20 à 90 ans à température ambiante, 2000 ans à 9.4°C et jusqu'à 2 millions d'années à -18°C.[14] Pour un accès plus fréquent (plusieurs fois par an), l'ADN lyophilisé reste la meilleure option, même si son stockage en milieu aqueux présente également de bons résultats.[49] Toutefois, ces méthodes de stockage ne permettent de préserver l'information que pendant quelques années (3 à 4 ans).[50] Il est nécessaire de souligner que ceci reste des estimations sur les bases de ce qu'il a pu être observé en laboratoire. En effet, pour tester la robustesse de ces échantillons, des vieillissements artificiels sont réalisés. Par conséquent, des modèles doivent encore voir le jour pour corréler de façon précise ces résultats à la réalité, dans des conditions optimales.[48]

Au-delà des problèmes environnementaux et techniques, la durée des moyens de stockage actuels est une autre difficulté à surmonter. Alors que celle-ci ne dépasse pas 50 ans,[51] l'ADN se présente quant à elle comme une solution efficace pour stocker de l'information à long terme. Comme souligné par les découvertes d'Otzi, l'Homme glace âgé de 5300 ans :[9] son génome a pu être séquencé à 100% permettant de révéler son groupe sanguin, la couleur de ses yeux, et son intolérance au lactose.[52] Ces précisions ont pu être obtenues grâce à son ADN stocké de façon intacte dans son corps pendant toutes ces années. Cela témoigne qu'un isolement physique et chimique de l'ADN empêche toutes contraintes de dégradation (ici par le froid), et permet de conserver l'intégrité des informations pendant des milliers d'années.[53]

1.2.3. Stockage de données dans l'ADN, différentes preuves de concept

Cette idée de stocker de l'information numérique dans l'ADN et plus généralement à l'échelle moléculaire ne date pas d'hier. Cela a été suggéré pour la première fois en 1959 par l'américain Richard Feynman (Prix Nobel de Physique en 1965) lors de sa conférence visionnaire « *There's Plenty of Room at the Bottom* ».[54] C'est presque 30 ans plus tard (en 1988), avec *the Microvenus project*, que le premier message stocké dans de l'ADN a vu le jour.[55] En 2012, le groupe de Georges Church a réalisé une prouesse encore plus aboutie en stockant pour la première fois un livre contenant du texte et des images dans différentes séquences d'ADN. Les informations stockées ont pu à posteriori être séquencées et identifiées avec un taux d'erreur minime (10 bits sur 5.27 millions).[8] En 2013, Goldman et ses collègues ont réussi à repousser les limites en stockant 739 kb de fichiers informatiques dans différents fragments d'ADN. Cela est possible grâce au recouvrement de biopolymères de 100 unités, qui se chevauchent sur 75 unités, créant une redondance quadruple.[56] Parallèlement à ces études, l'entreprise Microsoft a également effectué des recherches autour de ces nouvelles stratégies de stockage. En 2018, Karin Strauss, Luis Ceze et leurs collègues ont démontré qu'il était possible d'effectuer un accès aléatoire à grande échelle, en stockant et récupérant individuellement 35 fichiers distincts (200 Mb de données) dans plus de 13 millions d'oligonucléotides d'ADN.[57]

Dans la continuité de cette technologie jaillissante, afin de faire preuve de concept et démarrer une révolution dans le stockage d'information, le CNRS, en collaboration avec les archives nationales de France, a stocké les déclarations des droits de l'Homme et du citoyen de 1789, puis celles des femmes de 1791, dans de l'ADN. C'est sous forme de capsules de métal que ces deux documents historiques figurent depuis décembre 2021 dans la célèbre armoire de fer aux côtés d'objets inestimables.[6] Pierre Crozet et Stéphane Lemaire ont développé une technologie brevetée appelée *DNA Drive* qui repose sur une « stratégie bio-inspirée d'archivage, en rupture avec les technologies existantes, puisqu'elle utilise des technologies biologiques et n'est pas basée sur des oligonucléotides ». Sachant que cette technologie peut être *in fine* décodée par des séquenceurs portables,[12] ce succès semble être une solution non négligeable pour le stockage du futur.

Bien que ces différentes preuves de concept montrent à quel point le stockage dans des macromolécules peut-être une solution révolutionnaire pour le futur, il est important de souligner que ces méthodes ne sont pour le moment adaptées qu'à un seul type de stockage : les données « froides ». Contrairement aux données « chaudes », les données froides font références aux données qui ne sont que très rarement consultées. Là où les données chaudes doivent être accessibles de manière fréquente et efficace pour être lues, copiées ou modifiées, les données froides sont généralement consultées toutes les dizaines d'années pour lecture. Sachant que ce type de données représente environ 60% des données dans le monde,[58] une alternative aux moyens de stockage conventionnels orientée sur la lecture exceptionnelle de ces données reste néanmoins un challenge louable. La prochaine partie détaillera les avancées réalisées dans la synthèse de

macromolécules biologiques ainsi que leur séquençage, pour une potentielle application dans le stockage de données numériques.

2. Stockage d'information dans des macromolécules biologiques naturelles

Depuis la mise en évidence de la structure de l'ADN par Watson et Crick en 1953,[47] de nombreux travaux ont vu le jour pour tenter de reproduire ce que la nature a réussi à réaliser dans les systèmes biologiques depuis des millénaires.[59] En utilisant des voies de synthèse biologique ou chimique, des brins d'ADN peuvent être créés, copiés et lus en quelques heures seulement. L'ADN permet donc de stocker de l'information de manière naturelle ou non, tout comme l'ARN et les protéines. Dans cette deuxième partie, les moyens de stockage dans les macromolécules biologiques naturelles (2.1), leurs synthèses (2.2), leurs caractérisations (2.3) et leurs limites (2.4) seront discutées.

2.1. Stockage et synthèse d'information dans des systèmes biologiques

2.1.1. L'ADN : structure et biosynthèse

Depuis que la vie a apparue sur Terre, l'information génétique n'a cessé d'être présente dans les êtres vivants. Il a fallu attendre le 20e siècle (1953) et les travaux de Franklin,[60] suivit des travaux de Wilkins[61] pour observer les premières photos à rayon X de l'ADN et faire la première hypothèse d'une structure hélicoïdale.

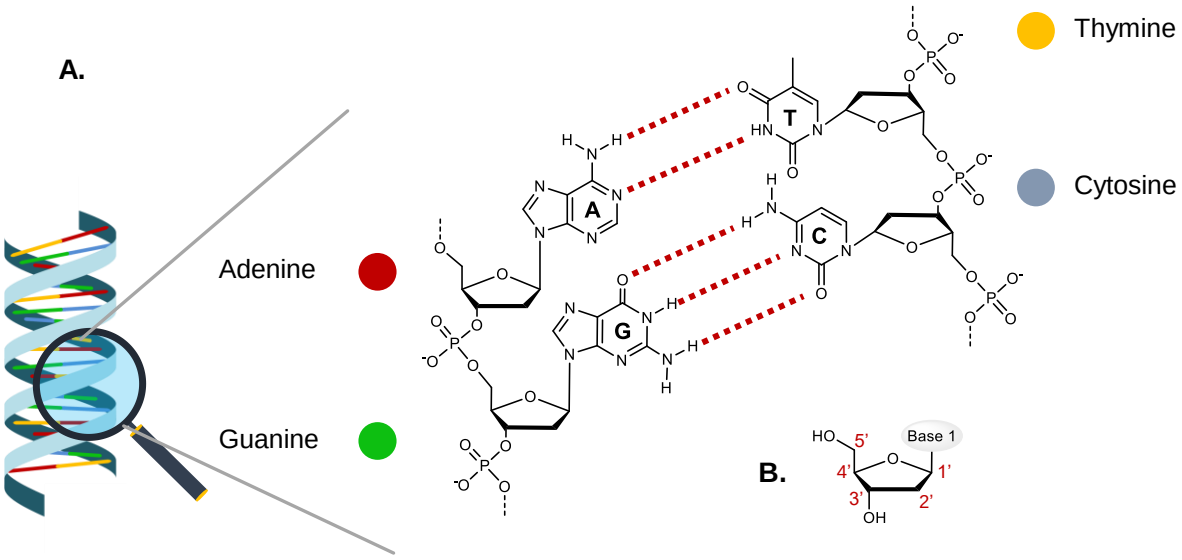


Figure I. 2 : Structure de l'ADN.
***A.** Structure primaire : liaisons phosphodiesters. Structure secondaire : liaisons hydrogènes (rouge).*
***B.** Nomenclature définissant les différentes positions des carbones de la base nucléique.*

La même année, Watson et Crick ont proposé la fameuse structure en double hélices de la macromolécule.[47] L'ADN est un biopolymère à séquence définie constitué d'un enchaînement de nucléotides (Adénine, Thymine, Guanine, Cytosine) permettant de stocker l'information génétique. Des affinités particulières entre A-T puis G-C émanent par

la formation préférentielle de liaisons hydrogènes entre elles (Figure I. 2). Ceci détermine la structure secondaire de l'ADN.

Malgré cette découverte, ce n'est que quelques années plus tard (1958) que cette dernière a réellement été confirmée grâce aux travaux de Meselson et Stahl.[62] Leur expérience décrite comme étant la plus belle de l'histoire dans le domaine de la biologie[63] a pu mettre en lumière la réplication de l'ADN. En assumant que lors de la division cellulaire, chaque cellule fille doit contenir la même information génétique que celle codée dans la cellule mère, lors de la réplication, une copie parfaite du complément génomique de l'ADN devrait être générée. En utilisant la différence de densité entre l'isotope N^{14} et N^{15}, l'ADN marqué à l'azote lourd peut-être distingué de celui non marqué par technique de sédimentation. En faisant profit de cette technique, Meselson et Stahl ont pu conclure que « l'azote d'une molécule d'ADN est réparti à parts égales entre deux sous-unités qui restent intactes pendant de nombreuses générations »[62] (Figure I. 3).

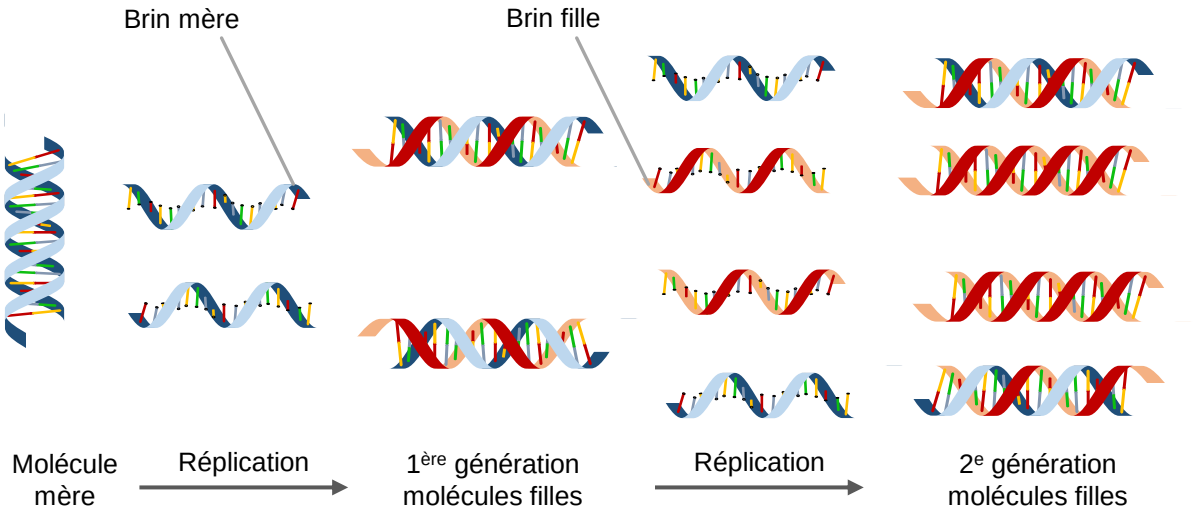


Figure I. 3 : Principe de conservation d'information lors de la réplication de l'ADN.

Lors de cette même période (1956), le biologiste A. Kornberg a démontré pour la première fois qu'il est possible d'isoler une enzyme capable de répliquer l'ADN et ainsi de le synthétiser : l'ADN polymérase. Après avoir extrait cette enzyme de la bactérie *Escherichia Coli*, Kornberg et ses collègues ont développé un protocole rendant possible la biosynthèse de l'ADN.[64] Le processus de réplication consiste à reproduire deux molécules d'ADN identiques à partir d'une molécule mère (Figure I. 4).[65] Premièrement, la double hélice s'ouvre grâce à une protéine (hélicase) permettant ainsi de séparer les 2 brins : un brin direct et un brin indirect. Pendant ce temps, une autre enzyme (primase) s'attache brièvement sur chaque brin et assemble une base sur laquelle la réplication peut commencer (amorçage) : l'ADN polymérase intervient en s'enroulant autour du brin direct et en catalysant la formation de liaisons phosphodiesters entre ses bases complémentaires dans le sens 5' → 3'. Le brin indirect quant à lui ne peut pas bénéficier normalement de l'ADN polymérase puisqu'il se trouve dans le sens inverse. La primase amorce alors la synthèse à plusieurs reprises en ajoutant quelques bases sur ce dernier

permettant ainsi la réplication dans le bon sens. Ce processus doit se produire de façon répété donnant ainsi naissance à des fragments : les fragments d'Okasaki.[66]

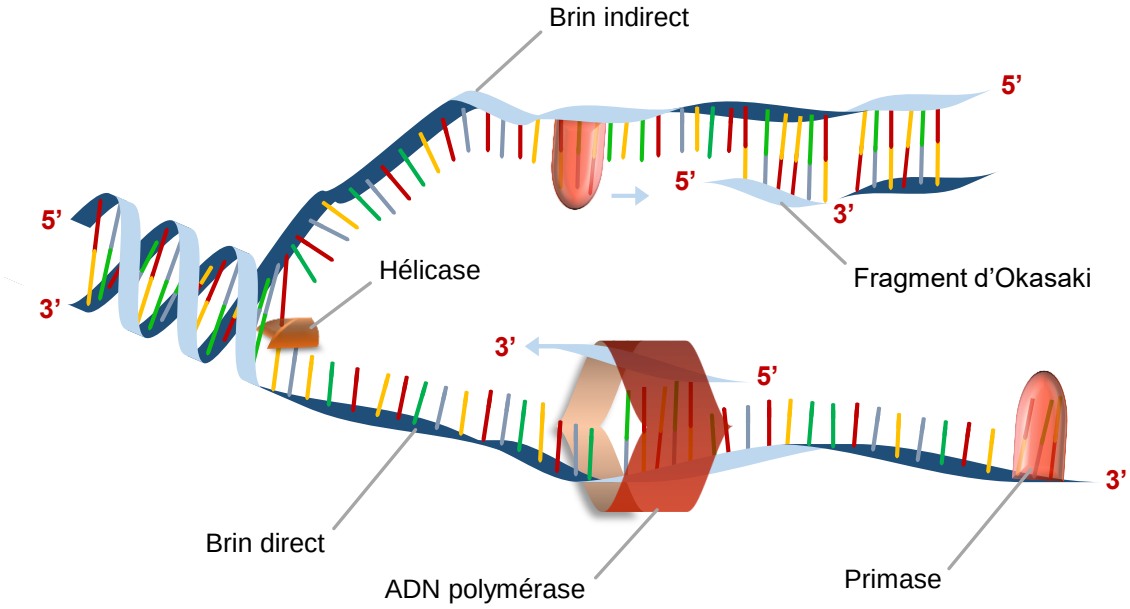


Figure I. 4 : Biosynthèse de l'ADN par l'ADN polymérase.

L'ADN est aussi connu pour pouvoir être amplifié, permettant ainsi d'obtenir des millions de copies en quelques heures seulement.[67] Ce processus a été découvert et développé par K. Mullis en 1983 avec la technique PCR (*polymer chain reaction*).[68] Ce procédé biochimique est composé de 3 étapes : (1) la dénaturation, dans laquelle le double-brin modèle d'ADN va être chauffé pour séparer les 2 brins ; (2) la recuisson, lorsque les chaînes courtes qui servent alors d'amorceurs se lient à des chaîne latérales voisines et, (3) l'extension, au cours de laquelle l'ADN polymérase étend la chaîne de chaque modèles précédemment formés. Cette technologie qui lui a valu le prix Nobel de Chimie en 1993[69] a grandement été utilisée pour produire de l'ADN dans les domaines de la chimie et biologie.

Depuis cette découverte, différentes méthodes de synthèse biologique ont émergé.[10] La synthèse enzymatique utilise le principe de synthèse sur support solide : une courte chaîne d'ADN est synthétisée sur le support et peut ensuite être étendue avec l'ADN polymérase en utilisant les nucléosides 5'-triphosphates (NTPs). Ce processus est efficace pour amplifier de l'ADN déjà existant, mais est inutile pour fabriquer de nouvelles chaînes. C'est pourquoi, une autre enzyme doit servir d'alternative lorsque le modèle de synthèse (généralement implanté par la primase) est absent. Une telle polymérase a été identifiée par Bollum en 1959 : *terminal deoxynucleotidyl transferase* (TdT).[70] TiEOS (template-independent enzymatic oligonucleotide synthesis) est l'acronyme donné au type de synthèse exploitant l'enzyme TdT.[71] Elle permet de catalyser l'addition de nucléotides sur l'extrémité 3' d'une molécule d'ADN sans nécessité de modèle (Figure I. 5). Cette technologie est l'une des plus utilisées aujourd'hui, notamment à l'échelle industrielle (*DNA script*,[72] ANSA Biotechnologies,[73] etc.).

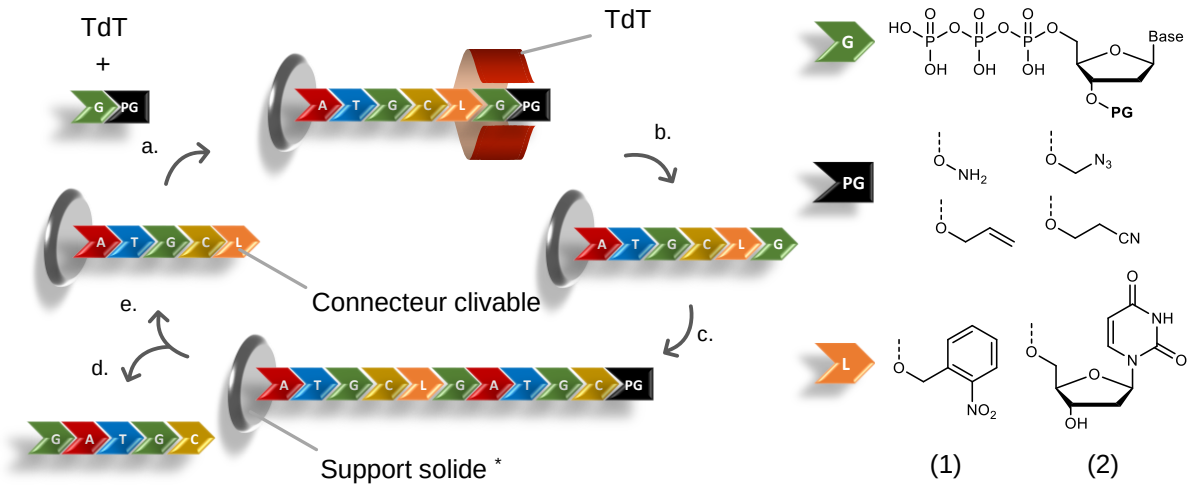


Figure I. 5 : Principe de synthèse enzymatique de l'ADN par TdT.
** Support solide (résine) fonctionnalisé d'un brin d'ADN servant d'initiateur. Un espaceur clivable en bout de chaîne (brique orange « L ») permet de récupérer le polymère désiré en fin de cycle après clivage. La synthèse s'effectue en plusieurs parties : (a) couplage avec un 3'-protégé (brique noire « PG »)[74] NTP (dans cet exemple, brique verte « G ») en présence de TdT ; (b) déprotection de PG ; (c) étapes a. et b. répétées n fois ; (d) clivage (peut généralement être clivé photo-chimiquement[75] en utilisant un groupe nitrobenzyl (1) ou bien enzymatiquement[10] à l'aide de l'uracile ADN glycosylase lorsque la deoxyuridine (2) est utilisée ; (e) l'initial brin d'ADN reste sur le support solide.*

Il existe également d'autres méthodes de construction d'ADN, cette fois-ci basées sur l'assemblage de molécules. Contrairement à la synthèse de simple brin qui est limitée à une longueur d'environ 200 nucléobases par chaînes,[10] l'assemblage permet en théorie d'accéder à des longueurs illimitées de chaîne d'ADN. Deux principaux types d'assemblage sont utilisés : l'assemblage de Gibson[76] et le PCA (*Polymerase Cycling Assembly*)[77] (Figure I. 6). Les deux méthodes sont deux approches enzymatiques complétant les méthodes de synthèse classique d'oligonucléotides par chimie phosphoramidite (qui sera discutée plus tard partie 2.2.3) et la méthode TIEOS. L'assemblage de Gibson, une fois les molécules d'ADN de taille moyenne obtenues (minimum 100 nucléobases), permet de fusionner deux fragments adjacents d'ADN partageant une même terminaison de séquence en une seule molécule d'ADN après une réaction isothermique (Figure I. 6 : L'enzyme T5 exonucléase dégrade chaque terminaison des brins à 50°C dans le sens 5' → 3' et, toujours à la même température, recuit ces mêmes fragments de sorte à les faire chevaucher. L'ADN polymérase et l'ADN ligase finissent le travail en fusionnant les fragments et ainsi former la nouvelle molécule d'ADN. Le PCA quant à lui est également un type d'assemblage qui repose sur la synthèse de plusieurs fragments d'ADN). Leur structure est réfléchie de façon à avoir un chevauchement répétitif entre plusieurs brins afin qu'après recuisson et l'action d'une polymérase, la molécule d'ADN puisse être reconstruite en intégrité.

En addition des techniques de biosynthèses discutées à l'instant, il subsiste une multitude d'autres méthodes plus ou moins utilisées dans la construction enzymatique de l'ADN : RCA (*Rolling Circular Amplification*),[78,79] jets d'encre sur micro-puces[80] (la partie 2.2.4 discutera cette technologie avec une approche chimique), etc.

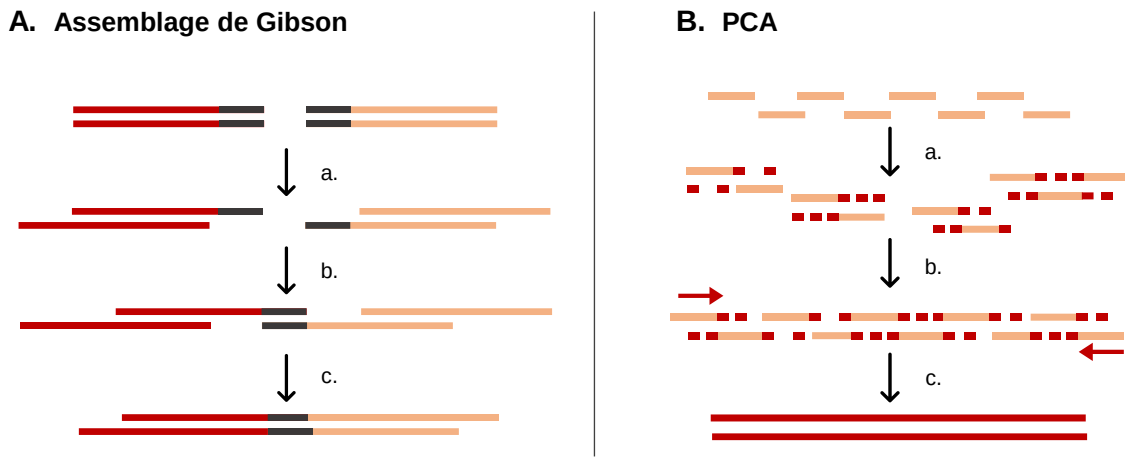


Figure I. 6 : Principes de synthèse par assemblage de Gibson et PCA. Adapté de la littérature.[10] ***A.*** *Assemblage de Gibson : (a) Digestion des chaînes terminales par exonucléase. (b) « Cuisson » des chaînes terminales complémentaires formées en (a). (c) Réparation par ligase et polymérase.* ***B.*** *PCA : (a) « Cuisson » des oligonucléotides se chevauchant (orange). (b) Extension de la chaîne modèle par polymérase. (c) PCR.*

L'ADN possède un éventail d'applications[59] biologiques ou chimiques, telles que la création de bibliothèques d'ARN,[81] l'ingénierie des protéines,[82] la synthèse de génomes entiers,[83] la thérapie génique, ou encore dans le stockage d'information.[7,8,31,84,85] Sa synthèse a, et suscite encore grandement de l'intérêt depuis le milieu du 20ème siècle. Nous verrons dans les deux prochaines sous-parties que d'autres macromolécules naturelles peuvent également être support d'information : l'acide ribonucléique (ARN) et les protéines.

2.1.2. L'ARN : structure et biosynthèse

L'ARN, également support d'information, représente l'intermédiaire dans le transfert du gène depuis l'ADN vers le ribosome pour la synthèse de protéines.[86] Il est chimiquement très proche de l'ADN, notamment par ses propriétés polymères issues d'un enchaînement de nucléotides. Cependant, l'ARN diffère de l'ADN par sa structure (généralement en simple brin, rarement en double brin à l'exception de certains virus)[87,88] et de sa composition : les sucres sont des riboses et la base azotée thymine est remplacée par l'uracile. Fruit de la transcription de l'ADN, l'ARN est formé à l'aide de l'ARN polymérase. Cette enzyme a pour rôle de dérouler un fragment d'ADN sélectionné servant alors de modèle pour synthétiser une molécule d'ARN. Comme dans la réplication de l'ADN, la séquence de nucléotides est déterminée par l'appariement des bases sur ce modèle.[89]

Moins stable que l'ADN du fait de ses résidus riboses (et non désoxyriboses, voir oxygène marqué en rouge sur la Figure I. 7),[90] l'ARN est un moindre candidat comme support d'information à long terme. Malgré des efforts dans sa synthèse enzymatique,[91] l'ARN suscite peu d'intérêt dans le stockage d'information de données numériques. Cependant, il est tout de même captivant du fait de sa capacité à transférer des données, et donc, ce genre de mécanisme pourrait être également utilisé pour développer des systèmes de stockage.[92]

2.1.3. Les protéines : structures et biosynthèse

Les protéines, biopolymères à séquence définie présents dans les cellules vivantes, permettent l'expression de l'information génétique. L'enchaînement d'une bibliothèque de 22 acides aminés liés par des liaisons peptidiques détermine la propriété du polymère biologique.[93] Leur structure peut synthétiquement être définie selon le dogme d'Anfinsen (Prix Nobel de Chimie en 1972) stipulant que la structure primaire d'une protéine est définie par l'enchaînement spécifique de ses acides aminés, et que sa conformation dans l'espace est régie par sa forme la plus stable thermodynamiquement (notamment grâce aux liaisons hydrogènes et interactions hydrophobiques formées).[94] De ce fait, sa structure secondaire émane de ces faibles interactions formant ainsi une forme hélicoïdale[95] (hélice alpha) ou planaire (feuillet béta). Sa structure tertiaire est décrite par son enroulement en 3 dimensions, également la conséquence d'interactions faibles entre les acides aminés. Enfin, sa structure quaternaire est définie par l'organisation spatiale des différentes chaîne protéiques (ou sous-unités) entre elles. Cette compréhension détaillée a été possible grâce aux techniques évolutives de caractérisation.[96] Par ailleurs, la diversité de structure des protéines implique de nombreuses propriétés physico-chimiques utiles au bon fonctionnement de la plupart des processus biologiques.[97] Notamment, la réplication de l'ADN catalysée par l'ADN polymérase comme évoqué précédemment.[98]

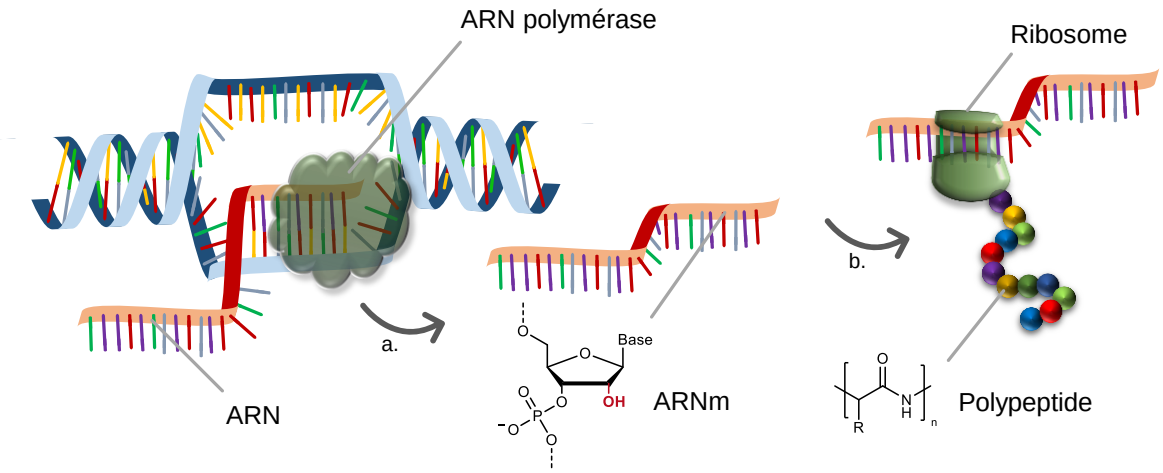


Figure I. 7 : Biosynthèse de l'ARN par l'ARN polymérase suivie de la biosynthèse de protéine grâce à la lecture de l'ARNm par le ribosome.

La biosynthèse des protéines est l'aboutissement du dogme défini précédemment (Figure I. 7). Après avoir été transcrit, l'ARN messager (ARNm) peut être traduit en se liant à un ribosome. Ce dernier permet d'effectuer la traduction des codons présents sur le brin d'ARNm. Un enchaînement de 3 bases nucléiques spécifiques détermine ainsi l'acide aminé correspondant. Etant stables et possédant des chaînes latérales différentes, les polypeptides sont intéressants pour le codage et stockage d'information.[99–101] Après avoir discuté des structures de différents systèmes biologiques ainsi que de leur biosynthèse (2.1), nous verrons dans la prochaine partie que le développement de leur synthèse par voie chimique permet une telle conservation de donnée.

2.2. Synthèse chimique de macromolécules biologiques

2.2.1. Synthèse de polypeptides (protéines)

2.2.1.1 Synthèse sur support solide : R. B. Merrifield

Depuis les travaux d'Emile Fischer sur la synthèse du tout premier polypeptide en 1903,[102] des efforts dans l'optimisation de la synthèse de peptide ont été nécessaire pour espérer de bons rendements. Pendant plusieurs décennies, la méthode Bergmann-Zervas a été utilisée pour réaliser une telle prouesse.[103] Une fois le groupe carboxyle d'un acide aminé et le groupe amine du second acide aminé protégés, par activation du groupe carboxyle libre, la liaison peptidique peut être formée. L'élimination sélective des deux groupes protecteurs conduit par la suite au dipeptide libre.[104] Ce procédé étant limité à la préparation de courtes séquences,[105,106] R. B. Merrifield a développé une méthode rapide, quantitative et automatique pour la synthèse de peptides à longue chaîne.[107]

En 1963, il a prouvé que cela était réalisable en démontrant un moyen efficace dans la préparation d'un tétrapeptide par synthèse sur support solide (SPS).[21] Désormais appelée synthèse peptidique sur support solide (SPPS), ce concept lui a valu le prix Nobel de Chimie en 1983 témoignant de sa grande contribution, notamment dans la préparation de macromolécules biologiques auparavant inatteignable (e.g. ribonucléase A pancréatique de bovin).[108] Contrairement aux réactions chimiques jusqu'alors effectuées en milieu liquide, ce procédé présente de nombreux avantages dont les principaux sont : simplifier et accélérer la synthèse multi-étapes (même vaisselle utilisée) et, obtenir de meilleurs rendements en minimisant les pertes de produits liées aux purifications intermédiaires.[104] Le principe général de SPS, et plus particulièrement SPPS, peut être résumé à l'aide de la Figure I. 8. Il implique l'ajout progressif d'acides aminés protégés à une chaîne peptidique croissante liée par une liaison covalente à une résine solide.

Cette stratégie de synthèse par protection/déprotection est facilitée par la possibilité de faire des lavages entre chaque couplage, éliminant ainsi les excès de réactifs utilisés et les sous-produits formés par filtration. Initialement développée à l'aide du groupe protecteur Boc (*tert*-butoxycarbonyle),[109,110] la stratégie de protection/déprotection a été optimisée en utilisant préférentiellement le groupe Fmoc (fluorénylméthoxycarbonyle).[111,112] Ce dernier permet un suivi de réaction plus simple et efficace, ainsi qu'une déprotection complétement orthogonale aux autres groupes présents sur la chaîne peptidique.[113] Faisant face au potentiel indéniable de cette technologie, R. B. Merrifield a également développé la chimie sur support solide à plus grande échelle. L'automatisation de ce procédé permet d'accéder plus rapidement et plus facilement à des macromolécules complexes, ce qui manuellement peut se révéler lent et fastidieux. Deux ans après avoir synthétisé son premier tétrapeptide par SPS, son nouvel instrument élaboré pour la SPPS automatisée a pu être inauguré avec un nonapeptide (bradykinin) obtenu après 32 heures de synthèse.[114] Depuis, de nombreux robots synthétiseurs ont vu le jour[115] et leur optimisation sont encore d'actualité, notamment pour améliorer leur efficacité en flux rapide,[116] ou encore en y couplant des réacteurs micro-ondes.[117,118]

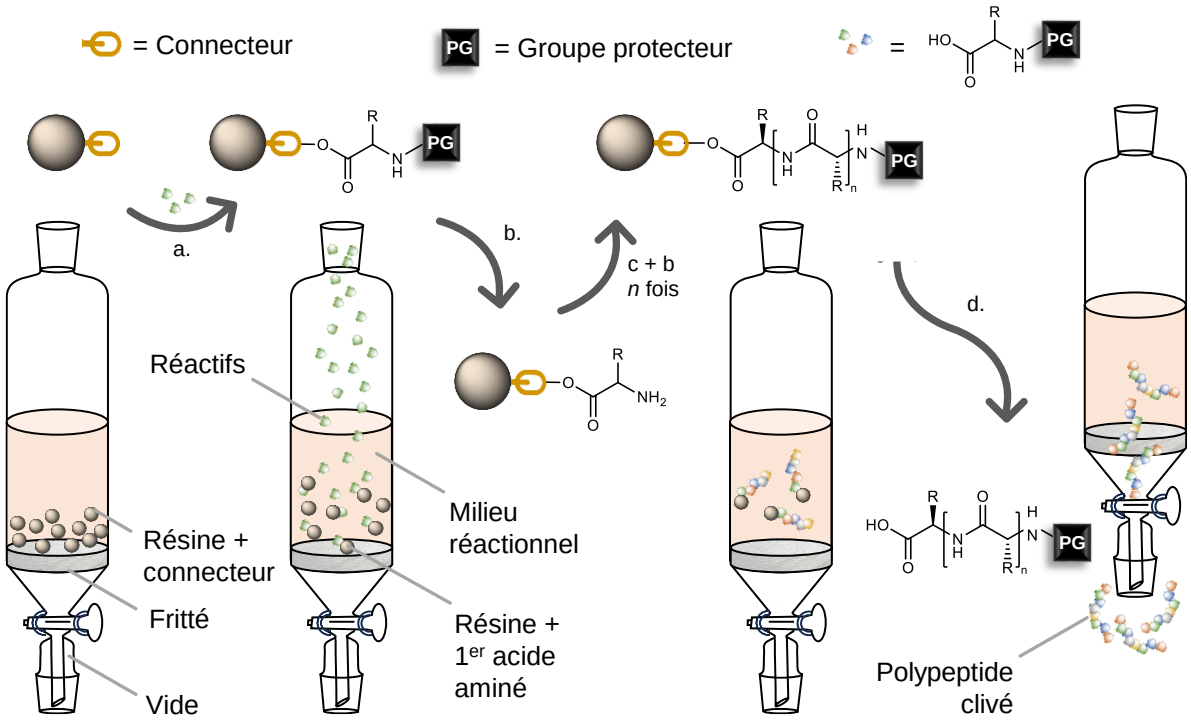


Figure I. 8 : Principe de la SPPS.
(a) 1er couplage sur la résine fonctionnalisée avec un acide aminé. (b) Déprotection du groupe protecteur (généralement Fmoc en présence de pipéridine). (c) Couplage peptidique généralement en présence de DCC. (d) Clivage de la résine (conditions selon le type de résine utilisée).

Traditionnellement, les supports solides utilisés dans la synthèse de tels biopolymères sont des résines réticulées à base de polystyrène (PS) comme initialement introduit par Merrifield.[21] En outre, ces supports peuvent bénéficier d'une large variété de groupes fonctionnels afin d'être versatiles et adaptables à d'autres types de chimie tels que la préparation des oligosaccharides,[119–121] polymères bioconjugués,[122,123] polymères synthétiques à séquence contrôlée[23,124] (discutés partie 3.2) ou encore oligonucléotides[22,125] (discutés partie 2.2.2).

2.2.1.2 Approche supramoléculaire

En 2013, le groupe de Leigh a publié le premier synthétiseur de peptide basé sur une machine artificielle moléculaire. Dans ces travaux, leur but a été de mimer ce que les ribosomes effectuent en construisant les protéines grâce à l'assemblage d'acide aminés dont l'ordre est déterminé par l'ARNm.[126]

La structure chimique du « moteur » utilisé est un anneau moléculaire enfilé sur un axe composé d'acides aminés (rotaxane). La machine décrite dans cette étude est un système comprenant un rotaxane dans lequel le macrocycle récupère séquentiellement les acides aminés présents sur l'axe moléculaire pour assembler un polypeptide dans un ordre bien défini. L'approche de cette étude est résumée avec la Figure I. 9 ci-dessous. La stratégie emploie une chaîne latérale du macrocycle rotaxane (bleu) fonctionnalisée avec un thiol permettant de catalyser l'ajout progressif d'acides aminés sur l'axe moléculaire (orange). Après transestérification impliquant le thiol préalablement déprotoné, un réarrangement S,N-*acyl* s'effectue entre l'acide aminé sur l'axe (vert) et celui attaché au macrocycle (noir)

(pour plus de détails sur les mécanismes impliqués, il est conseillé de consulter l'article originalement publié).[126] Malgré des rendements faibles rendant l'exécution de ce processus difficile et seulement applicable aux polypeptides, c'est une preuve de concept notable par sa complexité qui, avec de l'optimisation, pourrait bien être appliquée à d'autres types de chimie (ex. polymères à séquence contrôlée 3.2).

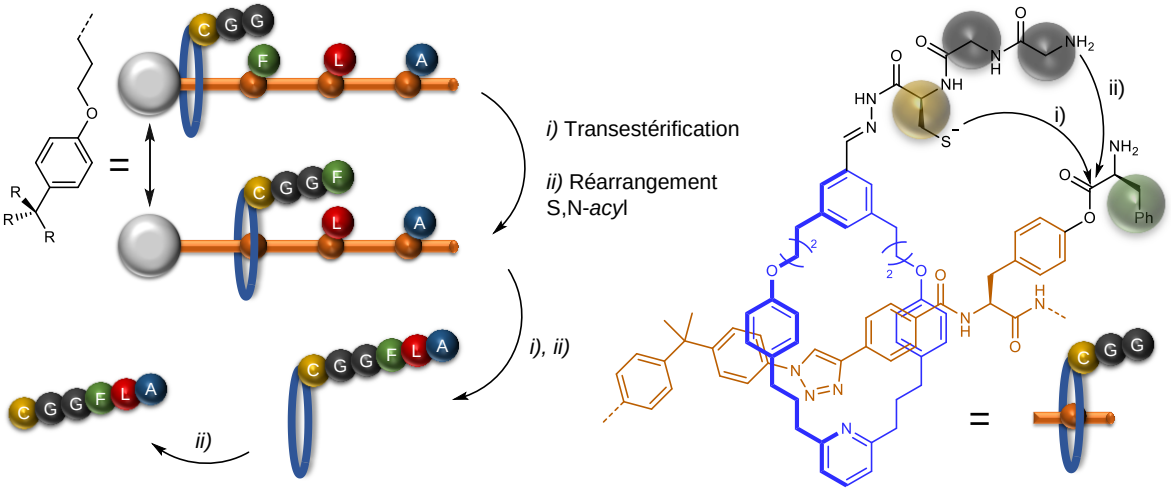


Figure I. 9 : Machine supramoléculaire pour la construction de polypeptides. Adapté de la littérature.[126] Anneau bleu : macrocycle rotaxane. Axe orange : motif répété n fois le nombre d'acides aminés à coupler. Boule grise : « Stoppeur » encombré (R = tert-butylbenzene). Acides aminés (C) = Cystéine, (G) = Glycine, (F) = Phénylalanine, (L) = Leucine, (A) = Alanine. (i) transestérification du thiol déprotoné sur le carbonyle de l'acide aminé. ii) Après catalyse acide, réarrangement S,N-acyl entre l'acide aminé accroché au macrocycle (noir) et celui attaché à l'axe moléculaire (vert). iii) Hydrolyse offrant l'hexapeptide.

2.2.2. Synthèse d'oligonucléotides naturels

La synthèse du premier oligonucléotide remonte à 1955 lorsque Michelson et Todd ont publié la préparation d'une dithymidine.[127] Dans cette étude, la 5'-O-thymidine protégée avec un groupe benzyle peut réagir avec l'oxygène 5' d'une 3'-O-thymidine protégée avec un groupe acyle. Après déprotection, le dinucléotide dT-dT peut être obtenu. Ceci représente la première synthèse par voie chimique d'un dinucléotide contenant une liaison inter-nucléotidique caractéristique des acides nucléiques. Depuis cette réalisation, de nombreuses optimisations à travers différentes stratégies dans la synthèse d'oligonucléotides naturels ont émergées.

2.2.2.1 Phosphodiester en solution : les travaux pionniers de H. G. Khorana

H. G. Khorana a grandement contribué au développement de la chimie des oligonucléotides.[128] En 1956, son groupe a publié une méthode efficace pour la préparation d'oligonucléotides en impliquant la phosphorylation d'un composé hydroxyle par un phospho-monoester d'acide phosphorique en présence de chlorure de *p*-toluenesulfonyle,[129] puis, l'année d'après, en présence de dicyclohexylcarbodiimide[130] (DCC), ce dernier offrant de meilleurs rendements en étant utilisé comme agent de couplage (Schéma I. 1). Ce procédé appelé par la suite « approche phosphodiester » présente tout de même des défauts puisqu'il contient un phosphate non protégé au niveau de la connexion inter-nucléotidique, aboutissant à des ramifications non voulues (Schéma

I. 1, rouge). Dans ce cas, plus la longueur de l'oligonucléotide souhaité est longue, plus la synthèse est difficile du fait des étapes répétées de purification.

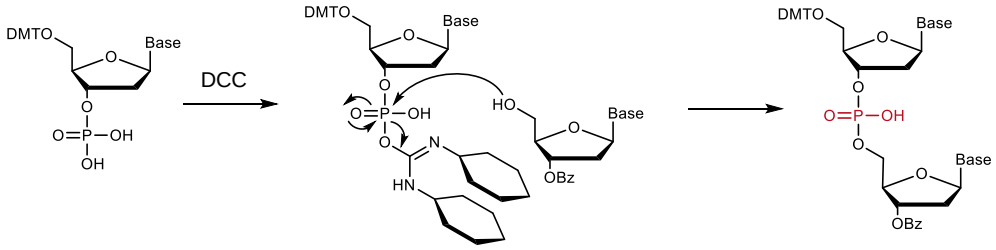


Schéma I. 1 : Couplage phosphodiester catalysé par DCC.
En rouge, phosphate non protégé.

Ces différentes problématiques l'ont amené à s'intéresser aux divers groupes protecteurs pouvant améliorer le schéma général de synthèse d'oligonucléotides. Premièrement, il s'est intéressé au groupe protégeant l'alcool secondaire réagissant sur le phosphate pour former la liaison phosphodiester. Ainsi, le groupe triphénylméthyle (trityl) a initialement été investigué.[131,132] Cependant, ce dernier n'a pas donné de résultats favorables dans la synthèse impliquant des purines, certainement dû à l'acidité labile de leur liaison glycosidique pouvant être fragilisée en conditions de déprotection acide.[133] C'est pourquoi, d'autres alternatives plus fragiles en conditions acides telles que les dérivés *p*-méthoxy des groupes trityl ont été étudiées, notamment le 5'-O-monométhoxytrityl (MMT) et le di-*p*-méthoxytrityl (DMT).[133] Le DMT s'est révélé comme étant un groupe protecteur de choix du fait de sa simple élimination en milieu acide, de son suivi rendu possible par sa couleur caractéristique après déprotection (orange) et de sa signature UV (Figure I. 10).

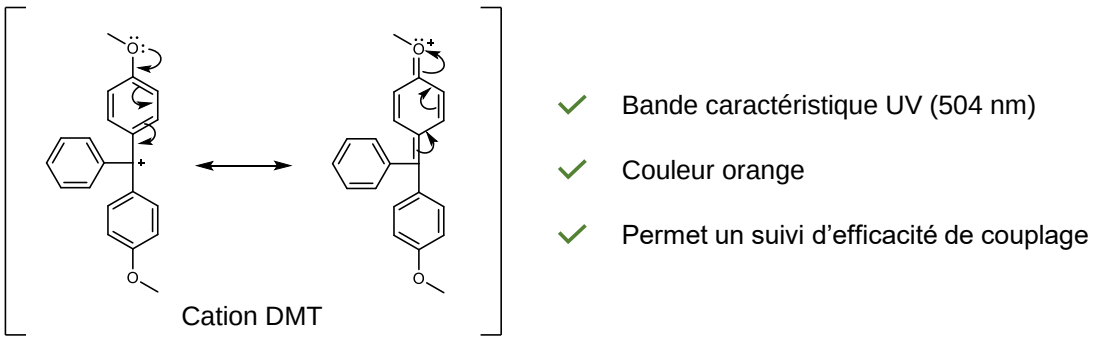


Figure I. 10 : Délocalisation des électrons du cation DMT impliquant une signature UV caractéristique.

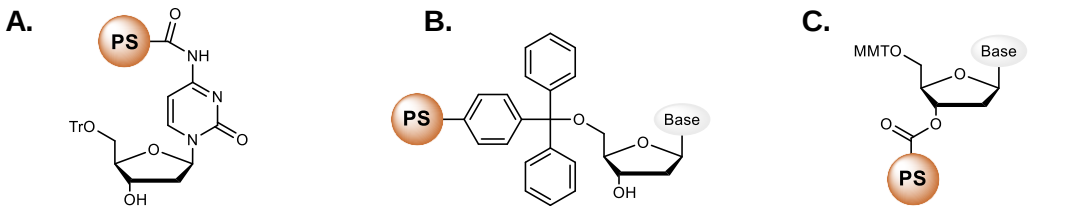
Parallèlement, la protection du groupe phosphate a aussi dû être étudiée. L'incorporation d'un groupe cyanoéthyle permettant son élimination dans des conditions alcalines douces et orthogonales au groupe protecteur DMT s'est avérée être une option efficace dans les couplages de nucléotides.[133] Khorana a également contribué à l'optimisation d'une chimiothèque de groupes protecteurs pour les nucléobases utilisées dans la synthèse d'oligonucléotides. Elles seront vues en détail plus tard dans la partie 2.2.3.4.

Finalement, grâce à ses travaux, le biochimiste a également contribué à la synthèse de polymères d'ARN en étudiant la connexion inter-ribonucléotidique,[134] ce qui lui a permis par la suite de réaliser le premier ARNt séquencé (77-nt),[135,136] puis la synthèse totale

d'un gène incluant différentes chaînes polymères (126-nt et 207-nt).[137] Sa grande contribution lui a valu le Prix Nobel de médecine en 1968.

2.2.2.2 Optimisation sur support solide et soluble : les travaux pionniers de R. L. Letsinger

R. L. Letsinger a également été très impliqué dans la synthèse d'oligonucléotides naturels.[138,139] Dans les années 1960, indépendamment de Merrifield[21] (2.2.1.1), il publie quelques mois plus tard la synthèse d'un dipeptide en utilisant un polystyrène di-vinylique comme support solide.[140] Parallèlement à ses études sur le comportement de ces supports en présence de différents solvants,[141] il a étudié le potentiel de cette technologie sur une autre chimie, notamment celle des oligonucléotides. En 1965, son groupe a publié le premier article impliquant une liaison inter-nucléotidique réalisée sur support solide à l'aide d'un copolymère « popcorn » composé de styrène (88%), d'acide *p*-vinylbenzoïque (12%) et de *p*-divinylbenzène (0.2%).[142,143] Cette résine ainsi formée est insoluble dans l'eau, les solutions alcalines et les solvants organiques classiques utilisés en synthèse oligonucléotidique. Elle peut être chargée avec la 5'-O-tritydeoxycytidine (Figure I. 11A), et par la suite phosphorylée en présence de DCC. Après filtration et lavage, le nucléotide peut réagir avec une thymidine en présence de pyridine formant la liaison phosphodiester. Après traitement basique en présence d'hydroxyde de sodium (NaOH), le dinucléotide est clivé du support et peut être déprotégé du groupe 5'-O-trityl en milieu acide pour donner le dimère désiré. Dans ses travaux préliminaires, Letsinger a utilisé une liaison amide sur la base nucléique pour attacher son support solide. Khorana a alors proposé une alternative avec les groupes protecteurs trityl dont son groupe à l'expertise, mais dans ce cas en utilisant un polymère soluble développé auparavant par Shemyakin et ses collègues.[144] Avec cette technique, il a initié la connexion 5'-O du nucléotide au support soluble chargé avec un groupe MMT (Figure I. 11B) pouvant être clivé du support en milieu acide[145] (Melby et Strobach ont étudié le même schéma de synthèse l'année d'après avec un support insoluble sans prouesse particulière).[146] Dans la même période, M. H. Caruthers, a également été impliqué dans le développement de ces supports et a proposé sa première optimisation à l'aide d'une liaison ester sur la position 3'-O d'une thymidine sur un support soluble (Figure I. 11C). Ses travaux ont alors démontré un procédé simplifié où le nucléotide attaché au support peut, après déprotection du groupe 5'-O-MMT, réagir avec la fonction phosphate d'un autre nucléotide. La méthode appliquée dans cette étude a également établi de meilleurs rendements comparativement aux stratégies par voie solide précédemment publiées par Khorana et ses collègues.[147]



Letsinger – PS insoluble	Khorana – PS soluble	Caruthers – PS soluble
Spécifique à la cytosine	Universel	Universel
Clivage basique	Clivage acide	Clivage basique
Attaché à la nucléobase	Attaché au 5'-O	Attaché au 3'-O

Figure I. 11 : Développement des supports solides.

2.2.2.3 Approche phosphotriester : la grande contribution de R. L. Letsinger

Après les études réalisées par le groupe de Khorana sur l'utilisation du groupe cyanoéthyle pour protéger le phosphate et ainsi éviter les hyper branchements indésirés (discuté précédemment partie 2.2.2.1), Letsinger a profité de leur optimisation pour d'abord l'appliquer à la synthèse d'oligoribonucléosides,[148] puis surtout à la construction d'oligonucléotides en phase solide. Le schéma de synthèse, étape par étape, ressemble à l'approche phosphodiester discutée auparavant puisqu'il commence par *i)* la phosphorylation d'un nucléotide mais dans ce cas en utilisant le phosphate 2-cyanoéthyle puis, *ii)* l'intermédiaire phosphodiester activé à l'aide du chlorure de mesitylenesuflonyle (MsCl) pour générer un meilleur groupe partant qui, *iii)* réagit avec la position 5'-O d'un autre nucléotide formant le phosphotriester. Après traitement basique faible, le groupe cyanoéthyle peut facilement être éliminé, tandis que le clivage du support se fait avec des conditions plus fortes[142,143] (Schéma I. 2).

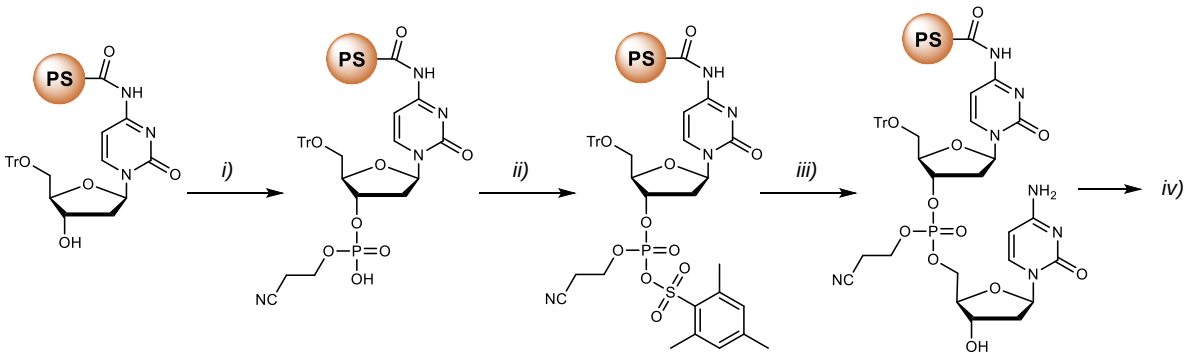


Schéma I. 2 : Approche phosphotriester.
i) Phosphate de ß-cyanoéthyle pyridinium, DCC. ii) MsCl, pyridine. iii) 5'-O-thymidine. Iv) Clivage basique (NH_3) donnant le dimère.[142]

Déjà utilisée précédemment sur support insoluble, cette méthode peut parfois demander une semaine d'agitation pour certains couplages.[142] Adaptée à leur technologie soluble, « l'approche phosphotriester » a par la suite présenté de meilleurs rendements et de favorables temps de réaction (ex. cycle passant de 44%[142] à 64%[147] et temps de

couplage inférieur à 2 jours).[149,150] Letsinger et ses collègues ont par ailleurs prouvé que cette approche peut être envisagée avec les différentes nucléobases classiques (ATGC).[151] Ils ont également démontré que l'utilisation du groupe benzoyle[152] pour bloquer l'amine primaire de la base nucléique est efficace du fait des conditions d'éliminations orthogonales aux deux groupes 5-O-trityl[153] et cyanoéthyle.[154]

Quelques années plus tard, l'agent activateur MsCl de l'intermédiaire phosphodiester discuté ci-dessus (étape *ii)* Schéma I. 2) a lui aussi fait l'objet d'optimisations et particulièrement par Reese et ses collègues.[155–157] L'utilisation des dérivés mesitylènesulfonyle, le 3-nitrotriazole (MSNT) dans un premier temps puis le 4,6-dinitrobenzyltriazole (DBMT), ont donné des efficacités de couplage atteignant les 95% et des temps de conversion à l'échelle de la minute, autant en employant les supports solubles que des résines solides.[157] Ces avancées ont notamment permis la synthèse de plus grands oligonucléotides (20 à 50 nt)[158–160] et la semi-automatisation de l'approche phosphotriester.[161]

2.2.2.4 Approche Phosphite Triester : la grande contribution de R. L. Letsinger et M. H. Caruthers

Dans les années 1970, toujours dans un but d'optimisation maximale du temps de conversion et d'efficacité de couplage dans la construction d'oligonucléotides, Letsinger et ses collègues se sont intéressés à une toute nouvelle approche. Basée sur l'utilisation d'un phosphite (III) plus réactif qu'un phosphate (V), l'originale méthode appelée « phosphite triester » nécessite une étape d'oxydation additionnelle aux procédés précédemment développés. Cependant, elle permet de diminuer de manière significative les temps de couplage dus à la nature hautement réactive de l'intermédiaire phosphomonochloridite, ce qui en fait une stratégie de choix pour la synthèse de longs polymères.[162,163] (Schéma I. 3 : X = Cl, DMT remplacé par MMT, et SiO_2 remplacé par PS). Néanmoins, la découverte de Letsinger reste perfectible puisqu'elle présente de faibles rendements sur résine (9% pour 5-nt),[164] posant une grande problématique pour son automatisation et son application à grande échelle.[163,164] Son ancien étudiant Caruthers, dirigeant à présent son propre laboratoire, a pris conscience de cette lacune en proposant un nouveau support solide fondé sur une matrice inorganique de silice : elle peut être fonctionnalisée par un groupe acide carboxylique, et éventuellement réagir avec une thymidine sur la position 3'-O en présence de DCC, laissant la position 5'-O-trityl disponible à une future déprotection et à un couplage nucléotidique.[165] L'application de ce support à la récente méthode développée par Letsinger permet la semi-automatisation de la synthèse d'oligonucléotides avec un cycle total inférieur à 2h30[166] (Schéma I. 3). Afin d'y parvenir, le support est compressé dans une colonne reliée à une pompe et une série de valves ; après déprotection, le nucléotide attaché au support peut réagir avec le groupe phosphite d'un autre nucléotide, puis, après une étape de piégeage des groupes hydroxyles supportés n'ayant pas réagis (non illustrée sur cette figure), le phosphite obtenu est oxydé en phosphate en présence d'I_2. Enfin, l'oligomère peut être clivé du support à l'aide du thiophénol et d'hydroxyde d'ammonium avant d'être purifié par chromatographie en phase inverse (rendement de 30% pour un déoxydécanucléotide).[166] Dans cette même étude, Caruthers a également montré que l'utilisation du tétrazole à la

place du chlorure permet d'obtenir une meilleure efficacité de couplage (95% au lieu de 90%) due à la formation d'un meilleur groupe partant lors de la condensation de l'alcool sur le phosphite (Schéma I. 3).

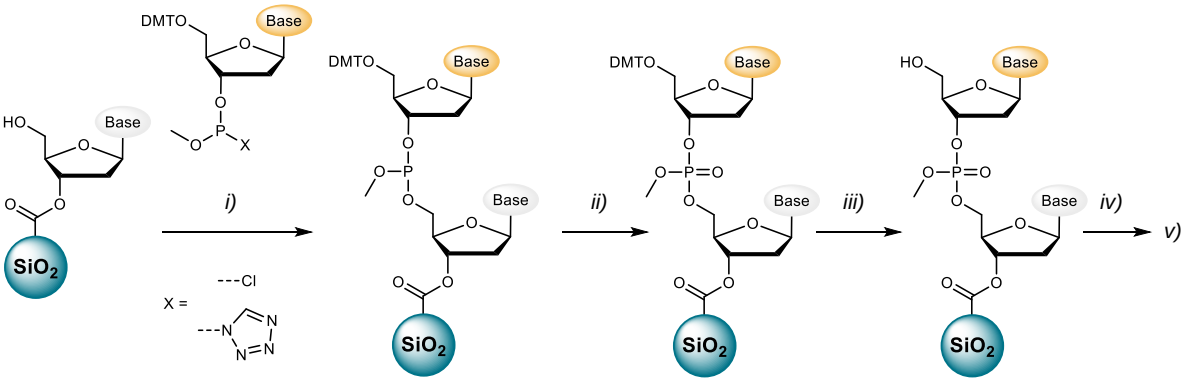


Schéma I. 3 : Approche phosphite triester en utilisant la matrice inorganique à base de silice de Caruthers.[166]
i) Formation de liaison phosphite triester : nucléobase, -78°C. ii) Oxydation / I_2, H_2O, t.a. iii) Déprotection du DMT / TsOH, t.a.. iv) Cycle itératif : i), ii) et iii) répétés n fois. v) Clivage du support (thiophénoxyde) offrant l'oligomère.

Malgré des développements encourageants, ce procédé présente toutefois des points négatifs. En outre, le protocole utilise le phosphomonochloridite qui doit être préparé à -78°C dans du tétrahydrofuran (THF) en présence de dichlorure de méthylphosphonique (CH_3OPCl_2). En addition de ces conditions de réactions peu souhaitables, cet intermédiaire activé ainsi que son dérivé tétrazole (obtenu en une étape à partir du phosphomonochloridite) doivent être stockés à basse température du fait de leur haute sensibilité à l'humidité et à l'oxydation. De ce fait, l'application à plus grande échelle ainsi que la commercialisation de cette technologie ne semble pas viable et nécessite encore d'être améliorée.

2.2.2.5 Approche Phosphoramidite : les travaux pionniers de M. H. Caruthers

Au début des années 1980, le groupe de Caruthers a instauré une nouvelle méthode de construction d'oligonucléotides en utilisant une approche phosphoramidite. Conscient des défauts de la méthode phosphite triester utilisant des réactifs non stables, il a étudié la construction d'intermédiaires plus stables. Inspiré par leur découverte sur l'augmentation d'efficacité de couplage en remplaçant le chlore par un tétrazole sur l'espèce phosphite (Schéma I. 3), il a publié la même année l'utilisation du *N,N*-diméthylaminophosphoramidite.[167] Alors que ses analogues chlorés et tetrazolés doivent être stockés sous forme de précipité à -20°C dans de l'hexane, ce nouveau réactif peut être conservé dans des conditions classiques de laboratoire sous forme de poudre et présente également de bons résultats de conversion. Même si la préparation de l'agent phosphorylant ($CH_3OP(Cl)N(CH_3)_2$) ainsi que sa condensation sur un alcool pour former le dialkylphosphoramidite se préparent toujours à -78°C dans du THF et dans l'hexane respectivement, le fait de pouvoir maintenir le phosphite généré de manière stable a représenté un progrès majeur puisqu'il facilite de manière significative son utilisation. Plus

tard, la formation du monomère phosphoramidite a connu plusieurs optimisations,[168,169] notamment sa formation *in-situ* catalysée à l'aide du *1H*-tétrazole à 25°C,[170] ou bien à l'aide de la base de Hunig (N-éthyl-N,N-diisopropylamine, DIPEA),[171] qui restent encore utilisées à ce jour (Schéma I. 4).[172]

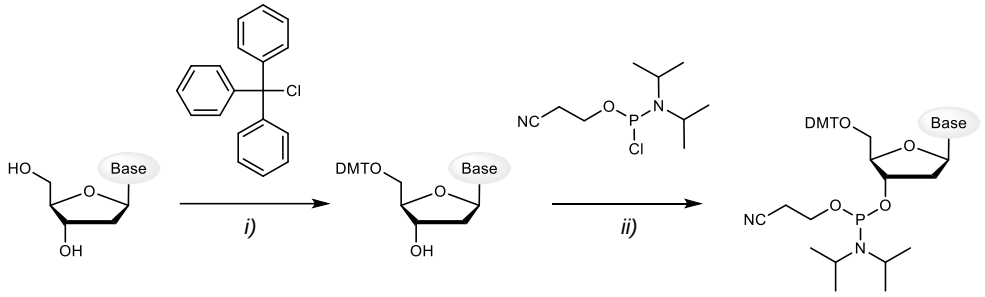


Schéma I. 4 : Synthèse du monomère phosphoramidite.
i) Protection : DMT-Cl, 0°C, pyridine. ii) Phosphorylation : DIPEA ou 1H-tétrazole, THF ou $CDCl_3$, t. a..[167,171]

Par ailleurs, après avoir observé que l'aminophosphoramidite peut être activé en condition acide avant un éventuel couplage inter-nucléotidique, des travaux permettant d'identifier le meilleur agent activateur ont été effectués. L'acide faible *1H*-tetrazole s'est révélé être le meilleur candidat en offrant des rendements quasi quantitatifs lors de la formation du phosphodiester. Des études supplémentaires ont démontré que l'exploitation d'un amine plus encombré sur l'intermédiaire phosphite donne de meilleurs résultats tout en restant stable pendant quelques semaines en solution d'acétonitrile, notamment les N-morpholino et N,N-diisopropylaminophosphoramidite,[173,174] dont ce dernier restera le meilleur choix postérieurement.[22]

Plus tard, l'approche phosphoramidite est devenue la meilleure stratégie dans la synthèse d'oligonucléotides, en particulier grâce à sa compatibilité avec le groupe protecteur cyanoéthyle discuté précédemment.[175] De plus, l'adaptabilité de ce procédé à de nouveaux supports solides (discutés partie 2.2.3.2) a apporté une révolution dans cette technologie.

Même si la chimie phosphoramidite fait encore l'objet d'études et d'optimisations (e.g. mécanosynthèse,[176] chiralité,[177,178] etc.), elle est encore aujourd'hui un premier choix pour la synthèse de longues chaînes d'oligonucléotides.[172] Cette technique de synthèse permet assurément d'obtenir des efficacités de couplage quasi quantitatives avec des temps de cycle à l'échelle de la minute, ce qui en fait un outil très utile pour la préparation de chaînes modèles d'ADN avant extension et/ou amplification (discuté précédemment partie 2.1.1.). Les bonnes performances de cette technologie ont notamment rendu possible son automatisation,[179,180,138] donnant ainsi un accès direct et simplifié à toute la communauté scientifique dans la production d'oligonucléotides. Les différents détails et aspects du processus de synthèse seront détaillés dans la sous-partie ci-après.

2.2.3. Principes de la méthode phosphoramidite

2.2.3.1 Les robots synthétiseurs d'ADN

En décembre 1980, l'entreprise Vega Biotechnologie a introduit le premier robot synthétiseur d'ADN avec le modèle 280 pouvant fabriquer une chaîne oligomérique de 15

unités maximum en utilisant les bases communes ATGC.[181] Deux années plus tard, M. H. Caruthers et ses collaborateurs ont créé leur entreprise Applied Biosystems (ABI) et ont commercialisé leur premier synthétiseur automatique en 1982 (modèle 380A).[138] À la fin des année 1990, la même entreprise a inauguré le modèle 3900, encore utilisé aujourd'hui par de nombreux laboratoires.[15,182–184]

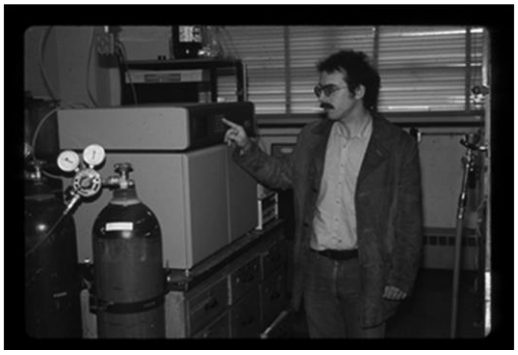


Image 1 : Caruthers avec son premier synthétiseur d'ADN robotisé.[138]

Image 2 : Synthétiseur automatique d'ADN ABI 3900.[185]

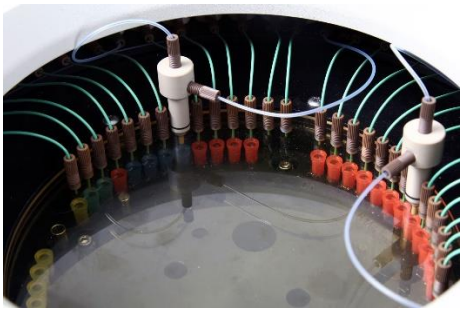


Image 3 : Valves permettant d'introduire les réactifs dans les colonnes supportées.[185]

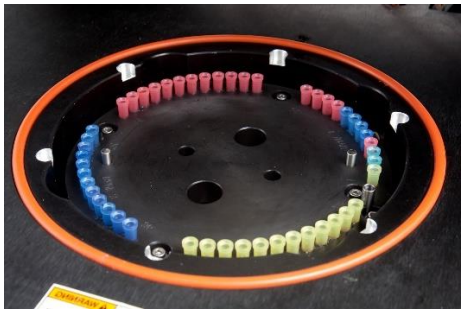


Image 4 : 48 positions disponibles permettant d'effectuer autant de synthèses simultanément.[185]

Dans un besoin de contrôler et suivre la synthèse automatique, d'autres robots ont été développés par cette même entreprise. C'est notamment le cas de l'Expedite 8909 qui permet de faire un suivi de l'efficacité de chaque couplage en quantifiant le taux de DMT déprotégé à partir d'une référence (discuté plus tard en détail dans le chapitre 3 partie 4.1). En outre, des protocoles détaillés peuvent être définis sur le robot, permettant un contrôle absolu des paramètres de synthèse en fonction des oligomères désirés. Sous atmosphère inerte, différentes valves permettent de délivrer la quantité voulue de différents réactifs placés au préalable dans des réservoirs, à travers une colonne chargée en support solide, où les temps de réaction peuvent être précisément ajustés en fonction du composé utilisé.

Aujourd'hui, les robots automatiques pour la synthèse d'ADN n'utilisent plus exclusivement la chimie phosphoramidite pour produire de tels biopolymères. Par exemple, l'entreprise *DNA Script* évoquée précédemment (2.1.1) utilise une « imprimante d'ADN » (*DNA printer*) en utilisant la voie enzymatique (TdT, 2.1.1) à partir de chaînes modèles synthétisées par la technique de Caruthers.[186] La chimie phosphoramidite automatisée reste donc indispensable et fait encore aujourd'hui l'objet d'innovation (e.g.

Agilent),[187,188] notamment la synthèse automatisée sur micro-puces (discutée plus tard partie 2.2.4).

2.2.3.2 Développement des supports solides

Depuis que Merrifield a initié la synthèse peptidique sur support solide (2.2.1.1), de nombreux efforts ont été investis dans le développement d'une stratégie de synthèse oligonucléotidique supportée (2.2.2.2). Différents types de supports ont pu être étudiés tels que les polystyrènes,[189,190] les polyamides[189,191] ou encore la cellulose.[192] Cependant, le gonflement de certains supports (notamment le polystyrène) au contact des solvants organiques empêche la diffusion rapide des réactifs à travers la matrice.[22] Ces limites ont poussé Köster et Cramer à rechercher des dérivés du polystyrène permettant d'éviter le gonflement.[193,194] Les résultats peu probants de polymères hydrophobes[194] les ont amené à s'intéresser aux supports hydrophiles. Particulièrement, aux billes de verres non poreuses qui théoriquement laisseraient les chaines d'oligonucléotides à l'extérieur des surfaces des sphères et éviteraient ainsi le problème de diffusion.[195] Le réseau rigide incompressible formé permettrait assurément l'application de débits élevés de solvants et de réactifs.[196] Cela rendrait également plus efficaces les différentes étapes de lavage nécessaires dans l'approche phosphoramidite.[197] Depuis, ce type de supports communément appelé CPG (*Controlled-Pore Glasses*) a été la principale source d'étude pour la chimie phosphoramidite supportée.[198–202] Leur préparation se fait traditionnellement par traitement de silice de qualité chromatographique en phase liquide à haute performance (*HPLC*) avec du 3-aminopropyltriethoxysilane ($(EtO)_3Si(CH_2)_3NH_2$). Cette silice aminoalkylée peut ensuite réagir avec un anhydride succinique, fonctionnalisant la matrice avec un acide carboxylique. Le couplage d'un nucléotide protégé sur ce dernier en présence de DCC achève la préparation du support insoluble.[22,165,166]

Le choix de *i)* la taille des pores ainsi que *ii)* la longueur de la chaîne entre la silice et le branchement nucléotidique se sont révélés être cruciaux dans l'optimisation des rendements et de la pureté dans les couplages phosphoramidites :

i) La taille des pores varie généralement entre 500 Å, 1000 Å et 2000 Å. Plusieurs études ont démontré que plus les pores sont grands, plus la chaîne polymère attachée peut être longue.[203] Comme le soulignent Katzhendler et ses collègues à travers d'élégantes images de microscopie électronique, la structure des oligonucléotides à la surface des pores devient plus organisée en augmentant la taille de ces derniers, résultant en une meilleure disponibilité des sites réactifs et par conséquent une efficacité de couplage accrue.[204] D'autres travaux confirment ce phénomène, notamment ceux de Kozlov et ses collègues qui, en comparant deux oligonucléotides de tailles égales (75-nt), ont démontré que le biopolymère obtenu en utilisant des pores plus grands (1400 Å contre 1000 Å) résulte en une meilleure pureté et un meilleur rendement.[205]

ii) Aerschot et ses collègues ont démontré que l'utilisation d'une chaîne d'au moins 25 atomes attachés sur la matrice de silice est optimale.[206] Ces résultats confirment l'étude préliminaire réalisée par Adams et ses collègues où ils ont utilisé

une longue chaîne alkyle amine (LCAA) pour former un polymère d'ADN de 51 unités.[207]

Par ailleurs, les CPG peuvent être fonctionnalisés à souhait en fonction des éventuelles post-modifications de l'oligonucléotide voulue. Entre autres, l'ajout de ponts disulfures stables en milieu acide ou alcalin permet de réaliser d'autres réactions chimiques à l'extrémité 5' des oligonucléotides : le clivage du pont disulfure en présence de dithiothréitol (DTT) dans l'ammoniac donne l'oligomère portant une 3'-amine libre pouvant réagir avec d'autres composés d'intérêts (e.g. entité fluorescente).[208,209]

Finalement, il est notable qu'en addition des CPG, différents types de support restent utilisés dans l'approche phosphoramidite avec des performances quasi similaires. C'est le cas par exemple de billes de polystyrène non gonflables qui ont permis la préparation d'un oligonucléotide de 150 unités.[210]

2.2.3.3 Le cycle itératif détaillé

Les différentes étapes du cycle itératif du procédé initié par Caruthers sont décrites dans cette section. Désormais principalement utilisé par chimie automatisée en phase solide, ce cycle s'applique également manuellement à la paillasse (Schéma I. 5).

i) **Déprotection (détritylation) :** le cycle commence par la déprotection du nucléoside 5'-O-DMT attaché au support solide en présence d'acide trichloroacétique (TCA) laissant par la suite la position 5'-O, prête pour réaliser le couplage inter-nucléotidique.

ii) **Activation et couplage :** un excès de monomère 5'-O-DMT phosphoramidite, en solution dans l'acétonitrile anhydre et activé à l'aide du *1H*-tetrazole (formé en atteignant la colonne simultanément), réagit avec l'alcool 5'-O déprotégé à l'étape précédente. Le couplage n'excédant généralement pas les 2min en incluant les étapes de lavage postérieures, assure une efficacité supérieure à 98%.

iii) **Piégeage (*Capping*) :** Malgré l'efficacité de couplage quasi quantitative, des chaînes attachées au support n'ayant pas réagi, et comportant donc un groupe 5'-O libre, sont susceptibles d'interférer lors du prochain cycle. Cette étape a par conséquent effet à piéger ces alcools à l'aide d'anhydride acétique activé en présence de *N*-méthylimidazole (DMAP). Ce processus évite la formation de chaînes tronquées et facilite de ce fait la purification du polymère désiré en fin de cycle.

iv) **Oxydation :** Le phosphite triester P(III) obtenu après couplage est oxydé en phosphate P(V) plus stable en présence d'I_2 et de pyridine dans l'eau. L'oxydation du phosphite est quantitative et ne dure qu'environ 30 secondes en incluant les lavages qui s'y accompagnent.

Finalement, après l'oxydation, le cycle peut se répéter *n* fois à partir de l'étape *i)* jusqu'à l'obtention de la séquence désirée.

v) **Clivage/déprotection :** Une fois le dernier cycle terminé, la colonne contenant le polymère attaché au support solide peut être enlevée du robot pour un

traitement post-synthèse. Généralement, le clivage du support ainsi que la déprotection des groupes protégeant le phosphate (cyanoéthyle) et les éventuelles bases nucléiques (discutés ci-après partie 2.2.3.4), interviennent simultanément puisqu'ils nécessitent tous les deux des conditions basiques (cependant, la déprotection de certaines bases comme la guanine requiert un temps plus long ou une température autour des 50°C selon le protocole pour être complète).[22,211]

Remarques sur le clivage : les conditions de clivages ont à travers différentes études été optimisées rendant le processus encore plus efficace et rapide, notamment avec l'utilisation d'un bi-composant $NH_3(H_2O)/MeNH_2(H_2O)$ (Ammoniac/MéthylAmine, AMA) facilitant l'attaque nucléophile sur la fonction ester du support.[212] D'autres protocoles ont également été développés, plus spécifiquement en y adaptant des conditions douces lorsque des dérivés nucléotidiques sensibles sont utilisés (*t*-butylamine/H_2O/MeOH,[213] K_2CO_3/MeOH[214] ou encore diisopropylamine/MeOH[215]).

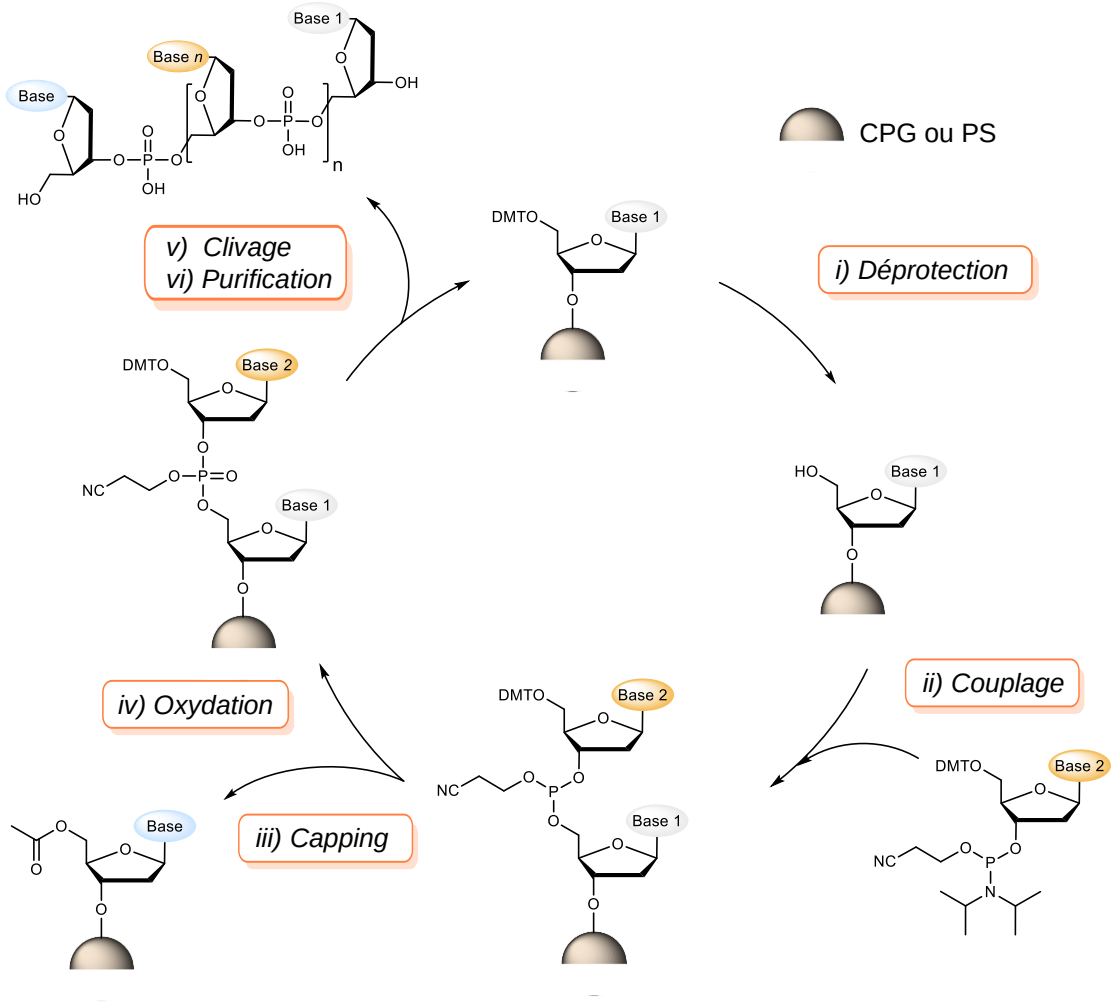


Schéma I. 5 : Cycle itératif pour la synthèse d'oligonucléotides par approche phosphoramidite.

vi) **Purification :** Différents protocoles de purification peuvent être appliqués suivant le mode de synthèse utilisé sur le robot. Habituellement, pour des aspects pratiques, l'oligonucléotide est produit avec le mode *DMT-ON* qui

signifie que lors du dernier cycle, le dernier monomère ajouté est volontairement laissé protégé avec le DMT en position 5'-O. Cette technique rend le polymère désiré chimiquement différent des séquences tronquées qui contiennent quant à elles une fonction ester. Cela permet en outre de séparer ces dernières par chromatographie en phase inverse. L'utilisation d'une colonne disponible commercialement et adaptable avec une seringue rend accessible ce processus à n'importe quel laboratoire (pas d'instrument spécifique requis). Il simplifie le lavage des groupes protecteurs éliminés lors du clivage et des chaînes non désirées, tandis que le polymère ciblé reste quant à lui sur la colonne. Après déprotection du groupe DMT attaché au polymère en présence de TCA, celui-ci peut être élué avec une solution basique (NH_4/OH) puis lyophilisé.

2.2.3.4 Les groupes protecteurs

Les groupes protecteurs pour la synthèse d'oligonucléotides jouent un rôle très important puisque le rendement d'un tel procédé est directement corrélé à l'efficacité de clivage et à l'orthogonalité de ce dernier en présence d'autres groupes fonctionnels. En outre, la protection de la fonction éxocyclique de l'amine des différentes purines classiques (Adénosine, Guanosine et Cytosine) est essentielle du fait des problèmes de dépurination récurrents en synthèse oligonucléotidique.[216] Des conditions acides ou basiques trop fortes peuvent résulter en une perte de la purine laissant le sucre sur la chaîne croissante non substitué[217] (Schéma I. 6).

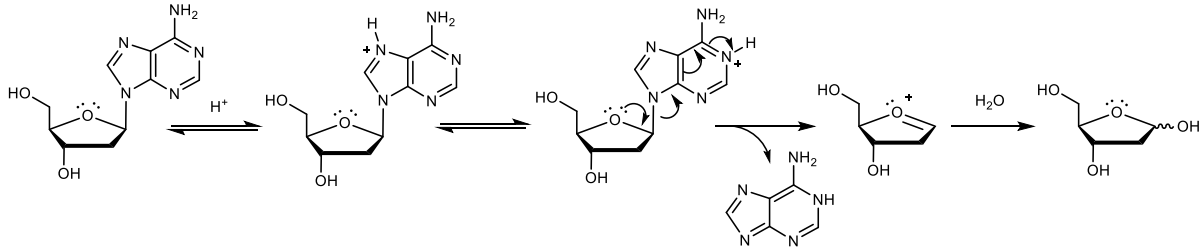
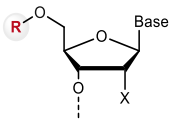
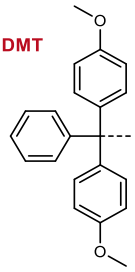
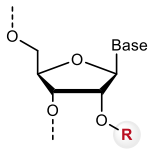
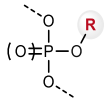
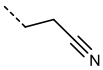
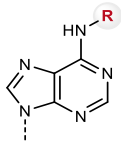
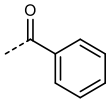


Schéma I. 6 : Exemple de dépurination avec la guanosine.[218]

Le choix d'un groupe protecteur clivable en conditions douces, mais orthogonal au groupe DMT utilisé en chimie phosphoramidite est par conséquent nécessaire. Khorana a grandement participé à leur optimisation comme discuté précédemment (2.2.2.1) avec les dérivés benzènes utilisés sur support solide[142,143,147,154,219] mais aussi les dérivés trityles, initialement utilisés en solution.[131–133]

Les méthodes classiques de protection dans la synthèse d'oligonucléotide sont détaillées dans le Tableau I. 1 ci-dessous. (À noter que chacune des nucléobases a fait l'objet d'une quantité considérable d'études et d'optimisations, détaillées par Beaucage et Iyer).[22,220] Sauf exception, qui dans ce cas sera discutée, les groupes protégeant les fonctions amines éxocycliques des nucléobases sont déprotégées simultanément lors du clivage du support solide en conditions basiques (NH_4/OH) à 55°C ou en conditions plus douces (AMA) à température ambiante. Par ailleurs, il est notable que seulement les purines bénéficient traditionnellement de tels groupes protecteurs.[220] La Thymidine et l'Uracile

ne possèdent pas d'amines éxocycliques et n'ont par conséquent pas la même sensibilité à l'hydrolyse pouvant générer une perte de la nucléobase.

Espèce protégée	Groupe protecteur	Commentaires
 X = H (ADN) ; X = OH (ARN)	R = DMT 	Discuté précédemment partie 2.2.2.1, après avoir étudié la réactivité du trityl et du MMT, Khorana a optimisé le groupe protecteur de la position 5'-O en utilisant le **DMT**, pouvant être éliminé en conditions acides douces (TCA) et permettant un suivi de réaction grâce au sel de DMT clivé, très actif en UV.
 Position 2'-O (ARN)	R = TBDMS	Le groupe *tert*-butyldiméthylsilyle (**TBDMS**) peut protéger la position 2'-O du ribonucléotide. Son avantage est de pouvoir être déprotégé orthogonalement des autres groupes protecteurs en présence de fluorure de tétrabutylammonium (TBAF). Utile puisqu'il doit être déprotégé le plus tard possible en raison de la possible migration de la liaison phosphodiester (3' -> 2').[221] D'autres dérivés silylés peuvent également être envisagés.[222]
 Phosphate ou Phosphite	R = Cyanoéthyle 	Également discuté précédemment partie 2.2.2.1, le groupe **cyanoéthyle** est indispensable pour éviter les hypers branchements sur le phosphate. Introduit avant le couplage inter-nucléotidique directement à l'aide de l'agent phosphorylant (Schéma I. 4), il est facilement déprotégé dans des conditions basiques douces et s'élimine généralement pendant le clivage du support.
 Adénosine	R = Benzoyle 	Le groupe protecteur **benzoyle** a été introduit par Schaller et Khorana en 1963 et est toujours utilisé aujourd'hui.[133] Il est apprécié pour sa résistance aux conditions de réactions employées en chimie phosphoramidite.[220] Des conditions de clivage plus douces peuvent être utilisées avec le groupe protecteur phenoxyacétyle (PAC) requérant moins de 4h à température ambiante dans du NH_3 aqueux (29%).[223]

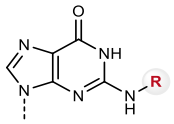
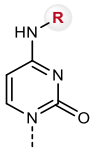
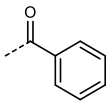
 Guanosine	R = *iso*-butyryle	Khorana et ses collègues ont également introduit en 1972 le groupe **isobutyryle**[224,225] qui s'est avéré être un groupe protecteur robuste pour la guanosine.[220] Autrement, Teoule et ses collègues publiaient quelques années plus tard le groupe méthoxyacétyle (MAC) nécessitant des conditions de clivage plus douces (les mêmes que son homologue PAC pour l'adénosine).[223]
 Cytosine	R = Benzoyle 	Le groupe **benzoyle** est aussi un premier choix pour protéger la cytosine du fait de sa robustesse.[220] Cependant, un groupe acétyle pour protéger la thymidine peut être nécessaire dans le cas où AMA est utilisé pour le clivage.[212] Par ailleurs, l'isobutyryle peut également être utilisé pour cette nucléobase en conditions plus douces (les mêmes que PAC et MAC ci-dessus).[223]

Tableau I. 1 : Les principaux groupes protecteurs utilisés par approche phosphoramidite.

2.2.4. Synthèse sur micro-puces utilisant la chimie de phosphoramidite

À partir des années 1990, lorsque l'approche phosphoramidite est établie comme étant la meilleure stratégie dans la construction d'oligonucléotides, la synthèse sur micro-puces a fait son apparition[226] en proposant deux principaux av antages :

i) Le moyen d'accéder à une grande quantité de différentes macromolécules simultanément : la microminiaturisation de la synthèse d'ADN combinée à son automatisation permet de réaliser en parallèle de nombreuses bibliothèques de différentes macromolécules biologiques.[227,228]

ii) Un coût de revient défiant toute stratégie : l'utilisation de micro-puces peut diviser le coût de revient par base nucléique jusqu'à 100 fois ($0.00001 par base contre $0.001 en synthèse classique sur colonne supportée).[59] Cette prouesse est possible grâce à l'optimisation de protocoles et la miniaturisation des matériaux nécessaires réduisant ainsi la quantité de réactifs utilisés (femtomoles contre des micromoles en synthèse classique)[229] et par conséquent le coût de synthèse.[230]

Cette technologie, développée initialement pour des applications biologiques et diagnostiques,[231] utilise un protocole modifié de l'approche phosphoramidite. Inspirés des travaux préliminaires d'Ekins et ses collègues sur le potentiel des marqueurs luminescents et fluorescents attachés sur les nucléotides,[232] Fodor et ses collègues ont développé cette stratégie en développant une chimie photo-active. Elle permet de contrôler spatialement différentes synthèses en parallèle à l'aide de techniques photo

lithographiques (dans ce cas, processus par lequel une entité est transférée depuis un photo masque vers la surface réactive du biopolymère), laissant place à la déprotection sélective de nucléotides photo labiles.[233]

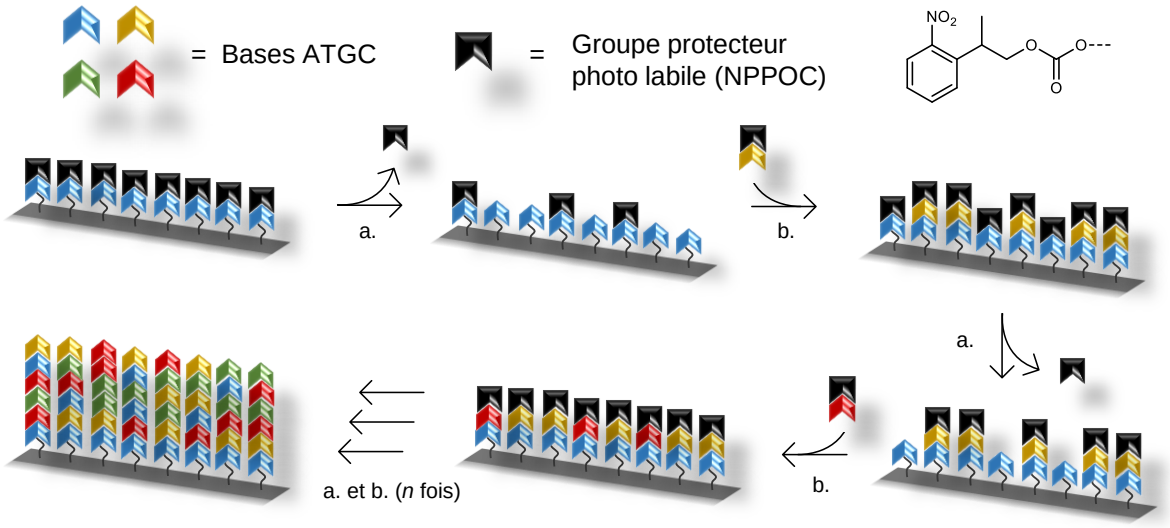


Figure I. 12 : Principe de synthèse sur micro-puces.
(a) : Déprotection (UV). (b) Couplage phosphoramidite miniaturisé avec des conditions classiques (2.2.3.3).

Techniquement, la position des trous du masque définie exactement l'endroit où le couplage phosphoramidite s'effectue, tandis qu'un couplage avec les positions dans l'ombre du masque est empêché. Alors que l'approche phosphoramidite traditionnelle développée par Caruthers utilise un groupe protecteur labile en condition acide (DMT), cette stratégie nécessite une entité pouvant se débloquer photo chimiquement. L'usage du 3'-nitrophenylpropyloxycarbonyl (NPPOC) a contribué de manière satisfaisante à cette fin donnant de très bons rendements[234] (Figure I. 12).

Par ailleurs, le recours à des photo masques coûteux, et parfois contraignant du fait de sa conception, a laissé place à de nouvelles optimisations s'affranchissant de tels supports, notamment des micromiroirs programmables,[235] des semi-conducteurs,[236,237] ou encore des jets d'encre.[238] Ces différentes approches ont par exemple été employées dans la synthèse totale de gènes incluant des séquences d'anticorps thérapeutiques (35 kilobases ont pu être encodées dans cette étude),[239] et ont également démontré leur potentiel dans la construction d'ARN.[240]

Dans cette section (2.2), nous avons pu voir que l'optimisation de la synthèse d'oligonucléotides à travers plusieurs décennies a permis de parvenir à un processus accompli avec des efficacités de couplage supérieures à 99%. Cela permet notamment à la communauté scientifique d'obtenir des chaînes polymères allant de 100 à 200 nucléotides.[8,56,57,187,241–244] Cependant, afin que ces macromolécules biologiques puissent bénéficier d'études supplémentaires (e.g. activités biologiques, candidats médicaments), il est nécessaire de pouvoir caractériser ces dernières de manière fiable, efficace et rapide. Dans la prochaine section, nous discuterons des différentes méthodes de lectures utilisées pour analyser ces biopolymères.

2.3. Méthodes de séquençage

Au même titre que sa synthèse, le séquençage d'ADN a été un sujet de recherche démontrant beaucoup d'intérêt depuis la découverte de sa structure en 1953, permettant notamment aujourd'hui de mieux comprendre certains génomes cancéreux,[245] ainsi que l'identité de maladies spécifiques.[246] Très récemment, la crise de la Covid-19 a également prouvé l'importance du séquençage de génomes avec plus d'un million de séquences du coronavirus partagées par la communauté scientifique dans le monde entier.[247] Dans cette section, les principales techniques de séquençage d'ADN seront discutées.

2.3.1. 1ère génération : W. Gilbert et F. Sanger

Les méthodes de séquençage parallèlement développées par Gilbert et Sanger dans les années 1970 sont considérées comme étant les techniques de 1ère génération et leur ont notamment permis de partager le Prix Nobel en 1980.

Maxam et Gilbert ont proposé en 1977 une procédure déterminant la séquence de nucléotides basée sur la labélisation de la chaîne terminale d'une molécule d'ADN pouvant être partiellement et préférentiellement rompue à chaque répétition de bases (typiquement un ^{32}P radioactif). L'ensemble de fragments imbriqués obtenu peut ensuite être résolu par électrophorèse à l'aide d'un gel polyacrylamide où leur taille révèle dans un ordre bien précis le point de rupture. La comparaison des résultats obtenus avec un modèle de bandes radioactives permet de retrouver l'ordre exacte de la séquence (environ 100 bases dans cette étude).[248] Malgré l'automatisation de cette technique,[249] quelques défauts tels que l'utilisation d'atomes radioactifs ou encore d'hydrazine rendent néanmoins difficile sa manipulation.

Dans la même période, Coulson et Sanger ont proposé une méthode de séquençage par « terminaison de chaîne ». Elle implique l'incorporation éparse d'un des 4 didésoxynucléotides (ddNTP) (Figure I. 13) par l'ADN polymérase lors de la réplication de l'ADN. Une fois placées dans la séquence, ces bases labélisées (marqueurs fluorescents ou radioactifs) empêchent l'extension de la chaîne par liaison phosphodiester générant par conséquent des fragments tronqués à la position 3' du ddNTP. Finalement, en fin de processus, l'échantillon contient un mélange complexe de ces fragments qui est analysée par électrophorèse. Alors que cette technique permet de séparer les différentes espèces par leur taille, en confrontant les données récoltées à des références bien décrites, l'ordre de la séquence peut *in fine* être retrouvé.[250,251] Comparativement à la technique développée par Gilbert, ce procédé reste moins dangereux lorsque des marqueurs fluorescents sont impliqués, d'où son automatisation quelques années plus tard (1986)[252] par ABI qui établit cette technique comme étant celle de premier choix dans le séquençage d'ADN à cette époque.

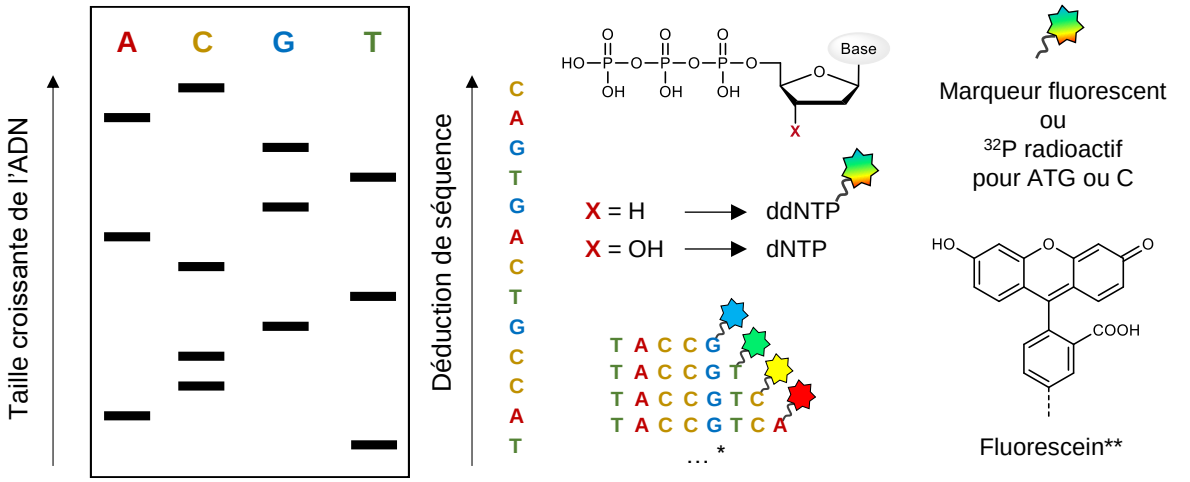


Figure I. 13 : Principe du séquençage d'ADN développé par Sanger et Guilbert.
** Mélange de séquences tronquées à la suite de l'ajout de ddNTP lors de la réplication. ** Exemple de fluorophore utilisé comme marqueur dans le séquençage d'ADN : la fluorescéine.*[252]

2.3.2. 2e génération : séquençage de nouvelle génération (*Next Generation Sequencing*, NGS)

À partir de la fin des années 1980, de nombreux efforts ont été investis pour explorer de nouvelles alternatives au séquençage par électrophorèse, qui restait alors coûteux et lent. Contrairement aux techniques de 1ère génération, NGS est une approche s'appuyant sur la multiplicité et impliquant un séquençage parallèle massif de mélanges complexes d'ADN.[253]

Une technique notable est celle du « séquençage par synthèse » (SBS) dont une principale stratégie peut être retenue : le pyroséquençage. Ce procédé repose sur la mesure luminescente de pyrophosphates générés lors de la réplication de l'ADN sans avoir recours aux labels fluorescents (ou radioactifs) précédemment employés. Il implique deux enzymes dans lequel l'adénosine triphosphate (ATP) sulfurylase catalyse la conversion d'un pyrophosphate en ATP. Cette dernière, digérée par la luciférase (enzyme bioluminescente), produit ainsi un signal lumineux proportionnel à la quantité de pyrophosphate initialement engagée.[254] Néanmoins, cette méthode présente des limites lorsque plusieurs bases similaires viennent à être pyroséquencées, du fait des signaux lumineux résultants quasi analogues.[255]

D'autres procédures dérivées de cette technique ont par la suite émergées, notamment une méthode d'amplification d'ADN par émulsion (emPCR) adapté à un support solide paramagnétiques,[256] ou encore l'incorporation progressive de dNTP labelisés avec des marqueurs fluorescents par ADN polymérase.[257] Alternativement, non plus basées sur des techniques de SBS utilisant l'ADN polymérase, certaines méthodes reposent sur l'hybridation d'un brin d'ADN attaché sur une surface solide (Par exemple, le séquençage par ligation et détection (SOLiD) développé par l'entreprise ABI).[258]

Finalement, la méthode la plus simple et indéniablement la plus utilisée encore aujourd'hui est appelée "polonies"[259] ou "amplification par pont" (*bridge amplification* en anglais). Principalement développée par Pascal Mayer, elle consiste à amplifier une série de

modèles à l'aide d'amorceurs immobilisés sur une surface, afin que les copies de chaque modèle restent étroitement groupées. La Figure I. 14 résume cette méthode.[260]

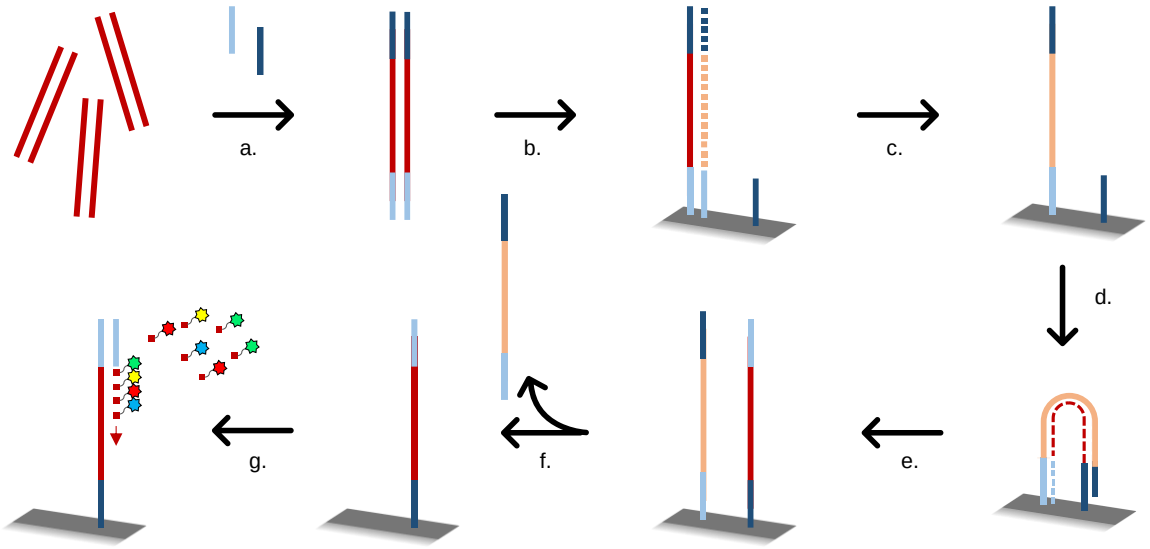


Figure I. 14 : Méthode de séquençage par « amplification par pont ».
Adapté de la littérature.[260] (a) Fonctionnalisation des bouts de chaînes d'oligonucléotides à l'aide d'adapteurs de séquences (différents pour chaque extrémité). (b) Copie d'un nouveau brin (pointillé) attaché sur une surface à l'aide d'adapteurs (bleu clair). (c) Dénaturation éliminant le brin initial (rouge). (d) et (e) Un second brun est copié (pointillé) grâce à la formation d'un pont à l'aide de l'autre adapteur de séquences (bleu foncé). Le processus est répété un très grand nombre de fois formant des colonies. (f) Clivage sélectif du brin complémentaire (orange) grâce à un adapteur différent (bleu clair). (g) La lecture du brin se fait par ajout progressif de bases nucléiques labelisées avec des fluorophores. Un détecteur permet de lire en temps réel les bases ainsi ajoutées.

Quelques décennies d'innovation ont donné naissance à de nombreuses solutions dans le séquençage d'ADN. À l'époque compétitrice avec les entreprises ABI, 454 (Roche) et bien d'autres, Illumina reste aujourd'hui le principal leader sur le marché en utilisant la méthode d'amplification par pont décrite ci-dessus (Figure I. 14). En utilisant celle-ci, plus d'un milliard de lectures indépendantes peuvent être générées en deux jours par un étudiant sur un seul instrument (Illumina NovaSeq) pour quelques milliers de dollars.[253]

2.3.3. 3e génération : séquençage en temps réel de molécules singulières

Les deux générations à l'instant susmentionnées nécessitent dans la plupart des cas une amplification de matrice par ADN polymérase. Cependant, cet usage présente des défauts tels que des erreurs de copies, pertes d'information (dépurination) ou encore le temps additionnel souvent fastidieux requis dans la préparation d'échantillons.[253,255] Idéalement, les biopolymères devraient être analysés de manière fiable, rapide, sans limite de longueur et sans digestion de protocoles préalables à leur séquençage. Cette 3e génération essaie de répondre à ces défis en proposant une lecture « brique par brique » et ce, en temps réel.

Deux principales approches peuvent être soulignées dans cette catégorie. Une première est basée sur l'observation optique de l'activité d'une ADN polymérase en temps réel.[261] La technique appelée SMRT (*Single-Molecule Real-Time*) développée par l'entreprise Pacific BioSciences capture l'information de la séquence durant la réplication du biopolymère d'intérêt à l'aide de quatre bases nucléiques contenant des marqueurs

fluorescents. Ces derniers génèrent ainsi des émissions de spectres distinctes lors de leur lecture.[262] Les brins d'ADN circulaires formés par ligase à l'aide d'adaptateurs sont guidés et piégés dans des micro puits où une polymérase est immobilisée. Lorsque la réplication commence, la séquence peut être lue en temps réel grâce à l'excitation des entités fluorescentes (Figure I. 15).

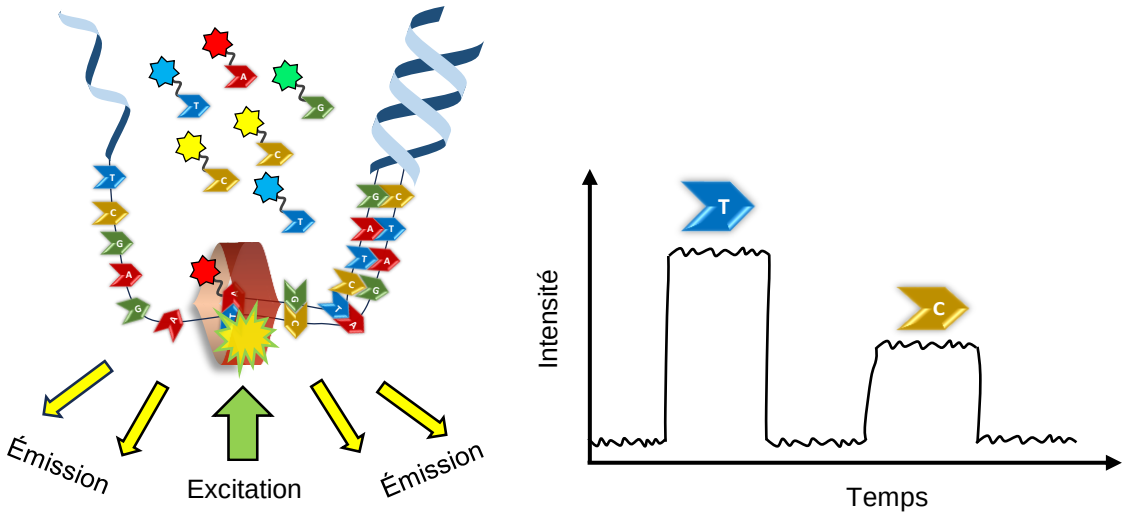


Figure I. 15 : Principe de séquençage par SMRT.

Une deuxième approche, admirablement plus efficace que celle juste discutée à l'instant, est le séquençage par nanopore. Dès les années 1980, plusieurs groupes ont identifié le potentiel d'utiliser les pores des protéines afin d'apporter une technique répondant aux défis de la 3e génération mentionnée plus tôt.[263] Plus tard, les travaux remarquables de Gouaux et ses collègues sur l'identification de la structure de la protéine α-Hemolysin[264] ont laissé place à la détection stochastique de molécule (en d'autres mots, l'observation d'événements individuels de liaison entre l'analyte et un récepteur),[265] et ont ainsi ouvert la voie à de nombreux chercheurs dans l'optimisation du séquençage par nanopore.

Alors que cette technologie peut être développée et appliquée en théorie à toute macromolécule d'intérêt biologique, c'est pour l'ADN qu'elle a suscité le plus d'intérêt. Dans ce cas, le principe réside sur la mesure des variations de courants électriques lorsqu'un acide nucléique passent à travers un nanopore : une tension constante produit un courant ionique à travers le nanopore qui va entraîner le biopolymère à l'intérieur de ce dernier dans un sens unique. L'action d'une hélicase attachée à l'entrée des pores de la protéine permet de dérouler le double brin d'ADN en simple brin à travers cette dernière de façon constante, nucléobase par nucléobase. La différence de courant ionique observée lors des différentes translocations des bases nucléiques avec les pores permet d'analyser en temps réel la séquence du polynucléotide (Figure I. 16).[263]

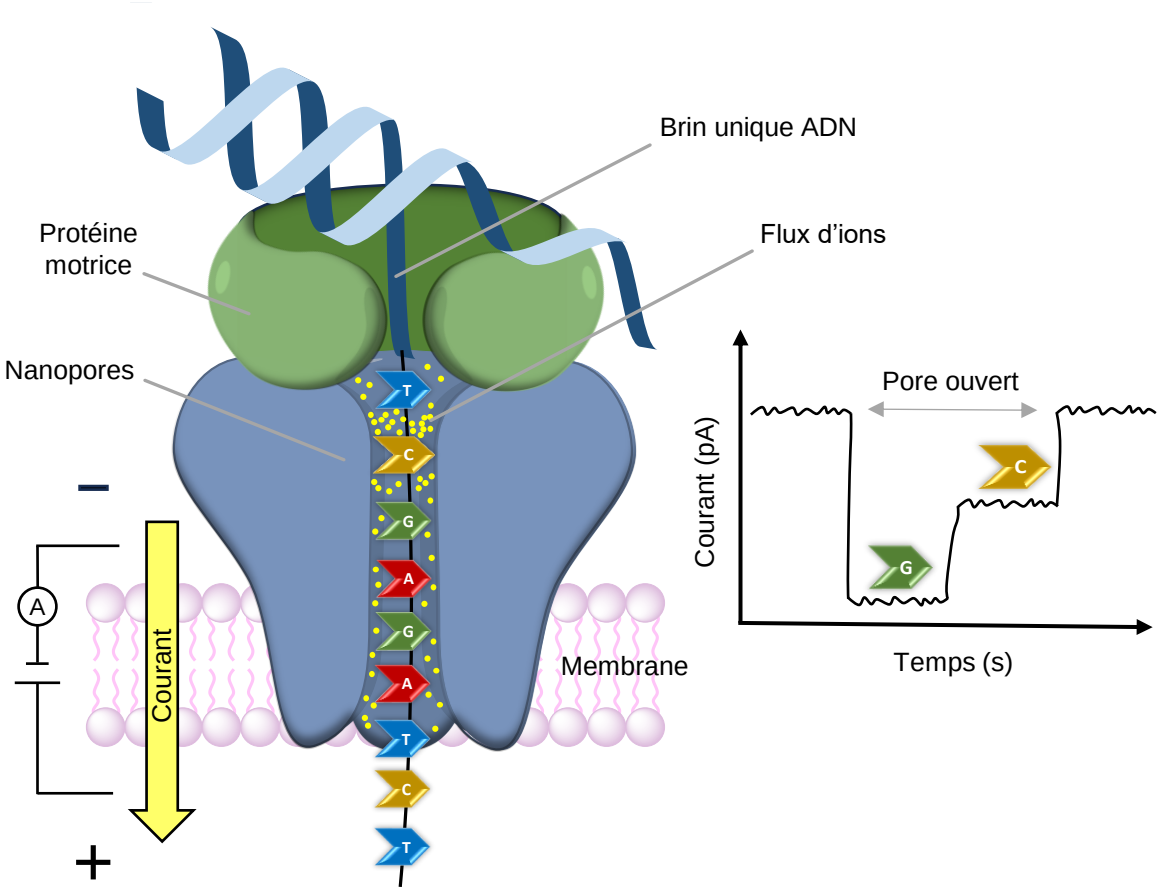


Figure I. 16 : Principe du séquençage par nanopores.

J. J. Kasianowicz et H. Bailey ont grandement contribué au développement de cette technologie.[266] Ce dernier a notamment exploité les découvertes de ces dernières décennies en fondant son entreprise Oxford Nanopore Technology (ONT) en 2005, qui, quelques années plus tard, a créé des séquenceurs automatisés pour l'ADN. Alors que l'entreprise Illumina reste commercialement l'entreprise la plus représentée dans le domaine du séquençage,[253] ONT se démarque par la portabilité de ses appareils. En effet, grâce à un processus pouvant être miniaturisé du fait de l'aspect électrique de l'analyse, la société de Bailey a mis au point le seul séquenceur portable permettant de lire l'information contenue dans l'ADN en temps réel (MinION).[267] Le dispositif de la taille d'une clef USB pèse moins de 500g et peut générer jusqu'à 50Gb d'information (en 72h, 420 bases par seconde).[267] Cet outil montre à quel point des progrès ont été réalisés depuis les travaux pionniers de Sanger et Gilbert dans les années 1970. Aujourd'hui, ce type d'équipement permet efficacement de séquencer des virus directement sur le terrain comme démontré par la lecture de génomes de l'Ebola présents chez 14 patients en Afrique, où les résultats ont été obtenus 48h après réception des échantillons.[268]

Dans la partie 2.3, l'historique des techniques de séquençage et leurs principales optimisations ont été détaillées. Lire l'information contenue dans les macromolécules biologiques s'est révélée essentiel dans la compréhension du génome humain ainsi que la caractérisation de certains virus. Effectivement remarquable par la technologie NGS utilisée dans le séquençage du coronavirus ces dernières années, cela a notamment

récompensé P. Mayer (un des pionniers du NGS avec la technique de génération de clusters et d'amplification)[269] du Breaktrhrough Prize 2022.[270]

Néanmoins, l'innovation des techniques de séquençage suscite encore de l'intérêt, et ce, spécifiquement pour une application dans le stockage de données numériques.[271–273] La prochaine sous-partie discutera des limites du stockage d'information artificiel dans des macromolécules biologiques.

2.4. Limites des systèmes biologiques dans le stockage d'information

Récemment, la combinaison de différentes technologies (e.g. synthèse sur micro-puces, synthèse enzymatique utilisant TdT, etc.) a permis d'encoder des dizaines de GB d'information comme des données *Wikipedia*, des films *Netflix* ou encore la musique d'introduction du jeu vidéo *Super Mario Bros*.[274,275] Le catalogue de méthodes de synthèse d'ADN aujourd'hui existant encourage la communauté scientifique à collaborer et encore plus innover pour trouver ensemble des solutions durables et bâtir un écosystème autour du stockage de données dans l'ADN.[274]

Cependant, il existe de nombreuses limites dans le stockage sur support biologique qui apparaissent difficiles à surmonter.[48,276] Premièrement, la synthèse doit se faire dans la plupart des cas sur de petits fragments (~100-200 nt) avant d'être digérés par des enzymes ayant pour rôle l'assemblage (e.g. assemblage de Gibson) ou l'amplification (PCR) jusqu'à atteindre la longueur et la quantité de séquence désirée. Dans ce cas, les erreurs de copies ou la perte d'information précédemment mise en relief dans le NGS (2.3.3) s'appliquent également au moment de la synthèse de ces polynucléotides. Alors que les avancées récentes démontrent l'utilisation d'appareils miniaturisés à des fins économiques et pratiques, le recours aux enzymes accompagnées de leurs défauts dans l'obtention d'une large quantité de biopolymères semble être irremplaçable.

Par ailleurs, l'ADN est connu pour être sensible à la lumière et/ou la chaleur résultant en une dégradation partielle ou complète selon les conditions appliquées (humidité, température).[14] Même si les observations du groupe de Grass amènent des solutions quant au stockage numérique dans l'ADN en le pérennisant dans des endroits stables à l'abri des UV et de la chaleur, certaines contraintes persistent, notamment la nécessité d'immobiliser le biopolymère dans une matrice inorganique de Silice afin d'éviter d'éventuelles dégradations liée à l'humidité.[14] D'autre part, l'encapsulation des polynucléotides dans cette Silice supprime un des avantages majeurs du principe de stockage à l'échelle moléculaire : le rapport densité/donnée. Effectivement, en considérant le système entier comprenant l'ADN et le support inorganique, Grass et ses collègues ont démontré que l'approche la plus fiable et robuste pour stocker l'information à long terme comprend seulement 0,7% de pourcentage massique du polynucléotide.[277] Malgré des efforts dans l'optimisation de cette technologie en utilisant un dispositif « couche par couche » (LBL, *layer by layer* en anglais) déposé sur des nanoparticules magnétiques, la proportion d'information par rapport au poids et la taille du matériel reste

plus faible que la densité théorique assumée précédemment ((1.2.2), 5 x 10^{11} GB par gramme contre 1 x 10^{10} dans cette étude).[278]

La lecture des données présente également certains inconvénients : comme vu précédemment, le séquençage de l'ADN nécessite dans la plupart des cas des fluorophores pouvant être coûteux et devant s'accompagner d'équipements optiques pour leur détection.[7] Même si les nanopores émergent comme une solution répondant à ces défis, d'autres contraintes liées à la pratique de cette méthode subsistent : lors de la lecture de longues chaînes d'oligonucléotides avec le MiniON, le taux d'erreur peut aller jusqu'à 15%[279] comparativement aux séquenceurs Illumina qui comportent une moyenne de seulement 0.25%.[280]

Finalement, une des principales limites dans l'usage de l'ADN comme moyen de stockage numérique reste le manque de flexibilité. Particulièrement, travailler avec des macromolécules biologiques impliquent de rester en conditions physiologiques, c'est-à-dire, majoritairement dans l'eau. Ceci amène éventuellement des contraintes si par exemple, de nouvelles méthodes de séquençage requérant des solvants organiques venaient à être développées. Par ailleurs, encoder de l'information dans un polynucléotide revient à être limité par un seul alphabet (ATGC). Cette barrière présente plusieurs inconvénients : premièrement, la densité d'information théorique contenue dans le biopolymère ne pourra jamais être excédée puisqu'elle dépend de ses quatre bases nucléiques où chaque base code pour 2 bits (Figure I. 1). Et deuxièmement, l'utilisation d'un alphabet restreint à quatre monomères peut manifester une augmentation de taux d'erreur. En outre, les séquences homopolymères où la même base se répète plusieurs fois à la suite résultent dans la plupart des cas en une non-concordance lors de leur lecture.[13]

Dans ce sous-chapitre 2, les différentes voies de synthèse et de lecture de macromolécules biologiques ont pu être détaillées. Malgré le développement d'un grand nombre d'approches autant dans la fabrication que dans le séquençage des biopolymères, des limites quant à leur application dans des systèmes de stockage numériques subsistent (2.4). Alors que les polymères d'origine naturelle sont limités par leurs propriétés biologiques et leurs structures non modulables, les polymères synthétiques demeurent une alternative intéressante : ils peuvent être élaborés à souhait avec comme seule contrainte, la physico-chimie. Le prochain sous-chapitre détaillera cette nouvelle classe de polymère, leur synthèse ainsi que leur lecture, dans l'application de stockage numérique.

3. Stockage d'information dans des macromolécules synthétiques abiotiques

Depuis que Staudinger a instauré le concept de polymère en 1920,[281] de nombreuses avancées dans la chimie des matériaux ont été réalisées. Ces décennies de recherches ont notamment ouvert la voie à de nouvelles catégories de polymères comme celle des

polymères à séquence contrôlée.[282] Ces derniers, intéressants par leur propriétés mécaniques, optiques, électroniques et/ou encore supramoléculaires (autoassemblage)[15] peuvent répondre aux défis persistants énoncés précédemment concernant les macromolécules biologiques. Dans ce sous-chapitre, les principes du stockage d'information sur support synthétique, les principaux polymères synthétisés à cette fin, leurs moyens de caractérisation, puis enfin la chimie de phosphoramidite comme outil de stockage introduisant les travaux de cette thèse seront détaillés.

3.1. Comment stocker l'information dans des macromolécules synthétiques ? Principes et avantages

Au même titre que les macromolécules biologiques, les polymères synthétiques peuvent être utilisés comme moyen de stockage.[283] Alors que l'ADN emploie ses quatre bases nucléiques pour constituer son alphabet numérique, une entité synthétique peut concevoir un alphabet à souhait. Le procédé reste le même que pour ses analogues biologiques (Figure I. 17) : l'information binaire est transcrite en langage moléculaire avec l'alphabet de monomères choisi puis, après synthèse du polymère, la chaîne obtenue peut finalement être séquencée pour retrouver le code binaire initialement encodé.

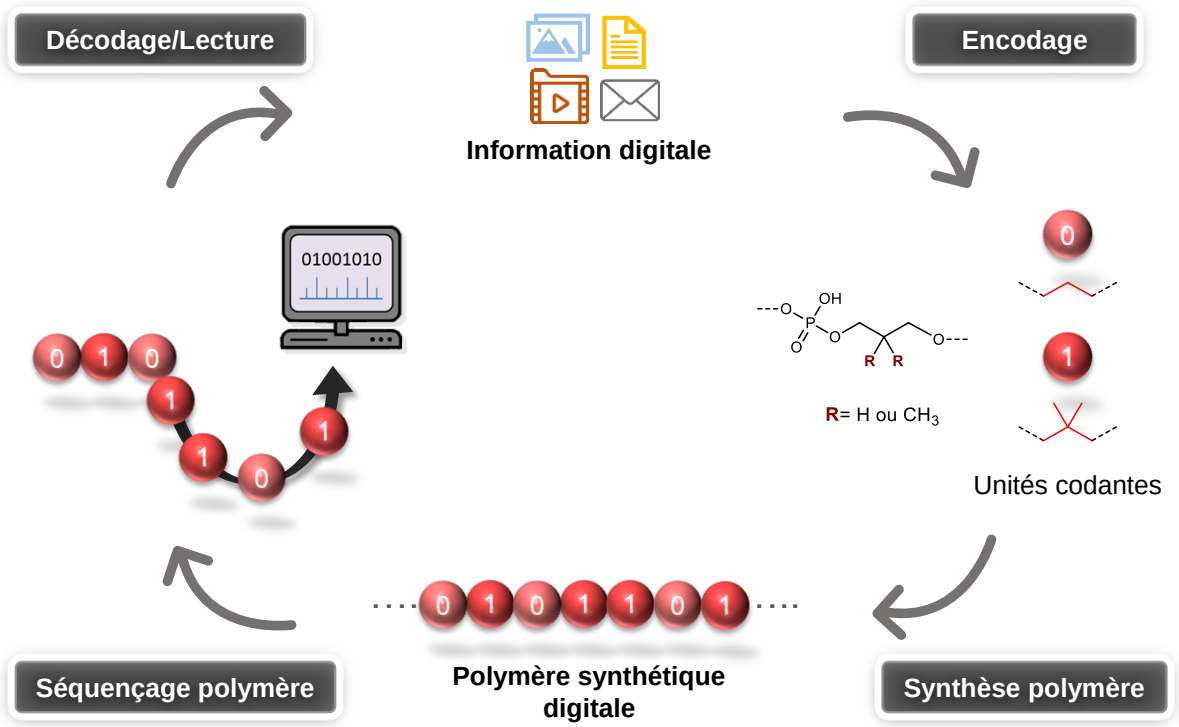


Figure I. 17 : Principe de stockage d'information numérique dans un polymère synthétique.

Remplacer l'ADN par les polymères synthétiques comme support d'information présente de nombreux avantages et répond à différentes problématiques énoncées précédemment (2.4). Alors que le polynucléotide est limité par l'utilisation de seulement quatre bases (ATGC), la diversité structurelle des macromolécules synthétiques offre une opportunité considérable pour augmenter la densité de stockage.[284] Effectivement, un alphabet de 2^n monomère encodera pour n bits. Par exemple, l'ADN encode théoriquement

2bits/nucléotides grâce à son alphabet de 4 monomères (Figure I. 1). Cette capacité tend vers l'augmentation en attribuant un *n* plus élevé à cette relation : lorsqu'un alphabet à 8, 16 ou encore 32 monomères est appliqué, un potentiel de 3, 4 ou 5bits/monomères respectivement en résulte.

De plus, outre la flexibilité intéressante pour encoder un alphabet augmenté, la polyvalence d'approches chimiques possible pour encoder de l'information reste un atout majeur. Effectivement, des enjeux importants dans le stockage d'information à l'échelle moléculaire sont, la vitesse d'écriture (cinétique de réaction) et l'utilisation de réactifs qui idéalement devraient être sans dangers et durables (ex. biosourcés). La grande variété de réactions chimiques existante laisse l'opportunité aux chercheurs, avec le temps, de trouver le meilleur compromis en répondant à tous ces critères, importants au développement d'un stockage pérenne.

Autrement, la considération de ces polymères non naturels pour un stockage d'information serait dans un premier temps, comme l'ADN, appliqué à des « données froides ». Tandis que leur synthèse et lecture ont déjà fait l'objet de nombreuses études,[23,285] leur moyen de conservation à long terme a très peu été étudié. Néanmoins, des recherches dans l'optimisation de préservation pour ce type de matériaux restent nécessaires et commencent à susciter de l'intérêt. En outre, l'assemblage LBL, dont Decher a été le pionnier dans les années 1990,[286] a récemment été intéressant dans l'organisation spatiale de polymères. Alors que les technologies modernes de stockage d'ADN utilisent des matrices inorganiques de silices unidimensionnelles (2.4), cette méthode de construction tend à croître la densité de stockage théorique puisqu'elle occupe éventuellement deux ou trois dimensions dans l'espace.[287]

Finalement, dans un contexte de stockage de données strictes, les polymères synthétiques ont également l'avantage de s'affranchir des contraintes structurelles et biologiques de l'ADN (double hélice et autoréplication respectivement).[288] Contrairement au biopolymère, la macromolécule non naturelle peut être conçue à la demande en s'appuyant sur les différentes méthodes de synthèses et de lectures déjà existantes. En effet, les innovations effectuées les dernières décennies dans la construction et le séquençage d'oligonucléotides profitent au développement de méthodes (nouvelles ou non) adaptées aux polymères synthétiques. Bien que les premières études sur cette nouvelle classe de polymère soient récentes, les progrès remarquables obtenus pour les macromolécules naturelles complexes, sur une ligne temporelle relativement courte, suggèrent la possibilité d'obtenir, avec du temps et de l'optimisation, des procédés performants.[288]

La prochaine sous-section détaillera différentes approches dans la synthèse de polymères à séquence contrôlée dans un but de stockage d'information.

3.2. Les polymères à séquences contrôlée pour le stockage d'information

Avant de discuter de la synthèse de tels polymères, il semble important de définir avec exactitude deux termes bien spécifiques. De l'anglais, « **un polymère à séquence**

contrôlée est une macromolécule dans laquelle les séquences de monomères sont contrôlées dans une certaine mesure. Ce contrôle peut être absolu mais n'est pas nécessaire (…). **Un polymère à séquence définie** est une macromolécule uniforme avec une longueur de chaîne exacte et une séquence de monomères parfaitement définie. En d'autres termes, chaque unité monomère se trouve à une position définie dans la chaîne (…) ».[23] les différentes méthodes de préparation de polymère à séquence contrôlée peuvent être regroupées en 3 catégories distinctes (Figure I. 18) : i) polymérisation par étape, ii) polymérisation en chaîne et iii) polymérisation multi-étape.[289]

i) Dans le cas des polymérisations par étape, des monomères multifonctionnels réagissent ensemble pour obtenir des structures linéaires, cycliques ou réticulées.[285] Les oligomères téléchéliques ainsi formés donnent après répétition du processus des polymères à séquence contrôlée avec une polymolécularité élevée Đ≈2. Cette classe de réaction rassemble divers types de réactions telles que les ADMET (métathèses de diène acycliques), les CuAAC ou encore les réactions impliquant des groupes ester, amide ou uréthane.[289]

ii) Lors des polymérisations en chaîne, au moins une espèce continuellement active réagit avec les autres monomères un par un et sont ainsi incorporés successivement.[285,289] Cette catégorie de polymérisation permet d'accéder à un meilleur contrôle de la séquence en offrant une polymolécularité comprise entre Đ≈1.01 et Đ≈2. De nombreux types de réaction ont été utilisés dans la construction de tels polymères regroupant principalement les « non contrôlés » donnant accès à des copolymères statistiques, alternés ou en blocs, puis d'un autre côté les « vivants » (en anglais *living polymerization*). Ces derniers offrent une meilleure maîtrise en utilisant un amorceur faisant croître la chaîne simultanément. Cette catégorie regroupe principalement les réactions ioniques, radicalaires (NMP, ATRP et RAFT) ou encore d'ouvertures de cycles (ROMP et ROP).

iii) Finalement, les polymérisations multi-étapes impliquent la formation de polymères uniformes, étape par étape. Les réactions successives entre des monomères bifonctionnels en utilisant une stratégie de protection/déprotection donnent des macromolécules parfaitement définies avec une polymolécularité Đ≈1. La chaîne croissante est généralement attachée à une résine afin de profiter des avantages de la chimie sur support solide développée par Merrifield discutée précédemment (2.2.1). Comme démontré à l'aide de la chimie de phosphoramidite par notre groupe, l'emploi de cette technique ne se limite pas aux oligomères mais peut donner des polymères à séquence définie d'une longueur de plus de 100 unités.[25]

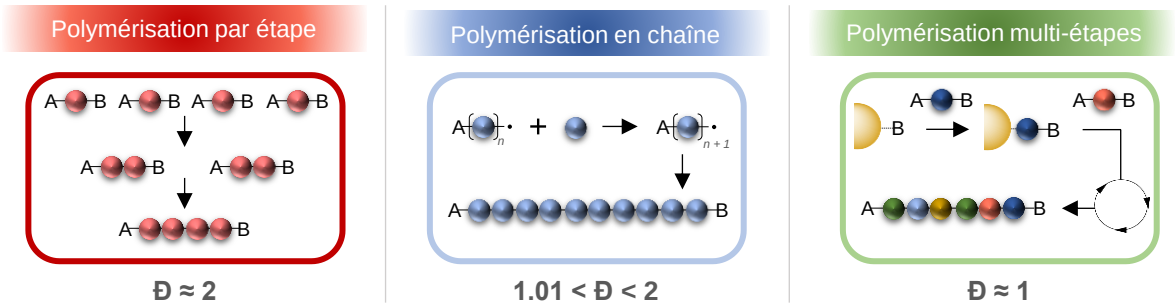


Figure I. 18 : Les trois différentes catégories regroupant les polymères à séquence contrôlée non naturels.
Adapté de la littérature.[285]

Les différentes catégories décrites ci-dessus laissent apparaître la possibilité de stocker de l'information dans des macromolécules synthétiques de deux façons : une en ayant un contrôle sur la séquence (polymolécularité comprise entre Đ≈1.01 et Đ≈2 en utilisant les polymérisations par étape ou en chaîne), et une autre en ayant un contrôle absolu sur chaque unité monomère (Đ≈1 avec les polymérisations multi-étapes). À l'instar des polynucléotides naturels où la séquence est définie à l'unité prêt (propriété essentielle pour y stocker de l'information génétique ou artificielle), la recherche d'alternatives synthétiques en considérant préférentiellement les macromolécules à séquence contrôlée uniformes (Đ≈1) paraît logique. Bien que des efforts aient été fournis pour obtenir une séquence parfaitement définie en employant les polymérisations en chaînes (par exemple notre groupe avec une polymérisation en chaîne radicalaire),[290] les résultats obtenus jusqu'à présents ne sont pas suffisants pour garantir un contrôle absolu.[11] Cette stratégie pouvait alors être comparée à de « l'écriture imparfaite »,[291] et n'était donc pas adéquate pour une application de stockage informationnel. C'est pourquoi, ce sous-chapitre discutera uniquement des synthèses multi-étapes pour la préparation de polymères à séquence définie.

Comme illustré sur la Figure I. 18 et expliqué précédemment, la polymérisation multi-étape repose sur l'ajout consécutif, étape par étape, de monomères bifonctionnels (AB). De cette manière, des macromolécules linéaires ou non-linéaires (dendrimères) avec une séquence totalement définie peuvent être obtenues. La construction de tels polymères utilise deux stratégies distinctes : une reposant sur des réactions orthogonales ne nécessitant pas de groupes protecteurs (« AA + BB » ou « AB + CD »), et une autre utilisant une stratégie de protection/déprotection (« AB + AB ») (Figure I. 19).

Dans ces deux approches, une chimie en phase solide est dans la plupart des cas utilisée. Aussi bien à l'aide d'une phase solide pour profiter des nombreux avantages de la SPPS (2.2.1.1), qu'à l'aide de phase soluble. Par ailleurs, notre groupe a démontré ces dernières années que cette approche est attractive par la possibilité de fabrication de supports sur mesure, avec des structures moléculaires soigneusement optimisées (longueur des chaînes, distribution du poids moléculaire, capacité de chargement et liens clivables).[292–295]

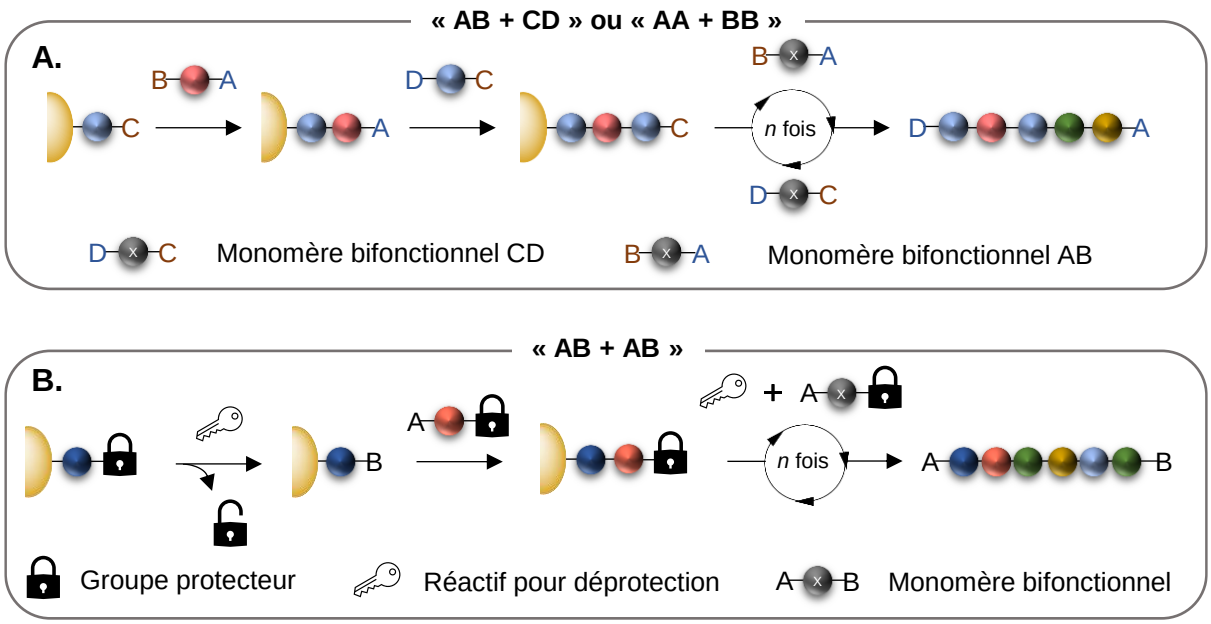


*Figure I. 19 : **A.** Stratégies « AB + CD » ou « AA + BB » et **B.** « AB + AB ».*
*Pour la stratégie « AB+CD » (**A.**), D réagit avec A et B avec C de manière orthogonale.*

3.2.1. Approches itérative « AB+AB » : la chimie de phosphoramidite

Au même titre que la synthèse d'ADN, les polymères non naturels sont capables d'être construits en utilisant une approche en phase liquide ou solide. Les nombreux avantages que proposent la synthèse sur support solide (2.2.1.1) ont naturellement amené certains groupes de recherches à obtenir de meilleurs résultats en employant cette stratégie, et particulièrement dans l'approche itérative AB + AB qui est la même que celle utilisée dans l'approche phosphoramidite. Dans cette méthode de protection/déprotection, la fonction B du monomère représente le groupe protecteur qui, inséré sur la chaîne croissante, peut être déprotégé et impliqué dans un nouveau couplage avec la fonction A. Manifestement, les conditions chimiques utilisées lors du couplage doivent être orthogonales au groupe protecteur et, dans le cas d'une synthèse sur support solide, au connecteur attaché sur la résine.

En 2015, notre groupe a réalisé la première avancée dans le développement de la chimie de phosphoramidite en tant que support de stockage en utilisant des monomères synthétiques à la place des nucléotides de l'ADN.[296] Ces travaux représentent les fondations d'une nouvelle stratégie qui sera par la suite, une des principales sources de recherche du groupe. Dans cette étude, afin d'incorporer un code moléculaire dans la chaîne polymère, deux monomères synthétiques tels que le propyl (2-cyanoéthyl (3-diméthoxytrityloxy-propyl) diisopropylphosphoramidite) et le 2,2-diméthyl-propyl (2-cyanoéthyl (3-diméthoxytrityloxy-2,2-diméthyl-prop-yl) diisopropylphosphoramidite) codant pour 0 et 1 respectivement ont été synthétisés. Leur préparation s'effectue en deux étapes à partir du propane-diol correspondant (Figure I. 20).

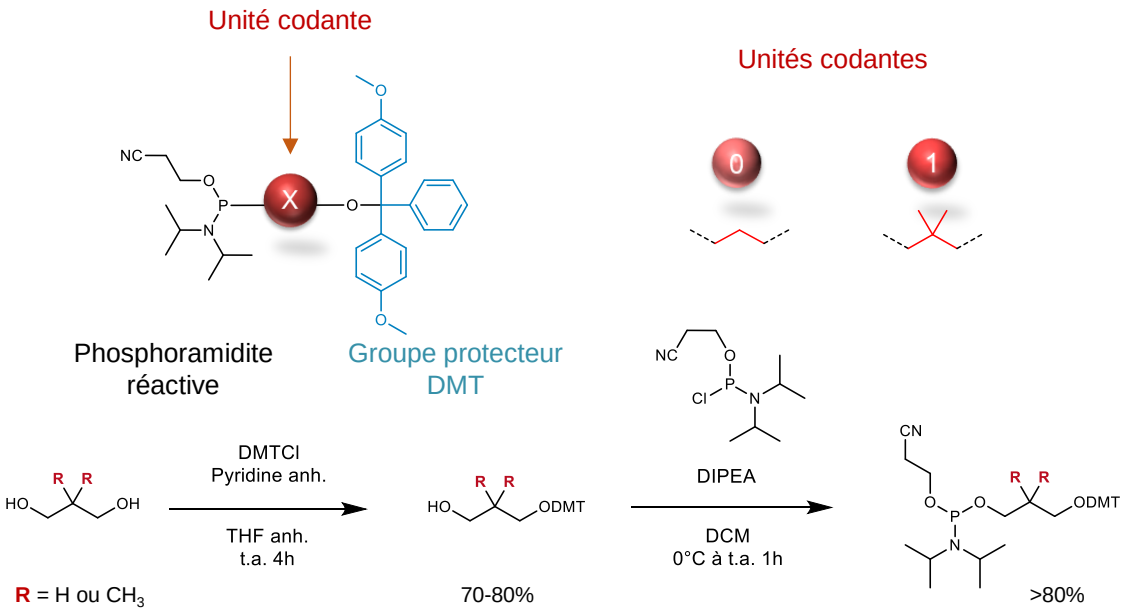


Figure I. 20 : Unités codantes utilisées dans les polyphosphodiesters numériques.

L'usage de l'approche itérative phosphoramidite sur un support solide fonctionnalisé (polystyrène) a permis d'obtenir des polymères à séquences définies d'une longueur allant jusqu'à 24 unités. Dans ce cas, les analyses ESI-HRMS, MALDI-HRMS, RMN et SEC ont indiqué la présence effective d'un polymère monodisperse, ainsi que la séquence initialement encodée. La même année, notre groupe a utilisé ces monomères synthétiques pour la préparation d'un long polymère par approche phosphoramidite automatisée. Ces travaux confirment la faisabilité d'obtenir une longue séquence définie sans l'usage de nucléotides naturelles et démontrent le premier mot encrypté dans une seule chaîne polymère (Figure I. 21). L'uniformité du polymère codant « Macromolécule » ayant un degré de polymérisation de 104 unités (DP = 104) a pu être confirmée par ESI-MS en mode négatif sous des états multichargés, par déconvolution du spectre donnant sa masse exacte, et également par HPLC grâce à 3 bases thymidines (actives en UV) incorporées en bout de chaîne.

La chimie automatisée utilisée dans cette étude repose sur le même principe que celle utilisée dans les robots synthétiseurs d'ADN puisque c'est un de ces instruments qui a en effet été utilisé (ABI 3900). Le cycle itératif phosphoramidite « synthétique » est résumé dans le Schéma I. 7 ci-dessous, où les étapes *i)* à *vi)* sont les mêmes que dans le cycle impliquant les nucléotides naturelles vu précédemment (Schéma I. 5).

Grâce à la chimie des polyphosphodiesters, d'autres études intéressantes ont pu être effectuées pour l'obtention de macromolécules informationnelles, ouvrant notamment la voie à de la modification d'informations. Les réactions post-polymérisations par CuAAC en sont un exemple. En incorporant des alcynes sur la chaîne polymère, des réactions click sélectives ont pu être effectuées sans altérer l'information contenue dans la chaîne. Cet outil servirait éventuellement à ajouter et/ou modifier des données numériques.[297] Par ailleurs, l'introduction de dérivés nitrobenzyles photo-clivables dans la séquence a permis la préparation de polymères numériques effaçables. Dans la même étude,

l'utilisation de deux composés isobariques dont un photo-clivable et l'autre inerte aux excitations UV a laissé la possibilité de modifier la séquence binaire initialement encodée.[298] Enfin, la synthèse photo-contrôlée de poly(phosphodiesters) numériques en utilisant le groupe NPPOC a également pu être investiguée. L'utilisation de ce groupe protecteur photo-labile généralement utilisé en synthèse d'ADN sur micro puces (2.2.4), a démontré la possibilité d'employer une stratégie similaire à la construction des biopolymères, mais pour les macromolécules synthétiques.[299]

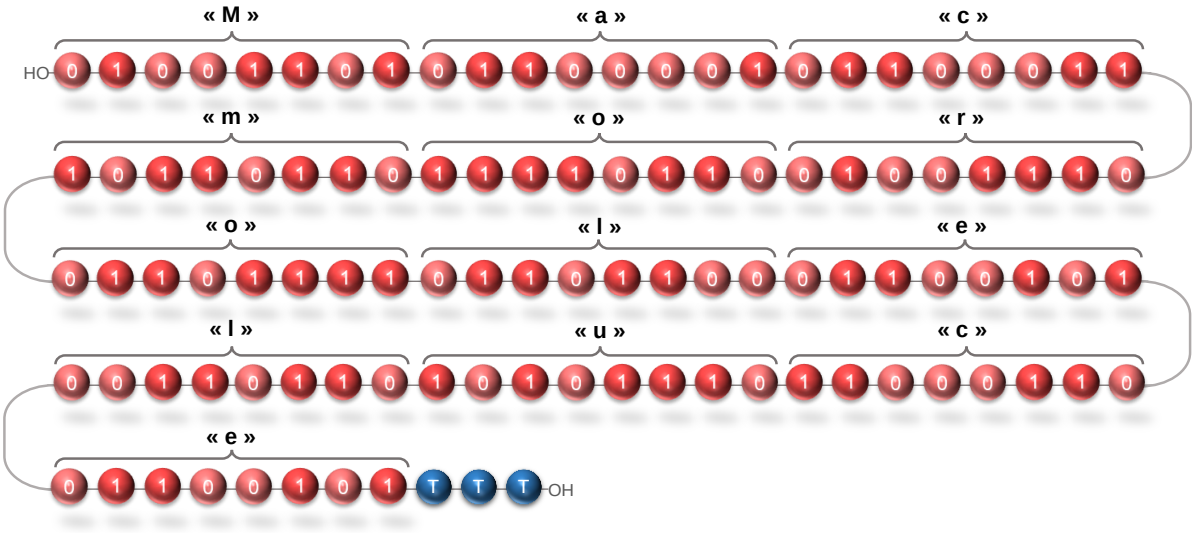


Figure I. 21 : Polymère à séquence définie non naturel encodant le mot « Macromolecule » à l'aide du code ASCII.

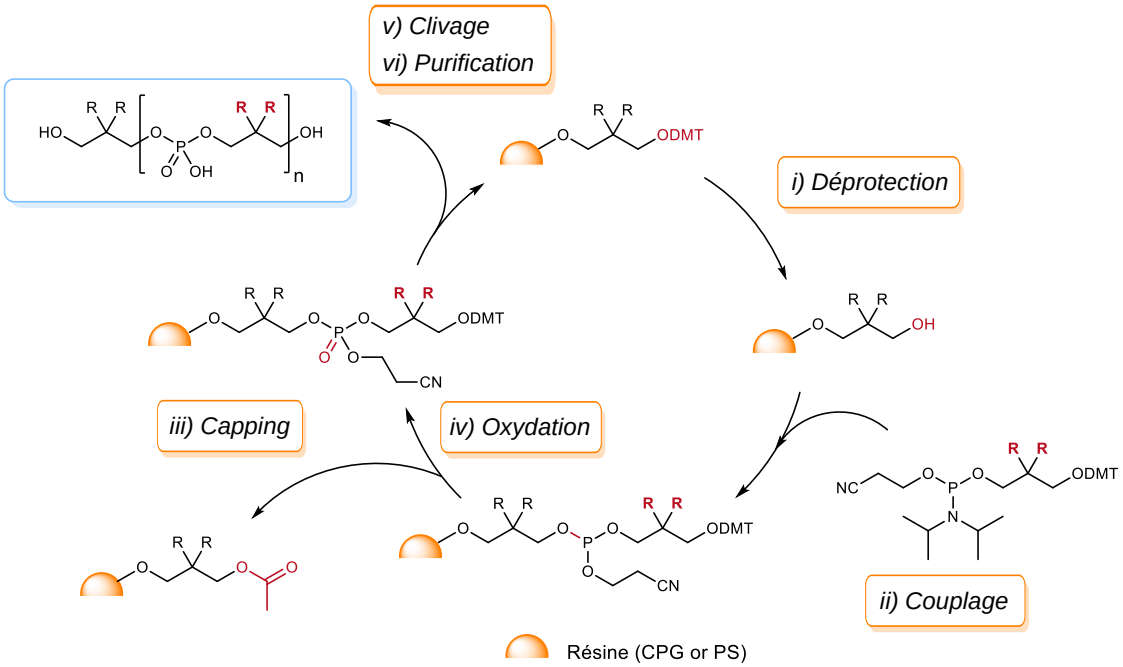


Schéma I. 7 : Cycle itératif pour la synthèse de polyphosphodiesters synthétique par approche phosphoramidite.

Cependant, malgré les innovations révélées dans ces exemples, des améliorations restent à faire pour la lecture de ces polymères. Particulièrement, les longues séquences

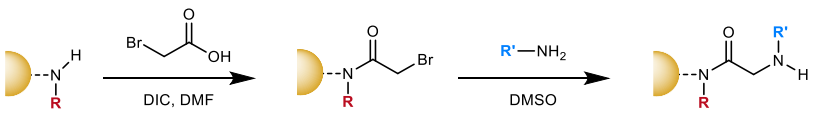
comme celle codant « Macromolecule » dont l'analyse MS/MS pour reconstruire bit par bit la chaîne informationnelle s'est révélée fastidieuse et longue. En effet, l'absence de liaison labiles permettant de sélectivement fragmenter le polymère complique et rallonge le déchiffrage de ce dernier. C'est pourquoi, les études parallèles effectuées avec les poly(alcoxyamines amides) (3.2.2.2) impliquant des liaisons alcoxyamines labiles pouvant sélectivement être fragmentées (3.3.1) ont été inspirantes pour la construction d'un nouveau design de polymères synthétiques : l'incorporation d'espaceurs clivables.

Cette nouvelle approche utilisant la chimie de phosphoramidite sera discutée plus en détail dans la partie 3.4, qui permettra notamment d'introduire les objectifs de cette thèse.

3.2.2. Approches itératives orthogonales

Une approche orthogonale signifie qu'aucun groupe protecteur n'est requis pour faire croître la chaîne polymère. Cette stratégie est intéressante puisque par la variété des différents groupes fonctionnels, elle permet de faire grandir rapidement la chaîne du polymère sans utiliser d'étapes additionnelles de déprotection. Deux principaux procédés peuvent être utilisés dans ce cas. AA + BB utilise deux monomères bifonctionnels alternativement ajoutés où A peut réagir avec B, tandis que l'approche AB + CD implique l'ajout successif de AB et CD où B réagit seulement avec C et D seulement avec A. D'autres variantes peuvent également être considérées avec par exemple « AA + BC ».

Zuckerman et ses collègues ont été à l'origine des premiers travaux dans la synthèse d'oligomères synthétiques à séquence définie en utilisant cette approche avec une stratégie de sous-monomère. En prenant exemple sur la SPPS qui repose sur le procédé AB + AB, ils ont pu mimer la synthèse de peptide mais dans ce cas, en impliquant des monomères *N*-substitués sans utiliser de groupe protecteur (Schéma I. 8). En utilisant i) un acide bromoacétique en présence de *N,N'*-diisopropylcarbodiimide (DIC), la fonction amine *N*-substituée attachée à la résine peut réagir avec la fonction acide carboxylique du dérivé bromé. Ensuite, ii) une autre amine libre également *N*-substituée peut simplement réaliser une substitution nucléophile sur le carbone en *α* du brome laissant une nouvelle amine libre sur la chaîne oligomère, capable par la suite de réitérer l'étape i). Cette méthode a effectivement été efficace pour obtenir des oligo(*N*-substituées glycines) (peptoïdes) d'une longueur allant jusqu'à 25 unités dans des rendements raisonnables.[300]



R, **R'** = alkyl, aryl, propylamine …

Schéma I. 8 : Principe de synthèse de peptoïdes par voie supportée (Zuckerman).[300] DMF, DMSO.

Depuis ces travaux, de nombreuses études ont été effectuées pour l'obtention de polymères à séquence définie en utilisant une approche itérative orthogonale. Les principales d'entre elles sont discutées dans les sous-sections ci-après.

3.2.2.1 La chimie click CuAAC : outil multifonctionnel

La chimie click introduite par Sharpless et Meldal (justement récompensés l'année dernière d'un Prix Nobel pour leurs travaux dans ce domaine) est un outil très utile dans le contrôle des séquences polymères. La cycloaddition 1,3-dipolaire de Huisgen catalysée par le Cuivre(I), plus couramment appelée la CuAAC, est une des réactions click les plus utilisées en chimie des polymères à séquence contrôlée ces dernières années. Cet outil permet non seulement le contrôle absolu de la séquence, mais aussi les modifications post synthèses pouvant être utiles dans l'obtention de macromolécules multifonctionnelles.[301]

Cette approche « AB + CD » nécessite un monomère portant un alcyne (A) et un acide carboxylique (B) ainsi qu'un autre monomère portant une amine (C) et un azoture (D). En l'occurrence, l'utilisation de groupes protecteurs n'est pas nécessaire puisque la fonction B ne réagit qu'avec la C, et la D seulement avec la A. Grâce à cette stratégie, un premier exemple de polymère numérique synthétique a pu être publié en 2014.[16] En variant la chaîne pendante en *α* du carbonyle sur le monomère AB, notre groupe a prouvé qu'il était possible d'encoder de l'information artificielle en utilisant cette stratégie (Figure I. 22).[16]

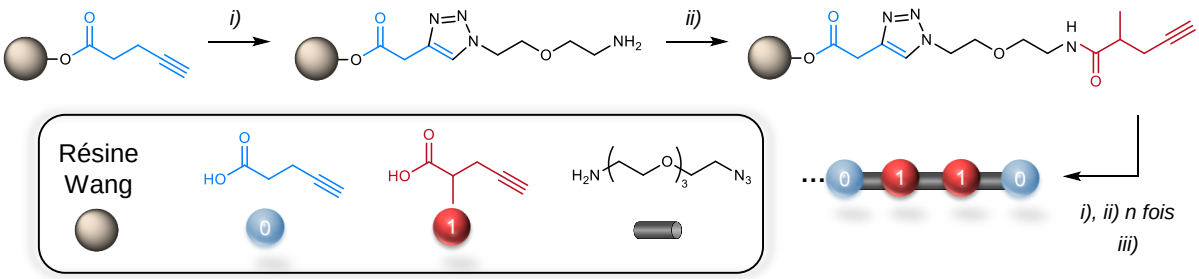


Figure I. 22 : Principe d'encodage en utilisant la chimie CuAAC.[16]
i) CuAAC : 1-amino-11-azido-3,6,9-trioxaundecane, CuBr, 4,4'-di-n-nonyl-2,2'-bipyridine, THF anhydre, 20h. ii) Amidification : Acide alcynoïque (« 0 » ou « 1 »), benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate, DIPEA, DCM, 3-4h. iii) Clivage : TFA, 2h.

Alors que cette étude a été réalisée sur support solide avec une résine polystyrène, notre groupe a également démontré la faisabilité d'appliquer cette chimie à un support soluble. Dans ce cas, le procédé a été automatisé et la densité d'information a été augmentée en utilisant un alphabet à 4 monomères codant chacun pour deux bits (dyades, non illustré ici).[302]

Néanmoins, malgré une chimie efficace dans la construction d'oligomères à séquence définie, leur lecture en masse laisse apparaitre un clivage compétitif entre la liaison amide désirée et la liaison éther (espaceur) non voulue, rendant complexe leur séquençage. Afin de pallier ce problème, le groupe de Laurence Charles spécialisé dans l'analyse en spectrométrie de masse, a démontré que la contrainte observée peut être évitée en remplaçant simplement l'espaceur glycol éther en une chaîne alkyle de même taille.[303] Cet exemple souligne l'avantage de la flexibilité dans l'élaboration de macromolécules synthétiques numériques.

3.2.2.2 Les poly(alcoxyamine amide)s

En 2015, notre groupe a montré une nouvelle stratégie pour encoder de l'information dans les polymères non naturels. La labilité de la liaison alcoxyamine incorporée entre différentes unités codantes dans une chaîne polymère définie, a permis de faciliter son séquençage par MS/MS (plus de détails sur l'optimisation de ces conditions analytiques sont données plus tard (partie 3.3.1). Cette stratégie repose sur l'utilisation de deux étapes de couplages chimiosélectives successives (stratégie AB + CD), à savoir la réaction d'une amine primaire (A) avec un anhydride d'acide symétrique (C) et le couplage d'un radical carboné (B) avec un nitroxyde (D) (Figure I. 23).[19] Alternativement, la même chimie peut être utilisée pour construire des poly(alcoxyamine amide)s de manière convergente en préparant quatre monomères codant chacun pour deux bits (dyades).[304]

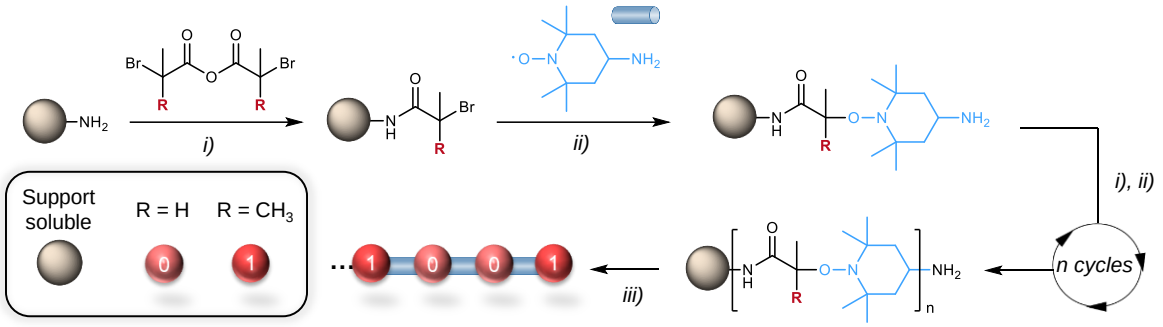


Figure I. 23 : Principe d'encodage en utilisant des poly(alcoxyamines amides).[19]
i) THF, DIPEA ou K_2CO_3. (ii) CuBr, Me6TREN, DMSO, (iii) TFA, CH_2Cl_2.

Par ailleurs, ces polymères présentent des propriétés physico-chimiques intéressantes grâce à la liaison alcoxyamine fragile. En chauffant le poly(alcoxyamine amide), une dégradation peut facilement être observée par MS et SEC. Dans la première étude énoncée ci-dessus, il a été démontré que cette décomposition de chaîne polymère peut être contrôlée. Cette caractéristique semble être directement applicable à de l'écriture d'information qui pourrait par exemple, après avoir été décodée, être sélectivement supprimée.[19]

3.2.2.3 Les dérivés phosphodiesters

Les phosphodiesters se sont révélés comme étant une alternative de choix dans le codage d'information numérique à l'échelle moléculaire comme introduit précédemment avec une approche itérative « AB + AB » (3.2.1). Dans cette section, il est démontré qu'une approche itérative orthogonale est également possible pour la construction de poly(alcoxyamine phosphodiester)s.[305,306]

Parallèlement aux études menées sur les poly(alcoxyamine amide)s (3.2.2.2), notre groupe s'est également intéressé à la combinaison des propriétés avantageuses de la liaison alcoxyamine et de la chimie des phosphodiesters.[305] Dans ces travaux, une stratégie AB + CD a été choisie : un alcool primaire (A) attaché sur un support solide peut générer une liaison phosphite avec un monomère phosphoramidite (C). Après son oxydation en phosphate, la deuxième fonctionnalité du monomère phosphoramidite (radical carboné caché) (C) peut réaliser un couplage radicalaire avec un nitroxide (D) (Figure I. 24). Les oligomères ainsi obtenus ont par la suite pu facilement être séquencés

par spectrométrie de masse en profitant de la fragmentation sélective NO-C de l'alcoxyamine.[305]

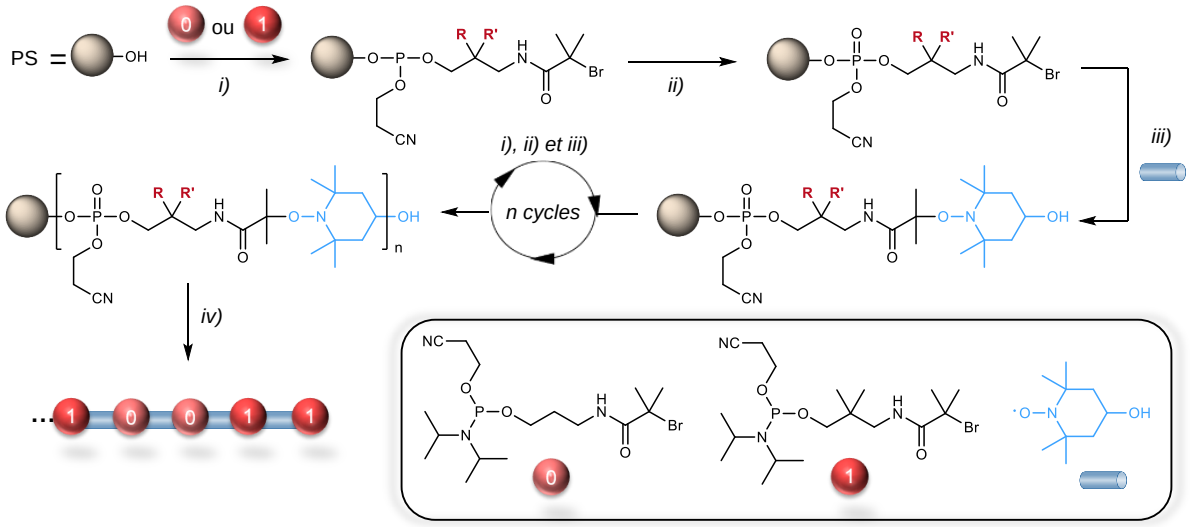


Figure I. 24 : Principe de synthèse d'oligomères à séquence définie combinant chimie phosphoramidite et couplages radicalaires.[305]
i) Couplage phosphoramidite : ACN, tétrazole, t.a.. ii) Oxydation : I_2, 2-6-lutidine, THF/H_2O, t.a.. iii) Couplage radicalaire : CuBr, Me6TREN, DMSO. iv) Clivage : piperidine, ACN, et $MeNH_2$, NH_4OH, H_2O.

Cette approche a également été appliquée à un alphabet augmenté en sélectionnant différents nitroxides, combinés à différents monomères phosphoramidites, portant chacun d'entre eux des chaînes alkyles différentes.[306]

3.2.2.4 Oligocarbamates (polyuréthanes)

Cette classe d'oligomère a montré ces dernières années son utilité pour des applications d'anticontrefaçons,[307] notamment pour des implants intraoculaires à base de méthacrylate.[308] Dans cette étude, notre groupe a utilisé une stratégie « AA + BC » sans l'usage de groupes protecteurs. Une résine Wang modifiée avec une fonction hydroxyle est capable de réagir *i)* avec le carbonate de *N,N'*-disuccinimidyl (DSC, monomère AA) pour former l'alcoxycarbonyle. Par la suite, ce dernier peut réagir *ii)* avec un aminoalcool (BC) portant une chaîne latérale modifiée (groupe R, Figure I. 25A) formant préférentiellement le carbamate correspondant.

La répétition du cycle itératif donne la séquence définie correspondante. En variant le groupe R de l'aminoalcool, l'oligomère peut effectivement contenir de l'information numérique, capable d'être séquencée *in fine* par MS/MS. En outre, l'utilisation de polyuréthanes codés est avantageuse pour une telle lecture grâce au groupe acide carboxylique à l'extrémité de leur chaîne α : cette fonctionnalité facilite la lisibilité du polymère par MS/MS en mode négatif puisqu'elle se déprotone toujours sur cette extrémité, induisant préférentiellement et prévisiblement la fragmentation C-O du carbamate.[307] Lorsque la macromolécule était finalement obtenue et caractérisée, son incorporation dans l'implant méthacrylate a été effectuée par polymérisation radicalaire *in situ* ou par stratégie de « gonflement/dégonflement » (en modifiant les paramètres de

solubilité du matériel). Enfin, le polyuréthane a facilement pu être extrait et le code retrouvé par ESI-HRMS.[308]

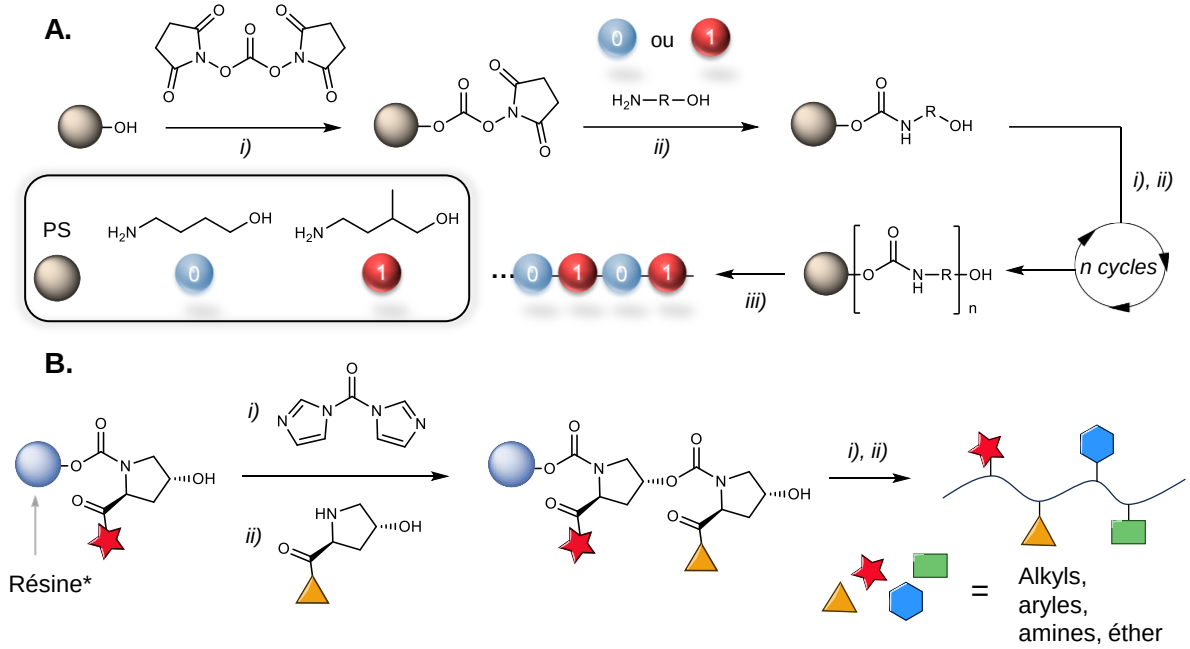


Figure I. 25 : Synthèse d'oligocarbamates à séquence définie.[308]
***A.** Encodage utilisant des dérivés aminoalcools. i) ACN, TEA, microonde, 60 °C. (ii) DMF, TEA, t.a.. (iii) Clivage : TFA/DCM. **B.** Synthèse d'oigocarbamates à l'aide d'hydroxyprollines.[309] *Matrice de silice fluorée. i) carbonyldiimidazole, DMF, t.a., 1h. ii) Monomère hydroxyproline, DMF, t.a., 17h.*

Subsidiairement, les oligocarbamates peuvent être modifiés en ajoutant des chaînes pendantes sur leur fonction amine pour ainsi former des poly(*N*-substitué uréthane)s.[310] Le polymère construit à l'aide d'un protocole itératif supporté par phase solide peut contenir de l'information numérique en ayant comme code des dyades (groupe R sur l'amine où R = Méthyle (00), Éthyle (01), Propyle (10) et Butyle (11)). Cette stratégie a par la suite été utilisée pour tester ses capacités de « dégradations et réparations ».[311] Effectivement, la dépolymérisation contrôlée en milieu basique (NaOH) sans trace de clivage aléatoire de la chaîne informationnelle a pu dans un second temps être « réparée » en présence de DSC et d'une amine secondaire. Cette dernière étude ouvre une voie intéressante pour la réparation à l'échelle moléculaire, potentiellement applicable à de l'édition d'information.[311] Par ailleurs, les oligocarbamates ont également été étudiés d'une manière différente en utilisant des dérivés hydroxyprolines, démontrant la variété de monomères disponibles en utilisant ce type de chimie (Figure I. 25B).[309]

Autrement, pour répondre à des besoins de quantité nécessaire à certaines applications (ex. marqueurs moléculaires pour des tonnes de plastiques), le groupe de R. Szweda a récemment utilisé une approche itérative en un seul pot (en anglais, *one pot synthesis*) permettant de générer jusqu'à 50 g d'oligomères à séquence définie. Cependant, la stratégie reste perfectible puisque la longueur maximale atteinte dans cette étude n'a été que de 5 unités.[312]

3.2.2.5 Les thiolactones

La chimie des thiolactones est très intéressante puisque par sa polyvalence, elle permet d'offrir un outil performant en ingénierie macromoléculaire. Grâce à son ouverture de cycle en présence de groupe amine ou hydroxyle libérant ainsi un thiol, elles sont capables de réagir avec une grande variété de groupes fonctionnels.[313] Le groupe de F. Du Prez a beaucoup été impliqué dans l'utilisation des thiolactones pour la préparation de polymères à séquence définie ces dernières années.[313,314] En addition des propriétés de post-polymérisation intéressantes que ces composés peuvent apporter, les thiolactones constituent des monomères efficaces dans la construction de chaînes polymères définies.[17] Un exemple récent a été publié par le même groupe. Il repose sur la synthèse de séquences d'oligo(thioéther-uréthane)s codées pour une éventuelle application en stockage de données.[315] La première étape consiste à ouvrir le cycle thiolactone en présence d'amine primaire. Le thiol ainsi libéré *in situ* peut directement réagir avec un époxyde laissant un alcool libre, prêt à réagir avec une nouvelle thiolactone pour étendre la chaîne (Figure I. 26A). En utilisant deux types de monomères, une chimiothèque de 12 amines et 7 époxydes a pu être préparée prouvant ainsi la faisabilité d'un alphabet moléculaire augmenté. Après séquençage par spectrométrie de masse en tandem (MS/MS), les différentes unités composant l'oligomère ont pu être attribuées laissant la possibilité d'appliquer cette chimie à du stockage d'information.

L'année d'après, le même groupe a utilisé cette même stratégie pour répondre à un des défis importants mais difficile à surmonter dans le domaine des macromolécules numériques : la réécriture d'information. En effectuant un clivage orthogonal d'un groupement thioester par aminolyse (Figure I. 26B), ils sont parvenus à déplacer un morceau de la séquence résultant en une modification du code initialement encrypté.[316]

Inspiré par ses premiers travaux sur la réaction des thiols avec des alcènes (réaction thiol-ene)[317,318] et des développements du groupe d'Alabi en utilisant cette même chimie,[319] Du Prez et ses collègues ont aussi démontré qu'un protocole automatisé peut facilement être développé. Dans ce cas, le thiol libéré après ouverture de cycle par un aminoalcool peut réagir avec un acrylate ou acrylamide.[320] La chaîne carbonée modifiable de l'aminoalcool laisse ainsi la possibilité de réaliser un alphabet moléculaire à souhait.

Malgré l'outil intéressant que représente l'utilisation de la thiolactone pour apporter un contrôle absolu dans la séquence, certaines contraintes dans la stratégie utilisée dans ces études subsistes. Premièrement, la préparation du composé isocyanate thiolactone utile à l'extension de la chaîne requiert l'utilisation du réactif cancérigène triphosgène. D'autre part, des temps de couplage trop longs ont été indispensables dans la formation de l'oligomère des premières études discutées ci-dessus (Figure I. 26A, 1h pour couplage amine et 1h pour l'extension la chaîne). Des défauts qui peuvent être problématiques en considérant une application pour un stockage d'information comme détaillé précédemment (3.1).

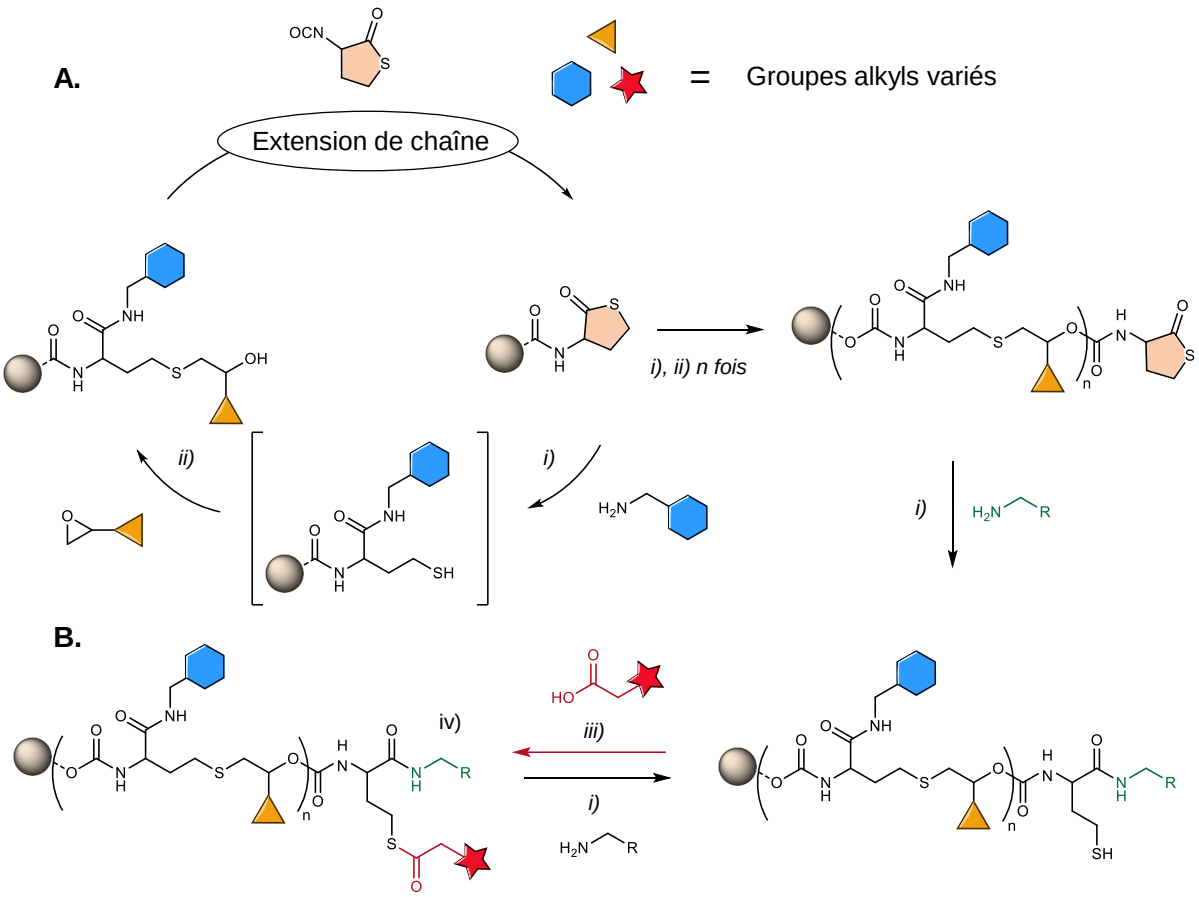


Figure I. 26 : Utilisation des thiolactones pour la synthèse de polymères à séquence définie. ***A.*** *i) Ouverture de cycle suivi de ii) la réaction thiol-époxy (1 seul pot) : Amine, époxyde, DBU, Diméthylphenylphosphine, THF (H_2O), ≈ 5min. Extension de chaîne : $CHCl_3$ anhydre, α-isocyanato-γ-thiolactone, $Zr(O_2C_5H_7)_4$, agité 1h.* ***B.*** *Modification de la chaîne. Après aminolyse du thiolactone, iii) thioesterification : DIC, DMAP, DMF, agité 30min.*

3.2.2.6 Réactions de Passerini et Ugi

Les réactions multi-composantes représentent un outil intéressant pour la construction d'oligomères à séquence définie. Particulièrement, les réactions d'Ugi et Passerini qui ont été les plus utilisées ces dernières années.

La réaction découverte par Mario Passerini en 1921 est une réaction à trois composants utilisant un acide carboxylique, une cétone ou aldéhyde, et un isonitrile pour former un *α*-acyloxycarboxamide, alors que celle découverte par Karl Ugi en 1959 est une réaction à quatre composantes avec une amine primaire en plus des trois autres à l'instant énoncées.[321]

Récemment, plusieurs exemples d'applications de tels procédés ont été publiés par le groupe de Meier avec la réaction de Passerini pour stocker de l'information dans des oligomères synthétiques.[322,323] Ils ont effectivement démontré qu'en utilisant une série de différents monomères aldéhyde ou cétone, il était possible de construire une chaîne croissante polymère en y contrôlant parfaitement la séquence. La structure de la macromolécule résultante a pu *in fine* être confirmée par spectrométrie de masse, RMN et GPC indiquant une potentielle application pour du stockage d'information. La Figure I. 27 illustre le principe de cette méthode en utilisant la réaction de Passerini, qui peut être

appliquée avec une stratégie orthogonale couplée à une réaction thiol-ene,[324] mais aussi une stratégie de protection/déprotection (ester de benzyle protégeant l'acide carboxylique bleu, non illustré ici) qui a notamment permis la synthèse d'un oligomère de 20 unités à l'échelle du gramme (44% de rendement).[325] Cependant, comme souligné précédemment avec la chimie des thiolactones, les temps de couplage restent pour le moment trop longs pour une réelle application d'écriture de données.

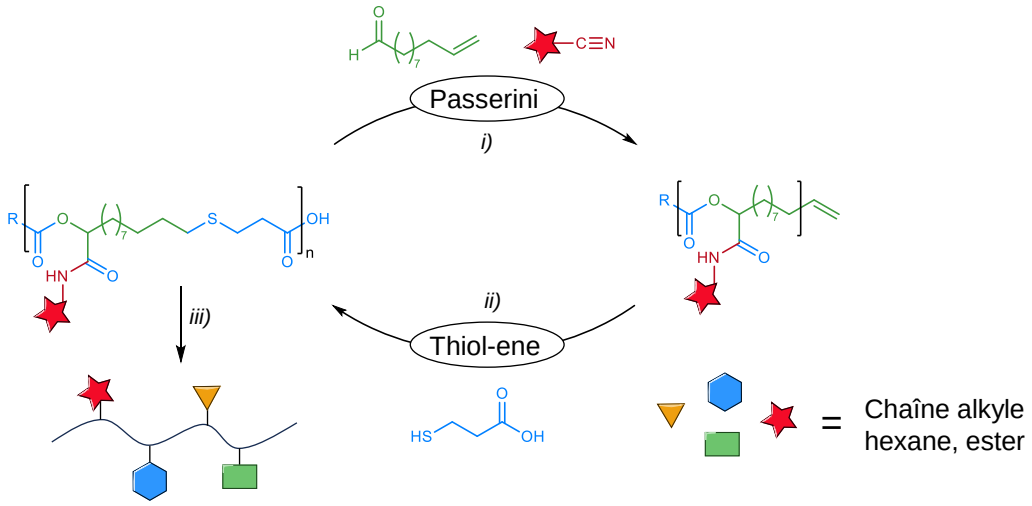


Figure I. 27 : Réaction de Passerini couplée à la thiol-ene pour la formation d'oligomères à séquence définie.
i) Aldéhyde et isonitrile, DCM, 2 à 24h. ii) Thiol, 2,2-Diméthoxy-2-phenylacetophenone (initiateur photochimique, 5 mol%), UV, 2h. iii) Cycle répété n fois donnant une séquence définie.

3.2.2.7 Polyesters

Les polyesters sont également de bons candidats pour un stockage à l'échelle moléculaire. Le groupe de Kim a récemment publié la synthèse de copolymères d'acides α-hydroxylés construit par combinaison de méthodes divergentes et convergentes à partir de quatre dyades de monomères (poly(acide phényllactique-co-lactique) (PcL) Figure I. 28). Le séquençage par MALDI-TOF du polymère a facilement permis de retrouver le code binaire initialement encodé (« SEQUENCE »).[326] La stratégie de cette étude repose sur la protection/déprotection orthogonale des groupes alcool et acide carboxylique présents sur le monomère d'acide lactique (0 et 1, Figure I. 28) (cette méthode n'est donc pas basée sur une chimie orthogonale contrairement aux autres approches discutées dans cette partie 3.2.2). *i)* Une fois la dyade établie après couplage entre 0 et 1 en présence de EDC·HCl et DMAP, une étape divergente peut prendre place impliquant *ii)* l'élimination sélective du groupe TBDMS par du $BF_3.Et_2O$ libérant d'un côté un alcool, et parallèlement *iii)* le déblocage sélectif du groupe ester de benzyle en présence de Pd/C et H_2 libérant un acide carboxylique. Une nouvelle étape convergente *i)* peut alors être effectuée entre les deux produits ainsi obtenus. La répétition de ce processus donne *in fine* un polymère à séquence définie avec un contrôle absolu sur la séquence.

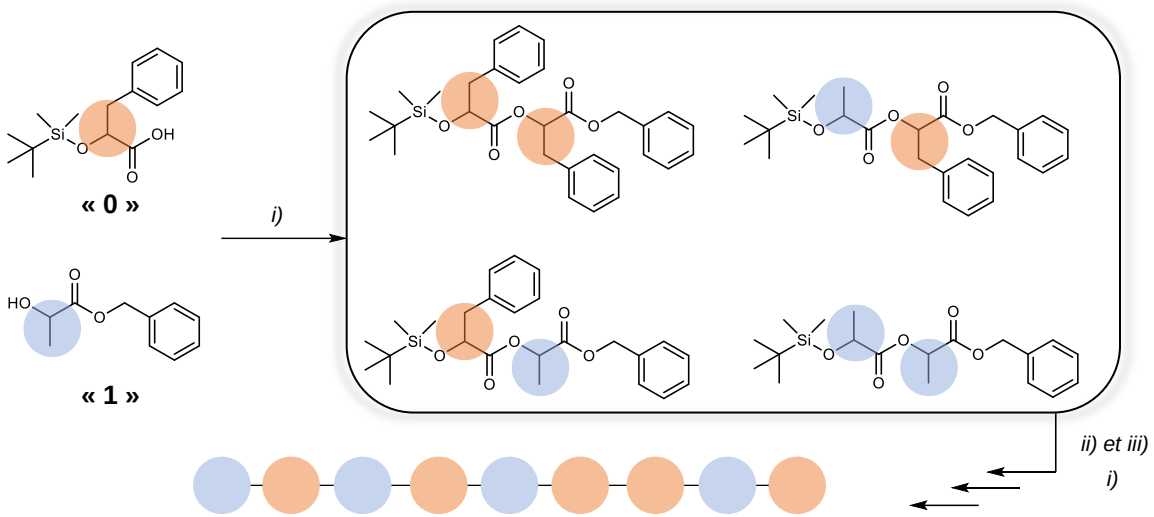


Figure I. 28 : Synthèse de polyester par méthode convergente pour du stockage d'information. i) EDC·HCl, DMAP, 8 h. ii) BF_3·Et_2O, t.a., 6 h. iii) Pd/C, H_2, t.a., 8h.

Récemment, le même groupe utilisait cette stratégie en semi-automatisant leur synthèse de polyesters, toujours en contrôlant parfaitement la séquence.[327] Ces deux exemples démontrent que la synthèse de polymères à séquence définie n'est pas forcément limitée à une approche étape par étape classique, mais peut aussi être améliorée en réfléchissant à une approche convergente, amenant de la complexité efficacement et rapidement. De plus, ces approches sont intéressantes par l'utilisation de dérivés d'acides polylactiques (PLA), principalement pertinents par leur propriétés biosourcés, qui ont pu être modifiés afin d'y encoder de l'information.

3.2.2.8 Les dendrimères

Depuis les premiers travaux dans les années 1980 sur les dendrimères,[328] ces macromolécules ont récemment suscité de l'intérêt pour un stockage de données à l'échelle moléculaire. Ces systèmes complexes, non linéaires, sont formés par synthèse divergente ou convergente. De manière divergente, le dendrimère est construit à partir d'un noyau multifonctionnel, qui est étendu grâce à de multiple réactions. De manière convergente, le polymère est formé grâce à une série de réactions de petites molécules, formant petit à petit le système souhaité.

Une preuve de concept a récemment été démontrée par le groupe de Z. Zhang qui a exploité des couplages de Michael thiol-maléimide permettant l'arrangement précis des bits 0 et 1 dans un dendrimère uniforme.[329] Par ailleurs, le même groupe a cette année utilisé des polymères amphiphiles à séquence contrôlée, qui s'auto-assemblaient en micelles numériques.[330] La structure parfaitement définie des oligouréthanes est attachée à un dendron composé de PEG, facilitant sa désorption en MALDI. En outre, l'affinité des ions Na^+ avec le PEG et le clivage sélectif des liaison benzyles-carbamates a permis de retrouver sans difficulté l'information numérique encodée dans la micelle. De plus, cette étude démontre un potentiel intéressant pour une application biologique, en administrant les micelles codées *in vivo* à des souris. Cela pourrait notamment être exploité pour la nanomédecine de précision (incorporation de brin d'ARN, quantification cellulaire sans marquage spécifique, etc.) ou encore l'étude des protéines couronnées.

3.2.3. Autres approches de stockage moléculaire

3.2.3.1 Processus photochimiques

Finalement, d'autres approches itératives orthogonales moins classiques ont été explorées. Par exemple, le groupe de C. Barner-Kowollik a publié en 2016 la synthèse de macromolécules à séquence-définie par photo-ligation.[331] En utilisant une série de différents benzaldéhydes photo-réactifs portant conjointement une fonction diène et diénophiles, une réaction photochimique a pu induire, après chauffage, une réaction de Diels-Alder. La répétition de ce processus a permis de construire un polymère à séquence définie croissant de manière symétrique à partir d'un noyau diène (dans ce cas, le 1,6-hexylbismaleimide).

3.2.3.2 Processus d'autoassemblages

L'utilisation de processus d'autoassemblage peut également être utilisé pour préparer des polymères informationnels, comme démontré par le groupe de F. Scott, en impliquant des peptoïdes préparés par la méthode de Zuckerman discutée précédemment (3.2.2).[332] Alternativement, le groupe de Colquhoun a récemment considéré des foldamères pour un stockage d'information moléculaire. La sélection de différentes molécules « tweezer » (molécules contenant une cavité capable de lier d'autres entités par chimie non-covalente) riches en électrons π, permet une interaction spécifique entre ces dernières et un motif aromatique polyimide sulfone. La liaison supramoléculaire sélective de certaines molécules tweezer vers des sites spécifiques du copolyimide permet d'identifier, par RMN et cristallographie, l'ordre d'une séquence contrôlée.[333]

3.2.3.3 Mélanges de molécules

Enfin, contrairement à l'information contenue dans une chaîne polymère définie, une grande quantité d'information peut également être encodée dans des mélanges complexes de petites molécules. Notamment, l'année dernière, le groupe de Rosenstein a démontré qu'en utilisant la réaction d'Ugi automatisée, ils est possible d'encoder une large quantité d'informations en synthétisant une chimiothèque de composés hautement diverse pouvant par la suite être retrouvée par spectrométrie de masse.[334]

En utilisant cette même stratégie d'encodage, le groupe de Meier a démontré en 2018 que, par approche combinatoire, un mélange complexe de molécules peut servir de clé moléculaire pour la communication de messages secrets.[335] En utilisant une série de 130 réactifs commercialement disponibles impliqués dans une réaction d'Ugi, 500 000 clés moléculaires ont pu équivoquement être fournies. La stéganographie, pratique de dissimulation d'informations, leur a finalement permis de réaliser une preuve de concept concrète. Par ailleurs, le même groupe a montré qu'en utilisant une diversité de composés commercialement disponibles (ex. des solvant organiques), il est facilement possible de définir un code basé sur le déplacement chimique (ppm) retrouvé. En sélectionnant huit molécules différentes et en instaurant une règle de lecture allant des champs faibles vers les champs forts (de gauche à droite sur le spectre), ils ont pu définir un code tel que dans un mélange : molécules présentes codent pour « 1 », non présentes codent pour « 0 ». Ainsi, dix tubes RMN contenant chacun des mélanges différents peuvent encoder dix lettres au total. En augmentant le nombre de composés présents en solution (24

composés permettant de coder 3 bits par molécules), le procédé peut également être appliqué à un séquençage par chromatographie en phase gazeuse en se basant sur le temps de rétention.[336] Plus récemment en utilisant le même principe, le groupe de Kéki a utilisé la spectrométrie de masse par ionisation en temps réel (DART) pour décoder un code binaire encrypté à l'aide de dérivés de nicotine. Grâce à la volatilité évidente de ces acides, la DART permet d'identifier la présence ou non d'un composé dans un mélange, laissant la possibilité d'encoder jusqu'à 22 bits par solution.[337]

Finalement, le groupe de Whitesides a montré qu'il était possible d'utiliser une diversité d'oligopeptides pour encoder de l'information numérique. Attachés sur des monocouches autoassemblées à l'intérieur de micro puits, l'analyse des différents oligomères liés au support par spectrométrie de masse a permis de retrouver les masses correspondantes.[338] Ceci démontre qu'un mélange de molécule peut aussi être un moyen de stockage à long terme, grâce à l'immobilisation de différents fragments sur un support solide.

Ces derniers exemples illustrent la possibilité d'encoder de l'information non pas sur une chaîne linéaire polymère, mais dans un mélange de composés bien distincts. Par ailleurs, cette approche permet dans certains cas d'utiliser des outils analytiques différents de la MS/MS, ce qui peut ouvrir la voie à une lecture simplifiée.

3.3. Méthodes de séquençage

Traditionnellement, et au même titre que les polypeptides et autres macromolécules biologiques d'intérêts (outre l'ADN), la spectrométrie de masse est l'outil principalement utilisé dans le séquençage des polymères à séquence contrôlée. Également, d'autres techniques ont pu être utilisées ces dernières années pour décoder l'information contenue dans les polymères non naturels. Ce sous-chapitre détaillera les méthodes les plus pertinentes. D'autres méthodes impliquant le séquençage spécifique de polyphosphodiesters seront vues dans la partie associée (3.4).

3.3.1. Principe et optimisation du séquençage par spectrométrie de masse

« La spectrométrie de masse en tandem (ou MS/MS) est une technique qui fournit des informations structurelles sur les molécules ionisées en se basant sur la façon dont elles se fragmentent ».[339] Effectivement, les différents clivages compétitifs des groupes fonctionnels sur la chaîne polymère permettent d'identifier des fragments macromoléculaires par leur différence de masse. Avant la MS/MS, le polymère doit d'abord subir une analyse MS afin de déterminer le rapport masse/charge (*m/z*) des ions d'intérêts. Cette analyse permet en outre de valider la structure complète du polymère et de vérifier son uniformité. Après cette première étape, une analyse MS/MS sur cette même macromolécule peut être effectuée : en sélectionnant judicieusement un ion ciblé (déterminé lors de la MS), l'analyse garantit une distribution des espèces ioniques à partir de ce seul ion précurseur. Par ailleurs, puisque l'ionisation MS/MS dépend de l'ion préalablement choisi, cette technique ne se limite pas aux polymères monodisperses : des données analytiques intéressantes peuvent effectivement être obtenues à partir de

mélanges polydisperses, permettant par exemple de mieux comprendre certains phénomènes survenus lors de la synthèse du composé visé.

Dans le cas des polymères synthétiques, les groupes fonctionnels peuvent intelligemment être choisis afin d'éviter des clivages compétitifs simultanés non désirés rendant difficile la lecture du spectre. C'est un des avantages de la construction à souhait d'un composé non naturel : le polymère synthétique peut être adapté à l'outil analytique, contrairement à ses analogues naturels limité par leurs propriétés structurelles.[340]

Dans cette logique, les premières améliorations notables pour le séquençage de polymères numériques remontent à 2015 avec les travaux du groupe de Laurence Charles et le nôtre, en utilisant des poly(alcoxyamine amide)s (discutés précédemment 3.2.2.2). En raison de la labilité des liaisons alcoxyamines, il n'existe pas de réaction de dissociation alternative concurrençant efficacement le clivage homolytique à faible énergie de ces liaisons entre le monomère TEMPO et l'unité codante 0 ou 1. En optimisant les conditions analytiques d'ionisation par électrospray, la macromolécule peut d'abord subir une première analyse MS, puis après MS/MS, les liens inter-bits sont rompus afin de retrouver l'information contenue dans le polymère.[19] Une illustration de cette stratégie est démontrée ci-dessous dans la Figure I. 29 avec un tétra(alcoxyamine amide).

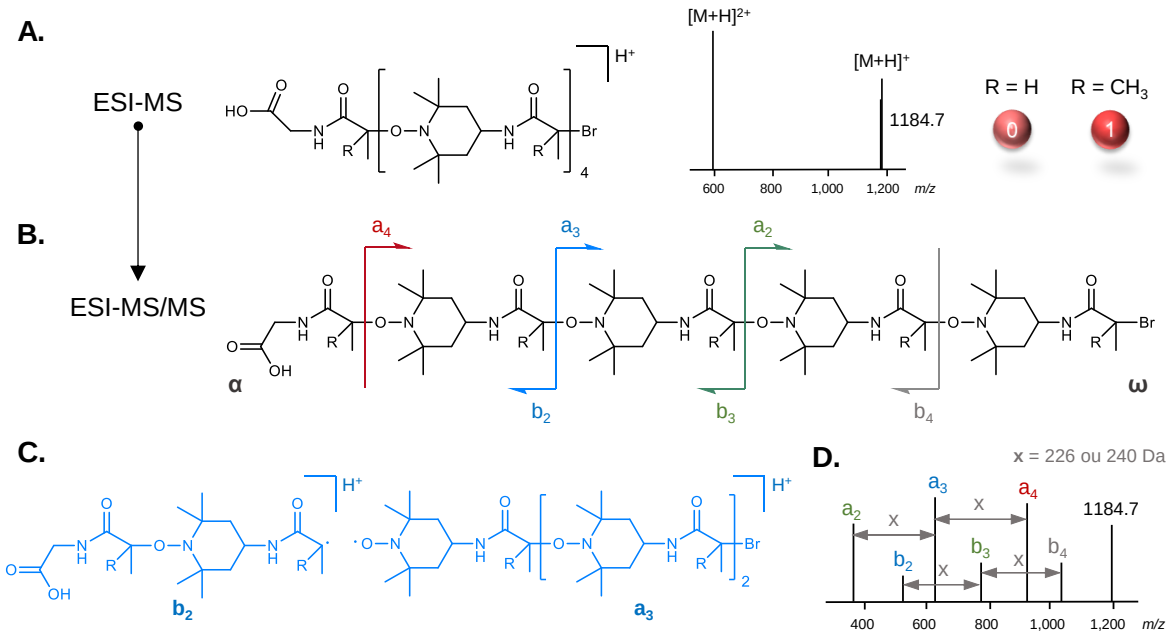


Figure I. 29 : Représentation du séquençage MS/MS d'un tétra(alcoxyamine amide).[19]
***A.** ESI-MS de l'oligomère donnant 2 états e charges sur le spectre de masse. **B.** MS/MS en sélectionnant l'ion précurseur à 1184.7 Da. **C.** Exemple de fragment retrouvé après analyse (ici, **b_2** et **a_3**). **D.** Représentation simplifiée d'un spectre obtenu après MS/MS.*

Les fragments générés après ionisation comportent des charges définies par la structure du composé analysé. La Figure I. 29 ci-dessus montre notamment deux bouts de chaînes α et ω différents. La différence de la valeur *m/z* entre deux fragment α ou ω, permet d'attribuer la valeur du bit encodé. Par exemple, dans ce cas, des intervalles de 226 ou 240 Da entre les fragments α ou ω indiquent la présence dans la séquence du motif 0 ou 1, respectivement. Finalement, la couverture des pics obtenus dans le spectre MS/MS

correspondant aux clivages inter-bits permet de retrouver le code initialement encrypté.[341] L'incorporation des liaisons alcoxyamines à souhait dans la chaîne polymère démontre l'intérêt de traiter le stockage d'information avec des macromolécules synthétiques. Le fait de pouvoir modifier aisément leur structure a notamment permis de faciliter leur stabilisation de charge : à l'instar des oligocarbamates (3.2.2.4), lorsque l'analyse est faite en mode négatif, l'acide carboxylique présent sur le bout de chaîne α facilite le séquençage de la chaîne macromoléculaire puisque ce groupe fonctionnel est le seul capable de stabiliser la charge négative, facilitant ainsi la lecture de ce fragment.[342]

D'autres exemples ont pu être considérés comme l'insertion de ponts disulfures pour contrôler le clivage entre chaque unité monomère codante.[343] Le thiol a également été utilisé pour l'insertion de liaisons C-S dans la séquence définie à l'aide de motifs dithiosuccinimides,[344] de réactions thiol-maleimide,[345] ou encore de couplage entre un dithiol et un allyle acrylamide,[319] permettant dans chacun des cas de décoder facilement les différentes unités par MS/MS. Pourtant, malgré l'existence d'une grande diversité de groupes fonctionnels en chimie synthétique, toutes les possibilités facilitant le séquençage de polymère à séquence contrôlée n'ont manifestement pas été explorées à ce jour. Cependant, les amides, inspirés par les polypeptides, et les alcoxyamines, introduites par les groupes de Laurence Charles et le nôtre, restent aujourd'hui les fonctions chimiques les plus courantes pour lire commodément les polymères numériques.

Alternativement, d'autres techniques de spectrométrie de masse existent pour identifier la structure de polymères synthétiques, et particulièrement, l'analyse MALDI-TOF. Cette méthode de désorption/ionisation laser assistée d'une matrice est réputée comme étant un processus d'ionisation doux, et fournit des spectres de masse simples qui montrent principalement des ions moléculaires à charge unique avec une fragmentation quasi inexistante.[346] De ce fait, cet outil est utile pour l'analyse de polymères à haut poids moléculaire, ou encore, l'identification de séquences comonomères dans les copolymères.[347,348] Cependant, même si certains groupes ont utilisés l'analyse MALDI-TOF pour déterminer la séquence d'un enchaînement de monomères dans des macromolécules synthétiques,[349] la MS/MS reste tout de même aujourd'hui, la méthode de spectrométrie de masse la plus utilisée dans l'identification de tels polymères.

Par ailleurs, la MS/MS peut être améliorée et couplée à certaines techniques de séparation, telle que la mobilité ionique. La spectrométrie de mobilité ionique (IMS) "à temps de dérive" mesure le temps nécessaire à un ion pour migrer à travers un gaz tampon en présence d'un faible champ électrique.[350] Ce procédé couplé à la MS permet une séparation préliminaires des ions moléculaires, avant de les analyser par masse (TOF par exemple), facilitant leur interprétation (Figure I. 30). À noter que dans ce cas, une seule analyse permet d'identifier les différentes espèces ioniques atteignant séparément l'analyseur de masse.

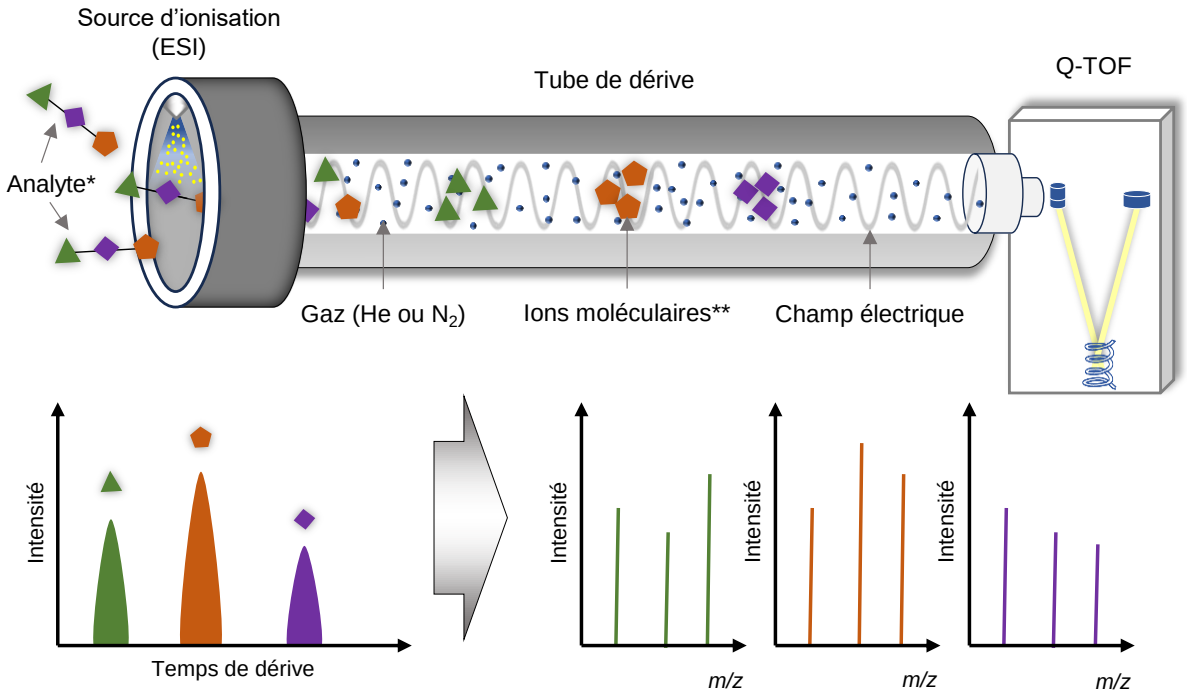


*Figure I. 30 : Principe d'analyse par spectroscopie de masse à mobilité ionique. *Par exemple, un mélange de composés organiques, ou bien un polymère. **Différents ions formés par ESI ayant un temps de dérive différent au contact du gaz et du champ électrique.*

En 2019, le groupe de Laurence Charles et le nôtre avons tiré profit de cette astuce pour révéler des données numériques encryptées dans des poly(alcoxyamine phosphodiester)s.[351] Pour ce faire, deux nitroxides différents ont été incorporés dans deux blocs distincts, chacun contenant un segment alkyle codant de même masse mais d'architecture différente. Par conséquent, la séquence numérique déterminée à partir des fragments primaires observés en MS/MS a dû être décryptée selon des règles appropriées, qui dépendent des temps de dérive mesurés par spectrométrie de mobilité ionique.

Pourtant, malgré l'efficacité et la fiabilité de la spectrométrie de masse dans l'analyse de séquence polymère définie, cette méthode est dite « destructrice » puisque l'échantillon étudié est fragmenté par ionisation lors de son analyse. Bien que les quantités requises pour détecter le polymère ciblé soient minimes (quelques milligrammes suffisent), la recherche d'alternatives non destructrices a naturellement suscité de l'intérêt ces dernières années. Celles-ci seront discutées dans la prochaine sous-section.

3.3.2. Autres méthodes pour décoder l'information dans les polymères numériques

3.3.2.1 Lecture Supramoléculaire

Récemment, le groupe de Leigh a publié un « cliquet moléculaire » pour lire de l'information encodée sur des « bandes moléculaires » à la façon des machines de type Turing (précurseur des ordinateurs).[352] En utilisant un éther couronne ("tête de lecture"), il est ainsi possible d'extraire l'information contenue dans une structure moléculaire codée ("bande"). Le processus est initié et guidé par des impulsions chimiques (dans ce cas,

changement de pH) permettant au macrocycle de se déplacer séquentiellement dans les différents compartiments de la bande. Pendant son mouvement directionnel, l'éther couronne change sa configuration en fonction de la stéréochimie des sites de liaison rencontrés. Cela lui permet de lire la séquence d'information stéréochimique programmée dans la bande, de manière non destructive, en produisant des variations dans la réponse de dichroïsme circulaire.

3.3.2.2 Séquençage par nanopores

Le séquençage par nanopore est également une des nombreuses pistes étudiées pour le séquençage de polymères synthétiques. À l'instar de la lecture des polynucléotides naturels par des nanopores contenus dans des protéines (2.3.3), les macromolécules non naturelles sont, avec une structure optimisée à cette fin, de bons candidats pour être analysées avec une telle technique.

En 2019, notre groupe en collaboration avec celui de M. Dal Peraro à Lausanne, a démontré une preuve de concept en utilisant les nanopores d'aérolysine pour précisément lire, avec une résolution bit par bit, l'information encodée dans un polymère numérique.[353] Cependant, cette étude nécessite des macromolécules biohybrides (mélange d'entités synthétiques et naturelles sur la même chaîne, dans ce cas, deux bases d'adénosine à chaque extrémité) afin de diriger la structure informationnelle vers les pores de la protéine. En restant dépendant de la chimie de l'ADN, la flexibilité avantageuse des polymères synthétiques n'est par conséquent pas totalement exploitée. De plus, d'autres défis doivent encore être appréhendés : lors de la lecture de l'ADN, une protéine motrice (hélicase) permet de contrôler la distribution des monomères à l'intérieur des pores de manière constante permettant une lecture précise et accrue. Dans le cas des polymères synthétiques, une autre stratégie doit être réfléchie pour espérer un déchiffrement fiable et ce, pour de plus longues séquences. Pour ces enjeux, plusieurs pistes sont à considérer : optimiser les séparations inter-bits pour éviter l'influence des unités voisines, augmenter la rigidité du squelette polymère pour réduire le bruit de signal, ou encore, réduire la densité de charge négative pour obtenir des temps de lectures plus longs.[353]

Par ailleurs, les principes de lectures par nanopores peuvent également être appliquées à des nanorubans de graphène pour crée un système analytique synthétique et modifiable. En outre, le groupe de Radenovic, également à l'EPFL, a démontré une preuve de ce concept en impliquant la translocation des bases nucléiques de l'ADN.[354]

Ce sous-chapitre (3.3) permet de comprendre plus aisément la manière dont les polymères à séquence contrôlée peuvent être séquencés. La MS/MS reste encore aujourd'hui la technique la plus utilisée pour retrouver l'information stockée dans les macromolécules informationnelles. Dans ce contexte, de nombreuses études ont été effectuées pour adapter la structure des polymères à cette technique analytique et ainsi en facilité leur lecture. Le sous-chapitre suivant détaillera entre autres le fruit d'une collaboration entre le groupe de Laurence Charles et le nôtre qui, sur quelques années, a permis d'optimiser les propriétés structurelles des polymères numériques établissant des méthodes de lectures quasi optimales.

3.4. La chimie phosphoramidite comme moyen de stockage synthétique : les dernières avancées

Ce sous-chapitre détaillera les dernières évolutions des polyphosphodiesters synthétiques pour un système de stockage de données numériques. Enfin, cette historique permettra d'introduire les objectifs de cette thèse.

3.4.1. Optimisation de séquençage à l'aide d'espaceurs

À la suite des premiers poly(phosphodiesters) synthétiques numériques introduits précédemment partie 3.2.1, des optimisations de design restaient nécessaires pour faciliter leur décodage par MS/MS. En 2017, notre groupe a démontré des travaux sur l'utilisation d'un « espaceur clivable » inséré entre les différents octets pour améliorer le séquençage du polymère.[27] Dans cette étude, les polymères ont été conçus pour subir des fragmentations contrôlées par MS/MS en conditions de dissociation induite par collision (CID). Chaque octet de la séquence a été labelisé à l'aide de marqueurs d'octets (Figure I. 31) facilitant la reconstruction de la séquence et évitant notamment les fragments isobariques (des bases nucléiques naturelles ou modifiées sont généralement utilisées à cette fin).

Parallèlement, une liaison alcoxyamine labile a été incorporée dans la séquence entre 2 octets résultant en une rupture contrôlée des liaisons NO-C lors d'une première activation et conduisant à une distribution d'octets intacts. Chacun de ces fragments peuvent ensuite être séquencés par pseudo-MS^3 pour retrouver l'information initialement encodée. Il est notable ici que la taille des blocs est choisie de manière à être suffisamment longue pour contenir 8 bits (pour se rapprocher des octets utilisés en langage binaire ASCII), mais également suffisamment courte pour éviter une charge importante du bloc pouvant compliquer la récupération de la séquence lors de l'analyse pseudo-MS^3.

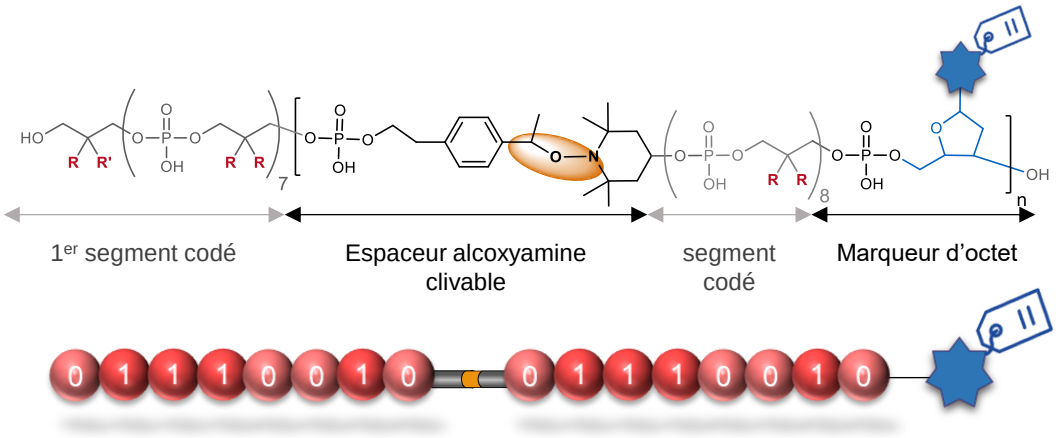


Figure I. 31 : Structure d'un octet fonctionnalisé d'un espaceur clivable et d'un marqueur de masse.

Afin de tirer profit de la labilité de la liaison NO-C, un processus de séquençage ordonné doit effectivement être respecté (Figure I. 32) : après l'analyse ESI-MS en mode négatif du polymère, un ion multichargé est intelligemment choisi pour subir une fragmentation MS/MS. Dans cette étude, le dodécanion $[M-12H]^{12-}$ a été sélectionné pour garder une distribution de charge homogène sur les 4 différents octets (3- par fragment). Par ailleurs,

l'analyse MS/MS de ce dodécanion n'est pas réalisée pour avoir une reconstruction directe de la séquence comme il a été discuté dans des exemples précédents (partie 3.3.1). L'activation de l'ion précurseur sert ici à dissocier préférentiellement les liaisons NO-C, avant de pouvoir sélectivement séquencer les octets résultants par pseudo-MS³. L'énergie d'activation doit par conséquent être rigoureusement sélectionnée pour ne pas rompre les liaisons phosphates, nécessitant quant à elles plus d'activation pour être dissociées.

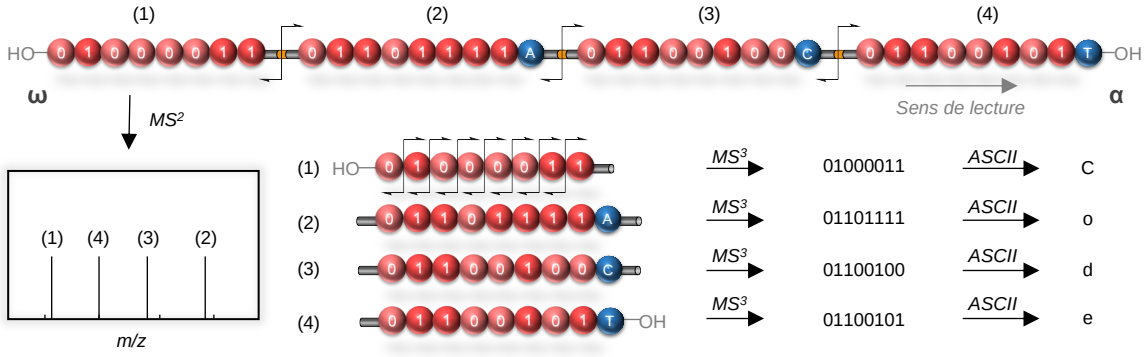


Figure I. 32 : Illustration simplifiée du principe de séquençage pseudo-MS³. Adapté de la littérature.[27]

En pratique, afin de pouvoir utiliser un spectromètre de masse classique équipé d'une cellule de collision, l'ion précurseur sélectionné est activé pour former une CID en source dissociant préférentiellement les liaisons alcoxyamines. Puis, les octets ainsi libérés sont sélectionnés dans le quadrupôle pour être excités et rompre les liaisons phosphates du fragment choisi. C'est pourquoi, en raison de l'absence de sélection des ions précurseurs avant la première étape d'activation, ce protocole de séquençage basé sur deux étapes CID est désigné comme une expérience pseudo-MS³.

L'augmentation de tension appliquée en pseudo-MS³ générant ces ruptures permet de retrouver l'ordre précis de l'enchaînement de monomères en vérifiant la couverture de chaque fragment tel que :

- Les fragments contenant la terminaison α (a^{z-}, b^{z-}, c^{z-}, et d^{z-}, avec z=1-4) sont utilisés pour reconstruire la séquence de la gauche vers la droite, tandis que les fragments contenant la terminaison ω (w^{z-}, x^{z-}, y^{z-}, et z^{z-}, avec z=1-4) permettent de reconstruire la séquence de la droite vers la gauche (Figure I. 33). Un exemple de reconstruction de séquence est donné en partie expérimentale (P.E., Figure PE.4. 55).

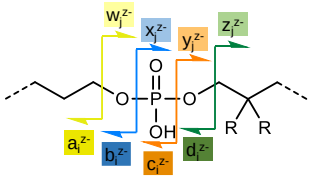


Figure I. 33 : Différents modes de dissociation du phosphate lors de l'analyse pseudo-MS³.

Malgré l'efficacité de cette méthode dans la reconstruction complète de la séquence, sa lecture automatisée par le logiciel MS-DECODER (logiciel développé pour déchiffrer automatiquement un spectre de masse en y renseignant préalablement les unités codantes)[28] n'est, avec ce protocole, pas réalisable. En effet, le manque d'abondance des ions formés en utilisant le ***SPACER1*** comme espaceur clivable compromet la lecture automatique des fragments (Schéma I. 9A). Cela peut notamment s'expliquer par la réactivité accrue du radical carboné lors de la pseudo-MS^3 induisant des dissociations radicalaires, alors compétitives avec la rupture désirée des liaisons phosphates.[355]

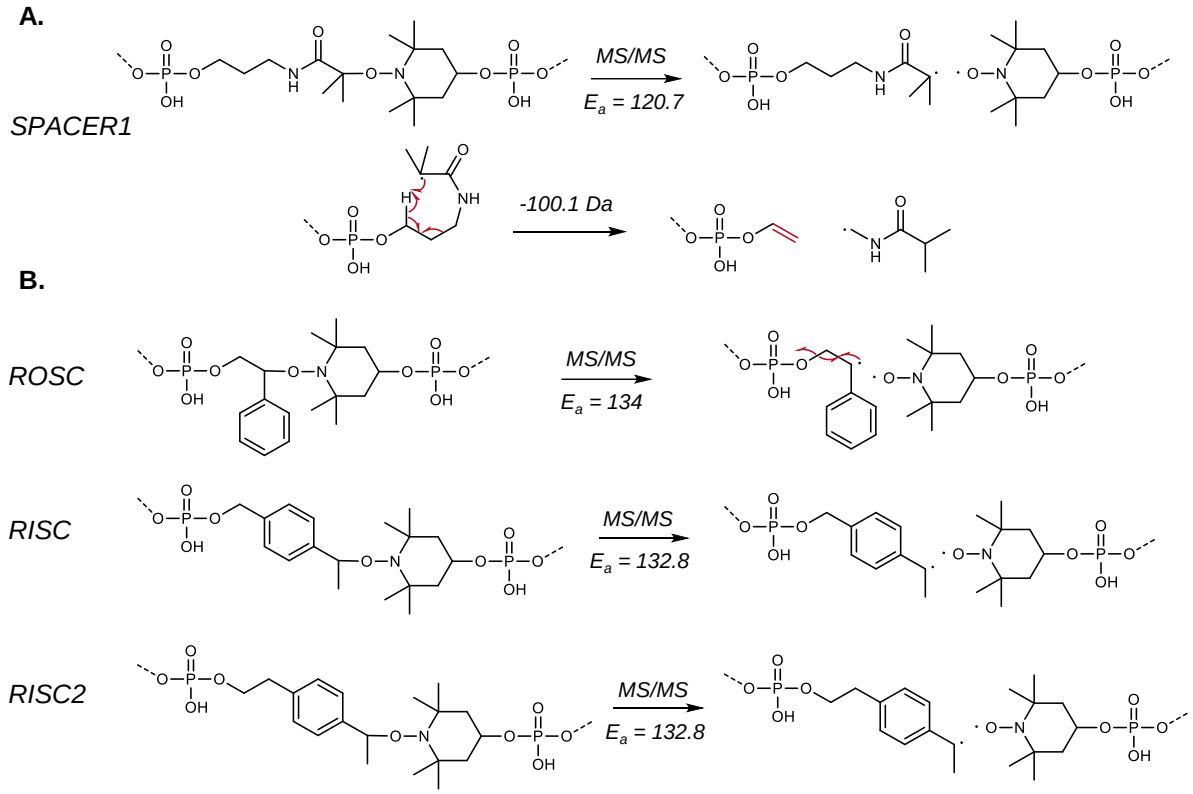


Schéma I. 9 : Différents espaceurs clivables pour optimiser l'analyse pseudo-MS^3.
***A.** Haut : première alcoxyamine utilisé. Bas : exemple de mécanisme proposé impliquant le radical carboné réactif. **B.** Evolution d'optimisation des alcoxyamines clivables. ROSC (Ring Outside Chain), RISC (Ring Inside Chain). E_a = Energie d'activation en kJ.mol^{-1} rapportée par notre groupe.[355]*

Pour y remédier, il a été nécessaire d'élaborer un nouveau design d'espaceur clivable en prenant compte d'un critère important : l'énergie de dissociation de la liaison alcoxyamine. Le ***SPACER1*** a déjà fait l'objet d'études pour obtenir un clivage sélectif sur l'alcoxyamine en présence de liaisons phosphates. En outre, son dérivé monométhylé (un méthyle en α de l'amide) contient un seuil d'énergie, pour l'homolyse de la liaison NO-C, proche de celui des liaisons phosphates, résultant en une pollution du spectre avec des fragments non désirés.[27] Il est notable qu'une attribution d'énergie de dissociation du phosphate dans le cas des phosphodiesters est difficile car le processus de rupture de liaison P-O (P=O et P-O) est certainement dû à la délocalisation d'un électron dans une orbitale non remplie, d'énergie nettement plus élevée qu'une liaison P-O classique.[356] Toutefois, l'énergie de dissociation des liaisons NO-C des alcoxyamines reste de manière générale plus faible que celle des liaisons P-O, et permet donc un clivage sélectif des liaisons NO-

C d'une alcoxyamine, si la structure de cette dernière est intelligemment choisie. (NO-C autour de 130 kJ mol^{-1} maximum dans Schéma I. 9 et la littérature[357] ; P=O autour de 500-600 kJ mol^{-1} et 368 kJ mol^{-1} dans le cas de la liaison P-O pour un phosphate $(MeO)_3PO$[358]).

Une première stratégie pour s'affranchir des réactions radicalaires compétitives a alors été de stabiliser le carbone radical en augmentant la délocalisation d'électron par l'ajout d'un cycle aromatique au lieu d'un carbone (***ROSC***, (Schéma I. 9B). Dans ce cas, les réarrangements radicalaires compromettant n'ont plus été observés. Cependant, l'ajout du cycle a ouvert la voie à une réaction de dépolymérisation par un processus de backbitting. En combinant rigidification et stabilisation avec un méthyle et un cycle aromatique respectivement, l'ensemble des problèmes a pu être résolu pour un séquençage optimal en pseudo-MS^3 (***RISC***, Schéma I. 9B). Pourtant, le design de l'alcoxyamine ***RISC*** pose un problème lors de la synthèse des polymères. Plus précisément, les réactifs (NH_3 et/ou $MeNH_2$) utilisés pour cliver la chaine du support solide induisent une substitution nucléophile en position benzylique et, de fait, conduisent à des échantillons polymoléculaires. Le développement de l'alcoxyamine ***RISC2*** (Schéma I. 9B) a permis de pallier ce problème.

Ces années d'optimisations ont finalement permis d'améliorer le séquençage de polymères à séquence contrôlée de manière fiable et efficace avec un protocole très bien décrit. Effectivement, l'utilisation d'espaceurs clivables et de marqueurs de masse a offert la possibilité d'encoder et de décoder parfaitement un grand nombre de données dans une chaîne informationnelle. Par ailleurs, la combinaison de ces outils avec un alphabet augmenté a considérablement fait croître la capacité limite de stockage dans un polymère non naturel. La prochaine sous-partie détaillera les résultats obtenus en associant espaceurs clivables, marqueurs de masse et alphabets augmentés.

3.4.2. Alphabets augmentés

Afin de repousser les limites de capacités de stockage dans des macromolécules synthétiques, notre groupe s'est intéressé à l'utilisation d'alphabets augmentés sur la base des monomères phosphoramidites vus précédemment,[26] mais aussi autrement, en variant les structures des monomères à l'aide d'une chaîne aminoalkyle (Figure I. 34B).[359]

Après avoir tester la viabilité de dyades en encodant des images à l'aide du code ASCII (Figure I. 34C), un alphabet à 8 monomères a été étudié dans un but d'obtenir non plus 2 bits mais 3 bits par monomère (Figure I. 34A).

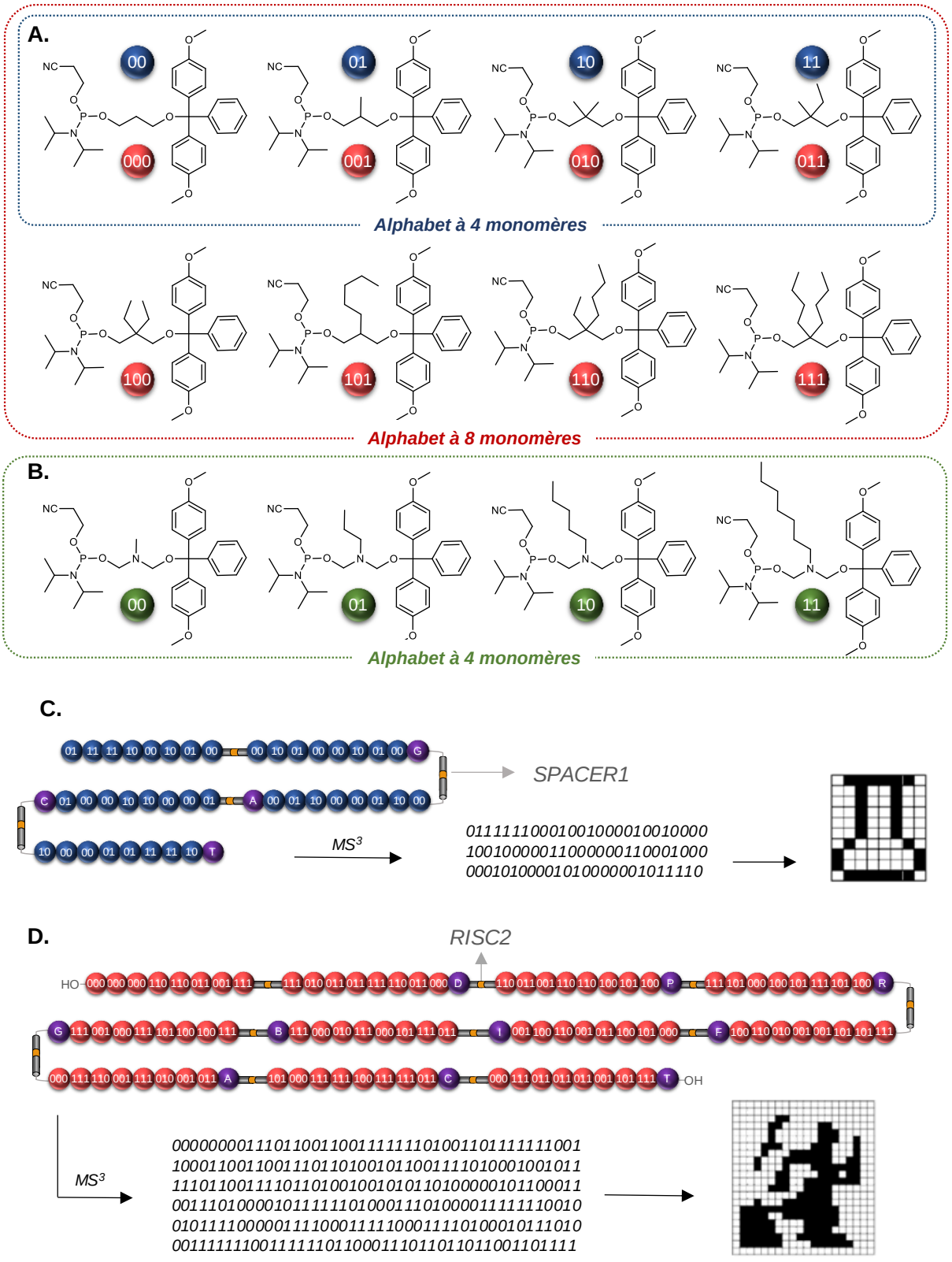


*Figure I. 34 : Alphabets augmentés à 4 et 8 monomères. Adapté de la littérature.[26,29,359] **A.** Monomères phosphoramidites classiques. **B.** Monomères aminoalkyles phosphoramiditse. **C.** Exemple d'encodage avec l'alphabet à 4 monomères. **D.** Encodage du portrait de Lavoisier à l'aide de l'alphabet à 8 monomères. Marqueurs utilisés : A (Adénosine), C (Cytosine), T (Thymidine), G (Ganosine), B (Bromouridine), (Iodouridine), F (Fluoroadenosine), R (Bromoadenosine), D (Fluorothymidine), P (desoxypurine).*

Comme preuve de concept, cet alphabet a été appliqué dans la construction d'un polymère à séquence définie avec un total de 108 unités monomère.[29] Dans cette étude, l'encodage du portrait d'Antoine de Lavoisier avec une résolution pixel a pu être encodée puis lu automatiquement à l'aide du logiciel MS-DECODER (Figure I. 34D). L'utilisation d'un alphabet augmenté permet entre autres d'augmenter considérablement la densité d'information à l'échelle moléculaire. En effet, en encodant 3 bits par monomère avec un total de 88 unités codantes, 264 bits ont pu être encryptés. Des techniques de compressions de données ont finalement permis d'augmenter la quantité de donnée à 440 bits dans une seule chaîne polymère, ce qui correspond au nombre de pixels du portrait du chimiste de renom illustré ci-dessus.

Cependant, la chaîne polymère est plus longue à cause des espaceurs clivables et marqueurs de masse insérés entre chaque octet. Dans ce cas, 10 espaceurs et 10 marqueurs de masse ont été utilisés : autant d'entités indispensables à la lecture mais inutile au stockage d'informations.

La prochaine sous-partie discutera des limites dans l'utilisation des outils de lecture susmentionnés, introduisant les objectifs de cette thèse.

3.4.3. Limites et objectifs de la thèse

Depuis la découverte de la synthèse sur support solide, et particulièrement l'approche phosphoramidite sur support solide, de nombreux groupes de recherche ont tenté de repousser les limites de longueurs pour la synthèse de polymères à séquence définie (naturels ou non). Le développement de protocoles automatisés et/ou miniaturisés (synthèse sur micro-puces) a permis d'obtenir des chaînes polymères allant autour des 200 nucléotides maximums. Effectivement, même avec un protocole établi et des efficacités de couplage >98%, le rendement d'un polymère tend inexorablement vers 0%. Par exemple, en ayant une telle réussite, un polymère de 200 unités aurait un rendement théorique finale de $0.98^{\wedge 200}$ = 18%, et encore plus long avec 250 unités, il serait alors seulement de 6%. C'est pourquoi, un des grands défis de la construction de polymères à séquence définie est l'élargissement des capacités de synthèse à l'instant énoncées.

Ce défi s'applique également aux macromolécules informationnelles. Etant à ce jour limité par la taille maximum de ≈ 200 unités pour une chaîne unique, des optimisations peuvent également être effectuées sur la constitution et la structure du polymère. En effet, la dernière sous-partie a démontré que l'utilisation d'un alphabet augmenté peut déjà améliorer les capacités de stockage en encodant 264 bits avec 108 monomères (un polymère avec un alphabet à 2 monomères aurait nécessité au minimum 264 monomères pour encoder la même quantité d'information).

Ces limites peuvent être encore plus repoussées en réfléchissant à un nouveau design pour les espaceurs clivables et les marqueurs de masse. En outre, ces entités inutiles au stockage d'informations mais importante pour séquencer facilement le polymère, pourraient être incorporées de manière intelligente dans la structure du polymère pour répondre à trois principaux besoins :

i) Réduire l'usage de monomères « non informationnels » afin d'améliorer la densité de stockage.
ii) Faciliter la synthèse en ayant recourt à une chimiothèque de monomères multifonctionnels.
iii) Faciliter la lecture de macromolécules informationnelles à l'aide de nouvelles méthodes pour lesquelles de nouveaux monomères peuvent être élaborés.

Cette thèse a tenté de répondre à ces besoins en réfléchissant à un nouveau design d'espaceur clivable en se basant sur les travaux effectués auparavant dans notre groupe, impliquant la conception de l'espaceur ***RISC2*** et son insertion en chaîne polymère.[355]

Pour ce faire, la structure du ***RISC2***, utile au séquençage pseudo-MS^3, peut être révisée pour y incorporer un marqueur de masse, nécessaire à l'identification d'octets (Figure I. 35). Cette nouvelle stratégie permettrait de repousser les limites de stockage établies avec le portrait de Lavoisier en nécessitant moins de monomères « non informationnels ».

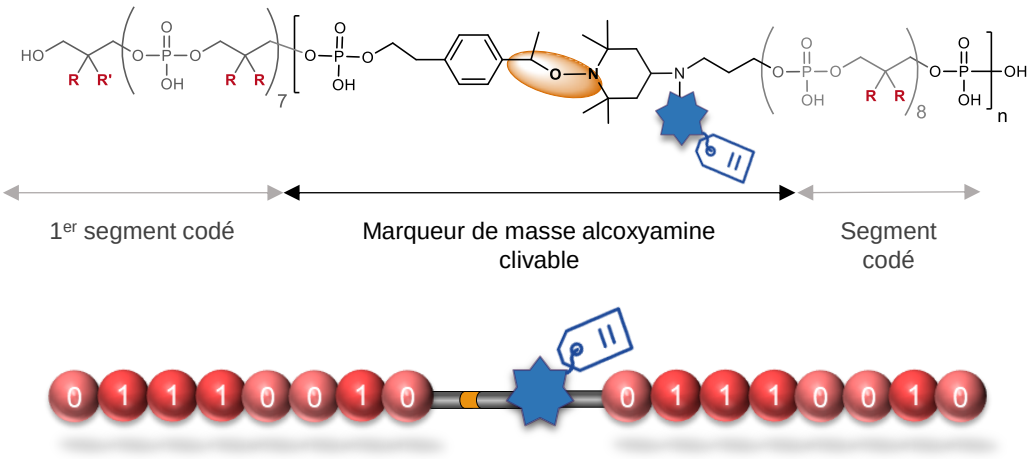


Figure I. 35 : Nouveau design pour un marqueur de masse clivable.
(R = H ou CH_3)

Par ailleurs, le développement d'une nouvelle série d'espaceurs pourrait être appliquée à de nouvelles méthodes d'analyses, en l'occurrence, la spectrométrie de mobilité ionique (IMS). Cette technique faciliterait le séquençage d'un polymère numérique en répondant à la contrainte suivante : le séquençage pseudo-MS^3 décrit précédemment (3.4.1) nécessite *(i)* une MS/MS pour sélectionner les ions précurseurs puis, *(ii)* autant d'expériences pseudo-MS^3 que d'octets contenus dans le polymère. Cette stratégie n'est donc, par la nécessité d'opérations individuelles répétées, pas viable pour un séquençage rapide d'une large quantité d'informations. Pour faire face à cette problématique, l'utilisation de la MS/MS couplée à l'IMS (MS/MS-IMS-MS/MS) permet de combiner ces deux étapes en profitant *i)* de la séparation d'ions basée sur leur mobilité puis, *ii)* le séquençage MS/MS de ces ions pour retrouver la séquence de chaque octet de manière indépendante. L'analyse effectuée en une seule étape est automatisable et offre des résultats en 2D donnant également la possibilité d'automatiser la lecture des octets horizontalement alignés sur le spectre.

4. Conclusion et perspectives

La nature ne serait pas aussi belle si l'ADN n'avait pas la possibilité de se diversifier dans la création et le stockage d'information génétique. Ce biopolymère est aujourd'hui le fruit de quelques millions d'années d'évolution offrant un moyen de stockage d'information naturelle (génétique) ou synthétique (numérique). Dans ce premier chapitre, l'état de l'art d'un tel stockage à l'échelle moléculaire a été discuté.

Les nombreuses optimisations dans la synthèse et le séquençage de l'ADN ont permis, après quelques décennies, de reproduire des séquences parfaitement contrôlées de génomes ou encore de séquencer des maladies telles que la Covid-19 ou l'Ebola. Parallèlement, le stockage d'informations synthétiques est devenu un des grands défis du 21e siècle. Devant les quantités énormes de données numériques produites par les humains et les ressources naturelles limitées pour les stocker, des alternatives doivent assurément être explorées. L'ADN en est une : en codant un code binaire à l'aide de ses 4 bases nucléiques, ce biopolymère peut servir de substitut aux moyens de stockage classique avec de nombreux avantages. Par exemple, le stockage étant à l'échelle moléculaire, la densité d'information est beaucoup plus importante dans un polymère que dans un disque dur.[31] Aussi, la durée de conservation d'une telle macromolécule est théoriquement beaucoup plus grande comparée aux moyens de stockages actuels. Ces différents atouts ont mené de nombreux chercheurs à étudier la chimie de l'ADN, ce qui a permis de démontrer des preuves de concepts prometteuses (ex. stockage de livre, vidéos, musiques, etc.).[8,274,275]

Cependant, l'ADN présente des limites puisque son utilisation restreint les possibilités de création que la chimie offre aux scientifiques. En effet, seulement les 4 bases ATGC du biopolymère peuvent être utilisées, ce qui peut également amener des contraintes structurelles et/ou biologiques (répétions de bases problématiques).[13] C'est pourquoi, les polymères synthétiques et particulièrement ceux à séquence définie ont été des macromolécules d'intérêt depuis le début du 21e siècle. Leurs propriétés sont flexibles à souhait et peuvent donc, avec de l'optimisation, être adaptées à de nombreuses applications (stockage de donnée,[16,336] anti-contrefaçon,[308,360] agent de transfert de médicaments,[361] photovoltaïque,[362] cryptage de donnée[335,351]). Ce chapitre a notamment discuté des différentes approches pour conserver les données numériques à l'aide de ces polymères, et a particulièrement détaillé l'approche phosphoramidite comme un moyen de stockage non naturel prometteur. Effectivement, les travaux des groupes de Laurence Charles et le nôtre ont démontré à travers différentes études qu'il est possible d'adapter les structures des polymères informationnels afin d'augmenter leur capacité de stockage mais aussi de les décoder plus facilement.[355,26,29]

En plus du développement d'un alphabet augmenté codant jusqu'à 3 bits par monomère et augmentant ainsi la densité d'information, l'étude d'un nouvel espaceur clivable a permis l'automatisation de lecture des polymères en utilisant un logiciel adapté (MS-DECODER). Pourtant, des optimisations restent toujours nécessaires dans la conception du polymère, notamment pour pallier les longueurs de chaînes maximales trop courtes (≈

200 monomères) imposées par les limites de la synthèse chimique. Dans ce contexte, les travaux de cette thèse essaient de répondre aux différents défis établis et listés dans ce chapitre. Particulièrement, l'amélioration de la densité de stockage en réduisant l'usage de monomères non informationnels. Avec un nouveau design, une nouvelle série de macromonomères pourrait servir à cette fin, tout en profitant à de nouvelles méthodes d'analyse : l'IMS.

Le prochain chapitre discutera d'une première approche pour la synthèse de nouveaux marqueurs de masse clivable : ***TISP*** (**T**ag **I**nside the **SP**acer).

Chapitre II.

Stratégies *TISP* et *TISP2* – Synthèse et incorporation dans une chaîne polymère contrôlée

1. Introduction

Dans le Chapitre 1, différentes approches dans la construction de macromolécules informationnelles ont pu être discutées. Particulièrement, l'utilisation de la chimie des phosphodiesters pour un stockage d'information artificielle qui s'est révélé efficace. Néanmoins, des progrès visant à augmenter la capacité de stockage de tels polymères restent nécessaires. En outre, réduire l'usage de monomères non informationnels (espaceurs clivables et marqueurs d'octets) est un des principaux axes d'amélioration. Un nouveau design combinant au sein d'une même molécule une liaison labile servant d'espaceur clivable (alcoxyamine) et un groupe fonctionnel variable servant à différencier sa masse répondrait effectivement à ce défi.

Dans ce contexte, les travaux de cette thèse ont visé à synthétiser un nouveau marqueur d'octet clivable, puis tester sa robustesse en synthèse phosphoramidite automatisée et séquençage pseudo-MS^3. Pour ce faire, la réflexion sur un nouveau design s'est appuyée sur la base structurelle préalablement établie grâce au développement du ***RISC2***. En outre, cela permet de profiter des optimisations accomplies sur cet espaceur clivable résultant en un séquençage MS^3 optimal.

Ce chapitre détaillera deux stratégies dans l'élaboration d'un tel marqueur d'octet reposant sur l'utilisation d'un alcyne : le ***TISP*** et son dérivé saturé, le ***TISP2***.

D'abord, un état de l'art sur la synthèse du ***RISC2*** sera établi et différentes possibilités de conception réfléchies en début de projet seront discutées. Ensuite, la stratégie alcyne choisie, ainsi que l'incorporation du monomère obtenu dans une chaîne polymère contrôlée sera détaillée. Enfin, La deuxième stratégie employant le ***TISP2*** sera également expliquée.

2. Nécessité d'un nouveau design à partir du *RISC2*

2.1. Synthèse du monomère *RISC2*

Le ***RISC2*** introduit précédemment (Chapitre 1, 3.4.1) est le fruit de quelques années de développement autour d'un espaceur clivable pour un séquençage pseudo-MS^3 facilité, permettant en outre une automatisation de la lecture du polymère numérique à l'aide du logiciel MS-DECODER. C'est pourquoi, les premières réflexions dans l'élaboration d'un nouveau design se sont inspirées de la structure de ce macromonomère. Lors des premières études, sa synthèse a pu s'effectuer en six étapes en commençant à partir de l'acide 2-(4-acétylphenyl) acétique **a** (Schéma II. 1). La réduction de la cétone en présence de $NaBH_4$ permet d'obtenir un alcool **b** avec un rendement quantitatif avant d'être facilement substitué par un chlore à l'aide d'HCl (aq.). L'acide **c** peut par la suite être réduit en utilisant un complexe de diméthylsulfure de borane et, l'alcool instable ainsi formé **d** peut être protégé avec du chlorure de DMT pour obtenir **e**. Finalement, une ATRA, suivit d'un couplage phosphoramidite sur l'alcool **fT** a permis de donner l'espaceur ***RISC2*** désiré.

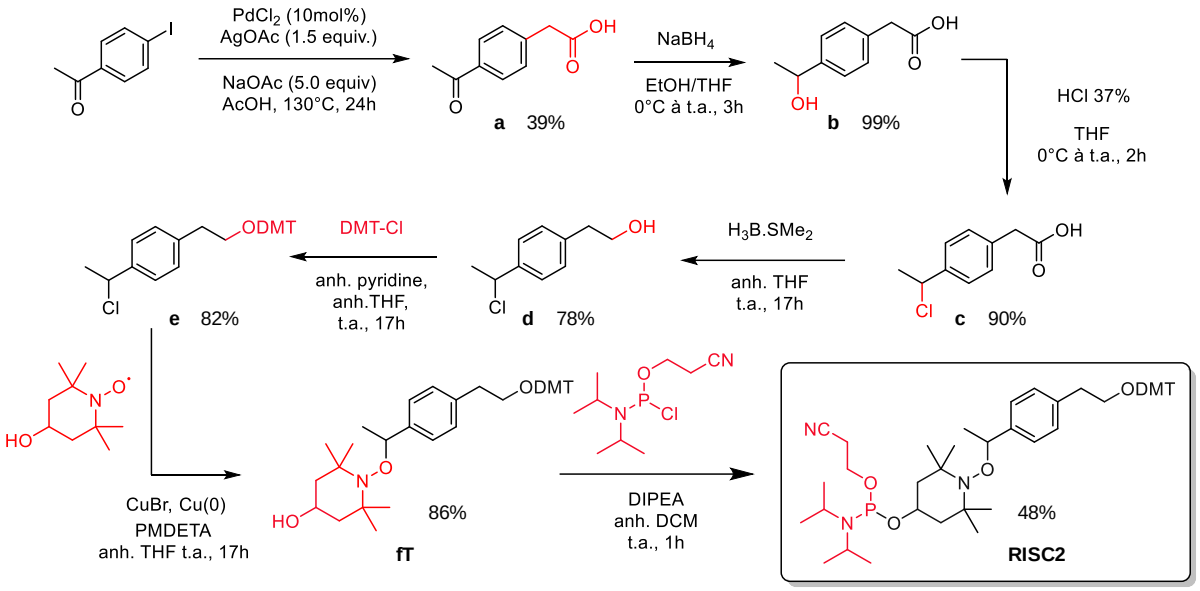


Schéma II. 1 : Synthèse du RISC2 adapté de la littérature.[355]

Quelques années après la première synthèse de ce composé, le réactif de départ **a**, n'était malheureusement plus disponible commercialement. Pour pallier ce problème sans revoir toute la stratégie, la préparation de l'acide **a** a été envisagée. Cependant, un couplage d'acide acétique sur le Iodoacetophénone en présence de Palladium a été la meilleure solution trouvée dans la littérature,[363] manifestement non optimale avec un rendement de seulement 39%. Par ailleurs, l'ajout de cette étape fait considérablement chuter le rendement global de la synthèse en passant de 24% à 9%. En commençant cette voie avec 15 g de produit, seulement 4 g d'un composé beaucoup plus lourd sont obtenus en fin de synthèse.

Par conséquent, l'élaboration d'un nouveau marqueur d'octet clivable a dû tenir compte des problèmes de rendements, en réfléchissant à une approche plus performante, tout en gardant le même squelette final.

Afin de définir une nouvelle approche pour la synthèse d'un tel espaceur, des règles imposées par le séquençage pseudo-MS^3 pour le choix d'un marqueur d'octet doivent être rigoureusement être suivies. Ces dernières sont détaillées dans la section sortante.

2.2. Règles imposées par le séquençage pseudo-MS^3 pour le choix d'un marqueur d'octet.

Afin de pouvoir décoder le polymère et attribuer tous les fragments issus des dissociations de liaisons phosphates, le choix de la masse moléculaire de chaque marqueur d'octet est très important. Effectivement, certaines règles doivent impérativement être respectées pour éviter les recouvrements de pics lors du séquençage. Tout d'abord, il est notable que ces règles sont dictées par l'utilisation de l'alphabet à 8 monomères introduit précédemment (Chapitre 1, 3.4.2). Alors que la différence de masse entre chaque monomère est de 14 Da (un CH_2 de différence), les règles à suivre sont les suivantes :

i) La masse molaire d'un marqueur ne doit pas être un multiple de 14. Cela permet de différencier les marqueurs des monomères codants.
ii) La différence de masse entre deux marqueurs ne doit être inférieure à 3, et ne pas être un multiple de 14, plus ou moins 3. Cela permet l'attribution des fragments par comparaison de masse en considérant les ions tri chargés.

Ces deux règles sont essentielles puisqu'elles permettent de décoder sans ambiguïté le polymère numérique lors de l'analyse pseudo-MS^3. En d'autres mots, aucun fragment généré en respectant ces règles ne cause une superposition de signal, pouvant compromettre l'interprétation de l'information. Par conséquent, ces principes ont été pris en considération dans le choix d'un nouveau design de marqueur clivable, qui est discuté dans la sous-section sortante.

2.3. Idées de conception pour un marqueur de masse clivable

Une des premières réflexions s'est tournée vers l'utilisation d'un couplage catalysé au palladium et plus particulièrement celui de Miyaura - Suzuki. Le rendement et la variété dans le choix des réactifs sont intéressants pour apporter de la diversité efficacement sur la structure du ***RISC2***. Cependant, après étude approfondie des réactifs disponibles commercialement, les composés phényles fonctionnalisés de groupes L dissemblables (bleu et rouge, Schéma II. 2) ne sont pas aussi facilement accessibles que souhaité.

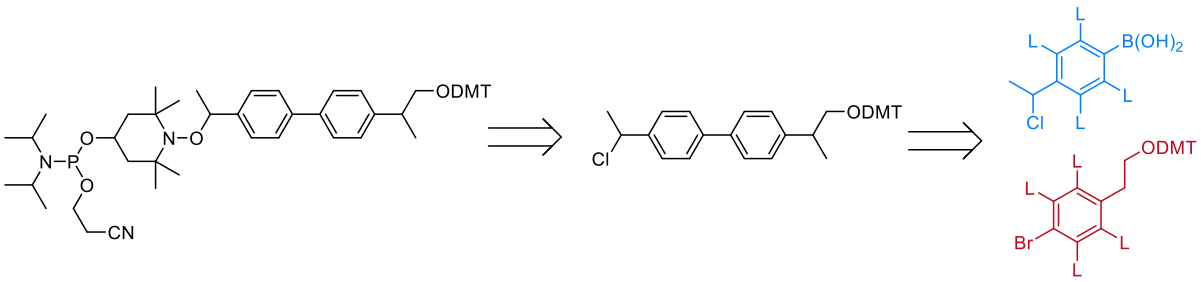


Schéma II. 2 : Proposition de synthèse d'un nouveau marqueur clivable à l'aide d'un couplage de Miyaura Suzuki.

En effet, pour construire une chimiothèque de marqueurs, des molécules de masse différentes doivent être utilisées. Or dans ce cas, peu de candidats sont disponible pour envisager une telle variété. Même si d'autres possibilités peuvent être réfléchies en prenant le couplage de Miyaura - Suzuki comme étape clef dans la synthèse, cette piste a finalement été abandonnée. Parallèlement, d'autre stratégies ont été étudiées et ont semblé plus prometteuses.

La modification du TEMPO a également été une voie de réflexion. Avant l'ATRA, une chaîne alkyle différente peut être incorporée sur le TEMPO laissant place à une variable élargie (Schéma II. 3). Pour cette stratégie, diverses options sont disponibles : en partant de l'amino-TEMPO *N*-substituée par un propanol (**T1**), l'amine secondaire peut théoriquement être impliquée dans plusieurs types de réactions pour amener de la diversité. Le choix s'est porté vers une substitution nucléophile en utilisant des dérivés bromés pour plusieurs raisons :

i) La très large variété de composés bromés permettant de construire une chimiothèque étoffée.

ii) La simplicité de la substitution nucléophile qui implique seulement deux réactifs.

De plus, en choisissant intelligemment la structure du dérivé bromé, des post-modifications peuvent être envisagées, apportant encore plus de complexité. Par exemple, avec un couplage catalysé au palladium en insérant un aryle fonctionnalisé d'un halogène (Schéma II. 3A, exemple avec un couplage Miyaura- Suzuki) ou encore, avec une réaction CUAAC (Schéma II. 3B). Autrement, une simple transformation de l'amine secondaire en amine tertiaire en utilisant une chaîne alkyle variée (Schéma II. 3C) peut également être suffisante pour obtenir un large choix de marqueurs d'octets. C'est en outre cette dernière option qui a été retenue, motivée par sa simple mise en œuvre.

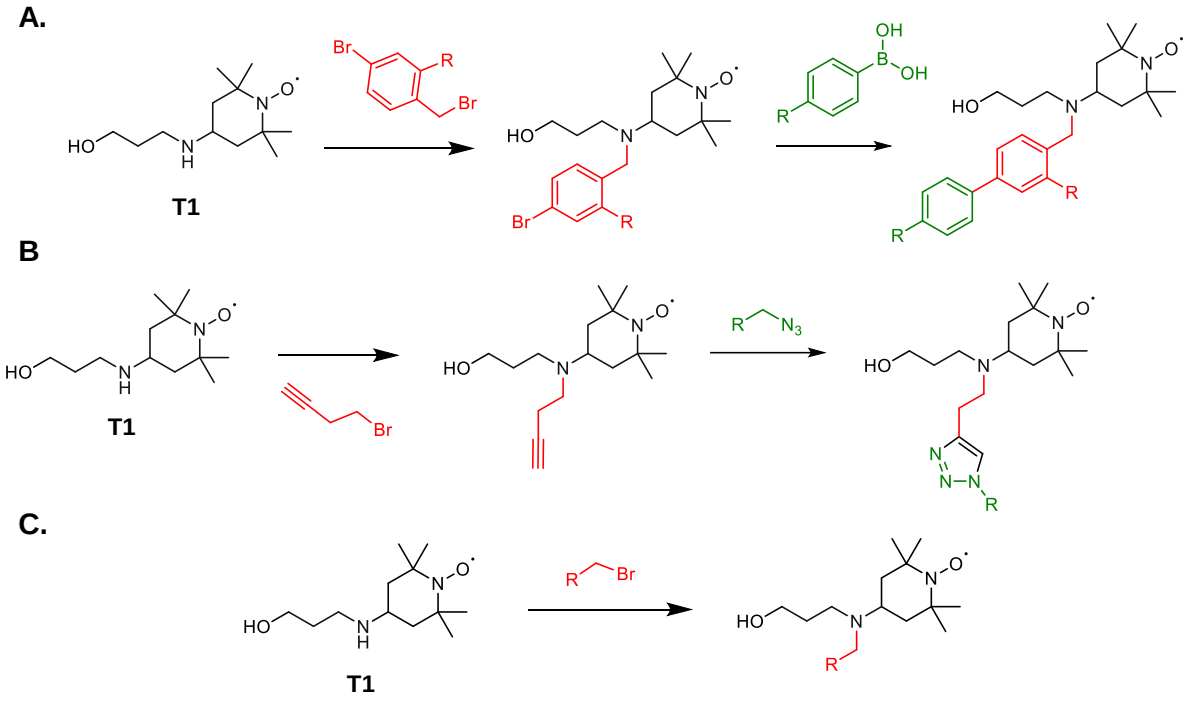


Schéma II. 3 : Idée de conception de marqueurs d'octets en variant le nitroxide TEMPO. ***A.*** *Post-modification pallado-catalysée (Miyaura – Suzuki).* ***B.*** *Post-modification par CUAAC.* ***C.*** *Diversité amenée par la chaîne alkyle. R = alkyle, aryle, halogènes, etc.*

Afin d'établir les premiers développements autour de cette synthèse, une voie permettant d'accéder à l'aminoalcool-TEMPO **T1** (Schéma II. 4) a dû être recherchée. Pour ce faire, deux approches ont été étudiées : la première, inspirée de la littérature,[364] repose sur une amination réductrice du dérivé céto-TEMPO **cT** (Schéma II. 4A) en présence de triacétoxyborohydrure de sodium et d'aminopropanol, tandis qu'une deuxième est basée sur une substitution nucléophile d'un dérivé amino-TEMPO **aT** sur un bromopropanol (Schéma II. 4B). Après avoir effectué des essais avec les deux types de réactions, l'aminopropanol-TEMPO désiré a pu être obtenu avec des rendements corrects dans les deux cas. Cependant, l'amination réductrice a été privilégiée du fait de sa facilité de purification et par conséquent sa pureté. De plus, le nitroxide obtenu comporte un radical sur l'oxygène pouvant être contraignant pour des premiers suivis de réactions par RMN.

Même si différentes solutions existent pour l'analyse de radicaux avec une telle technique (par exemple, réduction du nitroxide *in situ* à l'aide de phenylhydrazine dans du chloroforme deutéré),[365] un suivi par chromatographie sur couche mince (CCM) lors de la purification sur colonne chromatographique à base de Silice a été suffisant pour récupérer le produit pur.

Afin de suivre le schéma initialement proposé, l'étape suivante a consisté en une substitution nucléophile d'un dérivé bromé pour former une amine tertiaire **T2** (Schéma II. 4).

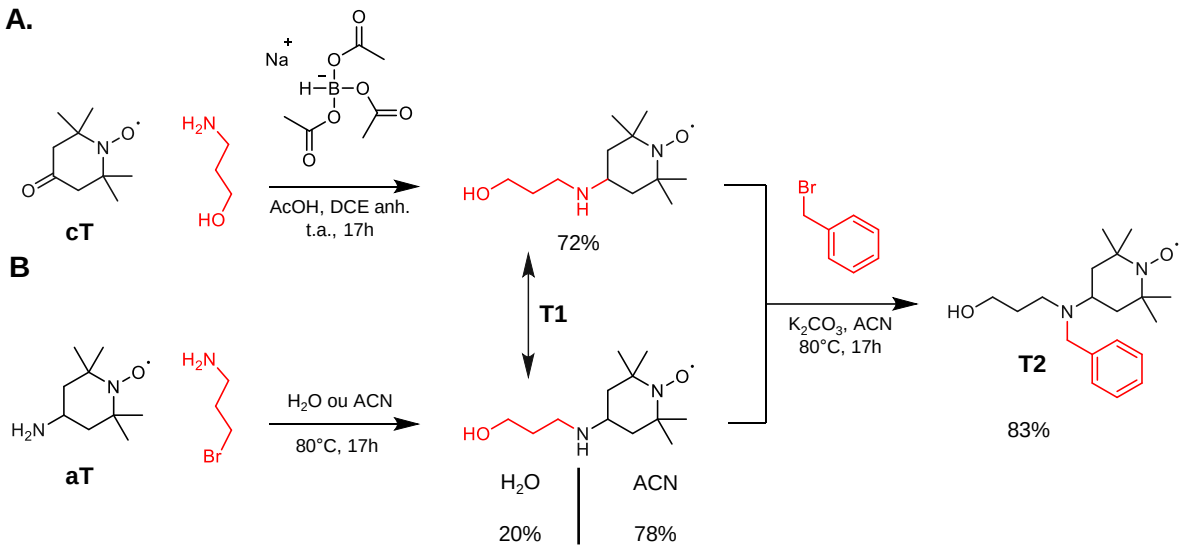


*Schéma II. 4 : Deux approches pour la synthèse de l'aminoalcool-TEMPO fonctionnalisé **T2**. **A.** Amination réductrice (DCE = 1.2 dichloroethane, AcOH = acide acétique). **B.** Substitution nucléophile.*

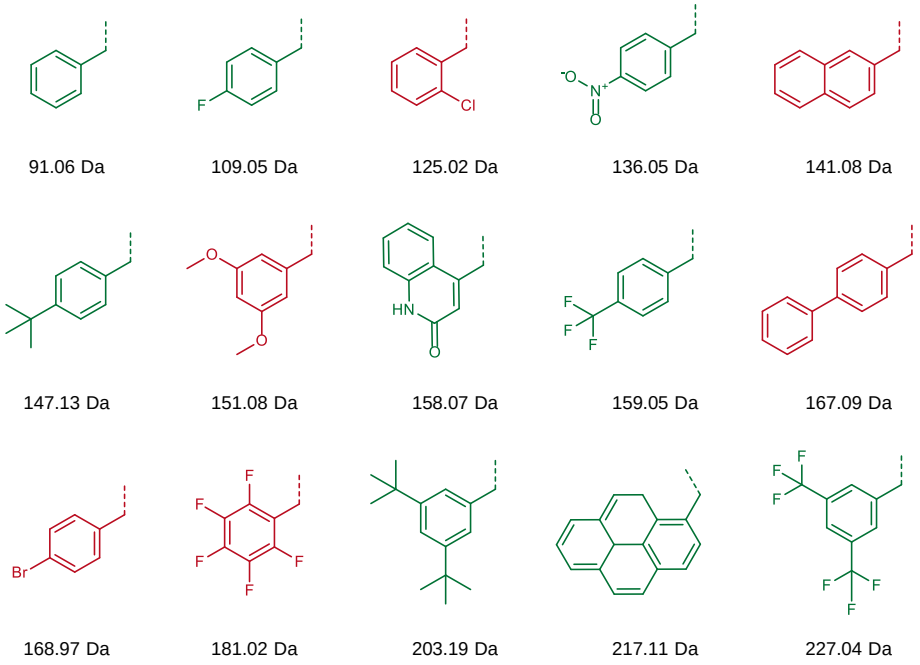
Au vu de la grande diversité de composés aryles bromés, un premier essai s'est établi en utilisant du bromobenzène. Inspiré des conditions recherchées dans la littérature,[366] même s'il était assumé que l'amine secondaire est plus nucléophile que l'alcool, une base faible doit néanmoins être utilisée pour favoriser l'attaque de l'amine sur le carbone électrophile en α du brome. Néanmoins, une base trop forte pourrait déprotoner l'alcool et une compétition entre ce dernier et l'amine en résulterait. C'est pourquoi, une première expérience a été effectuée avec K_2CO_3 dans de l'acétonitrile à reflux. Le premier essai dans ces conditions s'est avéré être un succès en donnant le dérivé amino-TEMPO modifié **T2** avec un rendement très correct. Ce protocole a été utilisé pour la synthèse de différents marqueurs d'octets qui seront discutés dans la sous-partie 2.5 de ce chapitre. Avant cela, il est important de vérifier si ce design respecte les règles définies précédemment (2.2) pour le séquençage pseudo-MS^3 du polymère numérique. La prochaine sous-section éclaircira ce point.

2.4. Vérification du respect des règles de séquençage avec le nouveau design amino-TEMPO

Grâce à la disponibilité commerciale des dérivés bromés aromatiques servant à fonctionnaliser l'amino-TEMPO **T2** (Schéma II. 4), une série de marqueurs a pu être recherchée telle que présentée dans la Figure II. 1.

Ces différents marqueurs résultent déjà d'une pré-sélection d'une grande diversité de dérivés bromés. Ces derniers ont été choisis en fonction de leurs groupes fonctionnels qui doivent êtres orthogonaux aux conditions employées en chimie phosphoramidite (sont à éviter par exemple : des groupes protecteurs déblocables en conditions acides, des esters clivables en conditions basiques, des composés oxydables en présence d'I_2, etc.). De plus, les dérivés alkyles variant seulement par un groupe méthyle ou CH_2 sont également à écarter puisqu'ils pourraient être confondus avec les monomères de l'alphabet à 8, qui se distinguent par une variation de masse Δm = 14 Da. En revanche, l'utilisation d'halogènes sur le groupe benzyle permet de varier la masse facilement en profitant d'une disponibilité commerciale assez élargie. La variation de cycles aromatiques avec les dérivés di phényles, naphtyles ou encore pyrènyles représente également des choix intéressants pour apporter de la diversité.

Par ailleurs, cette Figure II. 1 explique comment les règles introduites précédemment (2.2) sont vérifiées : le tableau de gauche permet de voir si le marqueur est utilisable en présence des 8 monomères de l'alphabet augmenté (réponse à la question : « est-ce un multiple de 14 +/-3 ? »), tandis que le tableau de droite détermine les marqueurs qui peuvent être utilisés ensembles (réponse à la question : « est-ce que la différence de masse entre 2 marqueurs est un multiple de 14 +/-3 ? »). Alors que la première règle à gauche est dans la majeure partie des cas respectée, le nombre de cases rouges sur le tableau de droite démontre la difficulté de trouver plusieurs espaceurs utilisables dans la même chaîne polymère. En effet, pour qu'ils puissent être employés dans la même macromolécule sans ambiguïté, une case verte pour chaque combinaison doit y figurer. Un exemple est donné dans le tableau de droite en entourant 3 marqueurs (bleu) qui peuvent être utilisés ensembles (91, 109 et 227). Dans ce cas, la différence entre 91 et 109 n'est pas un multiple de 14 +/- 3 Da, de même pour 91-227 et 109-227. Cependant, si par exemple un quatrième marqueur devait être rajouté tel que le 136, malgré le fait qu'il respecte les règles avec 91 et 227, il ne la respecte pas avec 109 (entourée en rouge). La combinaison 91, 109, 136 et 227 n'est donc pas possible.



Mass (Da)	Multiple de 14 +/-3 ?
91.06	**Non**
109.05	**Non**
125.02	Oui (14 x 9 = 126)
136.05	**Non**
141.08	Oui (14 x 10 = 140)
147.13	**Non**
151.08	Oui (14 x 11 = 154)
158.07	**Non**
159.05	**Non**
167.09	Oui (14 x 12 = 168)
168.97	Oui (14 x 12 = 168)
181.02	Oui (14 x 13 = 182)
203.19	**Non**
217.11	**Non**
227.04	**Non**

Différence de masse entre deux marqueurs est un multiple de 14 +/- 3 ?

	91	109	136	147	158	159	203	217	227
91		18	45	56	67	68	112	126	136
109			27	38	49	50	94	108	118
136				11	22	23	67	81	91
147					11	12	56	70	80
158						1	45	59	69
159							44	58	68
203								14	24
217									10
227									

Réponse aux questions des deux tableaux :

OUI = **ne respecte pas** la règle (cases rouges)

NON = **respecte** la règle (cases vertes)

Exemple de marqueurs pouvant être utilisés ensemble (91, 109 et 227)

Figure II. 1 : Tableaux expliquant la vérification des règles à respecter dans le choix d'un marqueur.

Ces contraintes peuvent également être illustrées à l'aide de la Figure II. 2 pour mieux comprendre les limites de cette stratégie. En respectant les règles imposées par le séquençage pseudo-MS^3, plus le nombre de marqueurs d'octets augmente, plus le choix d'un tel espaceur devient limité. Dans cet exemple, il est nettement remarquable que la marge disponible pour incorporer un nouveau marqueur diminue lorsque l'on passe de 1 (Figure II. 2A) à 2 (Figure II. 2B) espaceurs clivables.

L'attribution des combinaisons a été, au cours des recherches préliminaires, très fastidieuse du fait du contrôle manuel des possibilités. Grâce au développement d'un code python (réalisé par Isaure Carvin à l'ICR de Marseille), il a été possible de déterminer automatiquement les répartitions respectant toutes ces règles. Ceci a notamment permis d'établir d'un commun accord des propositions de tri-blocks séquençables discutés plus tard (Chapitre 3, 3, 4.3).

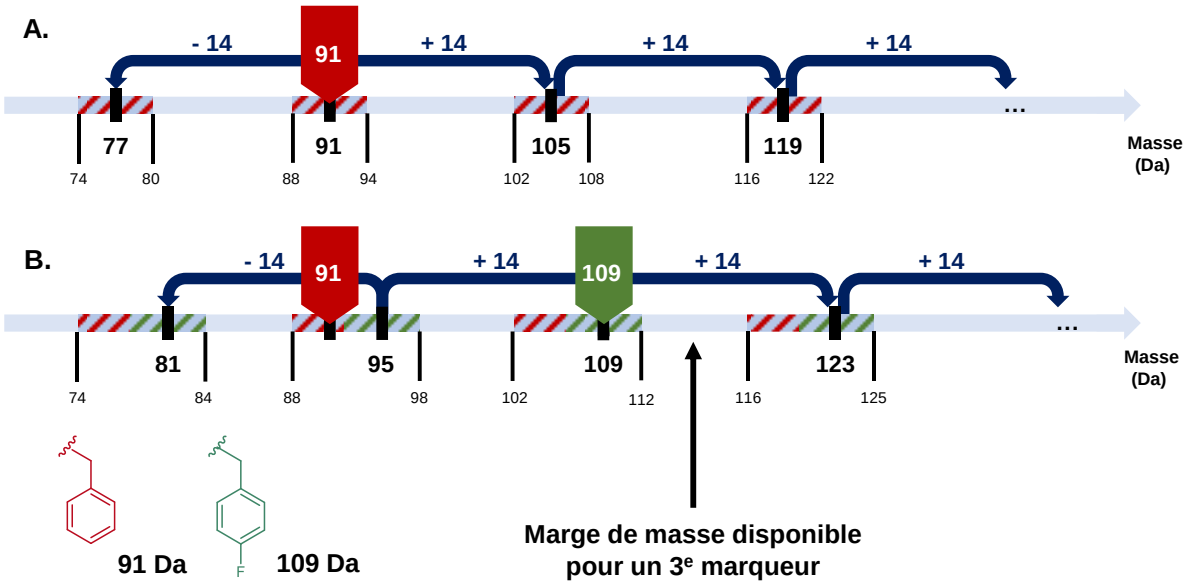


Figure II. 2 : Illustration des contraintes liées aux règles à respecter pour l'analyse pseudo-MS³.

Même si cette stratégie semble avoir des contraintes liées au choix limité d'une combinaison de marqueurs, l'automatisation du processus de sélection a permis d'identifier une série de 9 marqueurs maximums, utilisable dans une même chaîne polymère pour un séquençage pseudo-MS³ facilité (Schéma II. 5). Ce nombre peut facilement être agrandi puisqu'il prend compte d'une sélection préliminaire de dérivés bromés commercialement disponibles. En effet, en optimisant le design des marqueurs par différentes techniques de modifications (CuAAC ou couplage catalysé au palladium par exemple), une série d'espaceurs clivables peut être construite à souhait, en fonction des marges de masses disponibles pour incorporer de tels monomères, tout en respectant les règles dictées par le séquençage.

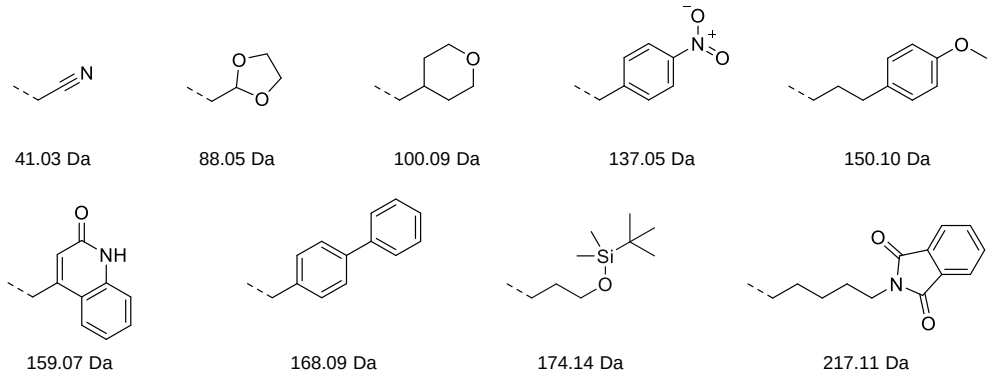


Schéma II. 5 : Exemple de série de marqueur utilisable dans une même chaîne polymère.

Le choix des marqueurs ayant été discuté, la prochaine sous-partie détaillera l'utilisation de ces derniers dans la synthèse d'un premier espaceur ***TISP***.

2.5. Le design *TISP* : optimisation et synthèse

Alors que le bloc amenant de la diversité a été obtenu avec l'amino-TEMPO **T2** modifié (Schéma II. 4), la structure globale du marqueur d'octet restait à concevoir. En effet, il a été discuté précédemment que les rendements de synthèse du ***RISC2*** chutent considérablement du fait de l'arrêt d'approvisionnement de l'acide 2-(4-acetylphenyl) acétique **a** (Schéma II. 1). Une autre voie de synthèse, plus efficace et offrant plus de quantités, tout en apportant les mêmes propriétés structurelles du ***RISC2***, a alors été réfléchie. Aussi, cette synthèse doit impérativement comporter l'étape clef d'ATRA permettant de coupler l'amino-TEMPO modifié **T2** obtenu précédemment.

Dans ce contexte, l'idée d'utiliser un couplage de Sonogashira en partant du même réactif de départ que celui utilisé dans la synthèse du ***RISC2*** (iodoacétophénone) a été tentant. Une nouvelle approche a alors été proposée et est illustrée dans le Schéma II. 6. Les deux premières étapes de *i)* réduction et de *ii)* chloration quantitatives sont, avec de l'optimisation, réalisables en un jour puisque la purification de ces intermédiaires n'est pas nécessaire au vu de leurs puretés identifiées par RMN. Le dérivé aryle iodo-chloré ainsi formé peut alors subir un couplage de Sonogashira sélectivement sur l'Iode en donnant l'alcyne **c1** (Schéma II. 6) avec un rendement supérieur à 90%. À partir de cet intermédiaire, deux options sont possibles :

i) Utiliser la voie 1 (Schéma II. 6A), intéressante puisqu'elle évite une étape de réduction supplémentaire. Aussi, cette structure donnerait un macromonomère légèrement différent du ***RISC2*** en apportant de la rigidité une fois incorporée sur la chaîne polymère.

ii) Utiliser la voie 2 (Schéma II. 6B), avec une étape de réduction additionnelle, menant aux propriétés structurelles quasi identiques au ***RISC2*** avec un seul carbone supplémentaire entre l'oxygène protégé par le DMT et le phényle.

La voie 1 a finalement été le premier choix pour réaliser les premiers tests. Après avoir protégé l'alcool avec le DMT, l'étape cruciale d'ATRA peut être effectuée. Cette dernière donne favorablement l'intermédiaire **e1** avec un bon rendement. Finalement, le couplage phosphoramidite sur l'alcool a été obtenu avec un rendement plus modéré de 50% (même ordre de grandeur observé avec le ***RISC2*** pour ce couplage) pour donner le macromonomère ***TISP***. Ce schéma de synthèse a également pu être appliqué au dérivé fluorobenzyle **T3** permettant d'obtenir **e1F** et ***TISP*-F** dans des rendements similaires, en apportant une première variation de masse à ce design.

Le macromonomère ***TISP*** ayant été obtenu, la prochaine sous-partie discutera de son incorporation dans une chaîne polymère.

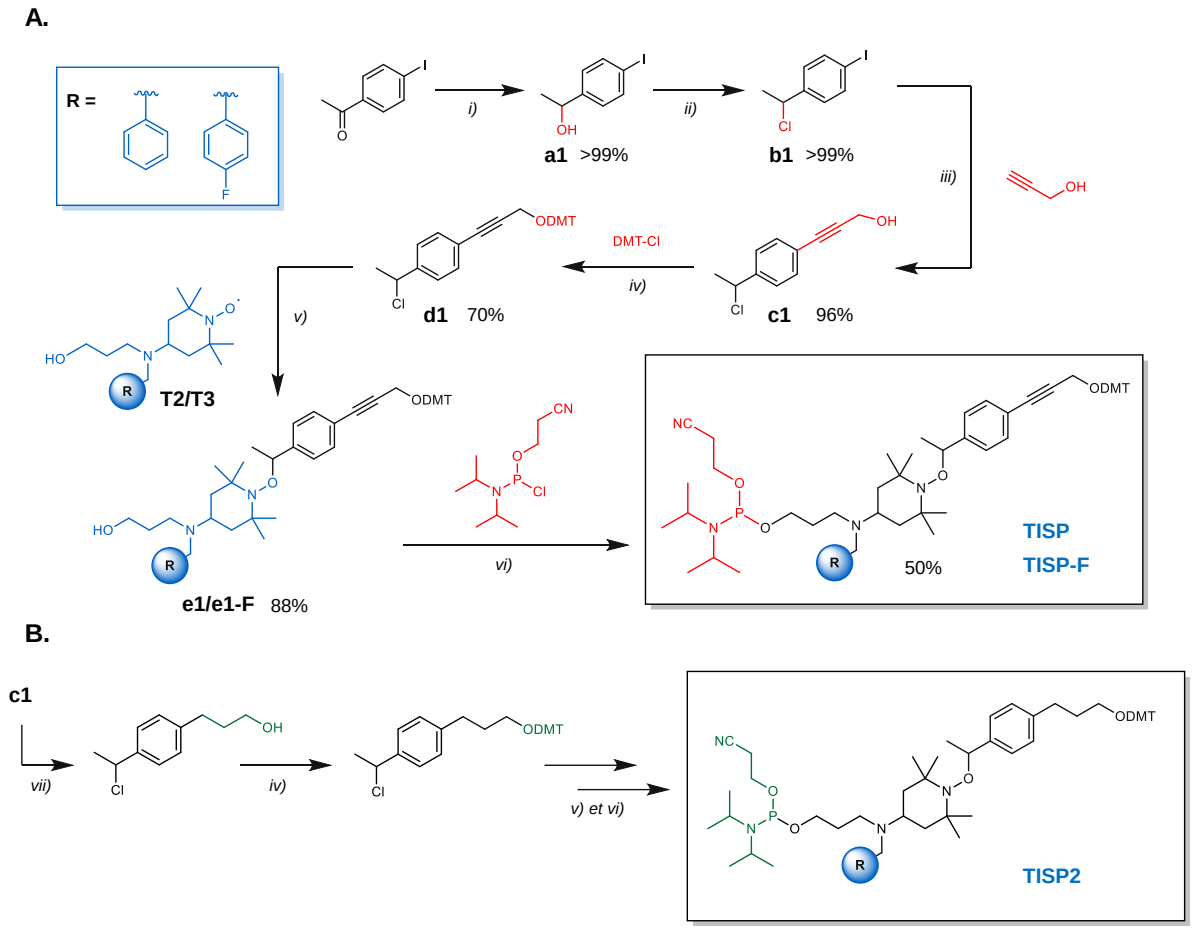


*Schéma II. 6 : Approche **TISP**/**TISP2** – Sonogashira.*
i) $NaBH_4$, EtOH/THF, 0°C à t.a., 3h; (ii) HCl 37%, THF, 0°C à t.a., 4h; (iii) $Pd(PPh_3)_2$, CuI, Et_3N, t.a., 17h; iv) Pyridine anhydre, THF anhydre, 0°C à t.a., 1h; (v) CuBr, Cu(0), PMDETA, THF anhydre, t.a ., 17h; (vi) DIPEA, DCM anhydre, 0°C à t.a., 1h. vii) Pd/C, H_2 (g), t.a., 1h.

3. Incorporation du *TISP* dans une chaîne polymère contrôlée

3.1. Méthodes pour la synthèse de polymères numériques

Afin de tester l'efficacité d'incorporation du monomère ***TISP*** dans une chaîne polymère contrôlée, le design présenté dans la Figure II. 3 a été choisi. Celui-ci utilise les monomères **0** et **1** décrits précédemment dans le Chapitre 1 (3.2.1) et par notre groupe.[296] Cet alphabet à deux monomères a été utilisé pour ces travaux car il donne une manière simple de coder de l'information binaire dans le polymère, ce qui facilite l'étude du monomère ***TISP*** au sein de la macromolécule numérique. Par ailleurs, les polymères sont synthétisés par chimie de phosphoramidite automatisée sur support solide chargé d'une thymidine. Ce dernier est simple d'utilisation et bien connu dans notre laboratoire, facilitant la construction et caractérisation des premiers polymères étudiés.

Enfin, différents segments d'information insérés de part et autres du marqueur clivable (Figure II. 3, haut) permettent de vérifier si l'espaceur a pu être incorporé correctement entre les blocs codants. De plus, ce design permet de contrôler l'enchainement des

monomères **0** et **1** par séquençage pseudo-MS^3, particulièrement pour vérifier le comportement du monomère ***TISP*** lors de sa fragmentation.

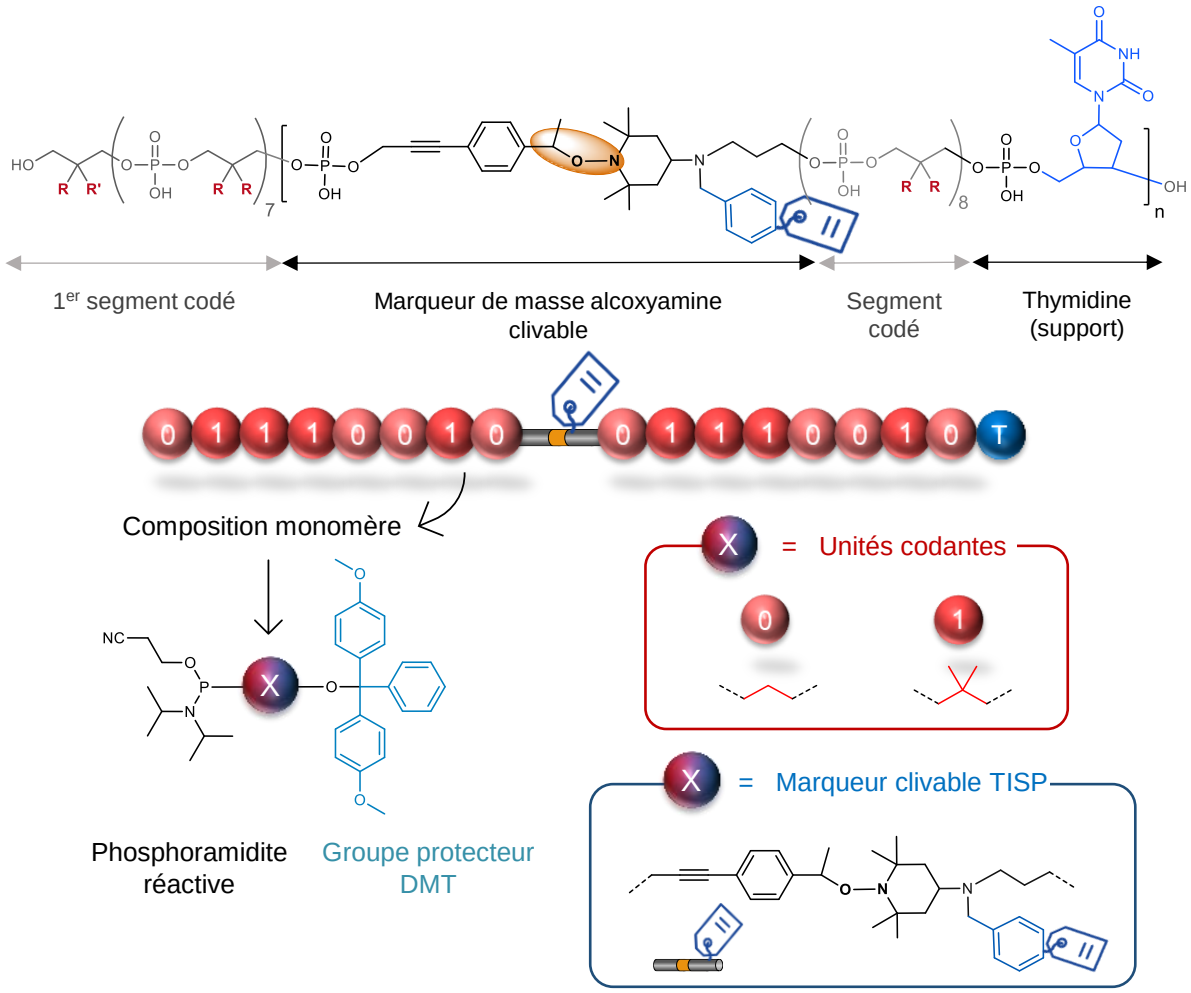


*Figure II. 3 : Design utilisé pour l'étude d'incorporation des monomères **TISP** en chaîne polymère contrôlée.*

3.2. Comment synthétiser et purifier les polymères ?

Il semble important d'expliquer la méthode de préparation des polymères afin de mieux comprendre l'étude de ces derniers. Les études effectuées au cours de cette thèse pour la construction de polymères numériques ont principalement été menées sur un robot synthétiseur d'ADN ABI 3900 (Figure II. 4) présenté précédemment (Chapitre 1, 2.2.3.1). Un tel instrument est équipé de différentes positions pouvant accueillir les réactifs phosphoramidites : 10 monomères différents peuvent être utilisés simultanément pour une même synthèse. Ces réactifs sont préalablement solubilisés dans l'acétonitrile anhydre à une concentration de 0.5 M et sont branchés sur le robot dans des fioles adaptées sous Argon. Parallèlement, l'instrument est doté de réactifs nécessaires aux différentes étapes de couplage phosphoramidite (TCA pour le déblocage, tétrazole pour l'activation et le couplage, anhydride acétique pour le piégeage et I_2 pour l'oxydation). Ces différentes opérations sont effectuées à l'intérieur d'une chambre de synthèse pressurisée sous Argon lors de la synthèse. À l'intérieur de celle-ci, jusqu'à 48 colonnes munies de

billes de supports solides peuvent être placées pour accueillir les chaînes polymères croissantes. Enfin, un ordinateur relié au robot permet de contrôler différents paramètres tels que les protocoles, les calibrations des lignes, et surtout l'ordre de la séquence à synthétiser.

Figure II. 4 : Robot synthétiseur ABI3900 utilisé dans ces travaux.
**Tétrazole, Anhydride acétique et I_2.*

Avec un cycle itératif de synthèse autour des 5 min, la construction d'un polymère de 27 unités nécessite environ 2h. Pour un polymère plus long de 175 unités par exemple, le temps sera approximativement de 12h. La répétition d'autant d'étape manuellement reste difficile voire impossible, spécifiquement avec un temps de synthèse aussi court.

Une fois la synthèse terminée, le polymère doit être clivé du support et purifié. Pour le clivage, la déprotection des groupes cyanoéthyles (protégeant le phosphate) et des éventuelles bases nucléiques, une solution de NH_3/$MeNH_2$ est généralement utilisée. Ce mélange est appliqué à l'aide d'une seringue à travers la colonne puis, après 45 min, le polymère est récupéré en solution. L'étape d'après consiste à purifier le polymère en lavant toutes les chaînes tronquées et piégées lors du cycle itératif à l'aide de colonnes de purifications de silice en phase inverse.

Note importante : le polymère est synthétisé avec le mode « DMT-ON » qui signifie que le dernier monomère ajouté ne sera pas déprotégé et gardera son groupe DMT. Ceci permet entre autres d'éliminer les séquences tronquées par simple lavage, tandis que le composé d'intérêt restera attaché à la colonne grâce au groupe protecteur apolaire.

Un protocole bien établi pour les oligonucléotides naturels[367] peut également être appliqué aux poly(phosphodiesters) synthétiques. Il peut s'effectuer avec des seringues ou bien à l'aide d'une pompe à vide. Cette dernière option présente plus de facilités dans le cas où plusieurs déprotections sont à effectuer simultanément (Figure II. 5). Le protocole pour une telle purification suit les étapes suivantes :

- La colonne de silice est conditionnée avec de l'acétonitrile et de l'acétate de triéthylamine (TEEA). L'acétonitrile élimine les résidus organiques de la résine et la mouille, tandis que la TEAA agit comme un réactif de pairage ionique pour améliorer la liaison de l'oligonucléotide DMT-ON à la résine.

- Le polymère clivé mixé à une solution saline (NaCl) est déposé sur la colonne de purification et reste attaché à la colonne grâce au DMT.
- Une solution de lavage (5% ACN dans H_2O(NaCl)) permet de rincer et éliminer les séquences tronquées.
- Le TCA permet ensuite de déprotéger le DMT, une ligne orange apparait en conséquence sur la colonne.
- Un rinçage à l'eau est par la suite effectué pour éliminer les excès de TCA.
- Le polymère peut enfin être élué dans différents *Eppendorf*^{TM} à l'aide d'une solution de 50% ACN dans l'eau en présence de 0.5% de NaOH. L'hydroxyde d'ammonium dilué est notamment utilisé pour neutraliser tout TCA restant.
- Finalement, les différentes fractions obtenues sont lyophilisées durant une nuit pour donner le polymère sous forme d'huile/solide dépendant de la pureté et de la séquence.

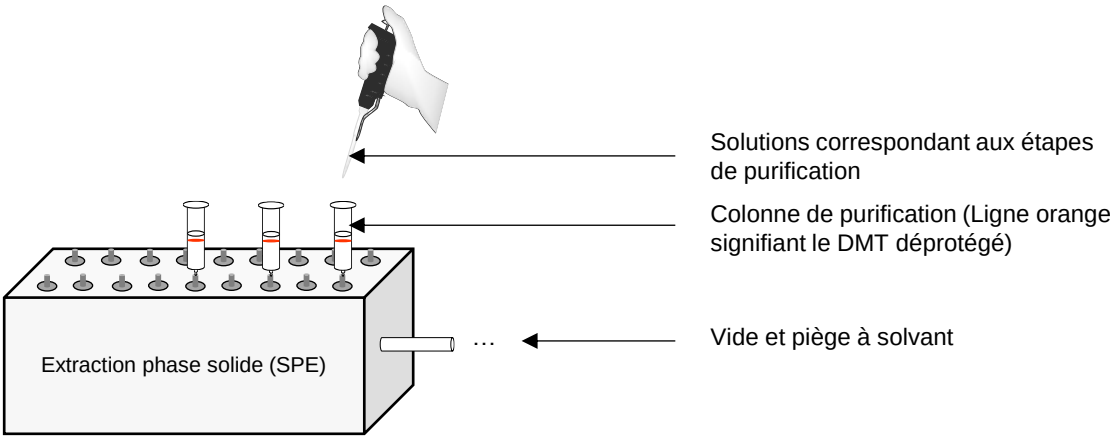


Figure II. 5 : Montage de purification des polymères synthétisés.

3.3. Incorporation du *TISP* dans la séquence

Afin de tester le comportement des deux marqueurs d'octets ***TISP*** préalablement obtenus (Schéma II. 6) dans une chaîne polymère informationnelle, des séquences modèles ont été construites en utilisant le robot synthétiseur ABI 3900 selon les procédures énoncées dans la sous-partie précédente (3.2). Deux premières macromolécules ont été synthétisées en choisissant des séquences très simples telles que :

- 0000 – ***TISP*** – 0000 – T (Figure II. 6, **P1**)
- 00000000 – ***TISP*-F** – 11111111 – T (Figure II. 6, **P2**)

Rappel : les monomères **0** et **1** font références aux monomères phosphoramidites précédemment introduits (Chapitre 1, 3.2.1 et Chapitre 2, 3.1), et T fait référence ici à la thymidine présente sur le support décrite également (Figure II. 3).

Ces deux premiers essais ont démontré que l'incorporation du ***TISP*** est possible dans une chaîne polymère puisque l'analyse ESI-HRMS des deux macromolécules ont indiquées la présence du polymère ciblé (Figure II. 6). Cependant, ces dernières ne sont pas uniformes du fait des impuretés visibles sur le spectre.

Dans **P1**, elles sont relativement mineures hormis l'impureté D qui correspond à une différence de -539 Da avec le polymère ciblé. Une analyse MS/MS de celle-ci permet de confirmer la présence du bloc se situant à droite de l'alcoxyamine. En s'appuyant sur ces données et la différence de masse observée, une structure a pu être proposée correspondant à la perte d'un bloc codant substitué par une amine (Figure II. 6, **P1**, structure grise). Une réaction non désirée lors du clivage hypothétiquement trop long avec une solution basique (NH_3/$MeNH_2$) a alors été considérée. Ces défauts de masses restent finalement négligeables puisqu'ils n'ont plus été remarqués sur des tests ultérieurs.

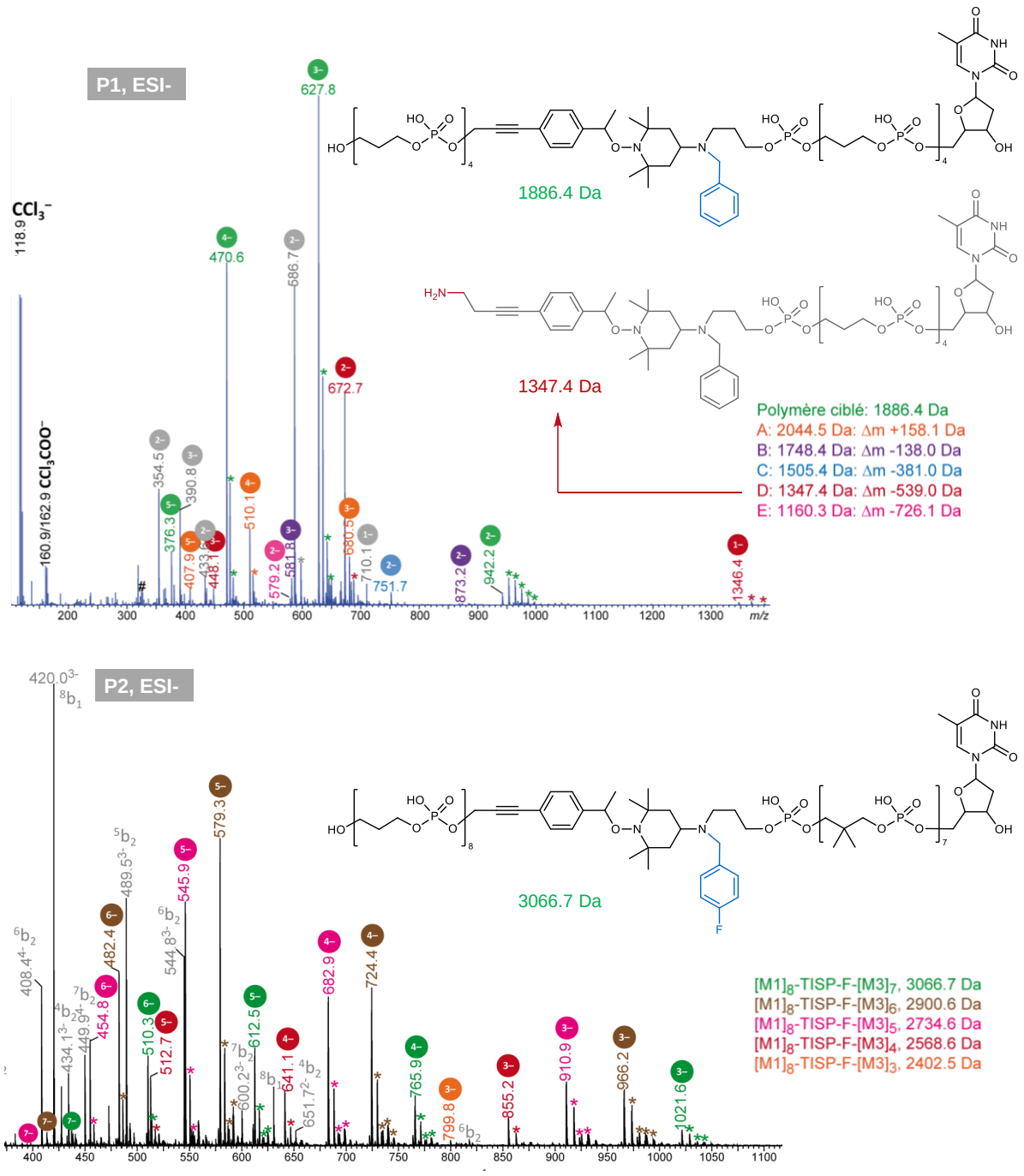


*Figure II. 6 : Spectres ESI-MS en mode négatif des échantillons **P1** et **P2**.*
(Gris = fragments en sources)

Sur ce même spectre, des traces de TCA sont également visibles. Ceci peut s'expliquer par un manque de lavage à l'eau lors de la purification, ou par les conditions de préparation de l'échantillon pour l'analyse en ESI-MS.

Dans **P2**, la plupart des impuretés sont majoritaires par rapport au polymère ciblé en vert. En l'occurrence, une différence de -166 Da peut être observée entre chacune des impuretés représentées. Ceci peut clairement être expliqué par le couplage non effectif des monomères **1** dans le second bloc (Figure II. 7).

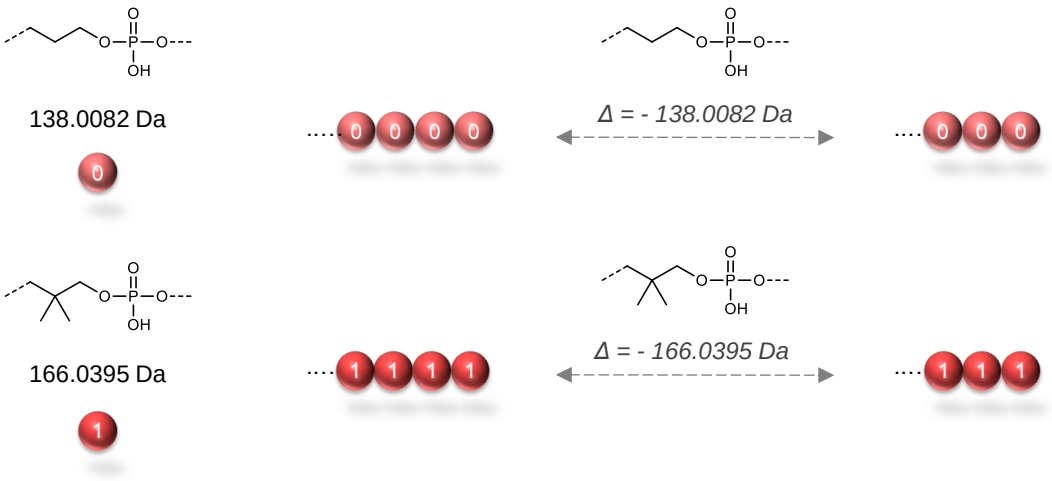


*Figure II. 7 : Défaut de masse lorsqu'une unité **0** ou **1** est manquante.*

Ce problème peut se manifester lorsque les conditions de synthèses ne sont pas optimales, et particulièrement avec l'usage de réactifs phosphoramidites trop peu anhydres. En effet, un des principaux obstacles dans l'obtention d'une efficacité de couplage >99% sont les traces d'eau qui peuvent être présentes dans certains réactifs utilisés dans la synthèse. Celles-ci peuvent diminuer l'efficacité de couplage de deux façons :

- Lors d'un couplage, l'eau peut réagir avec le monomère activé au tétrazole, empêchant ce dernier à réagir avec la position 5'-hydroxyl attachée au support, ce qui va donc réduire l'excès de monomère et baisser l'efficacité de couplage.
- L'eau peut tout simplement inhiber la fonction phosphoramidite en catalysant sa conversion en phosphonate, et ainsi réduire la concentration de phosphoramidite présente en solution.

Afin d'optimiser la synthèse, il est donc nécessaire de réduire un maximum la teneur en eau de chacun des réactifs impliqués dans l'étape de couplage (monomères phosphoramidites, ACN, *1H*-tetrazzole …). Idéalement, des solutions fraiches de monomères dans de l'acétonitrile strictement anhydre doivent être préparées avant chaque synthèse. Des petits sachets incorporés dans les fioles permettant de piéger les traces d'eau restantes peuvent également être utilisés à cette fin. Cependant, respecter des conditions anhydres strictes restent parfois difficile à accomplir lorsque la quantité de monomère engagée est limitée. Par exemple, dans les tests préliminaires impliquant le

TISP, les premières quantités obtenues de ce marqueur d'octet ont été de l'ordre de 100 mg. En appliquant une concentration de 0.05 M, seulement 2 mL ont été disponible pour des premiers essais. Sachant que les monomères doivent être calibrés sur les lignes du robot, et qu'un couplage lors du cycle itératif emploie 0.225 mL, peu de combinaisons sont alors possibles.

Néanmoins, les essais préliminaires effectués restaient à ce stade prometteurs et d'autres études ont été nécessaires pour confirmer la viabilité de ces espaceurs. Notamment, vérifier leur comportement lors d'un séquençage pseudo-MS^3 en impliquant une séquence modèle de 3 octets et les deux marqueurs d'octets clivables ***TISP*** et ***TISP-F***. La prochaine sous-partie discutera de ces évaluations.

3.4. Séquençage du polymère : résultat et discussion

Afin de vérifier la possibilité de retrouver la séquence initialement encodée dans le polymère, une séquence modèle a été construite telle que :

- 0000 – ***TISP*** – 1111 – ***TISP-F*** – 0000 – T (Figure II. 8, Figure II. 9, **P3**)

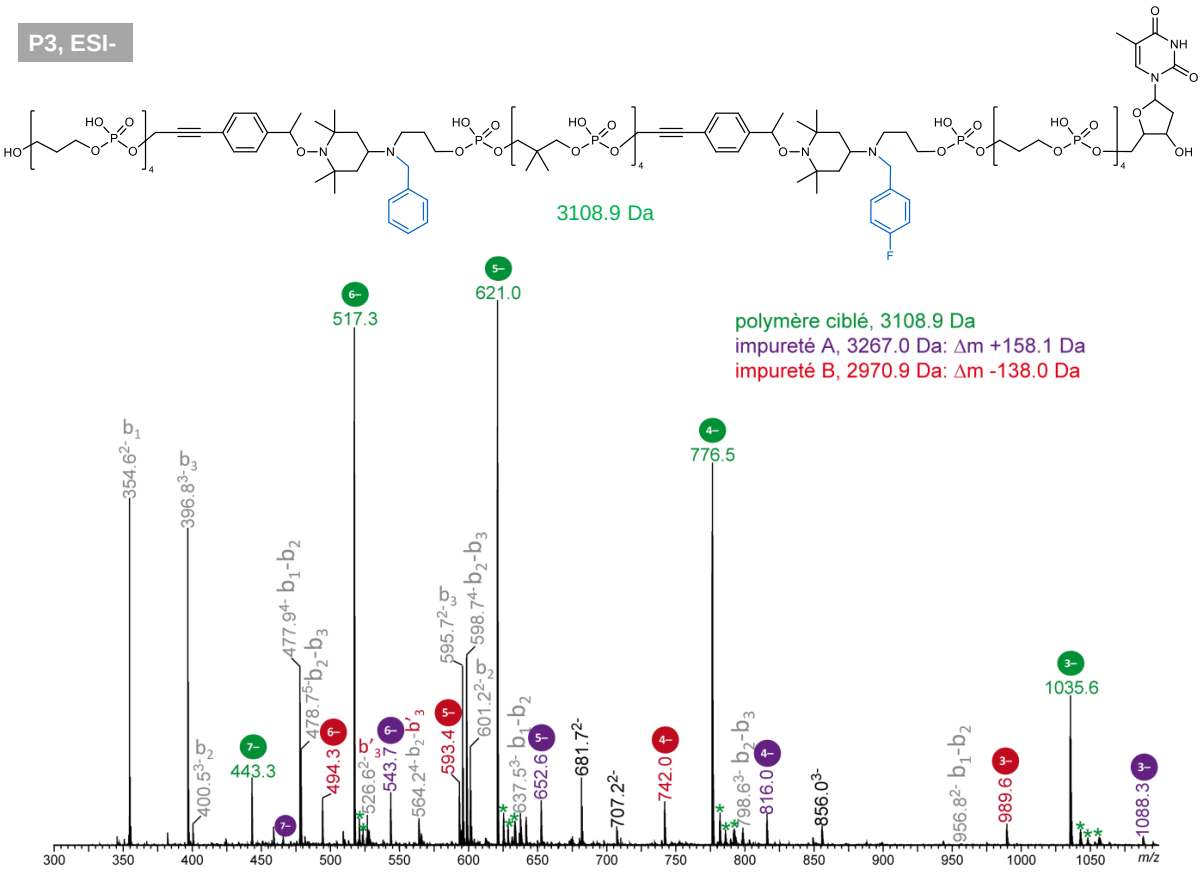


*Figure II. 8 Spectre ESI-MS en mode négatif de l'échantillon **P3**.*
Les pics repérés par un astérisque sont issus d'échanges H/Na et/ou H/K. Les fragments libérés en source sont notés en gris bi (ou b'i si défectueux) avec i le numéro du bloc qu'ils contiennent.

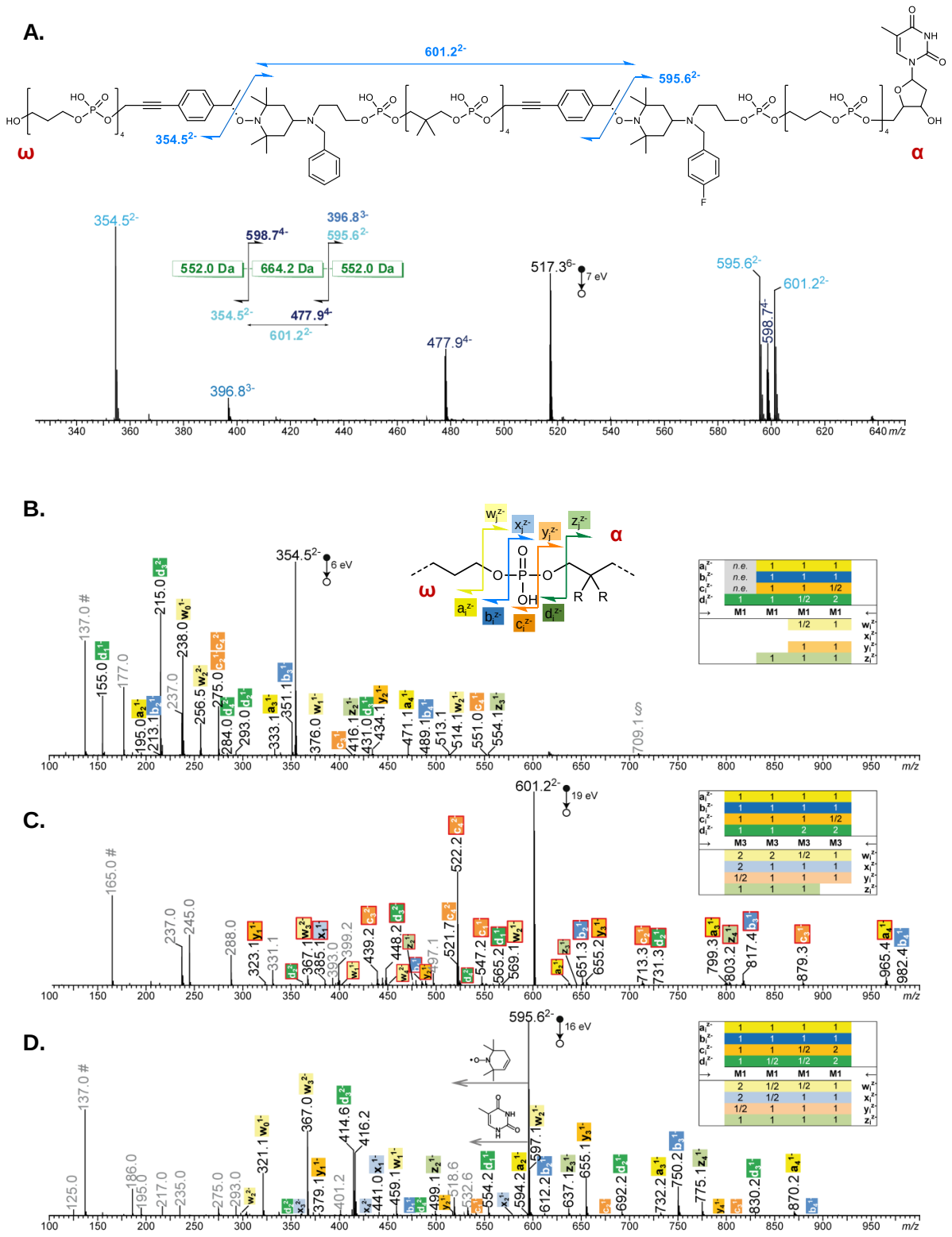


*Figure II. 9 : Spectre MS/MS et MS³ de **P3**.*

***A.** Spectre MS/MS de l'ion [M – 6H]⁶⁻ à m/z 517.3 et schéma de dissociation du polymère attendu dans l'échantillon **P3** (les valeurs encadrées correspondent aux masses des blocs codants). Spectres pseudo-MS³ des ions **B.** m/z 354.5²⁻ contenant le bloc 1, **C.** m/z 601.2²⁻ contenant le bloc 2, et **D.** m/z 595.6²⁻ contenant le bloc 3. Les annotations encadrées de rouge pour les fragments du segment central (**C.**) indiquent que les ions a/b/c/d sont majoritairement détectés avec un excès de masse de 1 Da tandis que les ions w/x/y/z sont majoritairement détectés avec un défaut de masse de 1 Da. Les fragments annotés en gris sont des fragments internes, parmi lesquels les monomères déprotonés sont repérés par #.*

Le spectre de masse enregistré en mode négatif pour l'échantillon **P3** montre nettement le polymère recherché sous les états de charges 3– à 7– (en vert, Figure II. 8). Cette attribution est également validée par le spectre MS^2 de l'ion $[M - 6H]^{6-}$ (Figure II. 9A) et le séquençage individuel des 3 blocs par pseudo-MS^3 (Figure II. 9B-D). Dans ce cas, seulement deux impuretés mineures ont été détectées : l'impureté A présente un excès de masse Δm +158.1 Da, identifié comme $C_{11}H_{10}O$ par mesure de masse exacte et localisée soit dans le segment contenant le bloc 1 soit dans le segment contenant le bloc 2 et, plus précisément, dans le "résidu alkyle" de l'alcoxyamine. L'impureté B correspond quant à elle à un manque d'unité **0**, certainement lié à une mauvaise efficacité de couplage. Malgré l'obtention du polymère ciblé en majorité, son séquençage MS^3 a laissé apparaître quelques problématiques.

Premièrement, la réactivité de la partie nitroxide avec l'élimination du TEMPO par transfert d'un proton sur l'amine tertiaire a été observée (Schéma II. 7). Toutefois, cette réaction n'a alors présenté aucune gêne dans l'interprétation des signaux.

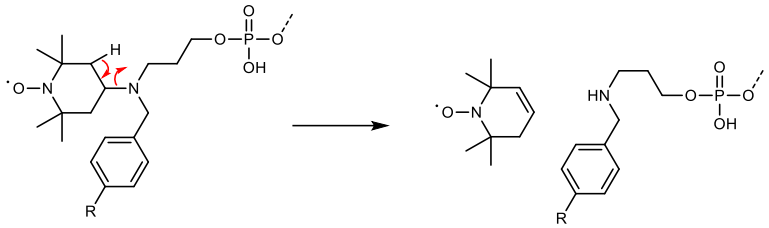


*Schéma II. 7 : Mécanisme conduisant à l'élimination du TEMPO de la partie nitroxide des alcoxyamines **TISP**.*

Autrement, alors que les données pseudo-MS^3 du segment central permettent bien de retrouver les fragments attendus pour assurer la couverture de la séquence (Figure II. 9C), ces fragments restent néanmoins minoritaires devant des ions détectés avec un excès ou défaut de masse Δm +/- 1 Da indiquant l'existence d'un transfert de radical H•. Cette réaction parasite ne se produit que dans les segments contenant à la fois le "résidu alkyle" et le nitroxide, suggérant un rôle de ce dernier dans le mécanisme, et particulièrement l'alcyne. En effet, la présence de la triple liaison renforce la stabilisation du radical dont la délocalisation est beaucoup plus forte que dans le cas de ***RISC2*** (Schéma II. 8). Ceci peut notamment être constaté expérimentalement puisque l'homolyse de la liaison NO-C en MS/MS nécessite des énergies d'activation sensiblement plus faibles que dans le cas de ***RISC2***. Dans la forme mésomère limite, l'atome d'hydrogène indiqué en rouge devient très facile à transférer sous la forme d'un H• car sa position ne bénéficie plus d'aucun effet de rigidification vis-à-vis de groupements situés à gauche, favorisant son départ.

En conséquence, les fragments de séquençage issus des ruptures des liaisons phosphates sont majoritairement détectés avec un excès de masse Δm +1 Da quand ils contiennent le groupement terminal α (a/b/c/d) et à un défaut de masse Δm -1 Da lorsqu'ils contiennent le groupement terminal ω (w/x/y/z). Néanmoins, les fragments théoriques attendus sont également observés mais avec des abondances beaucoup plus faibles, ce qui implique des problèmes de détections lorsque la taille des blocs codant augmente.

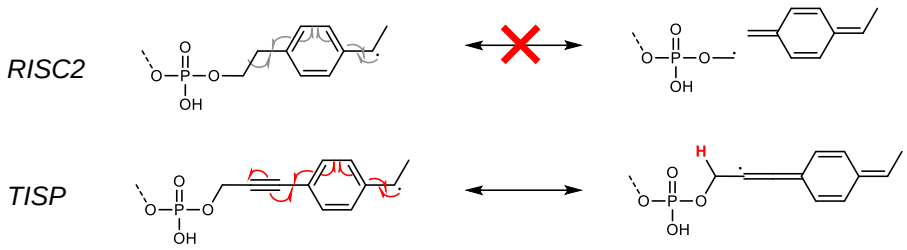


*Schéma II. 8 : Comparaison des délocalisations électroniques entre le **RISC2** et le **TISP**. Haut : délocalisation non favorable du radical dans le résidu alkyle **RISC2**. Bas : délocalisation favorable du radical dans le résidu alkyle **TISP** facilitant le transfert du radical H•.*

À la suite de ces observations, la stratégie alcyne a nécessité d'être réétudiée. Effectivement, en assumant que la triple liaison joue un rôle dans l'abstraction d'un H•, sa réduction totale devrait certainement résoudre le problème et éliminer les dissociations radicalaires non désirées lors du séquençage pseudo-MS^3. La prochaine sous-partie discutera de la synthèse d'un nouveau marqueur d'octet clivable correspondant au ***TISP*** saturé : le ***TISP2***.

4. *TISP2* – synthèse et incorporation dans la chaîne polymère

4.1. Réduction de l'alcyne

Les premiers essais de réduction de l'alcyne ont employé des conditions classiques d'hydrogénation avec du Pd/C et de l'hydrogène gazeux, auparavant décrites dans la littérature.[368,369] Cependant, comme il peut l'être remarqué sur la Figure II. 10 au milieu, cette réduction n'atteint pas 100% de conversion et donne, de manière surprenante, un mélange d'alcane et d'alcène (les ratios sont difficiles à quantifier étant donné la superposition des signaux).

En supposant que le manque de conversion de l'alcyne vers l'alcane est dû à la source d'hydrogène qui est apportée à l'aide d'un ballon, un montage pour faire buller l'hydrogène à l'intérieur de celui-ci a été élaboré. Pourtant, rien n'a changé à la réactivité observée.

La source du problème peut éventuellement provenir du Pd/C utilisé qui, en étant mal conservé, pourrait facilement être désactivé et non efficace pour catalyser une hydrogénation. Cette hypothèse a également été vérifiée en acquérant le réactif fraichement chez un revendeur, mais sans succès. Enfin, pour essayer d'optimiser cette conversion, d'autres conditions de réactions ont été testées en utilisant le Pd/C ou encore, le $Pd(OH)_2$ et ont été résumées dans le Tableau II. 1.

Après ces observations non concluantes, l'idée d'utiliser un autre catalyseur ayant un pouvoir réducteur plus important, tel que le Pt/C, a été envisagée. Effectivement, le premier test réalisé avec ce dernier a converti 100% de l'alcyne en alcane sans traces d'alcènes, confirmé par analyse RMN du produit brut après 1h de réaction (Figure II. 10, bas).

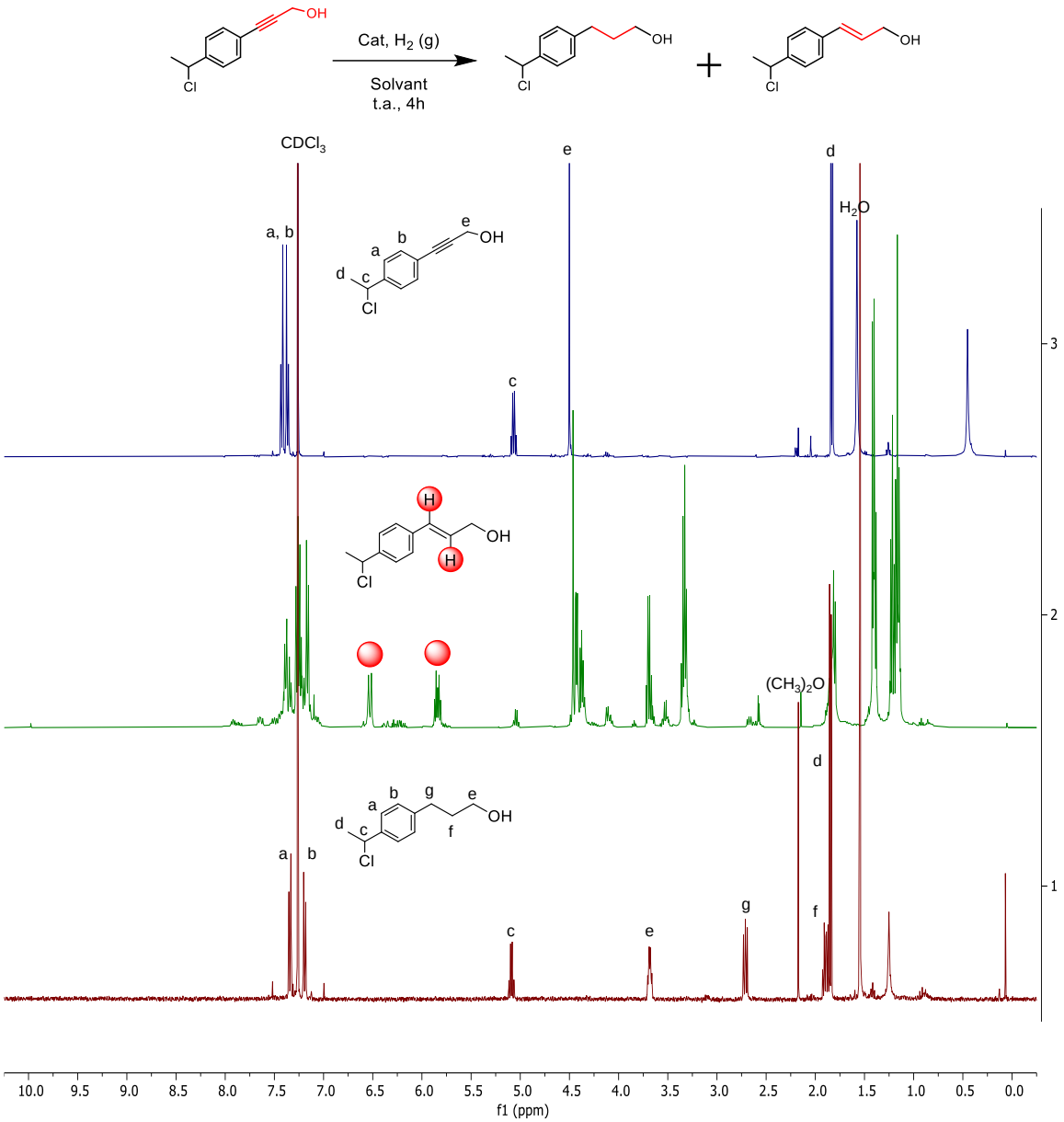


Figure II. 10 : Différents états de saturation du dérivé alcyne confirmé par RMN.

Catalyseur	Mol%	Solvant	Temps*	Résultat
Pd/C	0.3	EtOH	17h	Alcène
Pd/C	0.3	THF	17h	Alcène + Alcyne
$Pd(OH)_2$	0.25	MeOH (anh.)	3h	Alcène + Alcyne

Tableau II. 1 : Différentes conditions d'hydrogénation testées pour la réduction de l'alcyne.

Cependant, après purification du produit, le rendement observé a été inférieur à 15%. Même si le spectre RMN et la CCM du produit brut montrent des résultats encourageants avant traitement de la réaction, la purification du composé désiré est rendue difficile par les différents sous-produits avec des rapports frontales (Rf) quasi similaires. De plus, la principale problématique observée dans ces conditions est la formation d'un produit

secondaire majoritaire correspondant à une déshydratation (Figure II. 11). Un rendement de plus de 60% a pu être observé après purification de ce composé non espéré.

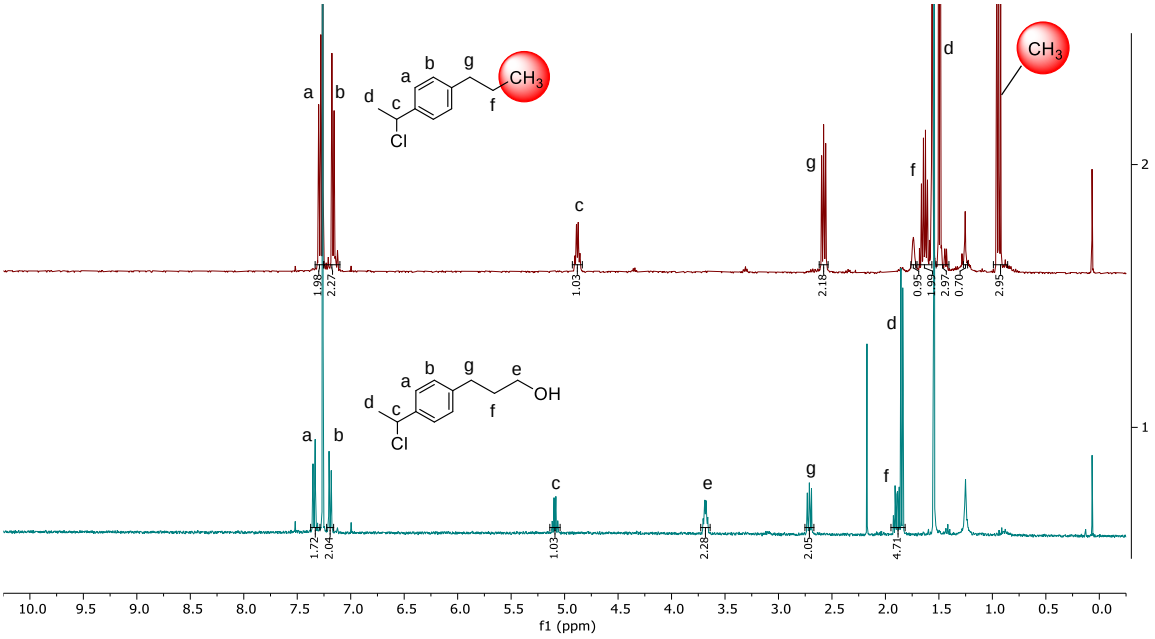


Figure II. 11 : Spectre RMN du produit secondaire observé (haut) correspondant à une déshydratation du produit désiré (bas).
Signal à 0.97 ppm correspondant au CH_3 (boule rouge).

En hypothétisant que cette déshydratation provienne des conditions anhydres pouvant parfois être mal appliquées, des conditions anhydres strictes ont été mises en œuvre (ballon chauffé et séché sous vide, nouvelle bouteille de THF anhydre, réactifs strictement séchés …). Pourtant, le dérivé alcane secondaire est toujours autant présent dans ces conditions.

Une autre théorie a suggéré que cette déshydratation se produit dans un second temps, une fois que la réduction de l'alcyne est terminée. De ce fait, des analyses RMN à différents temps de réactions (30min, 1h et 2h30) ont été effectuées et ont pu être comparées sur la Figure II. 12.

Il peut être remarqué que la formation du produit secondaire (signal correspondant au groupe CH_3 en rouge) et la consommation du réactif de départ alcyne (signal correspondant au groupe CH_2 en vert) évoluent avec une cinétique quasi similaire. Ceci indique que la déshydratation prend certainement place au moment de la réduction de l'alcyne, et donc, un temps de réaction réduit n'empêche pas la formation du dérivé alcane non désiré.

Finalement, des derniers essais d'optimisation ont été réalisés en testant différents solvants en pensant qu'une sélectivité peut être induite dans certaines mesures, comme indiqué dans la littérature[370,371] (Tableau II. 2). Il est important de noter que ces résultats ont fait l'objet de tests d'optimisation préliminaires. Dans ce cas, la conversion de l'alcyne et la formation du dérivé alcane non désiré ont été quantifiées de manière arbitraire en observant la disparition/formation des signaux correspondants (singulet à 4.7 ppm pour l'alcyne et triplet à 0.97 ppm pour l'alcane, Figure II. 12), sans l'utilisation d'étalon interne.

Le but a été, pour cette étude, de vérifier si certaines conditions permettent de s'affranchir complétement du sous-produit de réaction.

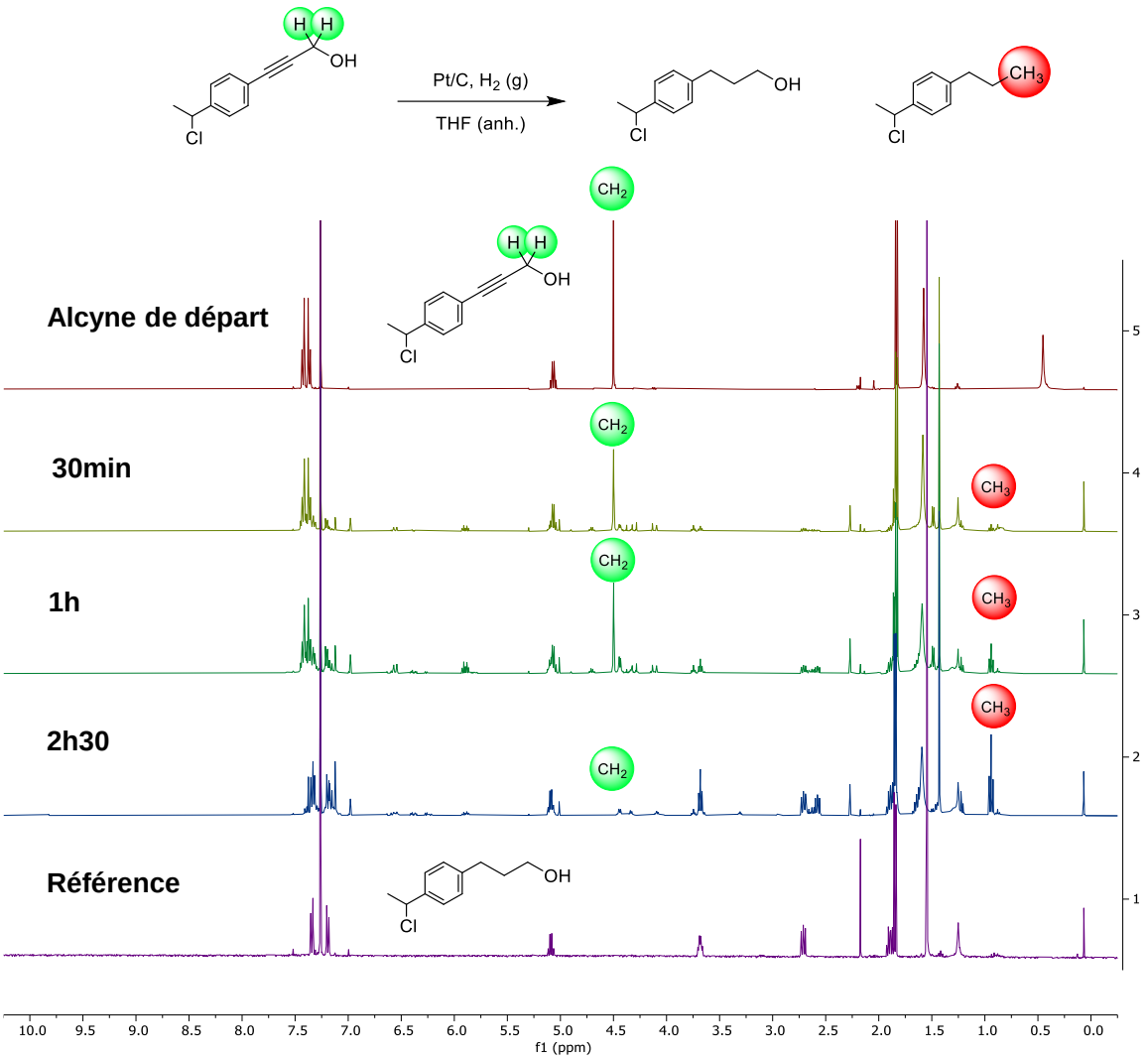
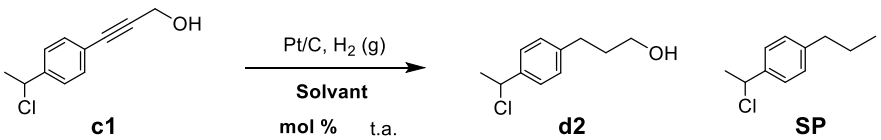


Figure II. 12 : Contrôle de la formation du produit secondaire par RMN à différents temps de réactions.
Triplet à 0.97 ppm correspondant au CH_3 du produit secondaire (boule rouge) et singulet à 4.4 ppm correspondant au CH_2 du réactif de départ alcyne (boule verte).

Autre remarque, les différents essais ont d'abord été effectués avec des réactifs de départ différents. Il a notamment été remarqué que la pureté du réactif est très importante pour réaliser cette hydrogénation. Effectivement, en utilisant des conditions similaires, dans certains cas, l'alcyne n'a pu être réduit en conséquence des impuretés présentes dans l'échantillon de départ. C'est pourquoi, tous ces essais ont été réalisés en choisissant un alcyne, dont la pureté a rigoureusement été contrôlée.



Entrée	mol%	Solvant	Temps	Résultat (produits presents)
1	0.3	THF	24h	Conversion **c1** 100% - Produits **d2** et **C**
2	0.2	THF	2h30	Conversion **c1**100% - Produits **d2** et **C**
3	0.2	THF (anh)	4h	Conversion **c1** 100% - Produits **d2** et **SP**
4	0.1	THF (anh)	24h	Conversion **c1** 100% - Produits **d2** et **SP**
5	0.2	MeOH (anh)	24h	Conversion **c1** 100% - Produits **d2** et **SP**
6	0.2	EtOH/H_2O	24h	**c1** non consommé totalement – Produits **d2** et **SP**
7	0.2	EtOH	24h	**c1** non consommé totalement – Produits **d2** et **SP**
8	0.2	EtOH/THF	2h30	Conversion **c1** 100% - Produits **d2** et **SP**
9	0.2	EtOAc	2h30	Conversion **c1** 100% - **d2** et **Grande** quantité de **SP**
10	0.2	*t*-BuOH	24h	Pas de conversion
11	0.2	THF/NEt_3	24h	Mélange non identifiable comprenant **c1**, **d2** et **SP**
12	0.2	THF/MeOH	24h	Pas de conversion
13	0.2	THF/H_2O	2h30	Conversion **c1** 100% - **d2** et **Petite** quantité de **SP**

*Tableau II. 2 : Comparaison de différentes conditions utilisées pour l'hydrogénation de l'alcyne **c1**.*

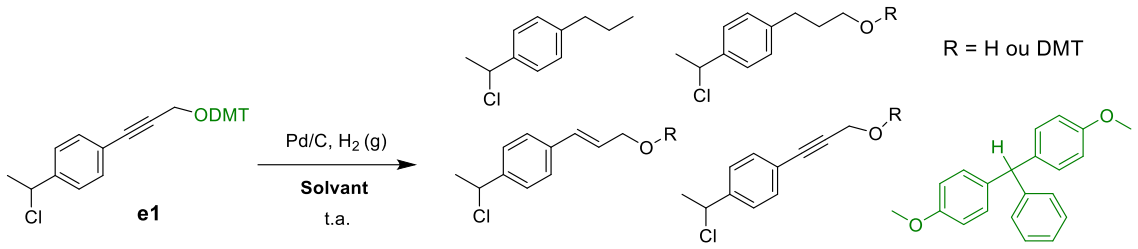
*Chacun des tests ont été effectué avec 0.025g d'alcyne **c1** pure en faisant buller de l'Hydrogène à l'aide d'un ballon, à l'intérieur du ballon réactionnel. Les données arbitraires récoltées ont été attribuées par intégration de pics, les produits n'ont pas été isolés pour déterminer leur rendement.*

Les entrées 1 et 2 ont été réalisées dans du THF non anhydre avec deux quantités de catalyseurs différentes. Ce dernier paramètre ne semble pas influer sur la formation du sous-produit **SP.** Les mêmes conclusions ont pu être établies en utilisant cette fois-ci du THF anhydre. Dans les entrées 5 à 13, différents solvants ont été testés afin d'essayer de limiter la production du dérivé alcane **SP**, mais sans succès. Dans certains cas, 100% de conversion de l'alcyne **c1** est observable mais la quantité du produit secondaire reste inchangée. Étonnamment, un des meilleurs résultats obtenus a été en employant un mélange THF/H_2O qui dans ce cas a montré une quantité moindre de **SP**. (Vérifiée en comparant l'intégration du CH_3 en bout de chaîne avec les autres tests effectués dans ce tableau).

Étant donné que les différents essais d'optimisations en variant les solvants et les quantités de catalyseurs n'ont pas été probants, une autre stratégie a été envisagée en protégeant l'alcool propargylique, discutée dans la sous-partie suivante.

4.2. Stratégie de protection/déprotection

Tout d'abord, afin d'éviter l'ajout d'étapes supplémentaires, l'hydrogénation de l'alcool propargylique protégé avec le groupe DMT a été testée. Cependant, cette stratégie n'a pas été concluante puisque ce groupe protecteur n'est pas orthogonal aux conditions de réductions en présence de Pt/C et par conséquent, son élimination a pu être observée, (Tableau II. 3, entrée 1). Effectivement, après purification, le sel de DMT ainsi que l'alcyne de départ ont pu être isolés sans traces de produits désirés.



Entrée	Solvant	Résultat après 1h	Résultat après 24h
1*	THF	Conversion **e1** 50%	Conversion **e1** 100% - Sel de DMT majoritaire
2	THF	Conversion **e1** 50%	Conversion **e1** 100%
3	ACN	Pas de conversion	Pas de conversion
4	MeOH	Conversion **e1** 50%	Conversion **e1** 100%

Tableau II. 3 : Différents tests d'hydrogénation réalisés sur l'alcyne protégé au DMT.
Pt/C utilisé démontrant une déprotection complète de **e1.*

En sachant que le Pt/C a un pouvoir plus réducteur que le Pd/C, il a été suggéré que l'utilisation de ce dernier peut réduire l'alcyne tout en laissant le groupe DMT. Cependant, les différents tests effectués dans ces conditions n'ont permis d'obtenir que des mélanges du dérivé attendu (Tableau II. 3, entrées 2, 3 et 4)

À la suite de ces résultats, il a paru évident qu'un groupe protecteur résistant aux conditions d'hydrogénation doit être utilisé. Le groupe TBDMS permet de répondre parfaitement à ces attentes puisqu'il a déjà été employé en présence d'hydrogène gazeux dans la littérature[372] (D'autres groupes tels que le tétrahydropyrane[373] a également été considéré, mais la disponibilité du TBDMS dans le laboratoire a facilité la mise en place des premiers essais).

Afin de tester sa robustesse, une nouvelle voie de synthèse a été élaborée en impliquant l'alcool propargylique protégé dans le couplage de Sonogashira (Schéma II. 9, étapes *ix* et *iv*). Grâce aux signaux des méthyles du groupe TBDMS facilement remarquables en RMN, l'efficacité de ce couplage a pu être vérifiée en isolant le produit **b2** avec un très bon rendement. Par la suite, l'hydrogénation de l'alcool protégé (*iv*) suivit de sa déprotection en présence de TBAF (*v*) a pu donner le dérivé réduit **d2** dans des rendements corrects.

Finalement, le schéma de synthèse initialement proposé a été repris à partir de l'alcool ainsi obtenu, en le protégeant avec un groupe DMT (*vi*), avant de réaliser une ATRA (*vii*) et un couplage phosphoramidite (*viii*).

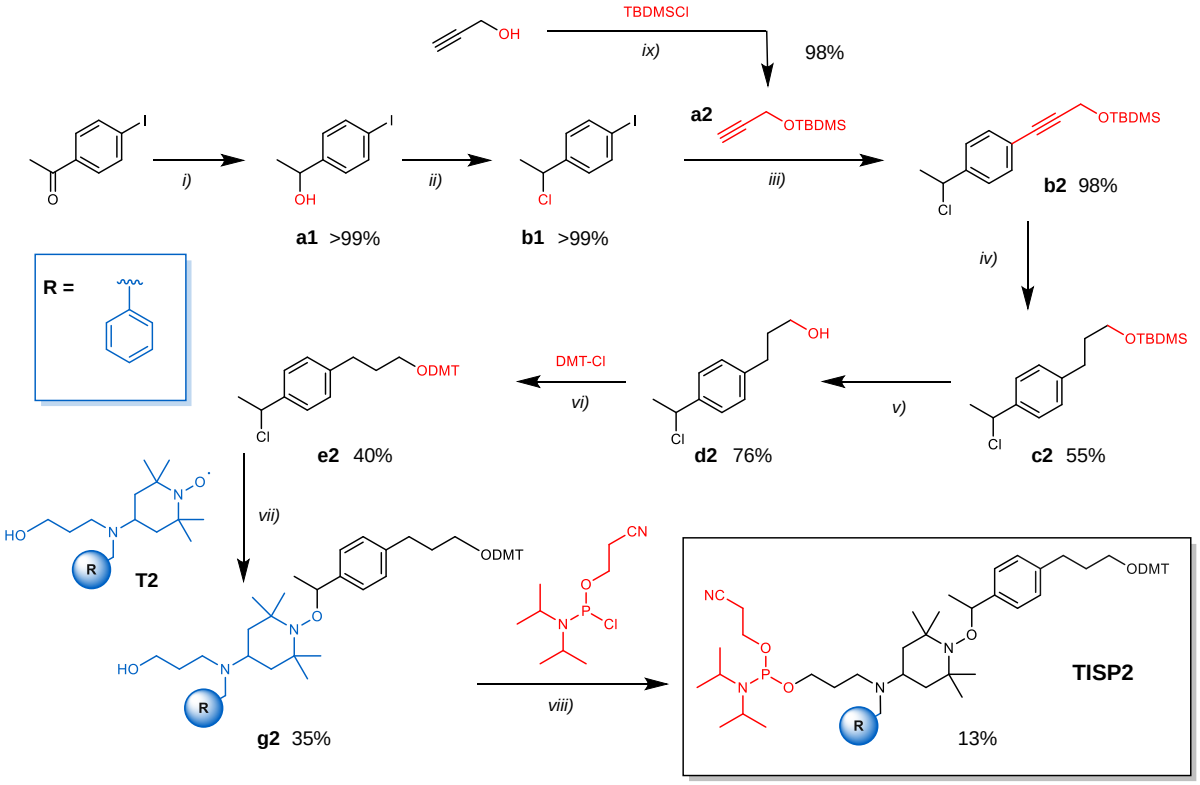


*Schéma II. 9 : Approche **TISP2** – stratégie protection/déprotection avec le TBDMS. i) $NaBH_4$, EtOH/THF, 0°C à t.a., 3h; (ii) HCl 37%, THF, 0°C à t.a., 4h; (iii) $Pd(PPh_3)_2$, CuI, Et_3N, t.a., 17h; iv) Pt/C, H_2 (g), t.a., 1h; v) TBAF (1M), THF, 0°C à t.a., 2h30; (vi) Pyridine anhydre, THF anhydre, 0°C à t.a., 1h; (vii) CuBr, Cu(0), PMDETA, THF anhydre, t.a ., 17h; (viii) DIPEA, DCM anhydre, 0°C à t.a., 1h; ix) Imidazole, DCM, 0°C à t.a., 1h.*

Les rendements obtenus sur ces premiers essais n'ont pas été probants, mais peuvent être expliqués et optimisés facilement par suite d'observations qui semblent importantes à détailler :

- Premièrement, il a été remarqué que l'intermédiaire alcool n'est pas stable et que sa protection doit être faite le plus rapidement possible. Lors des premiers essais, dans un souci de caractérisation détaillée, cet intermédiaire a parfois dû rester quelques jours sur la paillasse avant d'être parfaitement identifié. Même sous vide, ou dans le réfrigérateur, l'alcool ne semble stable et de ce fait, sa protection a par la suite été partiellement compromise. Effectivement, alors que l'analyse ESI-MS et MS/MS du macromonomère ***TISP2*** a permis de bien identifier le composé ciblé, ces mêmes résultats ont démontré l'existence d'espèces secondaires et une particulièrement majoritaire à 1147.7 Da (Figure II. 13). Ce sous-produit est en fait issu de réactions de dimérisation qui impliqueraient certainement l'intermédiaire alcool réduit (Figure II. 14).

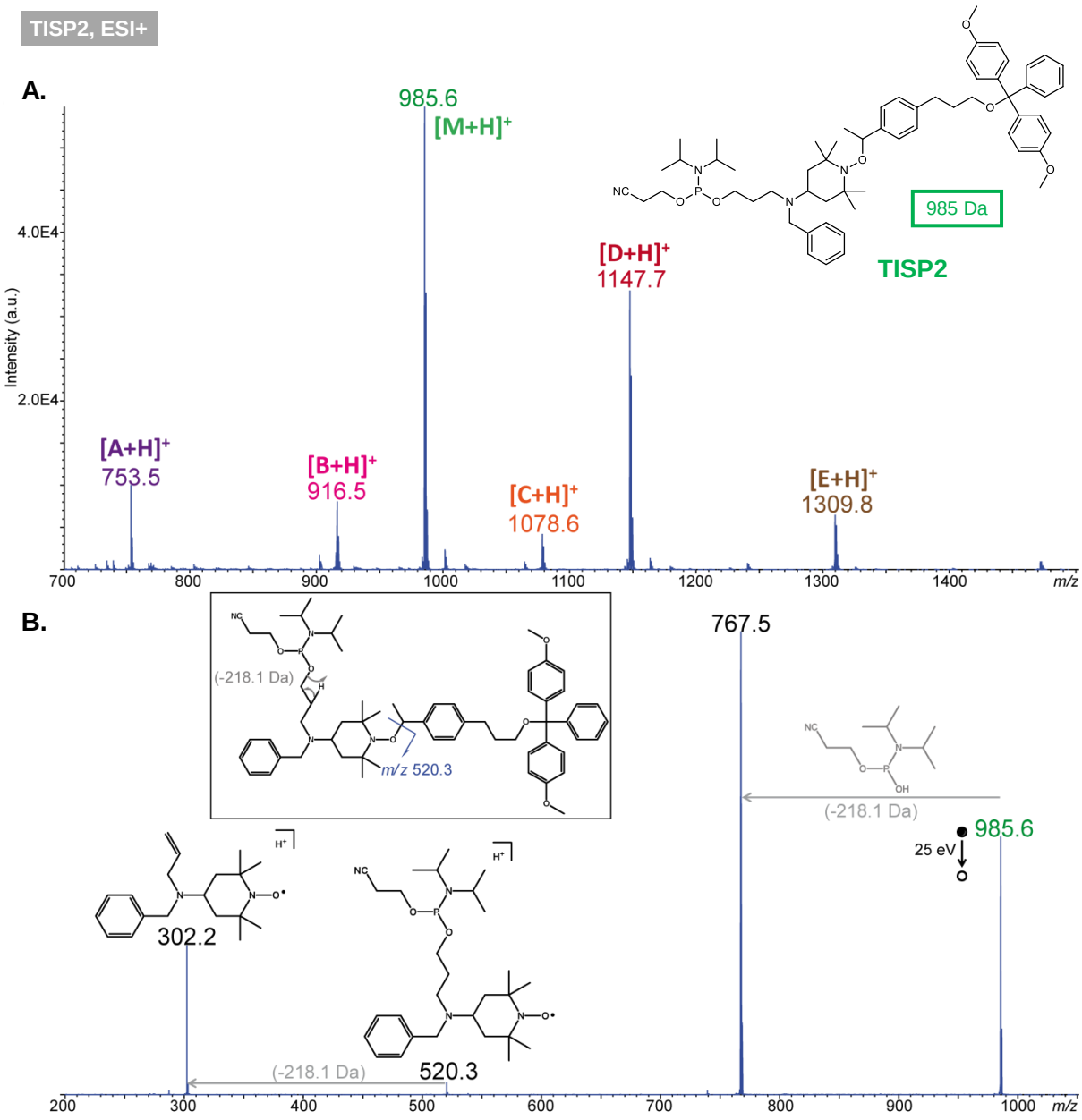


*Figure II. 13 : Spectre de masse validant la nature du **TISP2**.*
***A.** Spectre ESI(+)-MS de l'échantillon **TISP2**. **B.** Spectre ESI-MS/MS de l'ion [M+H]+ à m/z 985.6.*

En outre, le dimère formé est très difficile à séparer après l'étape de protection, comme suggéré par la présence d'un deuxième quadruplet visible en RMN correspondant au proton en α du chlore (Figure II. 14C). De plus, sur cette même figure, il est possible d'identifier le singulet à 3.5 ppm correspondant aux deux méthyles du DMT. Dans cet exemple, un signal superposé au même déplacement chimique annoté en noir indique la présence d'une espèce secondaire comportant également un DMT. Ces hypothèses ont pu être confirmées par MS/MS (Figure II. 14B) et soulignent deux points importants : la pureté des intermédiaires dans cette voie de synthèse est non négligeable, puis, l'alcool réduit est peu stable du fait de sa dimérisation favorable.

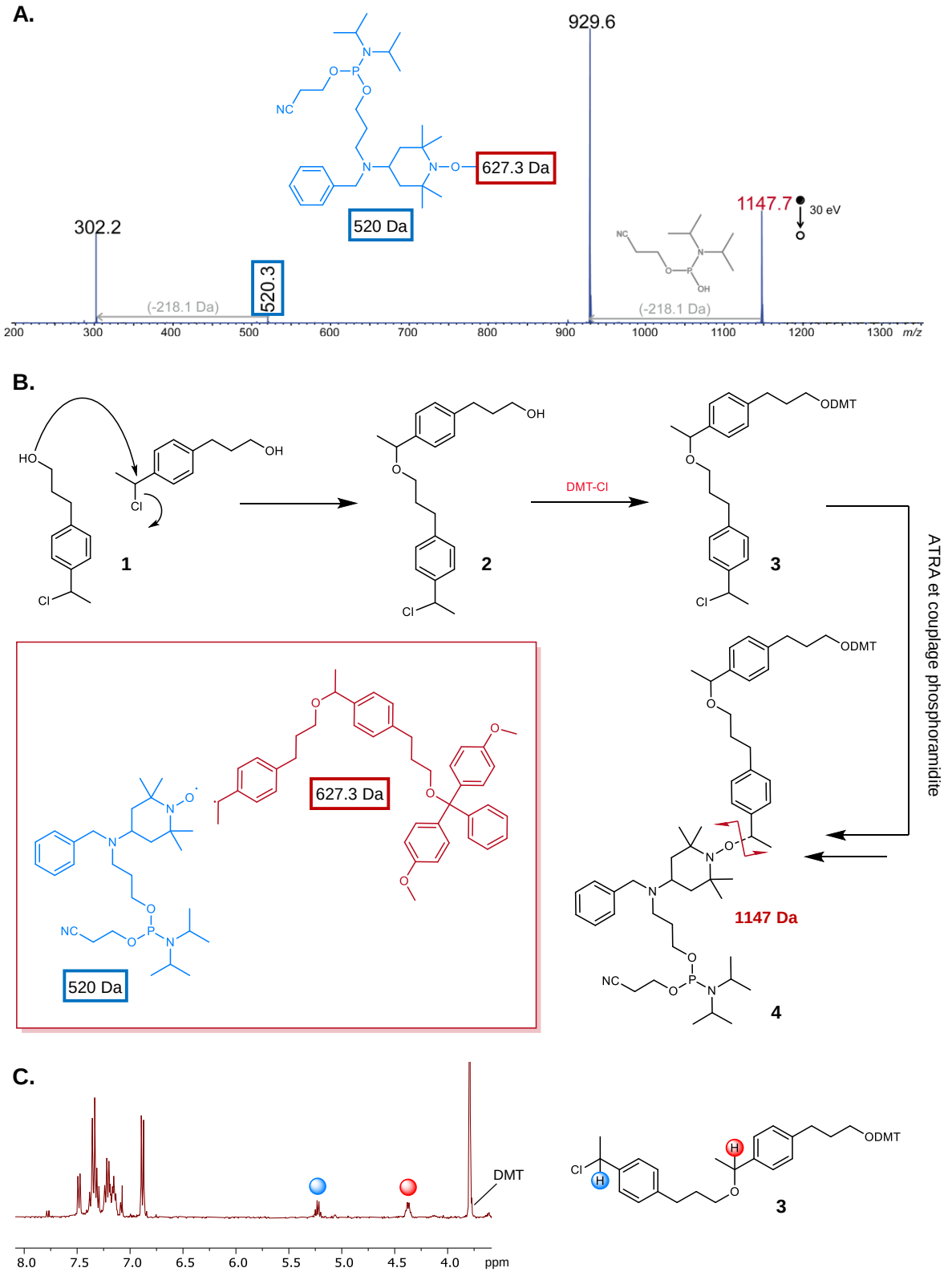


*Figure II. 14 : Interprétation de l'impureté majoritaire dans l'échantillon **TISP2**.*
***A.** Spectre ESI-MS/MS de l'ion [M+H]+ à m/z 1147.7 correspondant à l'impureté majoritaire dans l'échantillon **TISP2**. **B.** Interprétation de la nature de l'impureté provenant d'une dimérisation de l'alcool saturé **1** avant sa protection par le groupe DMT. **C.** Portion d'un spectre RMN correspondant au produit protégé issu de l'étape vi) Schéma II. 9.*

- Par ailleurs, les premières conditions de purifications utilisées lors de l'étape de protection du dérivé alcool saturé n'ont pas été appropriées. Effectivement, il a pu

être observé qu'en fonction du choix du solvant lors de la purification sur colonne chromatographique, différents comportements peuvent être remarqués. Notamment, en utilisant un gradient de cyclohexane/EtOAc, le produit désiré n'a pu être isolé qu'avec 10% de rendement. Après élution totale du produit restant sur la colonne en utilisant 100% d'EtOAc, un mélange correspondant en partie au composé désiré a pu être récupéré après vérification par analyse RMN. Il a été alors suggéré que le cyclohexane n'est pas un bon choix de solvant puisqu'il n'élue qu'en partie le produit d'intérêt. Certainement dû à la présence d'une chaîne alcane sur ce dernier, l'utilisation d'un gradient *n*-hexane/EtOAc a permis d'obtenir, de ce même mélange, 30% de rendement supplémentaire. Un choix de solvant approprié est donc particulièrement important pour cette purification.

Malgré ces observations laissant place à une optimisation de rendement non négligeable, le macromonomère ***TISP2*** a été obtenu en quantité suffisante pour tester son incorporation dans une chaîne polymère. Ces résultats sont discutés dans la partie ci-dessous.

4.3. *TISP2* est-il un bon choix ? Résultats et discussions

L'incorporation du macromonomère ***TISP2*** dans un polymère a pu être effectuée de manière satisfaisante, comme en témoigne le spectre ESI-MS ci-dessous (Figure II. 15). Le spectre de masse enregistré en mode négatif pour l'échantillon **P4** montre nettement le polymère recherché sous les états de charges 3– à 5– (en vert), comme validé par le spectre MS/MS de l'ion $[M - 4H]^{4-}$ à m/z 727.7 (Figure II. 15B). Au regard de la composition des blocs, leur séquençage pseudo-MS^3 n'a pas été réalisé. Une impureté mineure est également observée : notée X et de masse 3207.7 Da (en rouge, Figure II. 15), elle présente un excès de masse de 293.0 Da situé dans le 2nd bloc (Figure II. 15C). Cette impureté est due à l'utilisation d'un nouveau support solide et son mauvais clivage (support solide ne contenant pas de base nucléique, et laissant donc la terminaison α de la chaîne avec une fonction hydroxyle). Ce nouveau support sera discuté en détail dans le Chapitre 3 (3.2).

Malgré l'obtention d'un polymère monodisperse (sans tenir compte du problème de clivage du support), l'approche de protection/déprotection utilisée pour parvenir au macromonomère ***TISP2*** pure a finalement demandé beaucoup d'investissement, comme il a été vu dans la sous-partie 4.2 précédente avec les dimérisations occurrentes.

Afin de pallier ce problème et acquérir une meilleure stratégie de synthèse, d'autres conditions d'hydrogénation sur l'alcool propargylique initiale pourraient être testées, telles que l'utilisation de $NaBH_4$[374] ou triethylsilane[375] en présence de Pd/C, ou encore l'utilisation de nanoparticules de silice.[376] Cependant, même si beaucoup d'efforts ont été fournis pour optimiser cette hydrogénation, le principal objectif de cette thèse n'a pas été d'améliorer le rendement de cette synthèse, mais d'obtenir un marqueur d'octet clivable en quantité suffisante pour pouvoir l'incorporer dans une chaîne polymère.

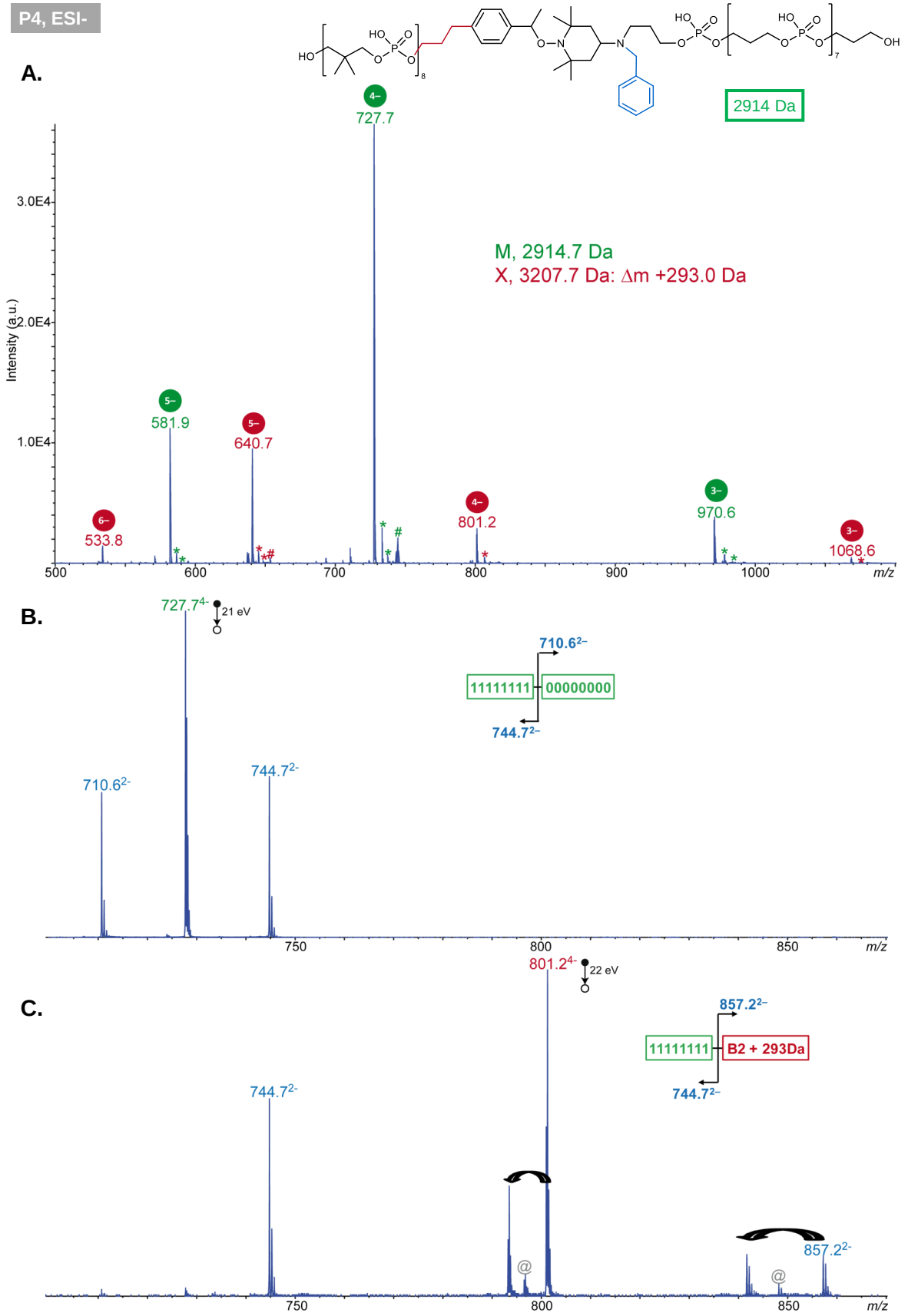


*Figure II. 15 : Analyses ESI-MS et MS/MS du polymère **P4** impliquant le **TISP2**.*
***A.** Spectre ESI-MS en mode négatif de l'échantillon **P4**. Les pics repérés par un astérisque sont issus d'échanges H/Na et/ou H/K et ceux annotés par # correspondent à un échange H/Cu. Les pics non annotés correspondent à des fragments formés en source. **B.** Spectre MS/MS des ions [M – 4H]4⁻ à*

*m/z 727.7 et **C.** [X – 4H]4⁻ à m/z 801.2 dans l'échantillon **P4**. Les flèches noires indiquent une perte de 31.0 Da à partir de tous les fragments contenant le bloc "Δm +293 Da". Des fragments secondaires sont annotés en gris (@).*

Dans ce contexte, l'intermédiaire du ***RISC2*** déjà étudié par notre groupe a permis de répondre à ce besoin, en l'utilisant dans une ATRA impliquant un des nouveaux amino-TEMPO **T2** modifiés.

De plus, en comparant les différentes stratégies de synthèses jusqu'alors étudiées (Figure II. 16), le rendement global permettant d'accéder à cet intermédiaire d'intérêt est quasi similaire en utilisant l'ancien protocole, mais nécessite moins d'étapes réactionnelles. Par ailleurs, le choix de réutiliser le squelette du ***RISC2*** est intéressant puisque ce dernier a déjà été optimisé pour un séquençage pseudo-MS³ performant.

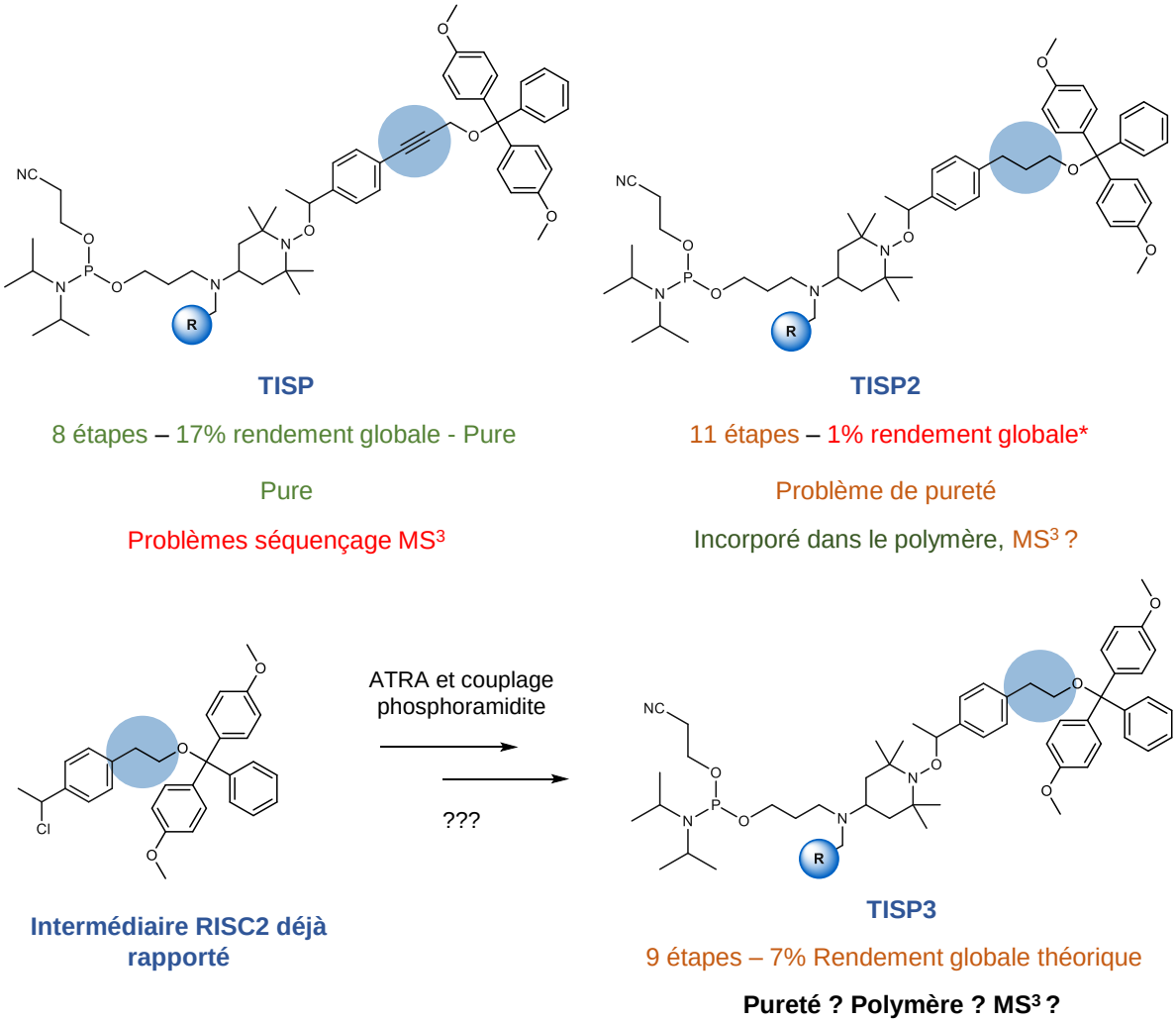


*Figure II. 16 : Récapitulatif des stratégies **TISP** et **TISP2** avec une perspective sur **TISP3**. *Rendement globale pouvant être amélioré comme discuté précédemment partie 4.2. **Rendement théorique en se basant sur la synthèse du **RISC2** déjà rapporté.*

5. Conclusion et perspectives

Dans ce Chapitre 2, différentes approches pour la préparation d'un marqueur clivable ont été détaillées. D'abord, les limites de synthèse pour la préparation de l'espaceur ***RISC2***,

décrit précédemment par notre groupe, ont pu être discutées. Par la suite, différentes idées de conception de marqueurs variables en gardant la même structure moléculaire du ***RISC2*** ont été détaillées. La modification d'un amino-TEMPO *N*-substitué par un dérivé benzyle, pouvant être incorporé grâce à une étape clef d'ATRA, a été la première source de recherche dans l'élaboration d'un espaceur clivable. Ce choix a particulièrement été motivé par la simple construction de cet amino-TEMPO **T1** en réalisant une amination réductrice, suivit d'une substitution nucléophile pour donner l'amine tertiaire désiré **T2**.

Par ailleurs, un rappel des règles à respecter dans le choix d'un marqueur pour que le séquençage pseudo-MS^3 par spectrométrie de masse reste optimale a été effectué. Ceci souligne l'importance de la sélection du groupe variable dans l'espaceur clivable, mais aussi les limites de cette stratégie puisque le nombre de marqueurs utilisés dans la même chaîne polymère est restreint. Cependant, un design approfondi du bloc variable, notamment en choisissant précisément la masse moléculaire à atteindre en fonction des marges de masses disponibles, permettrait d'optimiser ce nombre.

Des premiers essais pour la construction d'un espaceur clivable se sont orientés vers une stratégie employant un alcyne issu d'un couplage Sonogashira. Cette approche est très intéressante pour obtenir le marqueur de manière rapide et dans de bons rendements, contrairement à la stratégie ***RISC2*** auparavant étudiée. Cependant, la délocalisation électronique accrue du radical carboné, causée par la présence de l'alcyne et généré lors de l'analyse pseudo-MS^3, donne des défauts de masse de +/- 1 Da en conséquence d'une abstraction d'hydrogène radicalaire. Même si des efforts ont été investis pour pallier ce problème en recherchant des conditions d'hydrogénations de l'alcool propargylique attaché au groupe phényle, les rendements et la pureté des produits saturés isolés n'ont pas été pas satisfaisants. Une stratégie de protection/déprotection a permis d'éviter la formation de sous-produits majoritaires, mais l'ajout d'étapes diminue le rendement global de cette approche. Bien que des optimisations supplémentaires auraient pu être menées, notamment sur la purification des produits saturés en utilisant des solvant appropriés, ou encore en utilisant d'autres techniques de réductions avec du dihydrogène formé *in situ*, l'approche utilisant l'alcyne réduit semble compliquer l'obtention du marqueur clivable désiré.

Toutes ces études ont d'abord été menées pour trouver une alternative plus efficace, avec de meilleurs rendements, tout en gardant le squelette du ***RISC2*** pour la préparation de nouveaux espaceurs. Manifestement, l'approche utilisant l'alcyne permet de satisfaire ces critères, mais le design n'est malheureusement pas adapté au séquençage pseudo-MS^3. Bien que beaucoup d'efforts ont été investis dans la préparation d'un nouvel espaceur ***TISP*** ou ***TISP2***, il a été raisonnable de s'orienter vers un intermédiaire déjà connu dans notre groupe pour accélérer la construction d'une série de marqueurs. De plus, après comparaison avec la stratégie initiale déjà étudiée pour la préparation du ***RISC2*** et les approches développées ci-dessus, un rendement global quasi similaire peut être obtenu en impliquant moins d'étapes pour la préparation de l'intermédiaire ***RISC2*** d'intérêt (Figure II. 16). Par ailleurs, cela permet d'impliquer cette structure déjà optimisée pour un séquençage pseudo-MS^3, qui a déjà fait l'objet d'une thèse auparavant. C'est pourquoi,

le prochain chapitre discutera de l'utilisation de ce design pour la synthèse du ***TISP3***, et son incorporation en chaîne polymère contrôlée.

Chapitre III.

Stratégie *TISP3* – Synthèse et incorporation dans des chaines numériques

1. Introduction

Ce chapitre discutera d'un nouveau design de marqueur d'octet avec toujours un même objectif : incorporer ce nouveau macromonomère dans un polymère numérique et vérifier sa viabilité pour un séquençage pseudo-MS^3.

Le précédent chapitre a démontré que l'amélioration du rendement de synthèse dans la stratégie ***RISC2*** n'est pas aussi évidente qu'espéré. Pourtant, après avoir tenté d'autres approches, notamment celle du ***TISP*** et du ***TISP*2**, l'utilisation du design ***RISC2*** auparavant étudié semblé être le meilleur choix pour accéder rapidement et efficacement à une série de marqueurs.

C'est pourquoi, en prenant en compte des travaux déjà rapportés par notre groupe, ce chapitre discute de la synthèse d'un nouvel espaceur ***TISP*3**, ainsi que l'élaboration d'une chimiothèque étoffée utilisant ce design. Le macromonomère a également été incorporé dans une chaîne polymère contrôlée afin de vérifier son séquençage par pseudo-MS^3. De plus, de nouveaux supports solides ont été testés pour la synthèse phosphoramidite automatisée. Les optimisations de clivages permettant d'obtenir des polymères monodisperses sont décrites dans ce chapitre. L'efficacité de couplage des différents macromonomères synthétisés a aussi été quantifiée à l'aide de la spectroscopie UV. Ces résultats seront discutés et développés. Finalement, une recherche d'optimisation avec un nouveau design ***TISP*4** sera discutée.

2. Synthèse du *TISP*3

2.1. Schéma général

Afin de préparer le ***TISP*3**, une synthèse convergente a été établie, avec d'un côté la construction des TEMPO modifiés **T1-T11** servant à varier la masse sur l'espaceur (encadré gris, Schéma III. 1), et d'un autre côté, la synthèse de l'intermédiaire chloré **e** pouvant être impliqué dans une ATRA avec les nitroxides **T1-T11**.

Comme discuté dans le Chapitre 2, le réactif de départ préalablement utilisé pour la préparation du ***RISC2*** (l'acide 2-(4-acétylephényle) acétique **a**, Schéma III. 1) n'était plus disponible commercialement et a dû être synthétisé *i)* à partir du iodoacétophénone en obtenant un rendement peu probant de 39%, correspondant à la littérature.[363] Malgré le temps et la quantité de solvant nécessaire (purification) pour obtenir l'acide d'intérêt en bonne quantité, ce dernier a pu être caractérisé, confirmant sa pureté, et a pu être employé dans une étape de *ii)* réduction suivit d'une *iii)* chloration dans des rendements quasi quantitatifs. Ensuite, *iv)* la réduction de l'acide carboxylique pour donner l'alcool correspondant a été réalisée en présence de complexe de sulfurediméthyle de borane en donnant de bons rendements. L'alcool instable ainsi formé a pu, après purification, immédiatement être *v)* protégé avec du chlorure de DMT. L'intermédiaire chloré formé **l** est ainsi stable lorsqu'il est stocké au congélateur et a pu être engagé dans une *vi)* réaction d'ATRA avec un des dérivés TEMPO **T1-T11** préalablement synthétisés. La

dernière étape *vii)* correspondant à l'ajout de la fonction phosphoramidite sur le monomère permet d'obtenir le marqueur clivable d'intérêt avec un rendement modéré.

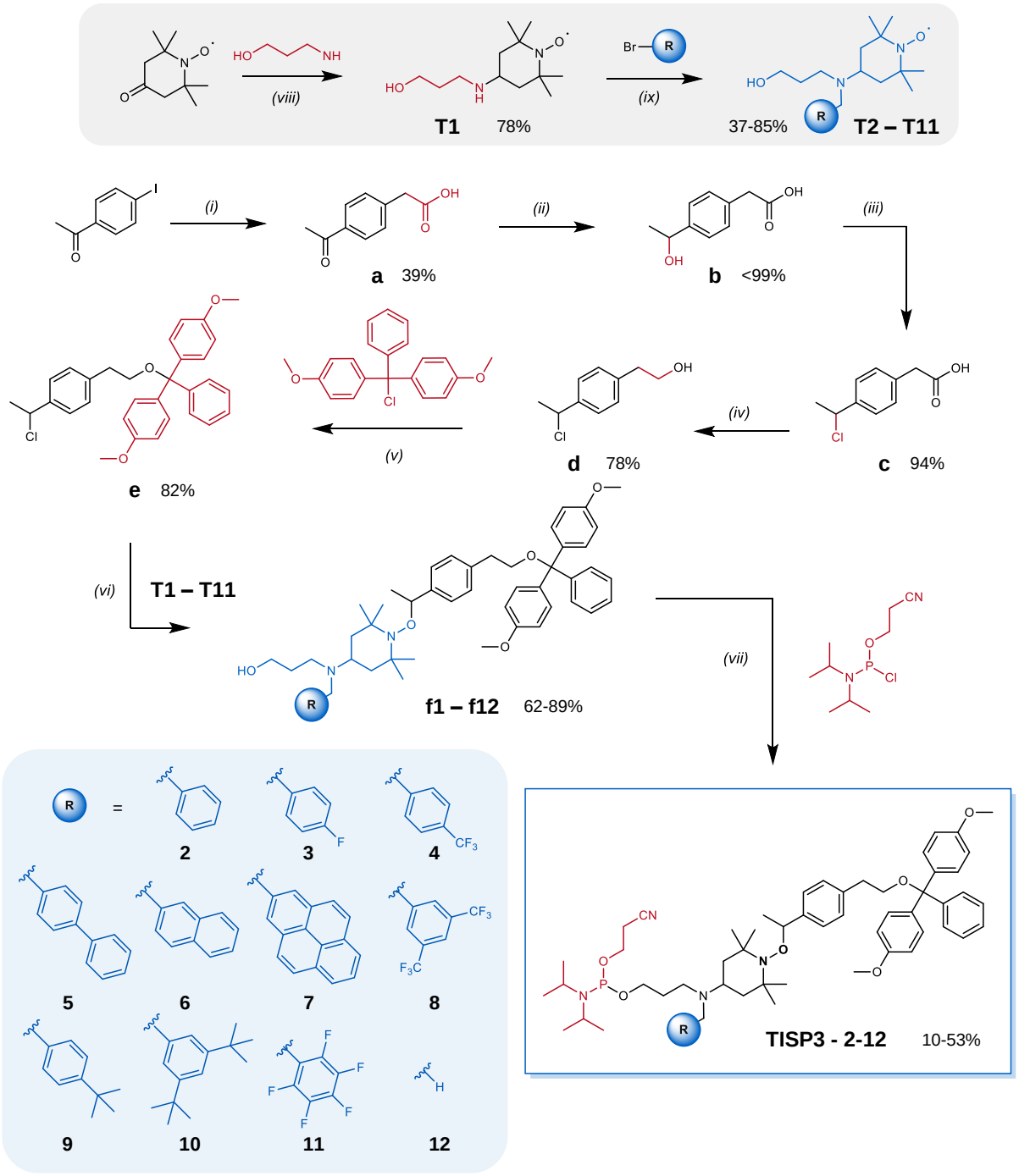


*Schéma III. 1 : Synthèse du marqueur clivable **TISP3**.*
i) $PdCl_2$, AgOAc, NaOAc, AcOH, 130°C, 24h. ii) $NaBH_4$, EtOH/THF, 0°C à t.a., 3h ; (iii) HCl 37%, THF, 0°C à t.a., 4h ; (iv) $H_3B.SMe_2$, THF anhydre, t.a., 17h ; v) Pyridine anhydre, THF anhydre, 0°C à t.a., 1h ; (vi) CuBr, Cu(0), PMDETA, THF anhydre, t.a ., 17h ; (vii) DIPEA, DCM anhydre, 0°C à t.a., 1h. viii) triacétoxyborohydrure de sodium, AcOH, DCE anhydre, t.a., 17h ; ix) K_2CO_3, ACN, 83°C, 17h.

Les différents nitroxides modifiés **T1-T11** ainsi que les différents macromonomères ***TISP3*** ont été obtenus de manière pure, dans la majorité des cas, comme confirmé par les différentes analyses RMN, ESI-MS et MS/MS de chaque composé (P.E., 3 et 4).

L'encadré bleu du Schéma III. 1 présente les différents dérivés benzyles utilisés pour la synthèse d'une série de ***TISP3***. Leur choix a été fait en fonction des critères décrits dans le Chapitre 2 (parties 2.2 et 2.4), mais aussi par rapport à leur diversité chimique permettant une étude approfondie de leur comportement par IMS.

Cependant, les réactions n'ont pas toujours donné de bons rendements, comme indiqué sur le Schéma III. 1 pour l'étape *ix)* correspondant à la synthèse des dérivés TEMPO modifiés **T1-T11**, et l'étape *viii)* faisant référence à l'ajout du groupe phosphoramidite sur le monomère. Ceci peut s'expliquer par la variation des groupes benzyles et de leurs encombrements stériques ou encore, leur différence d'électronégativité. La prochaine sous-partie tentera d'éclaircir ces hypothèses.

2.2. Quels sont les marqueurs problématiques ? Discussion des rendements obtenus

Alors que la plupart des techniques classiques de caractérisation à disposition (RMN, MS, MS/MS) permet d'identifier et de confirmer la pureté de chaque macromonomère ***TISP3***, les rendements sont, selon le dérivé benzyle utilisé, plus ou moins affectés (Tableau III. 1). Plusieurs hypothèses émanant des réactivités différentes observées ont été stipulées, en prenant compte des deux principales sources de variations dans la structure de ces dérivés : l'encombrement stérique et l'électronégativité.

D'abord, lorsque les groupes fonctionnels encombrés tels que le *tert*butyle et le di-*tert*butyle ont été utilisés dans la préparation des TEMPO **T9** et **T10** (**B9** et **B10**, Tableau III. 1), les rendements sont étonnamment bons et peuvent se justifier par le mécanisme de substitution nucléophile d'ordre 1 : le carbocation formé sur l'intermédiaire aromatique encombré est stabilisé par les groupements méthyles sigma-donneurs, résultant en une meilleure addition de l'amine secondaire et un meilleur rendement. Le scénario contraire peut être observé lorsque des groupements CF_3 électroattracteurs sont attachés au groupe phényle (**B4** et **B8**, Tableau III. 1). Ici, la position du groupe riche en électrons semble influer sur la réactivité puisque **B8** montre une meilleure réactivité que **B4**. Ceci est difficile à interpréter, mais peut potentiellement être lié à une délocalisation électronique différente entre un groupe CF_3 positionné en para ou en di-ortho du groupe phényle. Cependant, lorsque le groupe fonctionnel pentafluorobenzène est employé (**B12**, Tableau III. 1), il démontre une bonne réactivité pour la préparation de **T12**, laissant un doute sur l'implication des effets électroniques dans cette réaction. En outre, des effets de solubilité peuvent également jouer un rôle, comme avec les dérivés diphényles, naphtyles et pyrényles (**B5**, **B6** et **B7**, Tableau III. 1) ou leur interactions π exerceraient certainement une influence dans l'obtention de ces rendements modérés. Finalement, les dérivés benzyle et fluorobenzyle, présentent les meilleurs résultats (**B2** et **B3**, Tableau III. 1). À noter que **12** (dernière ligne du Tableau III. 1) n'est pas un dérivé bromé, mais utilise l'intermédiaire TEMPO **T1** décrit dans le Schéma III. 1. Dans ce cas, ce dernier a été utilisé pour la préparation de **f12** et ***TISP3*-12** sans coupler un dérivé benzyle. L'objectif a été de se servir de la masse du monomère ***TISP3*** sans modification de l'amine secondaire, pouvant également enrichir la chimiothèque de marqueurs clivables.

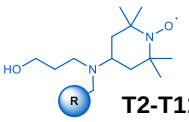
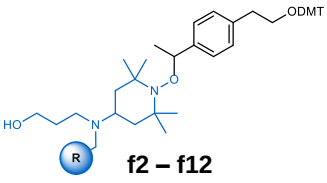
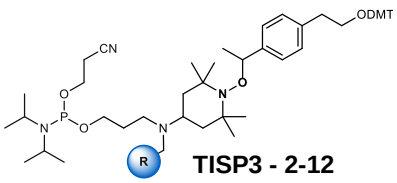
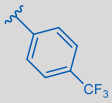
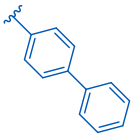
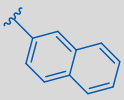
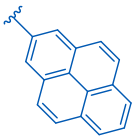
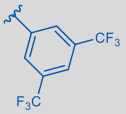
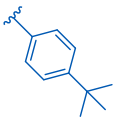
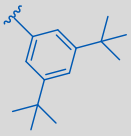
		 **T2-T11**	 **f2 – f12**	 **TISP3 - 2-12**
B2		85 %	82 %	45 %
B3		78 %	75 %	40 %
B4		37 %	89 %	46 %
B5		69 %	75 %	34 %
B6		53 %	69 %	53 %
B7		48 %	71 %	61 %
B8		49 %	84 %	27 %
B9		80 %	65 %	44 %
B10		82 %	41 %	24 %
B11		72 %	72 %	10 %
12		/	70 %	10 %

*Tableau III. 1 : Rendements obtenus pour la préparation des différents marqueurs **TISP3**. Rendements de **T2-T11**, **f2-f12** et **TISP3-2-12** correspondent aux étapes ix), vi) et vii) du Schéma III. 1 respectivement.*

Pour la préparation de **f2-f12** (Tableau III. 1) impliquant une réaction d'ATRA, les nitroxides électrodéficients présentent des bons rendements (**f3**, **f4**, **f8** et **f11**) alors que les dérivés électroniquement riches sont plus compliqués à obtenir (**f9** et **f10**). Cette

différence s'explique certainement par l'implication d'un nitroxide amoindri en électrons favorisant le couplage radicalaire avec l'intermédiaire chloré **e** (Schéma III. 1). À noter également que la purification des dérivés **f9** et **f10** est difficile du fait du rapport frontal quasi similaire entre le réactif de départ TEMPO (**T9** et **T10**) et le produit désiré. Bien qu'ils soient facilement distinguables en CCM, leur interaction avec la colonne de silice est voisine et une séparation lente n'est pas suffisante pour obtenir toutes les fractions du composé de manière pure. De ce fait, le rendement peut en partie être justifié par ces défauts de purifications laissant une portion impure, plus ou moins négligeable (environ 5% de rendement). Enfin, l'addition du groupe phosphoramidite correspondant à l'étape finale dans la préparation du macromonomère ***TISP3*** (étape *vii)*, Schéma III. 1) présente également différents rendements selon les marqueurs utilisés. Ces résultats sont dans l'ensemble similaires à ceux obtenus pour l'espaceur ***RISC2*** auparavant rapporté par notre groupe (rendement de 50%).[355] L'interprétation de ces derniers n'est pas aussi évidente que pour ceux obtenus avec **T2**-**T11** et **f2**-**f12**, mais quelques explications pouvait tout de mêmes être données. Premièrement, lorsqu'un intermédiaire encombré est impliqué dans ces couplages, le rendement n'est que modérément impacté avec le groupe mono-*tert*butyle (***TISP3***-**9,** Tableau III. 1), mais fortement en utilisant son homologue di-*tert*butyle (***TISP3***-**10**). La gêne stérique plus importante de ce dernier permettrait d'expliquer cette différence. Alternativement, et de manière surprenante, les dérivés naphtyles et pyrènyles donnent de très bons résultats (***TISP3***-**6** et ***TISP3***-**7**). Dans ce cas, l'encombrement stérique ne joue pas un rôle dans la réactivité, mais un effet de solubilité peut semblablement induire une proximité entre l'alcool et le réactif chlorophosphoramidite, favorisant le couplage. Par ailleurs, l'électronégativité des groupes fonctionnels attachés aux phényles semblent quant à eux, jouer un rôle dans l'efficacité de cette réaction (***TISP3***-**3**, ***TISP3***-**4**, ***TISP3***-**8**, et ***TISP3***-**11**). Le groupe **B3** ne démontre pas de réactivité particulière et est raisonnablement performant. Cependant, en comparant **B4** et **B8**, des rendements logiques ont été obtenus, notamment un plus faible en impliquant le dérivé plus électronégatif **B8**. Ce dernier induit une électronégativité plus faible sur l'alcool le rendant de ce fait moins nucléophile, comparativement à son analogue **B4**.

Finalement, ***TISP3***-**11** et ***TISP3***-**12** ont été les derniers synthétisés et présentent des rendements très faibles. Après études des conditions de purifications utilisées dans ces dernières synthèses, ces résultats ont pu récemment être interprétés. Il est tout d'abord notable que toutes les réactions discutées ci-dessus concernant l'ajout du groupe phosphoramidite ont été purifiées en utilisant un gradient *n*-héxane/EtOAc/DIPEA (89/10/1). Après répétition de ces synthèses et purification, la DIPEA a été identifiée comme étant difficile à séparer, même après plusieurs jours de séchage. L'idée d'utiliser la TEA comme substituant à la DIPEA pour pouvoir s'en affranchir plus facilement, tout en favorisant une bonne séparation sur la colonne a été testée. Un premier essai sur le dérivé benzyle (***TISP3***-**2**) a donné des résultats similaires à l'utilisation de la DIPEA, et semblait encourageant puisque la TEA est beaucoup moins problématique à évaporer. Cependant, des essais avec des groupes fonctionnels différents (***TISP3***-**8** et ***TISP3***-**10**) ont aggravé les rendements établis avec la DIPEA. Notamment, un dérivé phosphite a pu être isolé en quantité équivalente au produit désiré, lorsque la purification est faite en présence de TEA. Ce défaut, certainement accentué par l'utilisation de cette base et non

de la DIPEA n'a été remarqué que tardivement et sera notamment discutée plus en détail dans la partie 5.1**,** où les résultats obtenus avec un autre macromonomère a présenté le même type de problématiques. Toutefois, ces suspicions n'ont pas empêché d'obtenir chaque macromonomère présentés ci-dessus en quantité suffisante pour être incorporé en chaîne polymère contrôlée. La prochaine sous-partie discutera de ces applications.

3. Incorporation du *TISP*3 dans une chaîne polymère contrôlée

3.1. Essais préliminaires sur support solide d'ADN

Pour les premiers essais d'incorporation du marqueur ***TISP3*** dans une chaîne polymère en utilisant la chimie de phosphoramidite automatisée, une séquence test a été synthétisée telle que :

- 0101 – ***TISP3-2*** – 0101 – T (Figure III. 1)

Dans cet exemple, les unités « **0** » et « **1** » font références aux monomères décrits précédemment Chapitre 2 (partie 3.1). Le spectre de masse enregistré en mode négatif pour l'échantillon **P5** a nettement montré le polymère recherché sous les états de charges 2– à 6– (en vert, Figure III. 1A), comme validée par le spectre MS/MS de l'ion $[M - 4H]^{4-}$ (Figure III. 1B) et le séquençage de chaque bloc (Figure III. 1C et D). L'impureté mineure de masse 1822.5 Da (en rose, Figure III. 1A) a présenté un défaut de masse de 166 Da par rapport au polymère ciblé, suggérant l'absence d'une unité codante « 1 ». Les données MS/MS indiquent que ce défaut de masse est dans le bloc 2 (Figure III. 2A) et le séquençage pseudo-MS^3 de ce bloc indique la séquence 010 (Figure III. 2B).

Par ailleurs, la thymidine en bout de chaîne étant fortement sensible à l'UV, une analyse HPLC a permis de confirmer la distribution des pics du spectre de masse. En outre, deux signaux, dont un majoritaire ont pu être détecté, indiquant la présence du polymère ciblé en majorité, quasi monodisperse (vert, Figure III. 3).

Il est important de remarquer dans l'échantillon **P5** que des traces de TCA ont pu être détectées. Dans cet exemple, la quantité de ce sel n'a pas empêché l'attribution des pics sur le spectre de masse. Cependant, si cette quantité venait à être majoritaire, le sel rendrait difficile l'ionisation des polymères compromettant ainsi leur analyse.

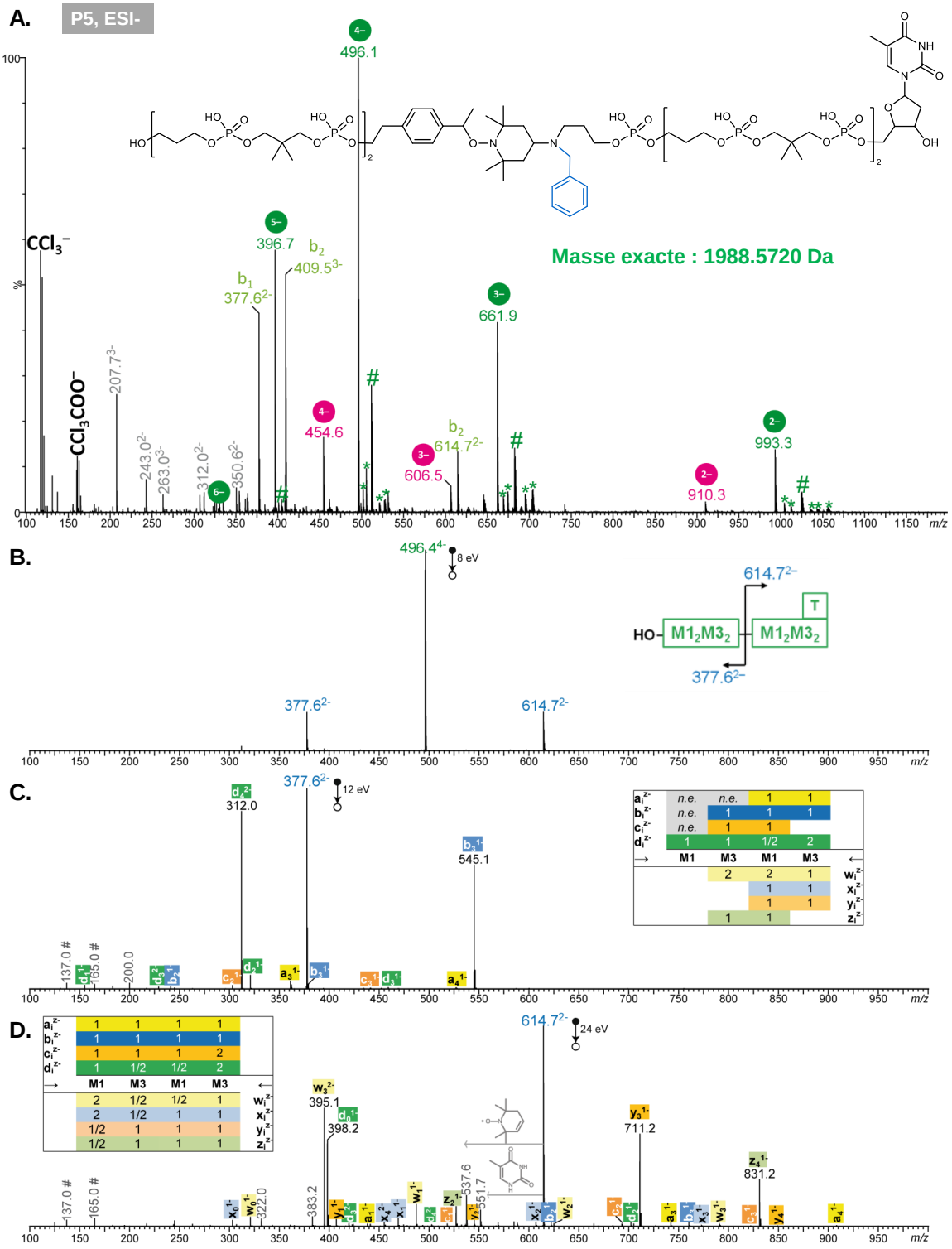


*Figure III. 1 : Analyse ESI-MS et MS/MS de **P5**.*

***A.** Spectre ESI-MS en mode négatif de l'échantillon **P5**. Les pics repérés par un astérisque sont issus d'échanges H/Na et/ou H/K et ceux annotés par # correspondent à un échange H/Cu. Les fragments libérés en source sont notés bj, avec i le nombre d'unités qu'ils contiennent. Seules des traces d'acide trichloroacétique (annotées en noir) ont été détectées dans cet échantillon**. **B.** Spectre MS/MS de l'ion m/z 496.4. Spectres pseudo-MS3 de **C.** l'ion m/z 377.62⁻ pour le séquençage du bloc 1 et **D.** l'ion m/z 614.72⁻ pour le séquençage du bloc 2, avec les tableaux de couverture de séquence en inset. Les ions en gris sont des fragments internes ou secondaires.*
M1 et M3 font références ici aux unités 0 et 1.

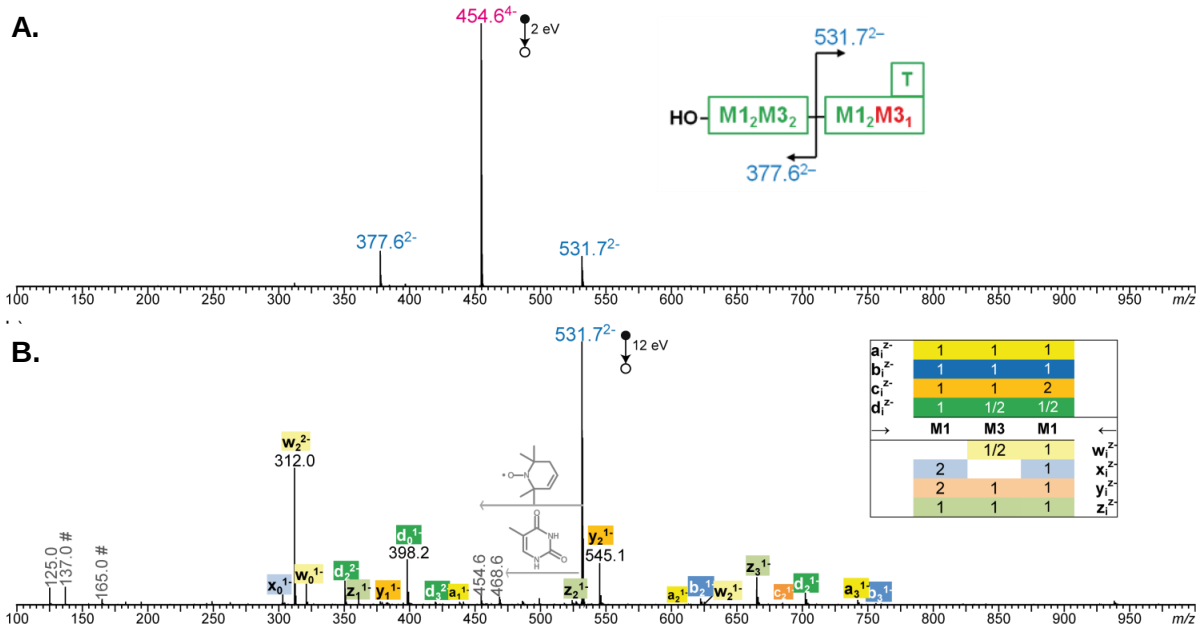


*Figure III. 2 : Analyse MS/MS et pseudo-MS³ de l'ion m/z = 1825.2 Da de **P5**.*
*Impureté rose Figure III. 1. **A.** Spectre MS/MS de l'ion m/z 496.4 dans l'échantillon **P5**. **B.** Spectre pseudo-MS³ de l'ion m/z 531.72⁻ pour le séquençage du bloc 2, avec le tableau de couverture de séquence en inset. Les ions en gris sont des fragments internes ou secondaires.* M1 et M3 font références ici aux unités **0** et **1**.

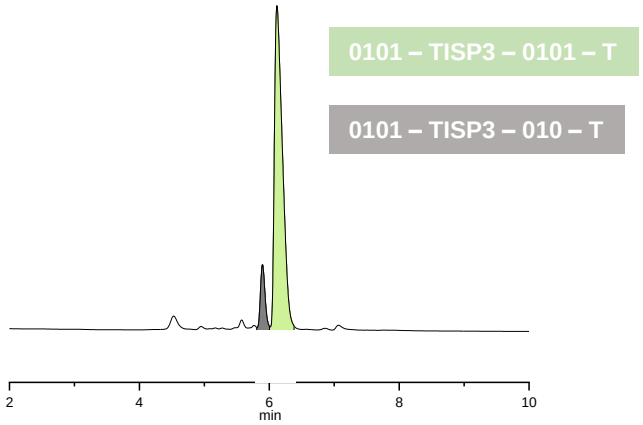


*Figure III. 3 : Analyse HPLC de **P5**.*
*Vert = polymère désiré ; gris = polymère avec une unité « **1** » manquante.*

Cette problématique a pu être observée lors de diverses synthèses au cours de cette thèse, et un exemple de spectre obtenu est donné sur la Figure III. 4, où la construction d'un pentabloc a été testée. Dans ce cas, le spectre de masse enregistré en mode négatif pour l'échantillon **P6** ne montre aucun des signaux attendus pour l'oligomère recherché : la forte concentration en acide trichloroacétique dont les agglomérats intenses (#) ont été détecté, empêche l'ionisation de toute autre espèce présente dans cet échantillon. L'explication la plus plausible reste le mauvais lavage à l'eau des polymères lors de leur purification. Même si ce paramètre a parfois pu être optimisé en observant moins de TCA sur les spectres grâce à des lavages supplémentaires, il semble que ces détections de sels peuvent également provenir de la préparation de l'échantillon et/ou de l'appareil utilisé pour l'analyse. En outre, un exemple permettant de confirmer cette hypothèse a été

réalisé en analysant un même composé, sur deux appareils différents : chez nos collaborateurs à Marseille, et dans notre nouvel institut à l'ISIS au service de spectrométrie de masse. Le polymère a facilement pu être identifié ici à Strasbourg, tandis qu'à Marseille, la trop forte concentration de TCA a empêché l'analyse du composé. Ceci peut provenir, soit de la dilution de l'échantillon : les spectromètres de masse étant différents, la géométrie des sources peut être plus ou moins propice à l'agglomération des sels ; soit de la technologie MS utilisée : une Orbitrap FT-ICR a été utilisée ici à l'ISIS, mesurant les ions après leur pré-accumulation (donc plus sensibles), alors qu'à Marseille, la technologie TOF a été employée.

Ces différentes remarques montrent que différents paramètres peuvent influer la détection qualitative du polymère en spectrométrie de masse. Ces variables, en fonction du composé à analyser, peuvent être efficacement adaptées, optimisant ainsi son interprétation.

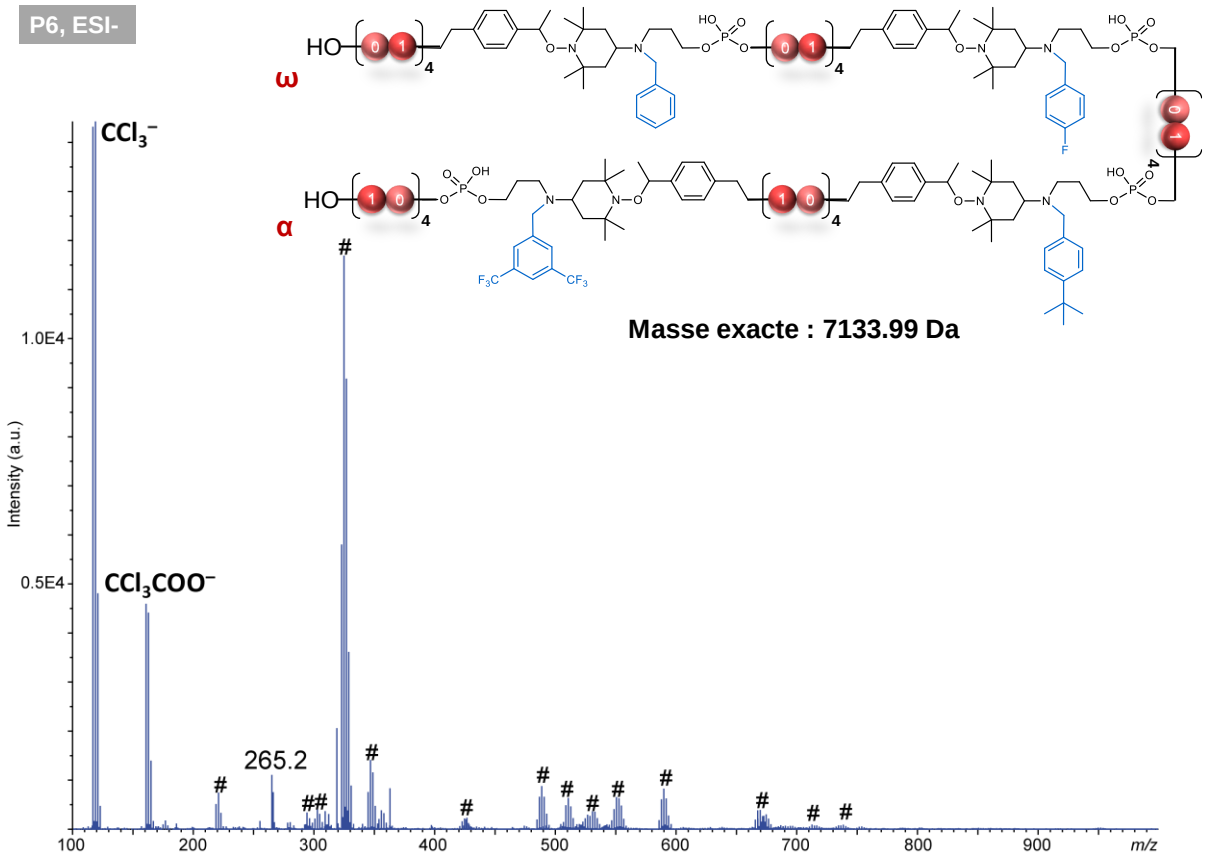


*Figure III. 4 : Spectre ESI(-)-MS de **P6** montrant des agglomérats de TCA (#).*

Par ailleurs, comparativement aux précédents exemples de polymères vus jusqu'ici dans le Chapitre 2 et ce début de Chapitre 3, L'échantillon **P6** (Figure III. 4) comporte une différence au niveau du bout de chaîne α, puisqu'il se termine par une fonction alcool et non par une base nucléique de type thymidine comme vu précédemment (Figure III. 1). Ceci est dû à l'utilisation d'un nouveau support solide universel permettant d'obtenir, après clivage et purification, le polymère désiré contenant uniquement la fonction alcool du monomère en bout de chaîne α.

Les causes du choix d'un nouveau support, ainsi que les optimisations quant à l'utilisation de ce dernier seront discutées dans la sous-partie suivante.

3.2. Utilisation d'un support solide universel

Dans la synthèse phosphoramidite automatisée, un grand nombre de variables existe pour le choix d'un support solide. Outre la taille des pores (de 500 à 2000 Å), le modèle de colonne selon le robot, le loading (de 0.2 à 10 µmol) ou encore le type de matrice utilisée (PS ou CPG), la sélection de la structure chimique attachée à cette dernière est nécessaire selon le dérivé polymère désiré. Dans l'exemple de l'ADN, les supports majoritairement utilisés sont des CPG contenant une des 4 bases nucléiques ATGC. Ainsi, la base attachée au support est prise en compte lors de la synthèse d'une séquence. Par exemple, dans le cas d'un enchaînement de type ATGCATGCT, la dernière base T correspondra à celle déjà présente sur le support. De cette façon, la séquence à définir sur le robot sera ATGCATGC, et lors du clivage du support contenant, ici, la thymidine, la séquence désirée ATGCATGCT sera obtenue (Figure III. 5).

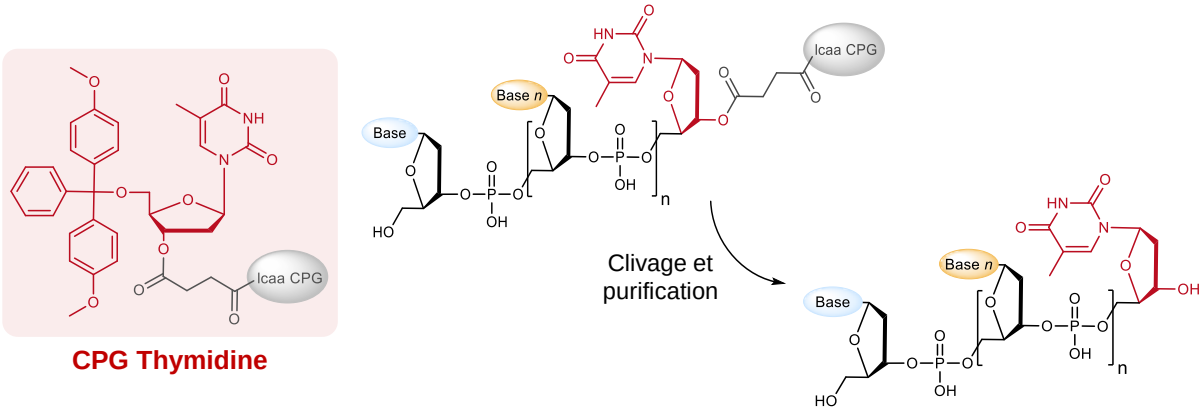


Figure III. 5 : Exemple de polymère d'ADN attaché sur un support solide CPG contenant une base nucléique Thymidine, avant et après clivage/déprotection.

Dans les premières études de poly(phosphodiesters) synthétiques menées par notre groupe, l'ajout de bases nucléiques en bout de chaîne a été utile pour une analyse HPLC, pouvant confirmer la monodispersité d'un échantillon.[25] Plus tard, l'utilisation de ces supports fonctionnalisés d'une nucléobase a servi de marqueurs de masse pour le dernier octet à l'extrémité α (bout de chaîne attaché au support).[27,29] Aujourd'hui, pour le design étudié dans cette thèse, où le marqueur de masse est incorporé dans l'espaceur clivable, l'utilisation de supports solides dérivés de bases nucléiques d'ADN n'est plus nécessaire puisqu'aucun marqueur indépendant n'est requis en bout de chaîne. C'est pourquoi, les études menées dans ce projet de thèse ont également eu pour objectif d'appliquer un nouveau support solide « sans résidu » à ce nouveau design de macromonomère ***TISP3***. « Sans résidu » signifie qu'après clivage du support, la dernière fonction alcool doit être retrouvée sans groupes fonctionnels complémentaires, dans ce cas inutile.

Le futur support solide qui répond à ces attentes doit être choisi selon plusieurs critères :

- Le support doit être facilement clivable, dans des conditions orthogonales aux groupes fonctionnels utilisés en chimie de phosphoramidite.

- Le support doit être utilisable sur l'ABI3900 et/ou l'Expedite 8909 et doit être commercial.
- Idéalement, le support doit, après clivage, permettre d'obtenir le polymère avec sa fonction OH terminale.
- Idéalement, le support doit être disponible en plusieurs loading et plusieurs tailles de pores.

En prenant compte de ces différents paramètres, une première sélection de support solides a pu être effectuée (Figure III. 6). À noter que les résines présentées ci-dessous proviennent de l'entreprise Glenresearch™, principalement impliquée dans le développement des réactifs et outils nécessaires à la chimie de phosphoramidite automatisée. Ces supports universels ont spécifiquement été développé pour obtenir le polymère, après clivage, avec sa fonction OH terminale (Figure III. 6).

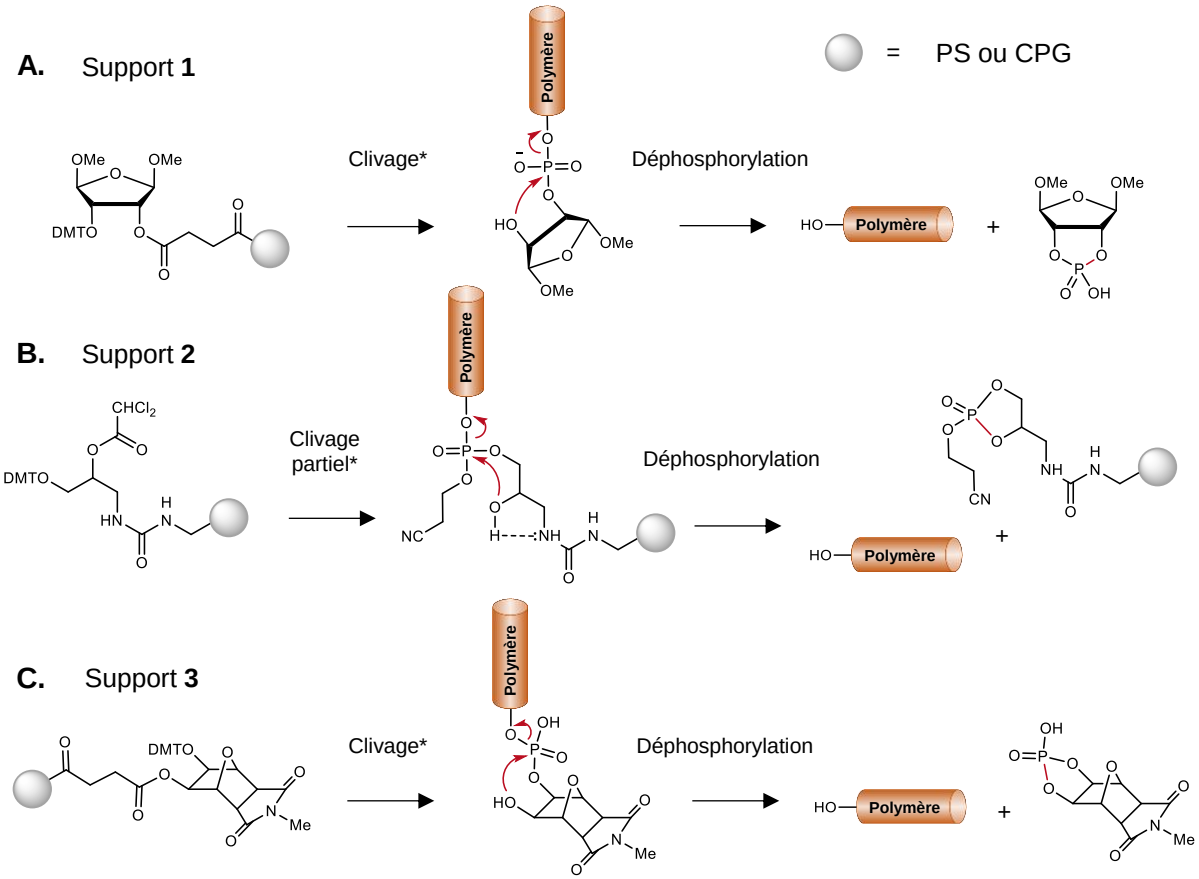


Figure III. 6 : Différents types de supports universels.
**Différentes conditions de clivage existe, la plus commune utilise AMA à t.a. ou à température (55°C à 80°C) pendant plusieurs heures selon les supports.*

Trois types de supports ont été élaborés depuis le début des années 2000 pour améliorer l'efficacité du clivage du polymère. L'originel (**1**, Figure III. 6A)[377] implique après clivage et déprotection, une réaction de déphosphorylation offrant l'oligonucléotide avec un alcool terminale. Cependant, ce dernier processus est une étape lente et demande parfois des conditions agressives pour obtenir le polymère détaché du support (températures >55°C pendant 3 à 17h selon les conditions utilisées). Un autre support a par conséquent été élaboré où la déphosphorylation prend place avant le clivage du support (**2**, Figure III.

6B).[377] Dans ce cas, des solutions douces impliquant une solution de NH_3/MeOH, AMA, NH_3/$MeNH_2$ ou bien encore plus douces avec du K_2CO_3 dans du méthanol réalisent la déphosphorylation suivie du clivage désiré. Pourtant, la difficulté d'adapter ce protocole à l'utilisation de gaz $MeNH_2$ est problématique pour les industriels exploitant cette méthode de clivage à grande échelle. En outre, l'usage de ce gaz déprotège dans un premier temps les groupes protecteurs cyanoéthyles, et ce, avant que la déphosphorylation ne se produise, amoindrissant ainsi l'efficacité de cette réaction et les rendements obtenus

Un dernier support a dans ce contexte été créé, et est encore aujourd'hui largement utilisé : l'Unysupport (**3**, Figure III. 6C).[378,379] En utilisant une molécule bicyclique rigide sur le support, il était espéré que la conformation préorganisé de la molécule faciliterait la formation de l'état de transition du phosphate cyclique, accélérant ainsi la déphosphorylation. Cela e effectivement été confirmé par plusieurs études.[378,379]

Notre choix concernant un nouveau support solide dans la synthèse de polymères synthétiques numériques réalisés dans cette thèse s'est porté vers l'Unysupport **3**. Cette décision n'a pas été prise en conséquence des optimisations effectuées sur l'adaptabilité du $MeNH_2$ gazeux sur ce support, puisque ce procédé n'est pas utilisé dans notre laboratoire, mais à cause des critères suivants :

- La disponibilité du support sous variante formes (CPG et PS) et donc sous différentes tailles de pores.
- Des conditions de clivages optimisées par rapport aux supports antérieurs.
- La nouveauté du support qui, certainement, fera disparaître ses anciens homologues.

Les conditions de clivage pour un tel support, préconisées par Glenresearch™, sont détaillées dans le Tableau III. 2 ci-dessous :[380]

Réactifs	Conditions
NH_3 (28% aq.)	80°C/2h
$MeNH_2$ (40% aq.)	55°C/8h
NH_3/$MeNH_2$ (AMA) 1v1	80°C/30min 65°C/1h 55°C/2h
$MeNH_2$ (gaz)	65°C/30min at 30 psi

Tableau III. 2 : Conditions de clivage pour l'unysupport PS.

Un premier essai a été réalisé avec un tribloc selon la configuration suivante :

- 111111111 – ***TISP3-2*** – 00000000 – ***TISP3-3*** – 11111111 – OH (Figure III. 7)

Pour ce support solide, les mêmes conditions de synthèse classique sur le robot synthétiseur d'ADN ABI3900 ont été utilisées (P.E., 1.2.7). Après synthèse, la colonne contenant le polymère attaché au support a été mise en contact avec une solution AMA (1mL) pendant 10 min pour opérer le clivage (étape rapide), avant d'être scellée dans un

flacon (3mL) et mise dans une étuve à 60°C pendant 2h (étape lente de déphosphorylation). La solution récupérée contenant le polymère peut finalement être purifiée, comme expliqué dans le Chapitre 2 (partie 3.2). Après lyophilisation du produit et préparation de l'échantillon pour analyse (P.E. 1.2.3), le spectre de masse enregistré en mode négatif pour l'échantillon **P7** a nettement montré le polymère recherché sous les états de charges 5– à 8– (M en vert, Figure III. 7), comme validé par le spectre MS/MS de l'ion $[M - 6H]^{6-}$ à m/z 795.4 (Figure PE.4. 37).

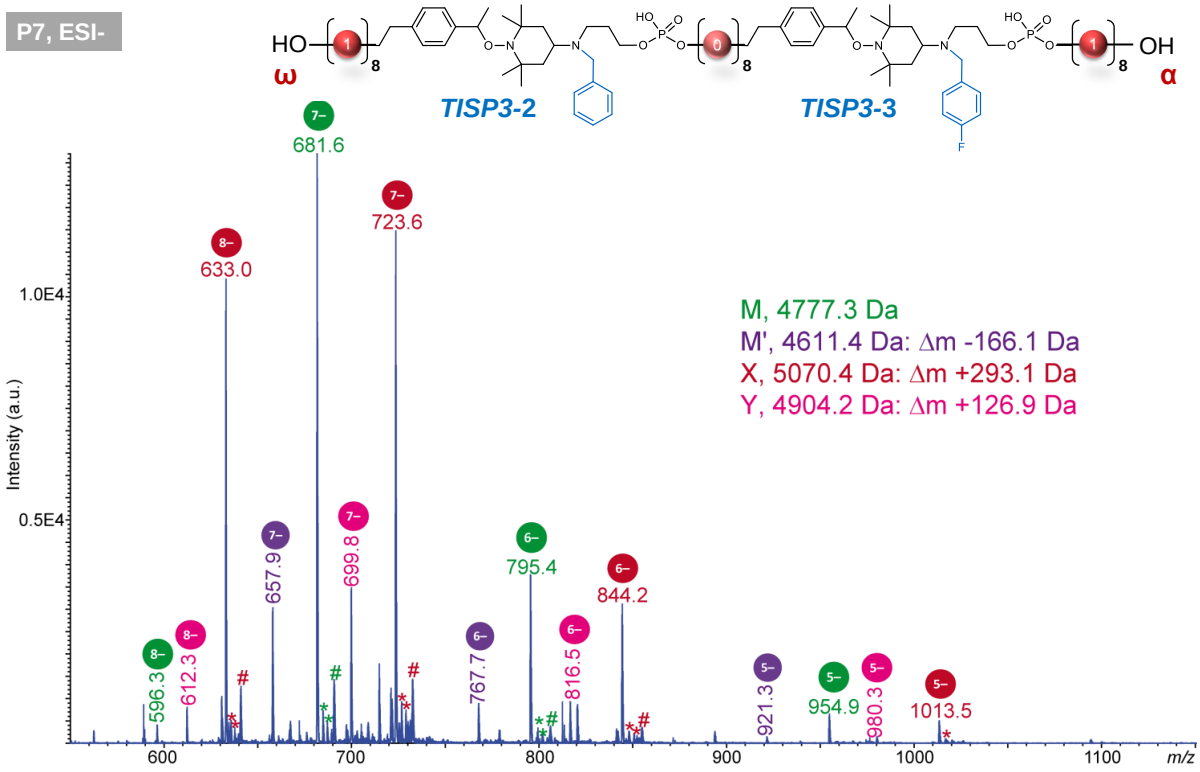


*Figure III. 7 : Spectre ESI-MS en mode négatif de **P7**.*
Les pics repérés par un astérisque sont issus d'échanges H/Na et/ou H/K et ceux annotés par # correspondent à un échange H/Cu. Les pics non annotés correspondent à des fragments formés en source.

Au regard de la composition des blocs, leur séquençage pseudo-MS^3 n'a pas été réalisé. Parmi les échanges cationiques observés pour cet oligomère, la prééminence d'un échange H/Cu (annoté par #) révélant la présence de cuivre dans cet échantillon est notable. Par ailleurs, trois impuretés sont également observées :

- L'impureté mineure M' de masse 4611.4 Da (en violet, Figure III. 7) présente un défaut de masse de 166 Da par rapport au polymère ciblé, suggérant l'absence d'une unité **1**. Les données MS/MS indiquent que ce défaut de masse se trouve dans le 3ème bloc (bout de chaîne α**,** (Figure PE.4. 37).
- L'impureté Y de masse 4904.2 Da (en rose, Figure III. 7) présente un excès de masse de 126.9 Da mais n'a pas pu être interprétée. Cependant, les données MS/MS indiquent que ce défaut de masse vient également du 3ème bloc (bout de chaîne α, Figure PE.4. 37)

- Finalement, l'impureté majoritaire X de masse 5070.4 Da (en rouge, Figure III. 7) présente un excès de masse de 293.1 Da situé dans le 3ème bloc au niveau du bout de chaîne α (confirmé par MS/MS**,** Figure PE.4. 37).

Cette dernière impureté X provient probablement d'un mauvais clivage du support, résultant en une masse supérieure (Schéma III. 2).

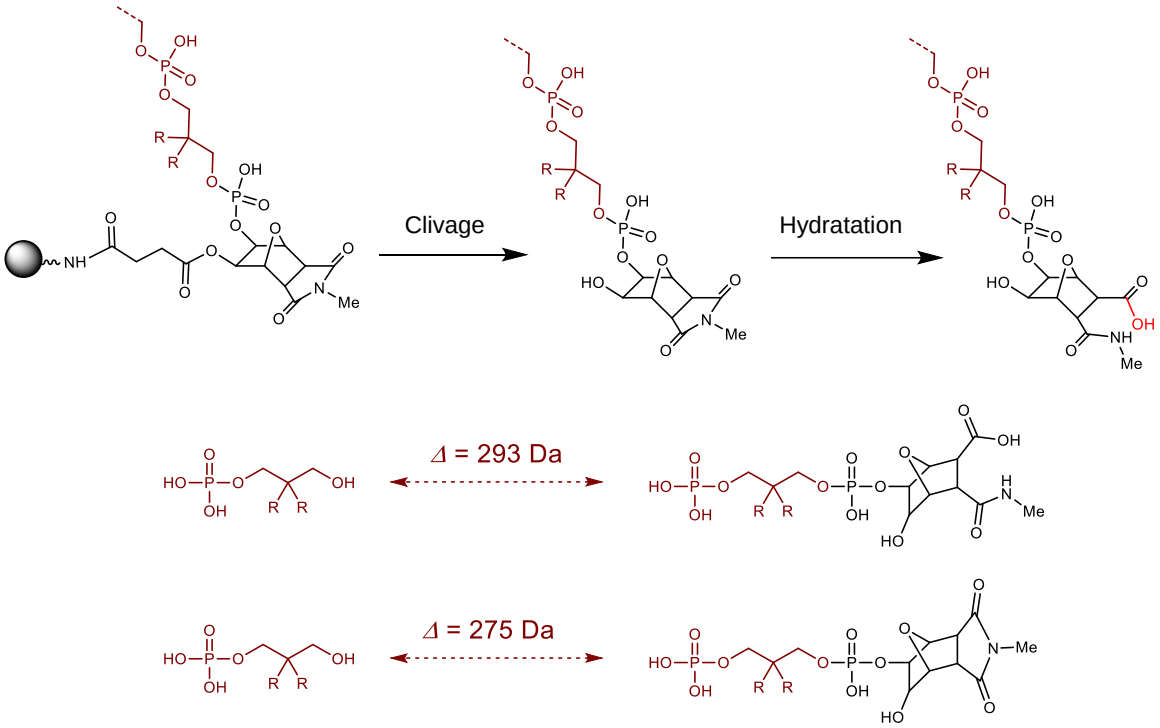
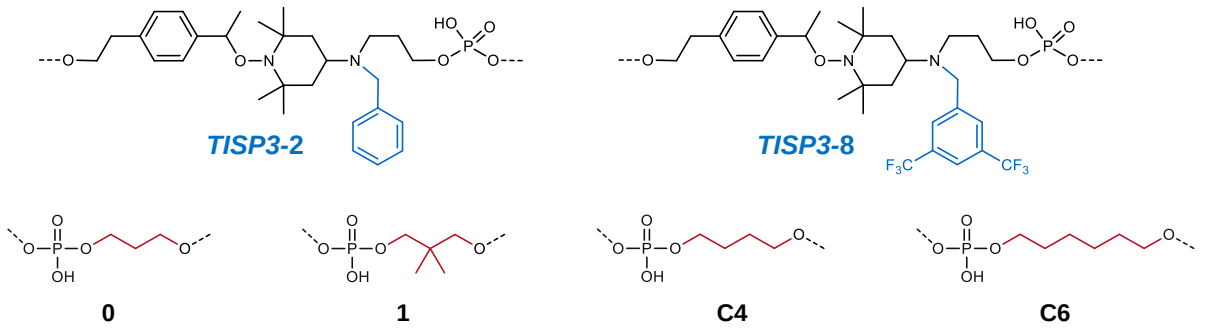


Schéma III. 2 : Impuretés identifiées en MS correspondant à une mauvaise déphosphorylation.

Effectivement, la déphosphorylation étant un processus lent, une température ou un temps de clivage mal choisi peut influer sur cette élimination. Par ailleurs, un monomère trop encombré attaché au support peut aussi compliquer le mécanisme de déphosphorylation (par exemple, si R = CH_3, Schéma III. 2) impliquant une élimination incomplète du phosphate.

D'après ces observations, plusieurs conditions d'optimisation ont été testées (Tableau III. 3). À noter que les spectres de chacune des expériences de ce tableau peuvent être retrouvés en partie expérimentale (4.6.2).

En comparant les premiers échantillons **P8** et **P9**, il semble que la présence d'un groupe plus encombré, dans ce cas diméthyle (**P8**, monomère **1** attaché au support dans le bout de chaîne α) favorise la présence de l'impureté à +293 Da comparativement à son analogue **P9** avec un monomère **0**. Cependant, en vérifiant cette hypothèse avec les monomères **C4** et **C6** (**P10** et **P11**, Tableau III. 3) qui comportent des chaînes alkyles étendues et favoriserait donc la déphosphorylation dans le cas où le problème venait de la gêne stérique, des résultats contradictoires ont été observés. Dans les deux cas, l'impureté à +293 Da est remarquable en quantité non négligeable. Ainsi, il a été conclu que le défaut de masse ne provient pas de l'encombrement du monomère attaché au support, mais certainement des mauvaises conditions de clivages.



Nom	Séquence*	Résultat**
P8	ω - $(1)_8$ – ***TISP3-2*** – $(0)_8$ – ***TISP3-8*** – $(1)_8$ - α	+293 Da
P9	ω - $(0)_8$ – ***TISP3-2*** – $(1)_8$ – ***TISP3-8*** – $(0)_8$ - α	Polymère désiré
P10	ω - $(0)_8$ – ***TISP3-2*** – $(1)_8$ – ***TISP3-8*** – $(0)_7$ – **C4** - α	Polymère désiré***
P11	ω - $(0)_8$ – ***TISP3-2*** – $(1)_8$ – ***TISP3-8*** – $(0)_7$ – **C6** - α	+293 Da

Tableau III. 3 : Tests effectués vérifiant l'implication d'un monomère trop encombré dans la mauvaise déphosphorylation du support.
L'impureté +293 Da correspondant à la mauvaise déphosphorylation se trouve dans le 3e bloc (à droite, bout de chaîne α). **Résultats quantitatifs de l'espèce majoritaire observés sur les spectres MS.* Polymère désiré/+293 Da (≈ 1/1)*

Afin de rechercher de meilleures conditions, des premiers tests ont été effectués avec la séquence ci-dessous, à 70°C, avec 2h (Figure III. 8A) et 3h30 (Figure III. 8B) de temps de clivage (comparativement à 60°C, 2h pour **P8**-**P11**).

- 10101010 – ***TISP3-2*** – 10101010 – ***TISP3-3*** – 10101010 – OH (Figure III. 8)

La comparaison des spectres de masse enregistrés pour ces deux échantillons (Figure III. 8) montre les mêmes espèces principales mais avec des abondances relatives différentes : l'espèce majoritaire est le polymère ciblé (M, en vert) dans **P12** tandis qu'un polymère où il manque une unité **0** dans le dernier bloc (M', en orange) est dominant dans **P13**. L'impureté à + 293 Da est également observée pour ces deux polymères, M+293 = X (en rouge) et M'+293 = X' (en rose), mais apparait moins abondante dans **P12**. Ces attributions structurales ont été validées par MS/MS (Figure PE.4. 43). Enfin, on peut noter deux impuretés supplémentaires dans **P13** : le composé Z (en bleu) a une masse qui suggère un manque d'une unité **0** et d'une unité **1** tandis qu'aucune hypothèse évidente n'a pu être proposée pour le composé Y (en violet). Ces deux impuretés sont trop peu abondantes pour être caractérisées par MS/MS. Ces observations démontrent qu'une température plus élevée semble jouer un rôle dans la déphosphorylation du composé bicyclique, favorisant cette dernière.

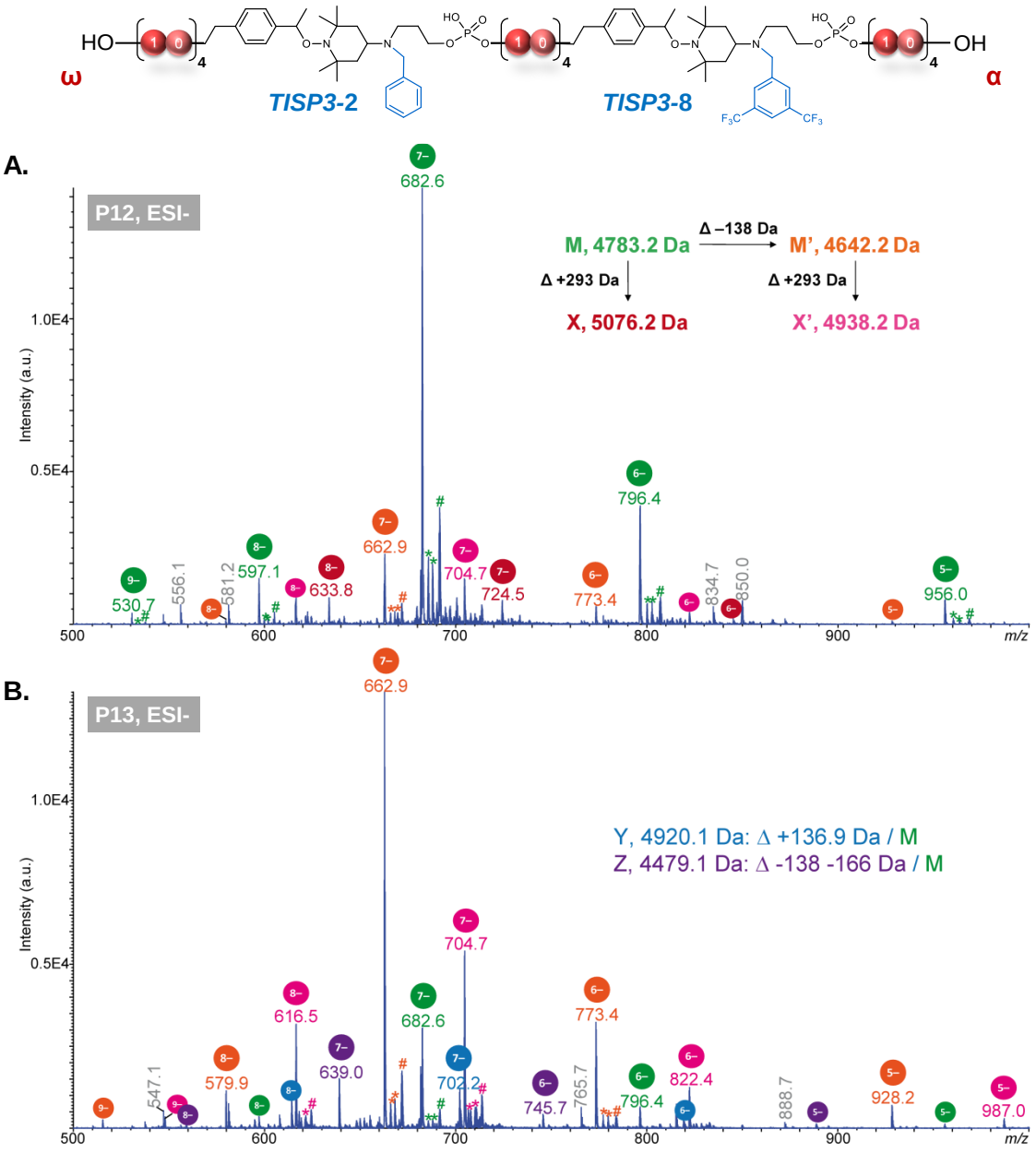


*Figure III. 8 : Spectres ESI-MS en mode négatif **A. P12** et **B. P13**.*
Les pics repérés par un astérisque sont issus d'échanges H/Na et/ou H/K et ceux annotés par # correspondent à un échange H/Cu. Les pics annotés en gris correspondent à des fragments formés en source.

Même si ces résultats étaient encourageants, l'objectif était de s'affranchir de cette impureté à +293 Da, tout en évitant une exposition trop longue à de hautes températures afin d'éviter de dégrader le polymère. Pour ce faire, une série de tests a été effectuée à différentes températures avec la séquence modèle ci-dessous :

- 00000000 – ***TISP3-2*** – 01010101 – ***TISP3-3*** – 11111111 – OH (Figure III. 9)
  - **P14** : 80°C pendant 0H30
  - **P15** : 65°C pendant 1H00
  - **P16** : 65°C pendant 1H30
  - **P17** : 65°C pendant 2H00

Les spectres de masse enregistrés en mode négatif pour tous les échantillons ci-dessus montrent la présence d'espèces polymères, avec les mêmes signaux dans des intensités variées (Figure III. 9 représente un exemple avec **P14**, les autres spectres correspondants à **P15**-**P16**-**P17** sont visibles Figure PE.4. 44). L'échantillon **P14** contient le polymère ciblé (M en vert), ce même polymère avec une unité **1** manquante (M' en orange), et l'impureté à +293 Da observée pour le polymère ciblé (X en rouge) ainsi que son analogue avec une unité **1** manquante (X' en rose).

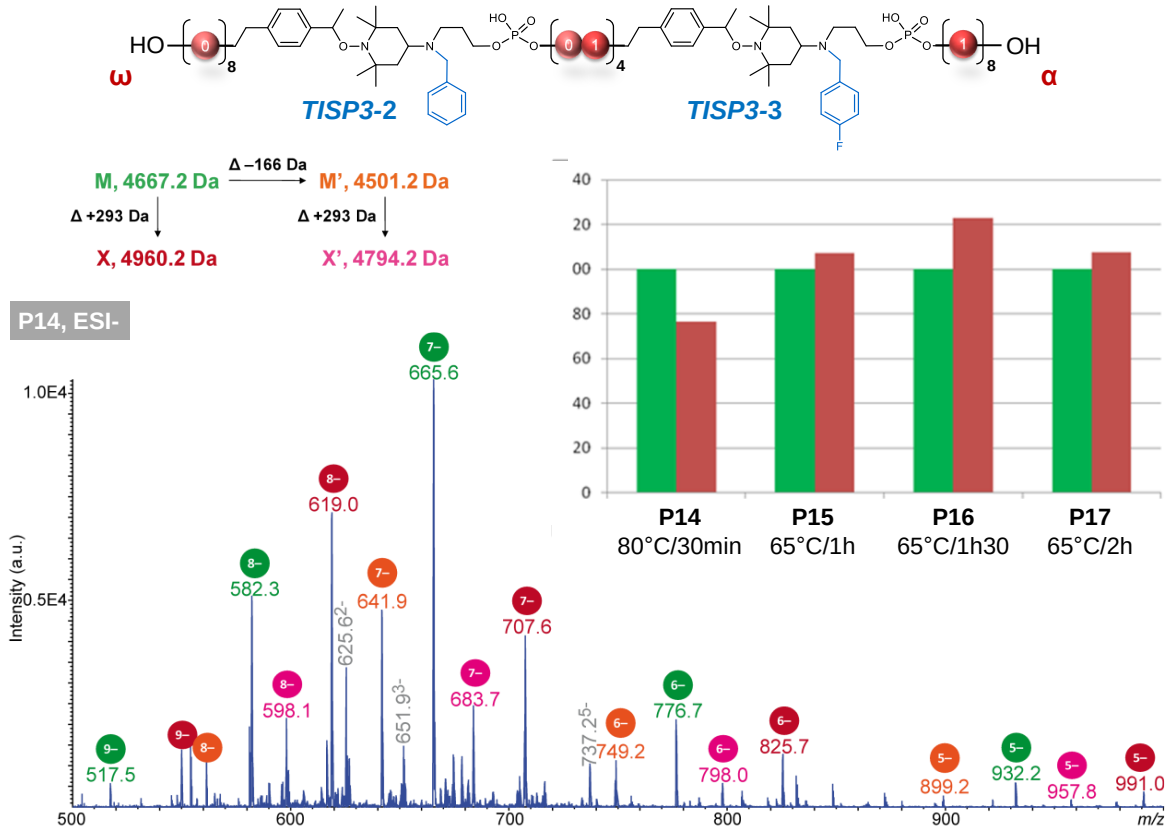


*Figure III. 9 : Spectre ESI-MS en mode négatif de **P14**, et comparaison semi-quantitative des intensités du polymère ciblé **P14**–**P17** et de l'impureté +293 Da.*
Diagramme : intensités de tous les états de charge du polymère ciblé sommées et normalisée à 100 (en vert) pour comparaison inter-échantillon de la proportion relative de l'impureté "Δ+293 Da" (en rouge).

Une comparaison semi-quantitative des intensités du polymère ciblé et de l'impureté +293 Da permet de remarquer un ratio quasi similaire pour tous les tests effectués (Figure III. 9, diagramme), avec une petite amélioration lorsque la déphosphorylation a lieu en chauffant légèrement plus (**P14**, 80°C/30min). Cependant, en comparant ces données avec les résultats obtenus dans l'échantillon **P12** (Figure III. 8A), où la température à 70°C semblait avoir quasiment éliminé toutes traces de l'impureté à +293 Da, des incohérences subsistent quant à l'influence de la température sur l'efficacité de la déphosphorylation.

Alors que cette impureté résulte certainement d'une mauvaise déphosphorylation liée à l'hydratation du composé bicyclique (Schéma III. 2), une autre possibilité peut également être l'utilisation d'une solution de clivage altérée. Hypothétiquement, elle pourrait ne plus être assez concentrée en NH_3 et $MeNH_2$, ce qui diminuerait l'efficacité de clivage (généralement, les solutions sont gardées quelques semaines dans le laboratoire dans

un tube de type Falcon vissé). Afin de vérifier cette éventualité, plusieurs essais ont été effectués avec des solutions de clivage faites le jour-même, et selon plusieurs facteurs, comme décrit ci-dessous (**P18**-**P20**, Figure III. 10). Même si des températures élevées semblent être efficaces pour cliver correctement le polymère du support, il est probable qu'une température trop haute dégrade les liaisons alcoxyamines du polymère. C'est pourquoi, les tests effectués ci-dessous ont été réalisés à 55°C, mais cette fois-ci, durant une nuit, en espérant qu'un temps rallongé, combiné à une température moins élevée, peut être performant pour cette déphosphorylation.

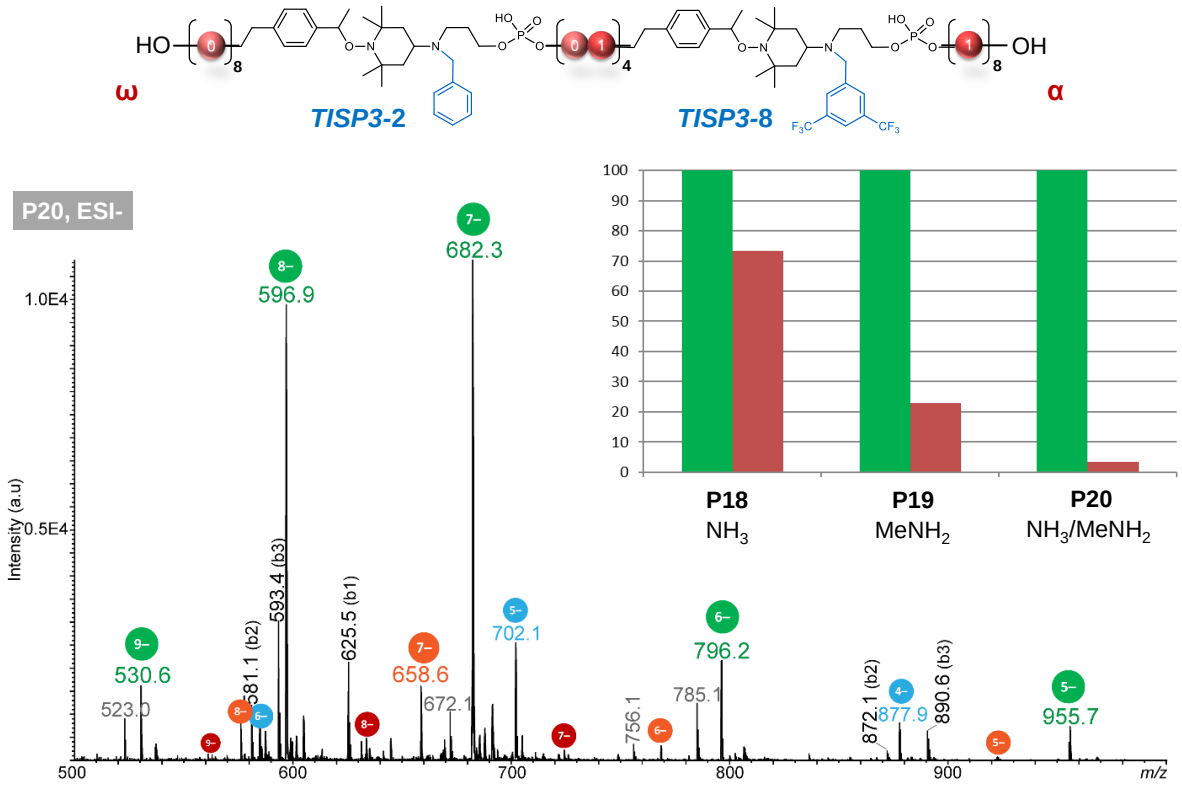


*Figure III. 10 : Comparaison semi-quantitative des intensités du polymère ciblé **P18–P20** et de l'impureté +293 Da.*
*Diagramme : intensités de tous les états de charge du polymère ciblé sommées et normalisée à 100 (en vert) pour comparaison inter-échantillon de la proportion relative de l'impureté "Δ+293 Da" (en rouge). Spectre : analyse ESI-MS en mode négatif de l'échantillon **P20**.*

Dans ces essais, trois différentes solutions ont été testées : NH_3 (28%) (**P18**), $MeNH_2$ (40%) (**P19**) et NH_3/$MeNH_2$ (AMA) (**P20**). Les résultats obtenus ci-dessus démontrent tout d'abord que la meilleure option reste l'utilisation d'AMA pour un clivage effectif (**P20**). Par ailleurs, les données recueillies avec ce même échantillon prouvent que, laisser la solution de clivage contenant le polymère pendant une nuit à une température amoindrie, est efficace pour cliver correctement le polymère du support.

À la suite de ces observations, des derniers essais ont été réalisés pour vérifier la dégradabilité ou non du polymère, dans le cas où ce dernier est clivé à 60 ou 70°C pendant une nuit. Pour ce faire, une séquence modèle comme ci-dessous a été préparée, où « RCS » a été encodée grâce au langage ASCII (petit clin d'œil à notre club de foot alsacien préféré) :

- 01010010 – ***TISP3-2*** – 01000011 – ***TISP3-3*** – 01010011 – OH (Figure III. 11)

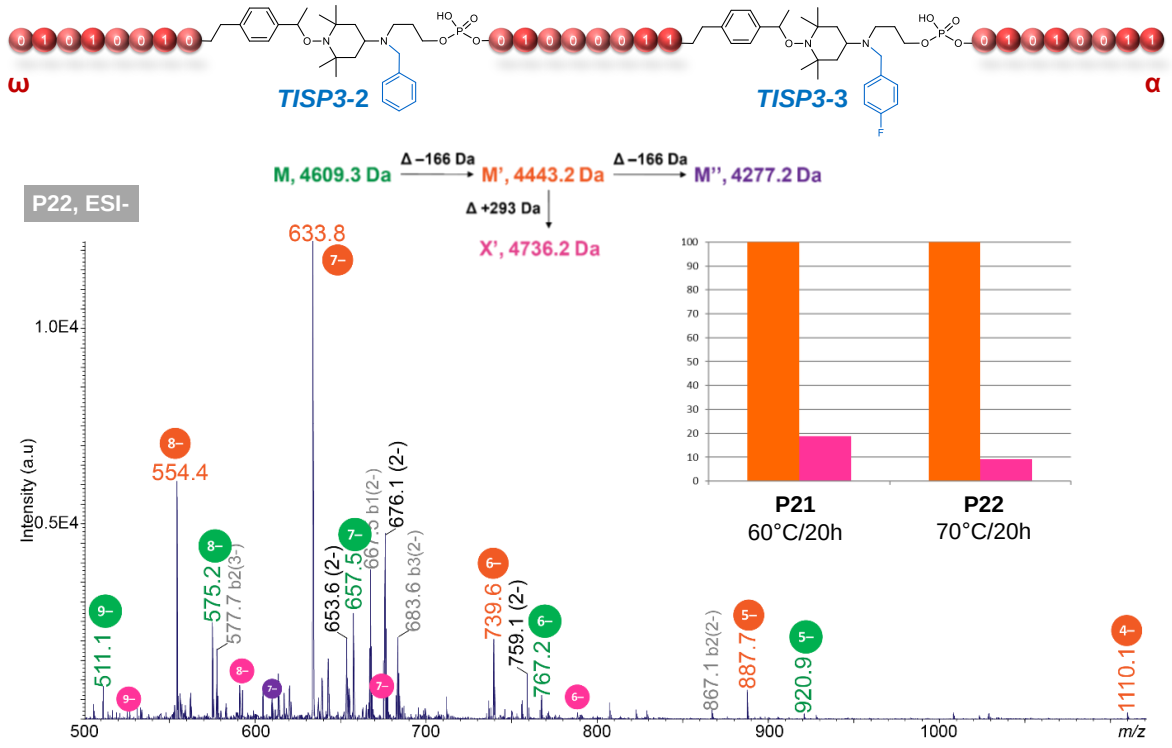


*Figure III. 11 : Spectre ESI-MS en mode négatif de **P22**, et comparaison semi-quantitative des intensités du polymère ciblé **P21**-**P22** et de l'impureté +293 Da.*
Diagramme : intensités de tous les états de charge du polymère ciblé avec une unité 1 manquante (le majoritaire dans les deux échantillons) sommées et normalisée à 100 (en orange) pour comparaison inter-échantillon de la proportion relative de l'impureté "Δ+293 Da" (en rose).

Les spectres ESI(–)-MS enregistrés pour ces deux échantillons (Figure III. 11, **P22** montré ici, **P21** sur la Figure PE.4. 46) présentent les mêmes espèces mais avec des intensités relatives différentes. Les signaux du polymère ciblé M (en vert) sont nettement plus intenses dans l'échantillon **P22** mais l'espèce majoritaire reste le polymère M' (en orange) où il manquait une unité **1** dans le dernier bloc. Dans le 3ème polymère détecté M'' (en violet), il manque 2 unités **1**. L'impureté "Δ+293 Da" est uniquement observée pour le polymère M' : M'+293 = X' (en rose).

Malgré la polydispersité du polymère ciblé, pollué par une espèce secondaire résultant d'une mauvaise efficacité de couplage du monomère **1**, le clivage à 70°C n'a pas montré de signe de dégradation particulier (**P22**). De plus, ce dernier présente moins d'espèce à +293 Da que son analogue à 60°C (**P21**). Par conséquent, afin de tester et d'incorporer chaque marqueur clivable synthétisé (Schéma III. 1), les conditions retenues, après tous ces essais d'optimisations, ont été : clivage à 70°C pendant une nuit. En couplant ces conditions à l'utilisation d'une solution strictement anhydre de monomère **1**, les efficacités de couplages et par conséquent les résultats devaient être améliorées

Une série de polymère, en accordance avec les règles définies précédemment dans le Chapitre 2 (partie 2.2), a été préparée pour vérifier l'aptitude de chaque macromolécule contenant les marqueurs clivables à être séquencée par pseudo-MS^3. Des triblocs codant

« PMC », faisant référence au nom de notre ancienne équipe à l'Institut Charles Sadron, ont été synthétisés (Tableau III. 4) :

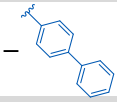
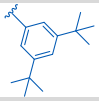
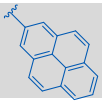
Nom	ω – Bloc1 –	– Bloc2 –	– Bloc3 – α
P23			
P24			
P25			
P26			
P27			

*Tableau III. 4 : Polymères à tribloc (**P23**-**P26**) et dibloc (**P27**) incorporant les différents marqueurs clivables **TISP3** (Schéma III. 1) préparés dans cette thèse.*

Le dernier polymère de ce Tableau III. 4 (**P27**) a été testé différemment, en conséquence d'une mauvaise solubilité du dérivé pyrènyle. Dans l'acétonitrile anhydre, ce dérivé conduit à des agrégats non solubles certainement dû à des interactions π. De ce fait, un test a été réalisé avec une solution de ACN/DCM anhydre (9/1), permettant de solubiliser le dérivé pyrènyle. Le polymère obtenu dans ce cas a conduit à de bons résultats comme illustré dans la Figure III. 12.

Cependant, et contre toute attente, aucuns des autres polymères n'ont conduit à des résultats satisfaisants. Chacun d'eux présentaient divers défauts tels que :

- Concentration trop élevée de TCA compromettant l'analyse du polymère (**P24** et **P26** visibles Figure PE.4. 47 et Figure PE.4. 49)
- Un mélange de polymères ayant pour certains, un marqueur manquant, ou encore, une ou deux unités **1** manquantes (**P23** et **P25**).

Un exemple de spectre contenant un mélange polydisperse dans l'échantillon **P23** est présenté dans la Figure III. 13.

Ces observations vont à l'encontre de celles observées pour l'échantillon **P22** précédemment (Figure III. 11). Alors que l'échantillon **P23** a également été clivé à 70°C pendant une nuit (mêmes conditions que **P22**), il est peu logique que les mêmes marqueurs ***TISP3*-2** et ***TISP3*-3** impliqués dans ces deux expériences, aient des comportements complétement différents. Une autre possibilité peut provenir de la déprotection au TCA lors de la purification du polymère : une concentration trop forte en TCA pourrait dégrader certains sites fonctionnels du polymère résultant en un mélange

polydisperse. Dans les essais **P23**-**P27**, un cycle de lavage TCA/H_2O avait été répété 3 fois dans le but de déprotéger un maximum de polymère et optimiser le rendement.

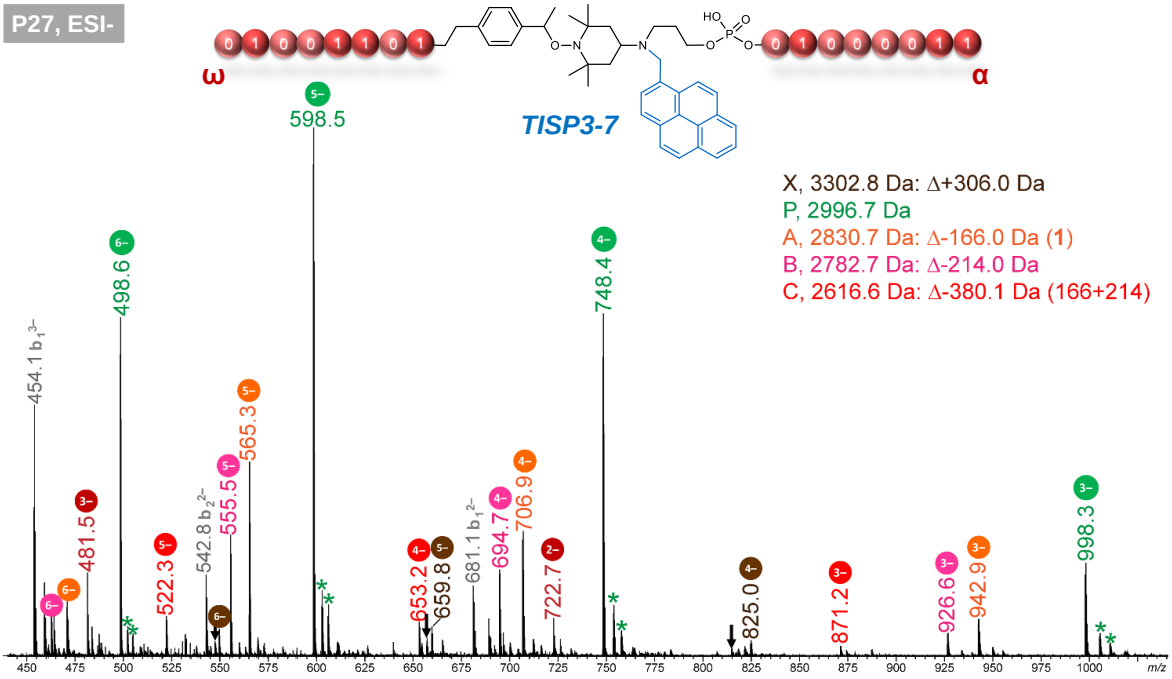


*Figure III. 12 : Spectres ESI-MS en mode négatif de l'échantillon **P27**.*

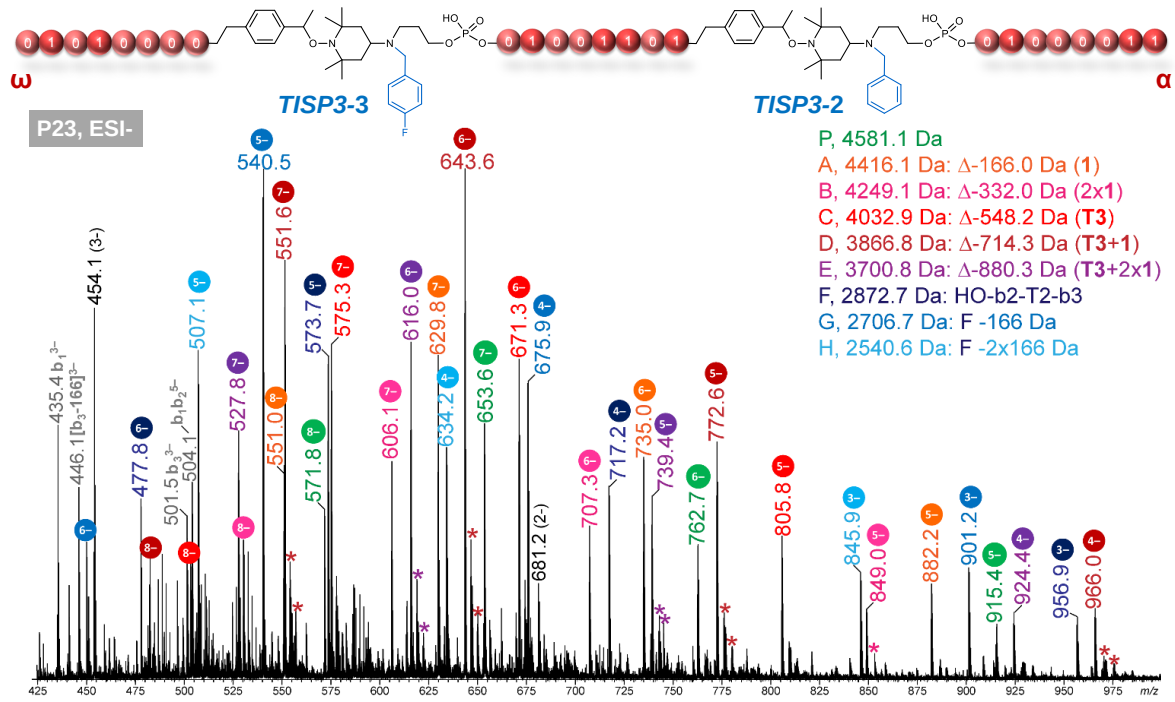


*Figure III. 13 : Spectres ESI-MS en mode négatif de **P23**.*

C'est pourquoi, pour vérifier l'hypothèse de l'implication du TCA dans cette polydispersité, un test a été réalisé en effectuant qu'un seul lavage TCA/H_2O, en rinçant avec 4mL H_2O au lieu de 2mL, pour éliminer de manière optimale les sels de TCA. En même temps,

deux séquences modèles ont été testées avec deux températures différentes : 50°C/21h et 70°C/21h.

- 01010000 – ***TISP3-5*** – 01001101 – ***TISP3-6*** – 01000011 – OH (Figure III. 14)
- 01010000 – ***TISP3-7*** – 01001101 – ***TISP3-10*** – 01000011 – OH (Figure III. 15)

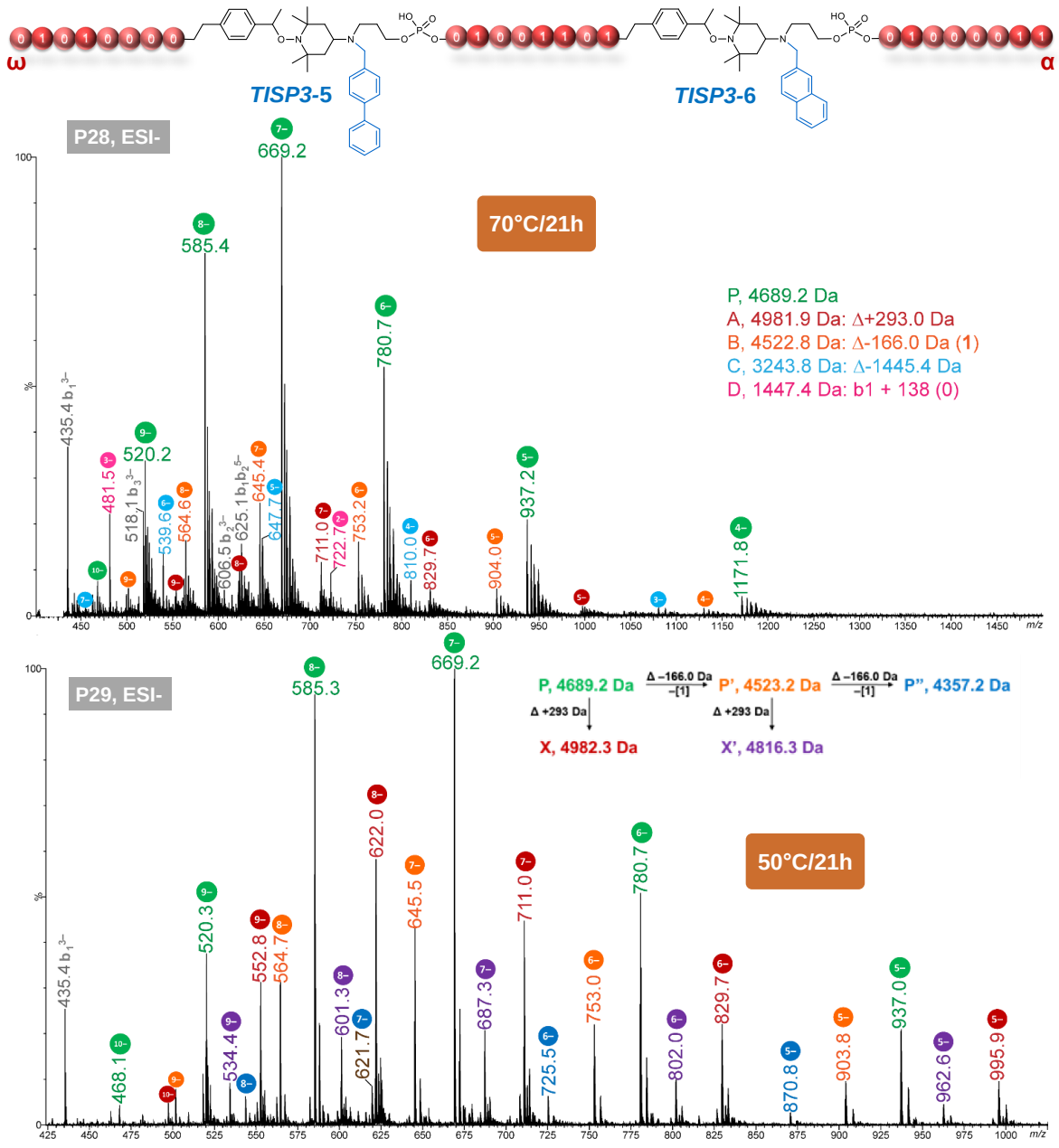


*Figure III. 14 : Spectres ESI-MS en mode négatif des échantillons **P28** et **P29**. **P28** (haut) clivé à 70°C/21h ; **P29** (bas) clivé à 50°C/21h.*

Le spectre de masse enregistré en mode négatif pour l'échantillon **P28** montre nettement le polymère attendu (en vert, **P28** Figure III. 14). La composition des 3 blocs attendus dans le polymère ciblé a été validée par MS/MS (Figure PE.4. 50). Les autres espèces inventoriées correspondent à l'impureté "+293 Da" (A, en rouge), à un polymère dans lequel il manque une unité **1** dans le bloc 3 (B, en orange), un polymère dibloc qui contient le bloc 3 et un autre bloc de masse 1684.6 Da (C, en bleu) et une espèce qui pourrait être

un bloc "bloc1+138" libéré en source (D, en rose) mais dont l'origine n'a pas pu être établie. Le même polymère clivé à 50°C, et analysé dans les mêmes conditions, montre aussi nettement le polymère attendu sous ses différents états de charges (en vert, **P29**, Figure III. 14). Il ne démontre pas de dégradation particulière par rapport à son analogue clivé à 70°C et contient principalement les mêmes types d'impuretés observées pour **P28**. Cependant, **P29** présente quant à lui une quantité plus importante de sous-produit à +293 Da signifiant que la température a permis, dans **P28** à 70°C, de diminuer la quantité de cette impureté.

Paradoxalement, le spectre de masse enregistré en mode négatif pour l'échantillon **P30** (Figure III. 15) n'a montré aucun des signaux attendus pour le polymère recherché mais une série de polymères sans espaceurs dans lesquels il manque plus ou moins d'unités codantes, comme inventorié sur cette même figure.

Par ailleurs, le spectre de masse enregistré en mode négatif pour l'échantillon **P31** montre quant à lui, pour le même polymère recherché mais clivé à 50°C, les signaux attendus (Figure III. 15) tel qu'indiqué par les données MS/MS (Figure PE.4. 52). Cependant, ces signaux restent minoritaires devant ceux d'un polymère ne contenant pas l'espaceur ***TISP3*-10** (P', en rose). D'autres impuretés ont également pu être inventoriées comme détaillé sur la Figure III. 15.

À la suite de ces observations, où **P31** présente un profil spectral moins polydisperse que son analogue **P30** clivé à 70°C, la température joue assurément un rôle dans la polydispersité observée jusqu'à présent, et ce, même si d'autres résultats semblaient démontrer le contraire (**P21**-**P22** et **P28**-**P29**). En effet, ces différences peuvent également s'expliquer par la sensibilité des espaceurs à la température appliquée pendant le clivage. Alors que l'échantillon **P31** montre une espèce majoritaire correspondant au polymère avec l'espaceur ***TISP3*-10** manquant, un autre exemple présenté précédemment, démontre également l'absence de certains espaceurs dans la macromolécule ciblée (**P23**, Figure III. 13). À ce stade, il a été préférable de vérifier un paramètre important : est-ce que les défauts de masses correspondants aux manques d'espaceurs proviennent de la température de clivage appliquée au polymère, ou bien à leur mauvaise insertion dans la chaîne comme il peut l'être fréquemment observé pour les monomères **0** et **1** ? Pour répondre à cette interrogation, des essais impliquant les espaceurs ***TISP3*-7** et ***TISP3*-10**, identifiés comme étant les plus problématiques jusqu'ici, ont été effectués avec la séquence modèle ci-dessous. Celle-ci utilise un support solide Thymidine **T** ne nécessitant pas de hausse de température durant le clivage :

- 01010000 – ***TISP3*-7** – 01001101 – ***TISP3*-10** – 01000011 – **T** (Figure III. 16)

Le spectre de masse enregistré en mode négatif pour l'échantillon **P32** ne montre pas les signaux attendus pour le polymère recherché (Figure III. 16A), mais deux autres polymères sont observés :

- P1 (en rouge) manquant l'espaceur ***TISP3*-10**, tel que confirmé par les données MS/MS de l'ion $[P1 - 6H]^{6-}$ à m/z 742.8 (Figure III. 16B).
- P2 (en orange) manquant l'espaceur ***TISP3*-10** et une unité **1**.

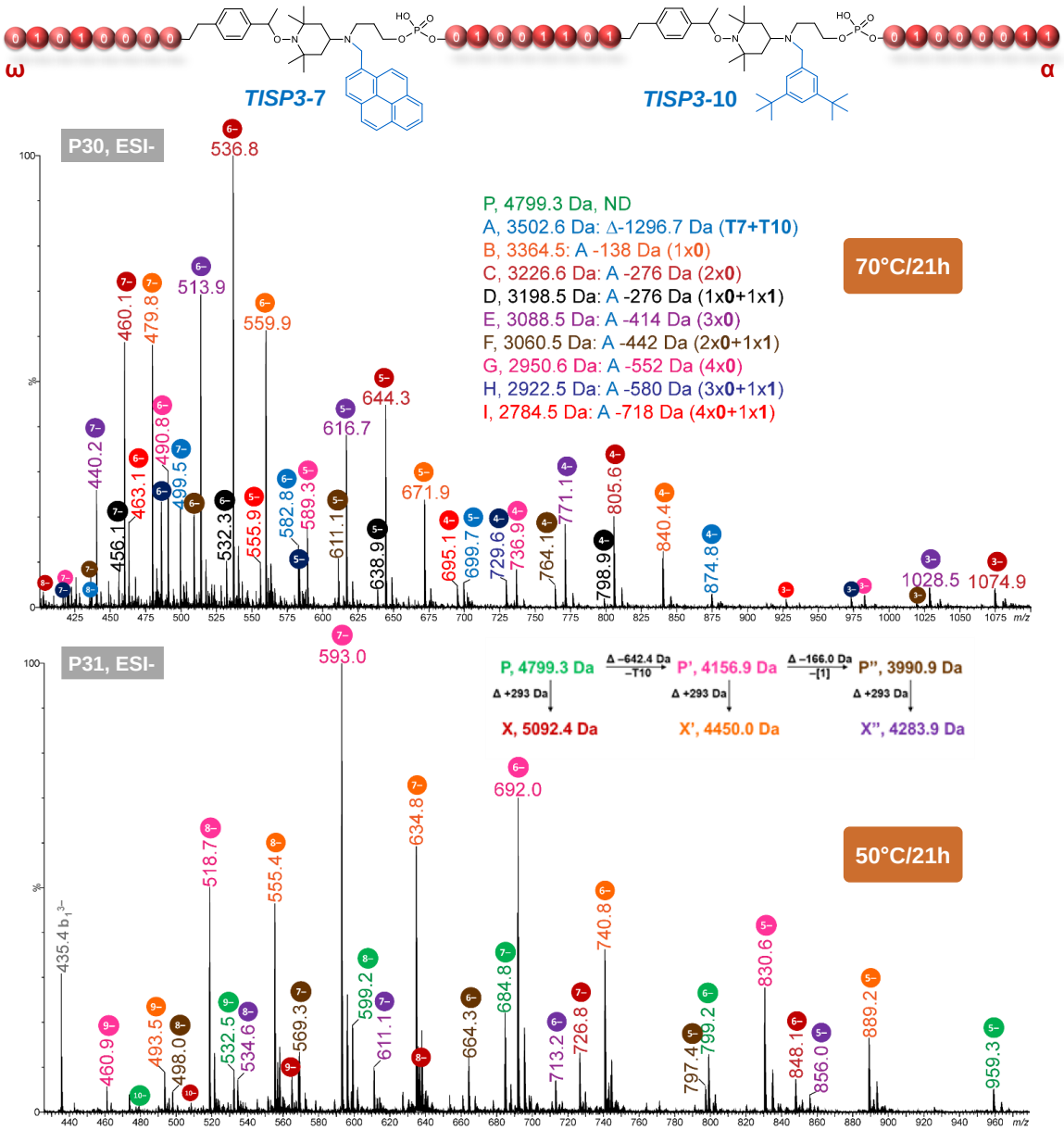


*Figure III. 15 : Spectres ESI-MS en mode négatif des échantillons **P30** et **P31**.*
***P30** (haut) clivé à 70°C/21h ; **P31** (bas) clivé à 50°C/21h.*

Plusieurs conclusions ont pu être tirées de cette expérience. Premièrement, l'utilisation d'un support solide CPG-Thymidine conduit à des résultats d'analyses beaucoup plus propres, où seulement un sous-produit majeur est détecté. L'utilisation des nouveaux supports solides Unysupports (Figure III. 6) nécessitent donc encore de l'optimisation. Par ailleurs, cet essai a également pu mettre en évidence que le défaut de masse correspondant au manque de l'espaceur ***TISP3*-10** n'est pas lié à la température de clivage appliquée au support, mais bien à la mauvaise incorporation du marqueur dans la chaîne polymère. Cette déduction pourrait avoir de lourdes conséquences quant à l'efficacité de couplage des différents marqueurs clivables synthétisés jusqu'à présent. Même si ***TISP3*-10** est un des seuls espaceurs avec lequel ce défaut a pu être relevé, les autres marqueurs pourraient très bien avoir été insérés dans la chaîne polymère, mais avec une très faible efficacité de couplage. Dans ce cas, la construction de longues

macromolécules utilisant plusieurs marqueurs de ce type serait compromise par un mauvais rendement de synthèse.

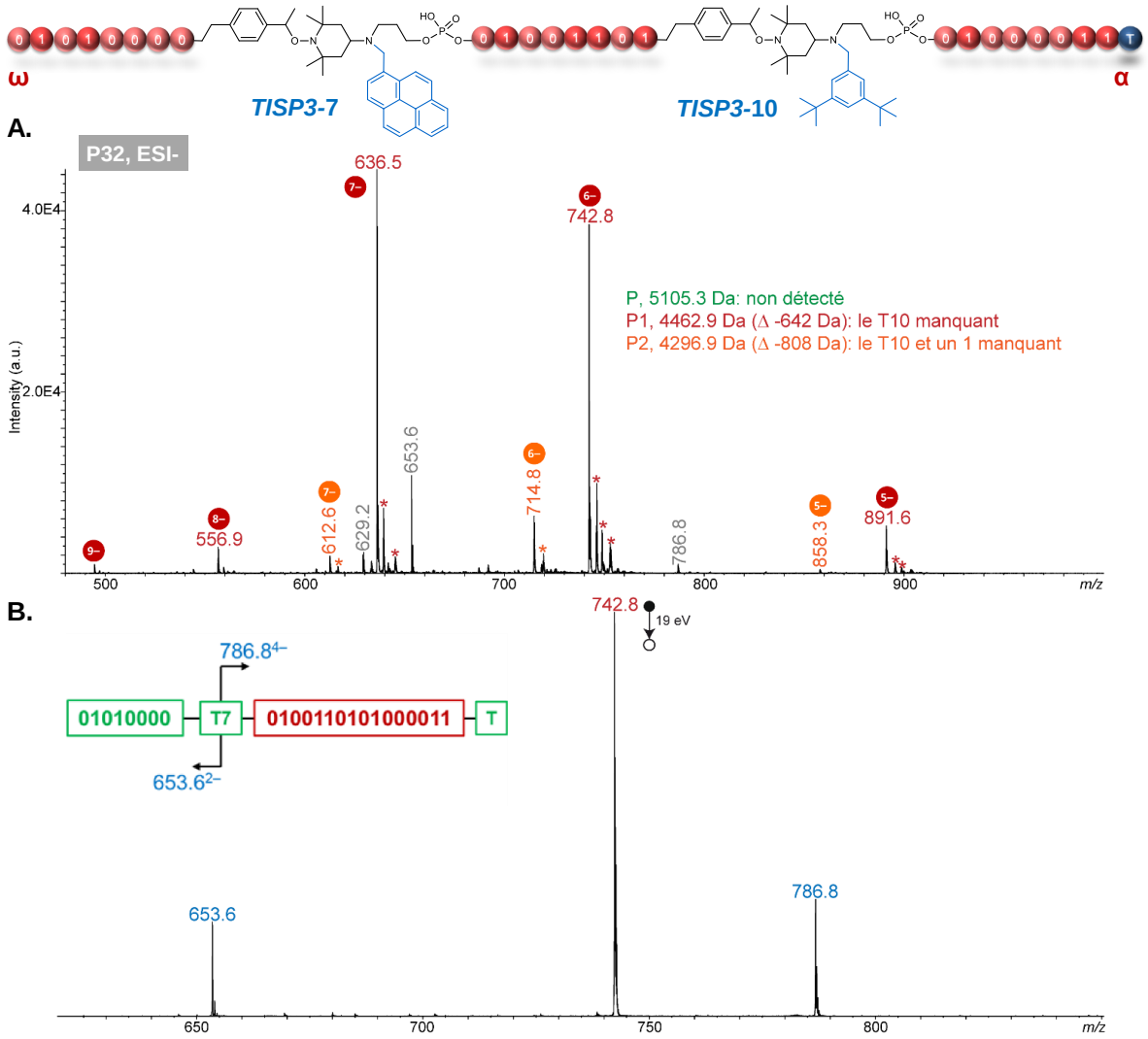


Figure III. 16 : ***A.*** *Spectres ESI-MS en mode négatif et* ***B.*** *MS/MS de l'ion [P1 – 6H]*$^{6-}$ *à m/z 742.8 de* ***P32****.*

Avant de continuer l'optimisation de ces nouveaux supports solides, il a donc été préférable, à ce stade, de vérifier l'efficacité de couplage de chaque marqueur clivable ***TISP3*** synthétisé. La prochaine sous-partie expliquera les méthodes et détaillera les résultats de telles études.

4. Contrôle de l'efficacité de couplage des monomères *TISP3*

4.1. Méthodes et principes : quantification du cation DMT

Comme discuté dans le Chapitre 1 (Figure I. 10), l'efficacité de couplage des monomères phosphoramidites peut être quantifiée grâce au sel de DMT déprotégé lors de la première étape de déblocage dans le cycle itératif décrit.[381,25] Plus précisément, en prenant en

référence la quantité de DMT déprotégé lors d'un couplage antérieur *n-1*, le taux de sel de DMT analysé après le couplage *n* permet, en comparant les 2 résultats obtenus, d'estimer l'efficacité de ce dernier couplage EC.

$$EC = \left(\frac{A_n}{A_{n-1}}\right) \times 100$$

Afin d'obtenir une telle quantification, les mesures d'absorbances de chaque fractions DMT sont faites à 504 nm, longueur d'onde à laquelle la délocalisation du sel émet de l'énergie. Enfin, les valeurs d'absorbance obtenues peuvent être comparées sur un diagramme tel que présenté sur la Figure III. 17.

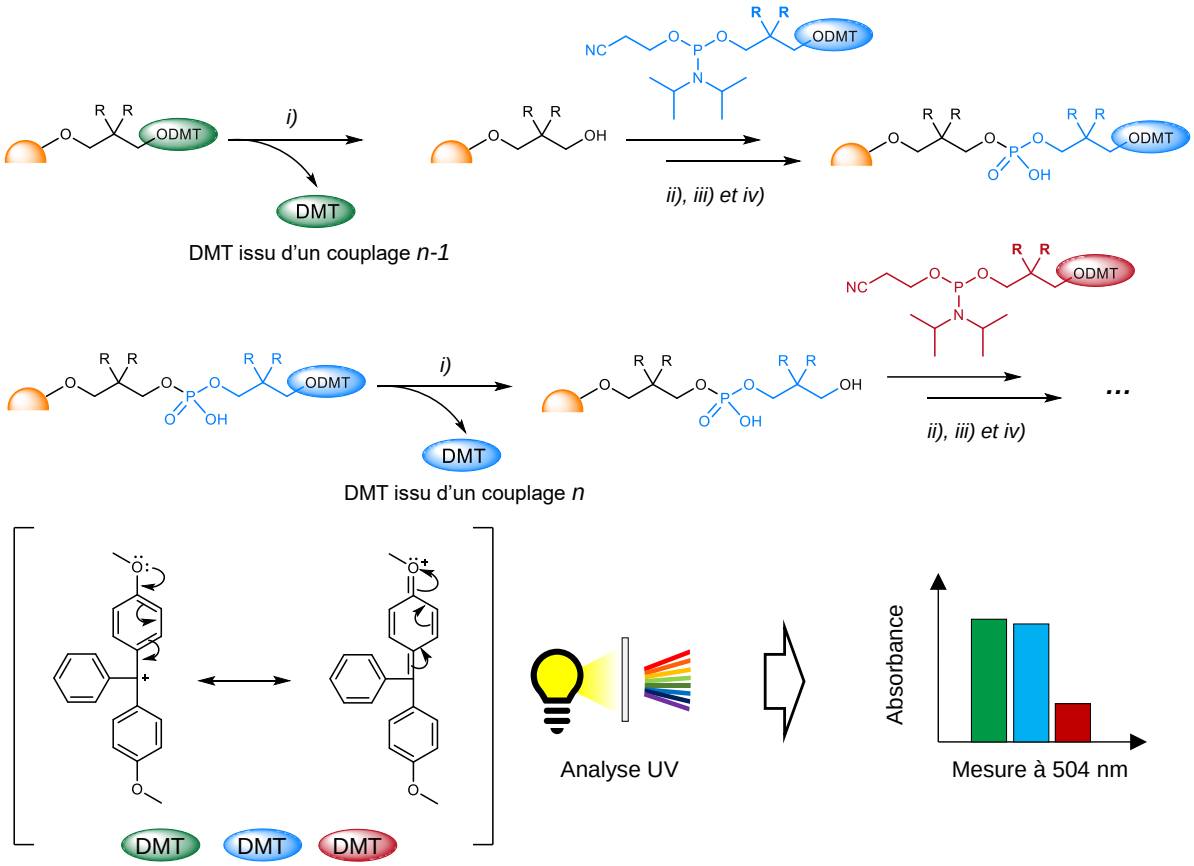


Figure III. 17 : Principe de quantification du sel de DMT par analyse UV permettant d'estimer l'efficacité de couplage.
i) déprotection ; ii) couplage phosphoramidite ; iii) piégeage ; iv) oxydation.

Dans cet exemple, l'absorbance du DMT bleu issu d'un couplage *n* peut être comparée au DMT vert issu d'un couplage *n-1*. En divisant A_n/A_{n-1}, un pourcentage d'efficacité de couplage peut être estimé. En comparant dans cette représentation le bleu au vert, l'efficacité serait proche de 100% puisque le taux de DMT déprotégé après le couplage *n* est quasi similaire à celui issu du couplage *n-1*. Paradoxalement, dans le cas où l'absorbance du DMT issu d'un couplage *n+1* viendrait à être plus faible (ici illustré en rouge) qu'un couplage *n* (bleu), l'efficacité serait amoindrie.

En pratique, le robot Expedite 8909 permet de réaliser de telles analyses. Grâce à un mode « *Hold* », le robot peut stopper la synthèse phosphoramidite avant de commencer

un nouveau cycle, c'est-à-dire, avant de déprotéger le monomère *n-1* fraichement ajouté. À ce stade, les injections de réactifs passés à travers la colonne peuvent être récupérées après chaque étape de déprotection dans des fioles jaugées. Après reformation de l'acide en solution grâce à une solution de TCA en utilisant des quantités bien précises (P.E., 1.2.2), les différentes fractions peuvent être analysées à l'aide d'un spectromètre UV classique et comparées, afin d'estimer l'efficacité d'un couplage *n*.

Généralement, cette technique est appliquée à la synthèse de longs polymères pour estimer l'efficacité de couplage moyenne globale issue de chaque cycle itératif. Si un seul monomère est utilisé pour une telle synthèse, cela permet de déterminer l'efficacité de couplage moyenne liée à ce dernier. Pour ce faire, ce n'est pas seulement 2 fractions *n* et *n-1* qui sont prises en compte mais toute la séquence : en comparant la troisième fraction (correspondant à la première issue d'un couplage entre 2 monomères sur la chaîne) et la pénultième, l'équation ci-dessous permet d'estimer l'efficacité de couplage EC, où ici, *n* = nombre total d'étapes :

$$EC = \left(\frac{A_{n-1}}{A_3}\right)^{\frac{1}{n-3}}$$

Le rendement théorique de la synthèse peut ainsi être retrouvé en appliquant la formule suivante :

$$(EC)^n$$

Cependant, dans le cas des espaceurs ***TISP3*** devant être étudiés, il est intéressant de quantifier l'efficacité de couplage entre ces derniers, et un monomère **0** ou **1**. Effectivement, déterminer l'efficacité de couplage moyenne en réalisant un homopolymère de marqueur clivable n'est pas judicieux, puisque le design étudié dans cette thèse, nécessite qu'un seul macromonomère incorporé à la fois dans la chaîne polymère. Paradoxalement, estimer l'efficacité de couplage à l'aide d'homopolymères **0** ou **1** est logique puisque dans certaines macromolécules numériques, une suite de plusieurs **0** ou **1** peut facilement être requise.

C'est pourquoi, la prochaine sous-partie détaillera les résultats obtenus des efficacités de couplages des marqueurs clivables ***TISP3*** en utilisant la première approche décrite Figure III. 17.

4.2. Résultats et discussion

Afin de tirer profit de cette stratégie pour quantifier les efficacités de couplages, les séquences modèles ci-dessous ont été réalisées :

0101 – **X** – **0**101T

X représente le marqueur clivable ***TISP3*** testé, tandis que T représente le support thymidine utilisé pour ces essais. Les fractions collectées pour comparer l'efficacité d'un couplage *n* (**X**) à un couplage *n-1* (**0**) ont été dans cette étude, les fractions F5 et F4 respectivement (Figure III. 18).

Avant de commencer les premiers essais, il a été important de définir un protocole pour être certain de collecter les fractions correspondantes au sel de DMT. Dans le cas du robot Expedite 8909, le nombre d'injections et/ou le volume nécessaire peuvent être modifiés pour chaque étape du cycle, afin de s'assurer que les réactifs atteignent la colonne en bonne quantité, et au moment opportun.

Ces caractéristiques propres au robot se sont révélées très importantes, comme illustré par le premier essai effectué avec le dérivé benzyle ***TISP3*-2** ci-dessous (Figure III. 18).

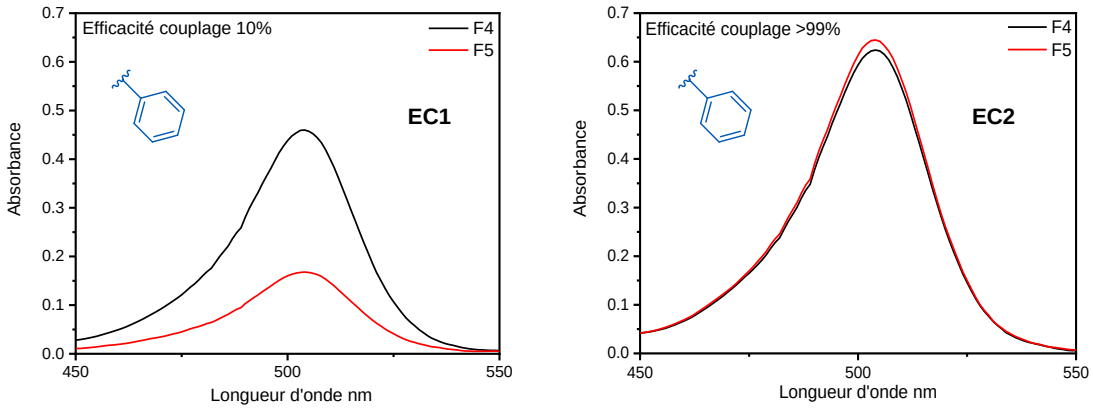


*Figure III. 18 : Comparaison d'estimation de couplage entre les positions A/T/C (**EC1**) et 5/6/7 (**EC2**).*

Alors que le premier test montre des résultats peu encourageants quant à la quantité de DMT mesurée après le couplage de ***TISP3*-2** en utilisant les positions A/C/T pour **0**, **1** et ***TISP3*-2** respectivement (**EC1**, Figure III. 18), ce même marqueur démontre une tout autre réactivité en utilisant d'autres positions du robot telles que 5/6/7 pour **0**, **1** et ***TISP3*-2** respectivement (**EC2**, Figure III. 18) (Ces résultats ont été obtenus à l'aide de Georgette OBEID, stagiaire M2).

Ceci peut s'expliquer par la différence d'injections imputée par défaut dans un protocole de synthèse d'ADN entre les positions A/C et T. Comme spécifié dans le manuel d'utilisation du robot, « chaque réactif est délivré sous forme d'injections à volume fixe par des injecteurs individuels. Par conséquent, des volumes de réactifs constants sont délivrés aux colonnes de réaction » (Figure III. 19 montre le diagramme fluidique du modèle Expedite 8909 utilisé et le Tableau III. 5 montre les différentes injections nécessaires pour atteindre la colonne en fonctions des positions). Pour le test **EC1**, la différence d'injection entre les positions A/T et C jouent à fortiori un rôle dans la quantification de DMT. De plus, il est possible que les protocoles définis pour les positions A/T/C/G ont été spécifiquement élaborés pour les nucléobases correspondantes, induisant des résultats erronés lorsque d'autres monomères sont utilisés à ces positions.

En conséquence de ces hypothèses, le test **EC2** a été réalisé en utilisant les positions 5/6/7 qui, par défaut, ont pour chacune des positions, les mêmes protocoles de synthèse et n'ont certainement aucun monomère (naturel ou non) approprié. Le résultat obtenu pour ce test est finalement très encourageant puisque l'efficacité de couplage a été estimée à >99% contrairement à **EC1** où elle n'était que de 10%.

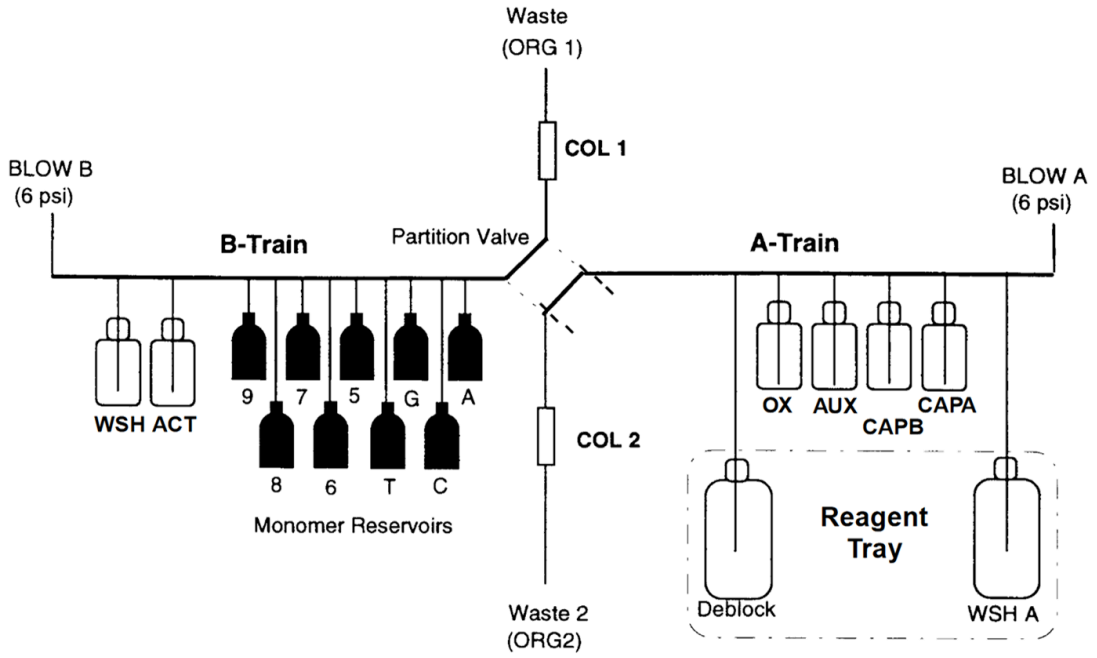


Figure III. 19 : Diagramme fluidique du robot Expedite 8909.[382]

Position du monomère	**A**	**C**	**G**	**T/U**	**5**	**6**	**7**	**8**	**9**
Injections vers la colonne	7	7	8	8	8	8	8	9	9

Tableau III. 5 : Nombre d'injections nécessaires pour atteindre la colonne en fonction des positions.

Il est important de rappeler que l'on parle « d'estimation de couplage » du fait de la méthode d'analyse arbitraire utilisée. Les résultats obtenus dans cette sous-partie font uniquement référence à la quantité de DMT déprotégée obtenue après un cycle itératif de couplage phosphoramidite, et peut par conséquent induire en erreur certaines estimations en fonction du monomère étudié. C'est notamment le cas des mesures que proposent le robot Expedite, en estimant l'efficacité de couplage directement sur l'interface de l'appareil grâce à un détecteur UV placé à l'intérieur de l'instrument. Un diagramme représentatif du test **EC2** est présenté dans la Figure III. 20 à titre d'exemple.

Ce diagramme présente un défaut principal : l'irrégularité des quantités de DMT observées. Particulièrement, la quantité de DMT plus importante pour un couplage *n+1*, qui est théoriquement impossible. Ce phénomène peut s'expliquer par plusieurs raisons, comme détaillé dans le manuel d'utilisation du robot :

- Un dégagement gazeux de la solution de déblocage dans la cellule d'écoulement au moment de la détection UV.
- Une anomalie de détection liée à un moniteur du trityle non performant.
- Mais surtout, les différences de taux de déblocage du DMT pour divers monomères. Par exemple, le groupe DMT d'un résidu G (Guanosine) est retiré en un laps de temps très court, résultant en une solution intensément colorée. À

l'inverse, le groupe DMT est éliminé d'un résidu T (Thymidine) sur une période plus longue, ce qui donne une solution moins colorée. Étant donné que l'intensité de l'absorbance n'est pas proportionnelle à la concentration de DMT, la quantité de DMT est sous-estimée pour un résidu G mais plus représentative pour un résidu T. Par conséquent, la barre pour un résidu G est un peu plus petite que la moyenne et la barre pour les résidus T est un peu plus grande que la moyenne.

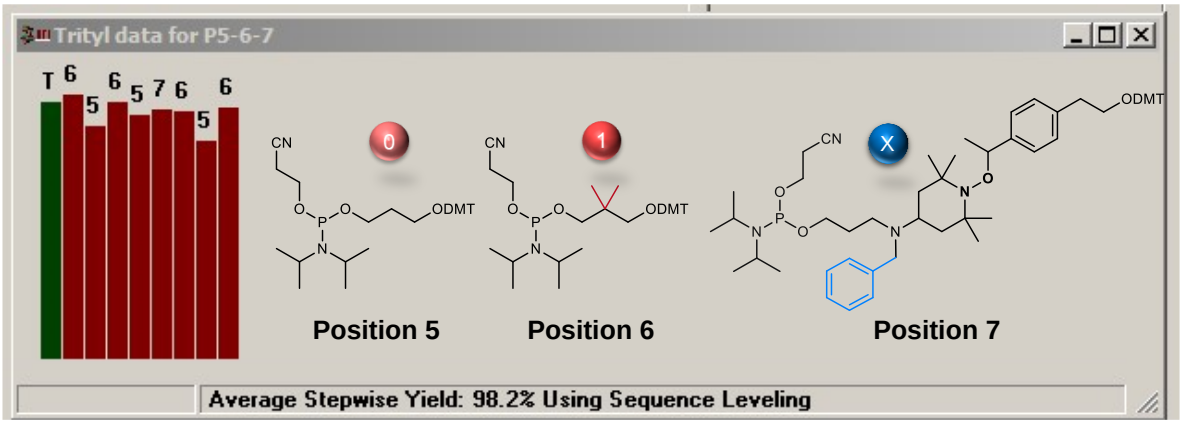


Figure III. 20 : Diagramme affiché sur le robot Expedite montrant les quantités de trityles estimées. T en vert représente la quantité de DMT déprotégé du support (1ère étape).

C'est pour ces raisons qu'une analyse rigoureuse de chaque échantillon a été effectuée à l'aide d'un spectromètre UV externe. En prenant soin de récolter les mêmes quantités de DMT pour chaque fraction par ajustement du protocole (Figure PE.1. 3), une meilleure estimation de couplage a pu être réalisée. Cependant, ceci ne reste tout de même qu'une estimation : la répétition d'essais similaires, dans de mêmes conditions, permet de confirmer la reproductibilité de l'analyse, à +/-2% prêt.

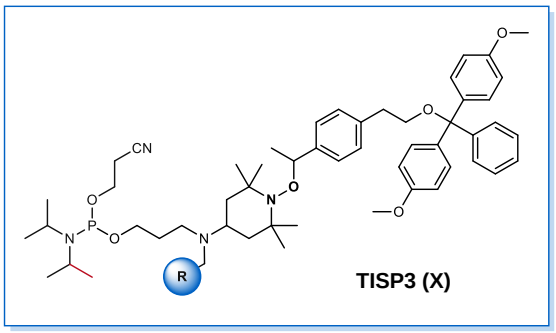
Après avoir considéré l'importance du choix des positions et défini clairement les méthodes d'analyses, l'étude de la réactivité de toute la série ***TISP3*** (Figure III. 21) a pu être réalisée. Le Tableau III. 6 répertorie les résultats obtenus.[383]

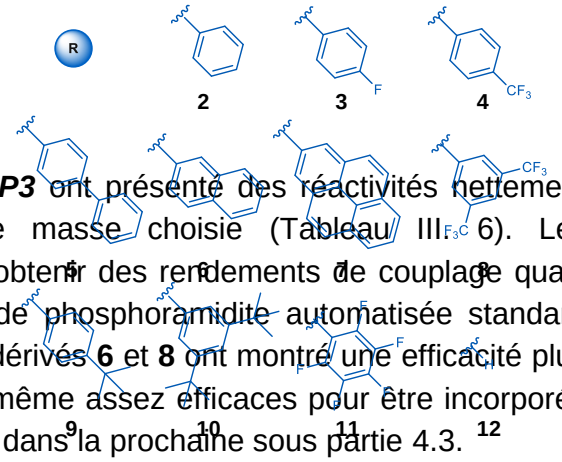
Les résultats les plus faibles ont été obtenus avec les dérivés **4**, **7**, **10, 11** et **12** (entrées 3, 6, 11, 12 et 13). Différentes raisons peuvent expliquer ces faibles estimations de couplage. Tout d'abord, l'effet stérique du marqueur **10** semble influencer l'activation du groupe phosphoramidite et donc le couplage de l'alcool sur ce dernier. Par ailleurs, la présence de groupes fluorés dans le dérivé **11**, ou de groupement CF_3 dans le dérivé **4,** semble diminuer l'efficacité de couplage, ce qui n'est pas autant remarquable pour leur analogue électrodefficient **8**. La raison d'une telle différence peut s'expliquer par la différence de position du groupe CF_3 sur le phényle, induisant hypothétiquement un effet électronique différent, et/ou de problème d'affinité avec le solvant. Entre autres, ce dernier facteur joue un rôle dans la faible solubilité du dérivé pyrènyle **7** dans l'acétonitrile, en raison de l'agrégation classique du pyrène induit par ses fortes interactions π. De ce fait, comme précisé plus tôt (**P27**, Tableau III. 4), un mélange ACN/DCM (9/1) a été utilisé dans ce cas. Cependant, même si des améliorations de solubilité ont été constatées à l'œil nu, l'efficacité de couplage est restée faible. Finalement, la faible efficacité de l'espaceur **12**, correspondant au ***TISP3*** sans marqueur de masse, peut s'expliquer par

l'amine secondaire du macromonomère, qui influe certainement l'activation du groupe phosphoramidite.

Entrée	*TISP*3[a]	Efficacité couplage (%)
1	**2**	> 99[a]
2	**3**	> 99[a]
3	**4**	7.5[a]
4	**5**	> 99[a]
5	**6**	89[a]
6	**7**	24[a]
7	**7**	91[b]
8	**7**	99[c]
9	**8**	78[a]
10	**9**	> 99[a]
11	**10**	9.5[a]
12	**10**	17[b]
13	**11**	9[a]
14	**12**	17[b]

*Tableau III. 6 : Efficacité de couplage des différents **TISP3** (Figure III. 21). Concentration espaceur : [a] 0.1 mol·l^{-1} ; [b] 0.15 mol·l^{-1} ; [c] 0.2 mol·l^{-1}.*





Dans l'ensemble, les macromonomères ***TISP3*** ont présenté des réactivités nettement différentes, en fonction de l'étiquette de masse choisie (Tableau III. 6). Les macromonomères **2**, **3**, **5** et **9** ont permis d'obtenir des rendements de couplage quasi quantitatifs dans les protocoles de chimie de phosphoramidite automatisée standard (entrées 1, 2, 4 et 10). En comparaison, les dérivés **6** et **8** ont montré une efficacité plus faible (entrées 5 et 9), mais restent tout de même assez efficaces pour être incorporés dans une chaîne polymère, comme il sera vu dans la prochaine sous partie 4.3.

Afin d'améliorer l'efficacité de couplage, plusieurs options sont possibles. Une première consiste par exemple à augmenter la concentration du macromonomère. Les entrées 6 à 8 du Tableau III. 6 comparent la réactivité de l'espaceur **7** à une concentration de 0,1, 0,15 et 0,2 mol·l^{-1}. En utilisant une concentration jusqu'à 2 fois plus élevée, une efficacité de couplage presque quantitative peut être observée. En utilisant une stratégie similaire, il a également été possible d'augmenter la réactivité du macromonomère **10** (entrée 12), mais avec un succès modéré. Par ailleurs, l'efficacité de couplage peut être améliorée en modifiant le protocole du cycle itératif du robot, en allongeant le temps de couplage. Ce test a été réalisé avec les positions A/T et C lors des premières études, démontrant des résultats encourageants (Figure PE.1. 4). Cependant, cette technique d'optimisation n'est théoriquement pas une option viable pour écrire de l'information numérique à l'échelle moléculaire rapidement. C'est pourquoi, dans l'ensemble, les marqueurs clivables **2-3** et

5-9 semblent être appropriés pour la préparation de polymères numériques, tandis que les dérivés **4**, **10, 11** et **12** sont moins convenables.

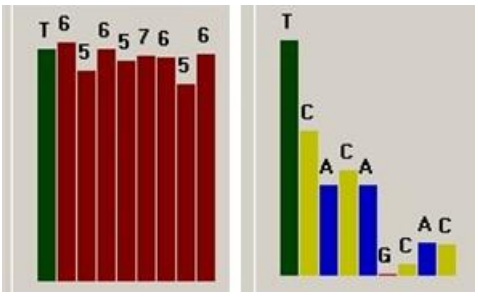


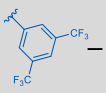
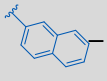
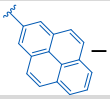
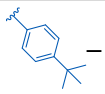
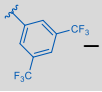
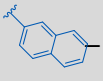
*Figure III. 22 : Diagrammes trityles **EC1** (gauche) et **EC2** (droite).*

Alors que les marqueurs clivables ***TISP3*** ont été fraichement préparés pour chaque essai discuté ci-dessus, les monomères **0** et **1** utilisés dans les séquences modèles ne l'étaient pas lors des premiers tests. De ce fait, un paramètre important a été remarqué sur la base d'estimation de couplage proposée par le robot synthétiseur : le vieillissement des solutions de monomères est en fait extrêmement important et semble être directement corrélé aux efficacités de couplages. Comme démontré sur la comparaison des diagrammes de la Figure III. 22, les monomères datant de plusieurs jours de **EC2** montrent des quantités de DMT considérablement plus faibles que ceux fraichement préparés dans **EC1** (comparaison avec la quantité de trityle déprotégée du support T en vert, et les couplages suivants). À noter que dans ce cas, les positions A et C ne jouaient pas un rôle dans la faible efficacité de couplage puisque ces deux positions sont équivalentes entre elles.

Pour donner suite à ces études d'efficacité de couplages, en considérant avec importance les conditions anhydres des solutions utilisées, des derniers essais ont été réalisé pour l'incorporation de ***TISP*3** en chaîne polymère. Ces tests concluants sont montrés dans la prochaine sous-partie.

4.3. Tests finaux pour l'incorporation des marqueurs clivables *TISP3* en chaîne polymère

Alors que l'efficacité de couplage des différents marqueurs clivables ***TISP3*** est connue, des dernières tentatives probantes ont permis d'obtenir les polymères informationnels désirés, dans la plupart des cas, monodisperses. Pour ce faire, différentes séquences ont été réalisées (Tableau III. 7, l'intégralité des spectres de masse des échantillons présentés ci-dessous sont visibles en P.E., 4.6.2).

Nom	ω – Bloc1 – – Bloc2 – – Bloc3 – α	Produit majoritaire
P33	ω – 01010000 –  – 01001101 –  – 01000011 – α	**Polymère** désiré*
P34	ω – 01010000 –  – 01001101 –  – 01000011 – α	**Polymère** désiré*
P35	ω – 00000000 –  – 11111111 –  – 00000000 – α	**Polymère** désiré

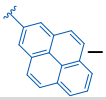
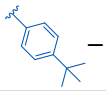
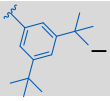
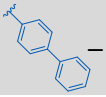
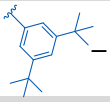
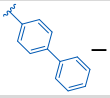
P36	ω – 00000000 –  – 11111111 –  – 00000000 – α	**Polymère** désiré
P37	ω – 00000000 – – 11111111 –  – 00000000 – α	**10** manquant
P38	ω – 00000000 – – 11111111 – – 00000000 – α	**4** manquant
P39	ω – 00000000 –  – 11111111 – – 00000000 – α	**5** et **10** manquant
P40	ω – 01010000 – – 01001101 –  – 01000011 – α	**10** manquant
P41	ω – 01010000 – – 01001101 – – 01000011 – α	**4** manquant
P42	ω – 01010000 –  – 01001101 – – 01000011 – α	**Polymère désiré****
P43	ω – 00000000 – – 11111111 – – 00000000 – α	Unité **1** manquante

*Tableau III. 7 : Séquences polymères incorporant les dérivés **TISP3**.*
*Tous les polymères ont été clivé à 50°C/20h. **X** = **4, 5** et **10** font référence à **TISP3-X** de la Figure III. 21. *Légère impureté +275 Da ou 293 Da. **À noter la présence intense du polymère sans **T5**.*

Dans l'ensemble, tous les espaceurs ***TISP3*** identifiés comme étant de bons candidats pour jouer le rôle de marqueurs clivables ont pu être incorporés dans des chaînes polymères. Les marqueurs ***TISP3*-4** et ***TISP3*-10**, sans surprises, n'ont pas été incorporés efficacement dans la macromolécule numérique (**P37** et **P38**). Par ailleurs, l'influence de la séquence ne semble pas jouer un rôle dans ces mauvais couplages (**P40** et **P41**). Il est notable que ***TISP3*-5**, malgré une efficacité de couplage estimée à >99%, n'a pas bien été incorporé dans la séquence (**P39** et **P42**). Ceci peut s'expliquer par le vieillissement du macromonomère qui était déjà vieux de plusieurs mois. En outre, l'échantillon **P28** montré précédemment n'a montré aucun signe de mauvaise incorporation pour cet espaceur (Figure III. 14).

Par ailleurs, en respectant des conditions anhydres strictes, l'impureté +293 Da discutée précédemment n'est quasiment plus visible dans les essais présentés dans le Tableau III. 7. L'implication de l'eau hydratant le support et affectant donc sa déphosphorylation est, après toutes les optimisations effectuées (3.2), l'hypothèse la plus réaliste. De ce fait, l'utilisation de conditions strictement anhydres dans la construction de tels polymères est extrêmement importante, puisque ces dernières sont directement corrélées à l'efficacité de couplage, et de clivage.

Cette sous-partie (4) a démontré que l'efficacité de couplage peut varier selon les macromonomères ***TISP3*** utilisés. Notamment, le dérivé di-*tert*butyle (***TISP3*-10**) montrant un des plus faibles rendements. Afin de pallier ce problème, certainement lié au gène stérique de ce marqueur, un nouveau design ***TISP4*** a été conçu et sera discuté dans la prochaine sous-partie.

5. Synthèse du *TISP*4 : Recherches d'optimisation d'efficacité de couplage

5.1. Nouveau design : *TISP* 4

Afin d'améliorer l'efficacité de couplage de composés volumineux tel que le dérivé di-*tert*butyle (***TISP3*-10**, Figure III. 21), l'allongement de la chaîne propyle de l'amino-TEMPO permet d'éloigner le site phosphoramidite réactif des gênes stériques et pourrait ainsi optimiser le rendement de réaction (Schéma III. 3). Pour ce faire, il suffit de remplacer l'amino propanol de départ par un analogue comportant une chaîne carbonée plus longue, tel que l'amino hexanol (Schéma III. 3B). Les procédures permettant par la suite d'obtenir le marqueur clivable restent les mêmes que celles vues précédemment (Schéma III. 1).

Les rendements obtenus dans la construction de l'amino-TEMPO modifié avec une chaîne allongée sont meilleurs que son analogue avec une chaîne propyle (93 % contre 67 %). Ensuite, l'ATRA a également donné de meilleurs résultats en impliquant ce nouveau dérivé TEMPO grâce notamment à une meilleure séparation sur colonne chromatographique (72 %). Alors que ces résultats étaient prometteurs, l'ajout du groupe phosphoramidite dans la dernière étape s'est avérée moins efficace (10 - 20 %). Dans cette réaction, un sous-produit majoritaire correspondant au phosphoramidite hydrolysé a par ailleurs été identifié. En effet, les analyses RMN ^{31}P, ^{1}H ainsi que les mesures MS ont permis de proposer la structure du dérivé phosphoramidite, avec le groupe cyanoéthyle substitué par une fonction hydroxyle (***TISP4*-E1**, Figure III. 23). De plus, les tables RMN ^{31}P disponibles dans la littérature confirment cette proposition de structure puisque la région comprise entre ≈ 0 et 20 ppm indique l'existence d'espèces $(RO)_2P(OH)$, alors que les phosphoramidites recherchés se situent autour des -150 ppm.[384]

Alors que la purification sur colonne chromatographique a été réalisée avec de la DIPEA afin d'éviter les problèmes discutés précédemment pouvant être liés à la TEA (2.2), la base ne semble pas jouer un rôle dans ce rendement faible obtenu (10%). Aussi, cette purification pourrait donner de meilleurs rendements si elle est faite rapidement, compromettant la dégradation et/ou l'hydrolyse du produit sur la colonne (ce qui n'a pas été réalisé dans ce premier essai). En appliquant ces paramètres après une nouvelle réaction, et en utilisant cette fois-ci la TEA au lieu de la DIPEA, le rendement n'a été que très peu amélioré (20%). Alors que le produit a été isolé pure comme confirmé par RMN du phosphore ^{31}P et de l'hydrogène ^{1}H (Figure PE.3. 81 et Figure PE.3. 85), la mesure MS de ce même échantillon (quelques jours plus tard) démontre une impureté majoritaire liée à la présence de TEA (***TISP4*-E2,** Figure III. 23A). Effectivement, le spectre de masse enregistré en mode positif pour cet échantillon montre bien le composé recherché sous sa forme protonée, $[M+H]^+$ à m/z 1125.8 (en vert), validée par MS/MS (Figure III. 23A). Mais, les signaux majoritaires sont observés à m/z 989.6 (Δm –136.2 Da, en rouge) et m/z 1090.8 (Δm –35 Da, en violet).

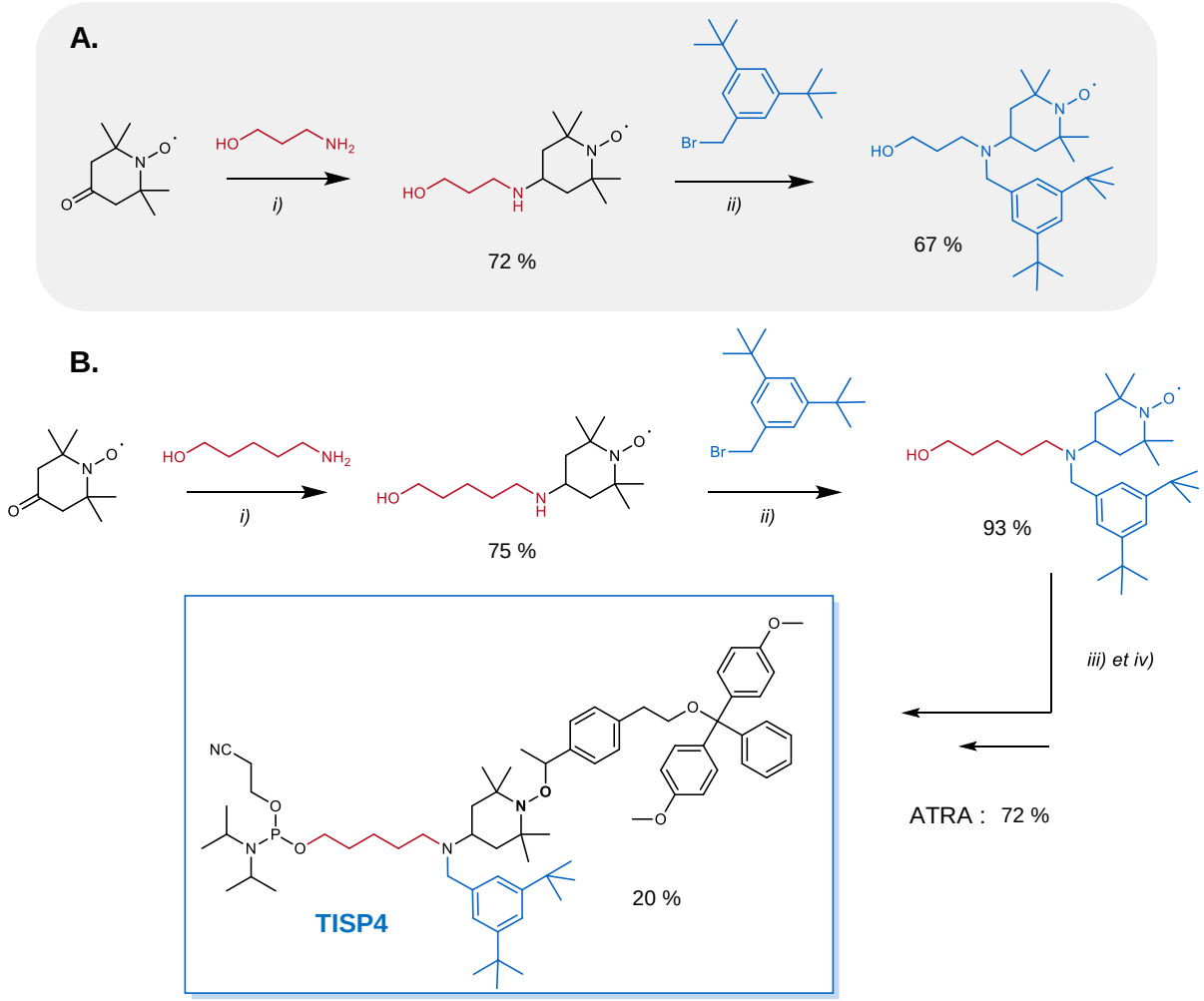
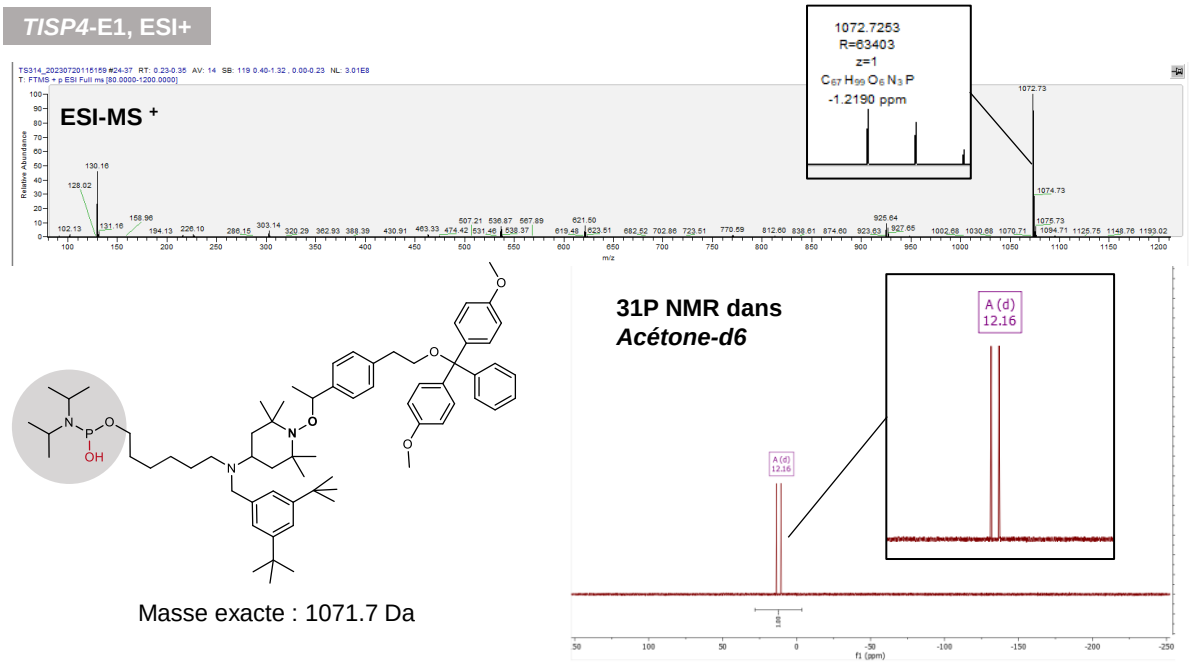


*Schéma III. 3 : Design et synthèse du **TISP4** comportant une chaîne alkyle rallongée.*
***A.** Stratégie **TISP3**. **B.** Stratégie **TISP4**. i) Triacétoxyborohydrure de sodium, AcOH, DCE anhydre, t.a., 17h ; ii) K_2CO_3, ACN, 83°C, 17h ; iii) ATRA ; iv) ajout du groupe phosphoramidite.*

Les données MS/MS (non montrées) indiquent que ces deux ions possèdent le groupement DMT (observé comme fragment à m/z 303.1) et que l'ion m/z 989.6 est un fragment de l'ion m/z 1090.8 (soit une perte de 101.2 Da). Ces observations confirment l'implication plausible de la TEA (102 Da) dans les défauts de masses identifiés.



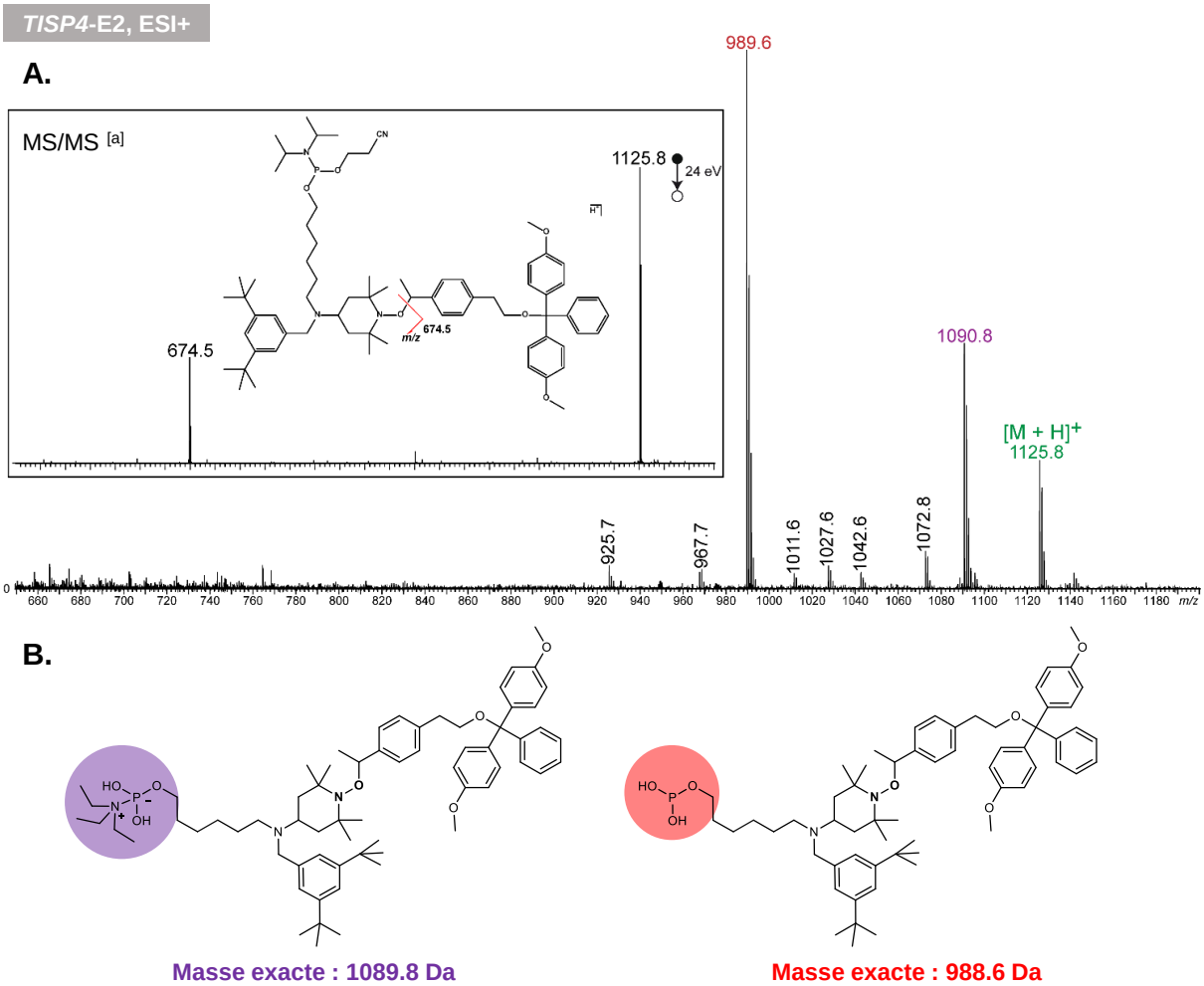
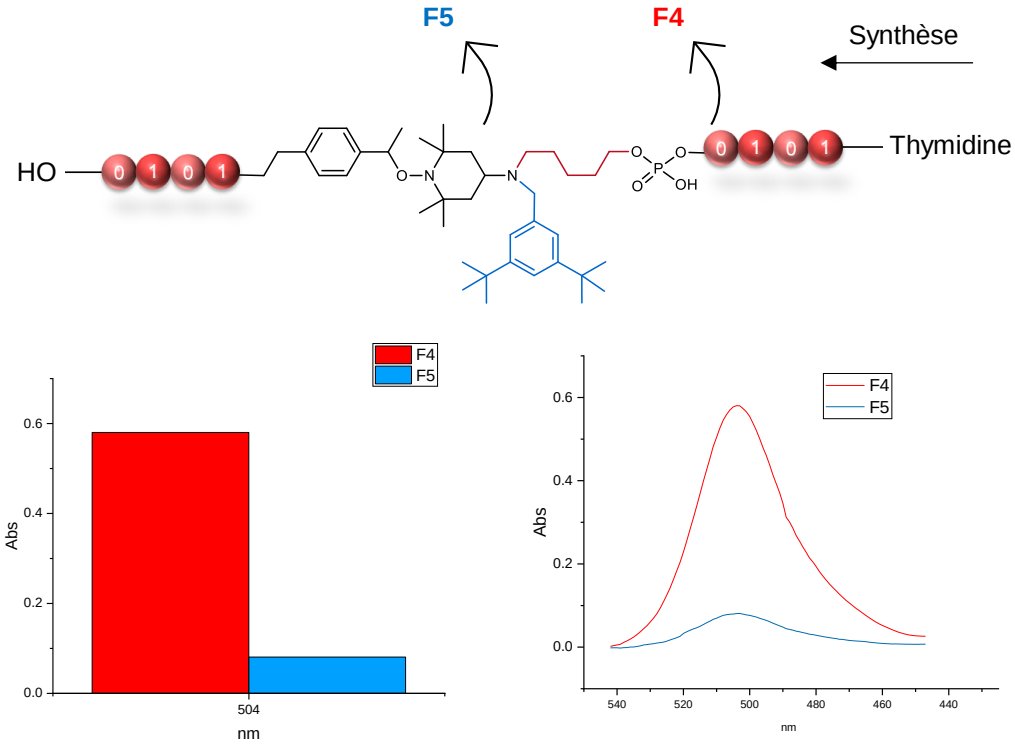


Figure III. 23 : Propositions de structures identifiées par analyses MS et RMN dans les échantillons ***TISP4-E1*** *et* ***TISP4-E2****.*
[a] MS/MS de l'ion à 1125.8 Da.

Même si les résultats en MS obtenus montrent des impuretés en quantité majoritaire, les RMN ^{31}P et ^{1}H semblent démontrer un produit extrêmement pur (Figure PE.3. 81 et Figure PE.3. 85). La préparation de l'échantillon en MS pouvant jouer un rôle sur la pureté et les éventuels sous-produits générés, l'efficacité de couplage de ***TISP4*** (échantillon ***TISP4*-E2**) a tout de même été étudiée, et sera discutée dans la prochaine sous-partie.

5.2. Efficacité de couplage du *TISP*4

Afin d'estimer l'efficacité de couplage du marqueur clivable ***TISP4***, une séquence modèle, telle que dans les essais effectués dans la partie 4, a été utilisée. La Figure III. 24 ci-dessous illustre les résultats obtenus pour ce test.



Efficacité de couplage **TISP4** : **14%**

*Figure III. 24 : Étude de l'efficacité de couplage de l'espaceur **TISP4**.*

L'efficacité de couplage pour le dérivé di-*tert*butyle (***TISP4***) n'a que peu été améliorée : 14% contre 9.5% pour son analogue propyle (entrée 11, Tableau III. 6). L'allongement de la chaîne n'a donc pas permis d'augmenter considérablement le rendement de ce couplage, signifiant que la gêne stérique n'est probablement pas la seule raison expliquant la faible efficacité de cette réaction. D'autres raisons peuvent expliquer ce défaut de réactivité, tels que les effets de solubilités observés avec le dérivé pyrènyle précédemment. Alors que les interactions π jouent assurément un rôle dans ce manque de solubilité, les groupes *tert*butyles peuvent également avoir une incidence sur l'homogénéité du soluté dans l'acétonitrile. Des tests vérifiant ces hypothèses sont en cours de développement, notamment en utilisant un mélange ACN/DCM 9/1. Ceci permettrait immanquablement de solubiliser de manière homogène, le dérivé ***TISP*4** étudié ici.

Dans le cas échéant, la stratégie ***TISP*4** ne semble pas être une bonne approche puisqu'elle n'améliore pas l'efficacité de couplage de l'analogue ***TISP3***, et augmente la masse du composé (3 carbones en plus). En outre, le design introduit avec les marqueurs de masses clivables a pour objectif de diminuer la quantité de monomère non informationnels, mais également d'effectuer une économie d'atome de la chaîne polymère formée. Ce dernier paramètre n'est pas forcément respecté en utilisant un design tel que le ***TISP4***, ce qui est démontré dans la prochaine sous-partie.

5.3. Design *TISP*4 ou *TISP*3 ? Gain de masse et économie d'atome

Alors que le design du marqueur clivable étudié dans cette thèse a pour but de réduire le nombre de monomère non informationnels dans la chaîne polymère, sa structure doit également être réfléchie en prenant compte sa masse moléculaire. En effet, l'objectif est également de réduire la masse totale de la macromolécule numérique formée, pour des raisons d'économies d'atomes et de facilité d'analyse.

Effectivement, les marqueurs clivables de type ***TISP*** (2, 3 et 4) permettent de s'affranchir des acides nucléiques utilisés auparavant dans notre groupe comme marqueur de masse.[27,29] De ce fait, il semble évident qu'en réduisant le nombre de monomères, la masse moléculaire totale du polymère diminuerait. Pourtant, cette hypothèse reste discutable, et particulièrement après comparaison avec l'ancienne stratégie. Lorsque la masse moléculaire de l'ancien espaceur ***RISC2*** additionnée à celle d'un marqueur de masse **M** est comparée à la masse moléculaire d'un dérivé ***TISP3***, la différence reste dans certains cas minimes (Figure III. 25).

C'est le cas lorsque les extrêmes sont comparés (**M** le plus léger comparé au ***TISP3*** le plus lourd Figure III. 25A). Paradoxalement, si le marqueur **M** le plus lourd est comparé au ***TISP3*** le plus léger, l'économie de masse est largement au profit de l'approche ***TISP3*** Figure III. 25B). Aussi, il est important de noter que l'utilisation d'un seul monomère dans cette stratégie réduit le nombre de couplage en chimie de phosphoramidite automatisée, et diminue donc considérablement la quantité de réactifs utilisée. À fur et à mesure que la séquence se rallonge, et donc que le nombre de blocs augmente, à plus cette stratégie aura de l'impact sur l'économie d'atomes et de réactifs.

Autrement, en confrontant le design ***TISP*4** aux deux stratégies à l'instant comparées, la substitution d'un groupe propyle par un groupe héxyle est moins bénéfique. Notamment, lorsque le ***TISP4***-di-CF_3 est comparé au marqueur **M** le plus léger, l'ancien design reste plus attractif en termes de masse moléculaire (Figure III. 25C).

Pour ces raisons, il semble que la meilleure stratégie reste celle du ***TISP3***. Cependant, des variantes pourraient permettre de construire une série de marqueurs étendue, tout en respectant l'économie du poids moléculaire recherchée. Particulièrement, en jouant sur la chaine propyle de l'amino-TEMPO, et des dérivés N-substitués moins lourds.

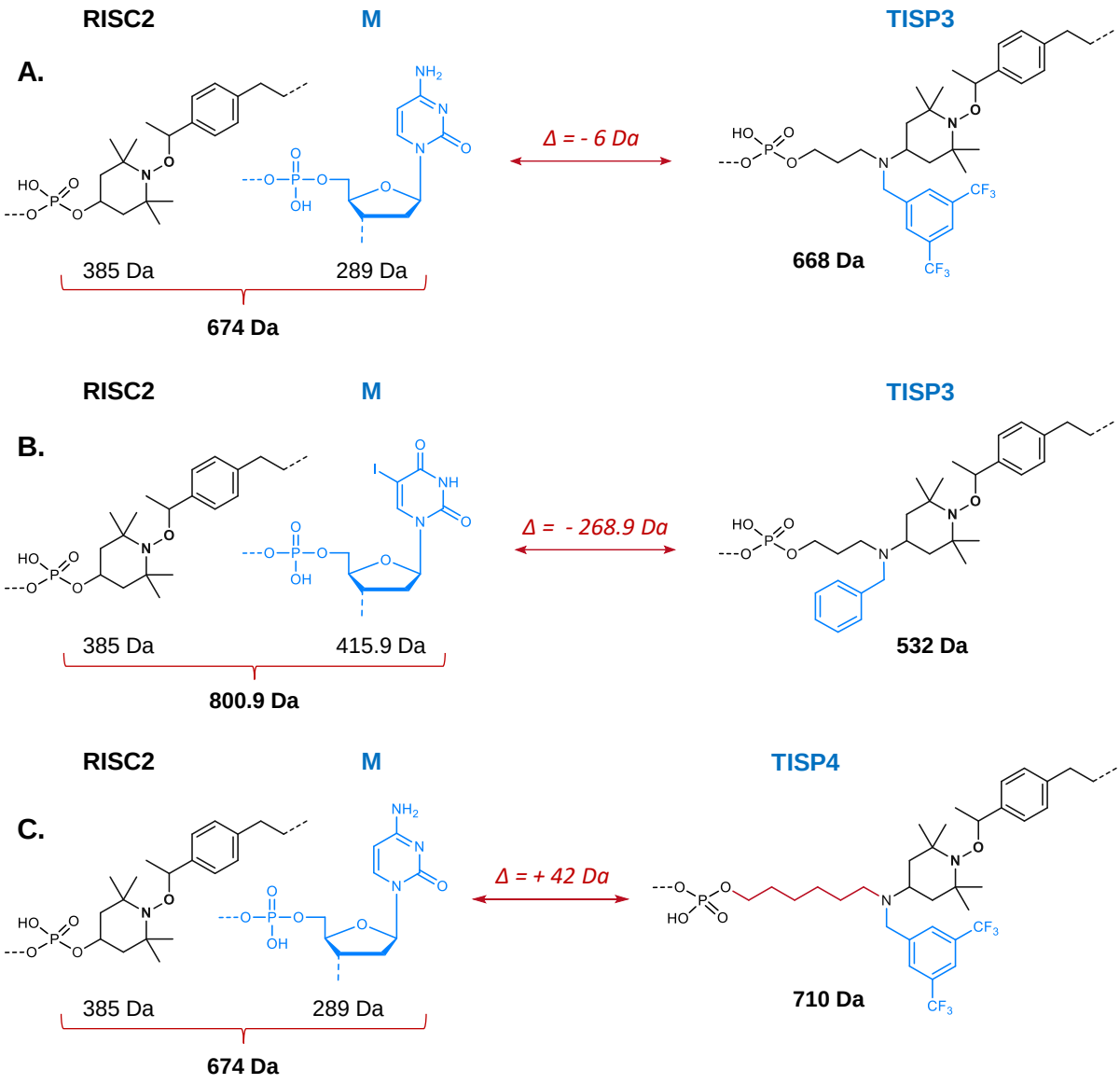


*Figure III. 25 : Comparaison des stratégies **RISC2**, **TISP3** et **TISP4** pour l'économie moléculaire massique.*
***A. M** le plus léger (désoxycytidine monophosphate) comparé au **TISP3** le plus lourd (**TISP3-8**). **B. M** le plus lourd (iodo-désoxyuridine monophosphate) comparé au **TISP3** le plus léger (**TISP3-2**). **C. M** le plus léger comparé au **TISP4** le plus lourd (non synthétisé dans cette thèse).*

6. Conclusion et perspectives

Ce chapitre a démontré que l'utilisation d'un nouveau design ***TISP3*** permet de répondre à l'objectif principal des travaux de cette thèse : élaborer un nouveau marqueur clivable pouvant être incorporé en chaîne polymère contrôlée pour un séquençage pseudo-MS^3 optimisé.

La synthèse d'un tel espaceur a pu être facilitée grâce au design ***RISC2*** déjà rapporté auparavant par notre groupe.[355] En outre, une chimiothèque de douze marqueurs clivables a pu être construite, dans l'ensemble obtenue avec des rendements corrects. Cependant, certains espaceurs n'ont pas été incorporés efficacement dans une chaîne

polymère. De ce fait, leur efficacité de couplage a pu être étudiée, donnant des résultats probants pour la majeure partie des ***TISP3***. Paradoxalement, certains espaceurs, comme les dérivés ***TISP3*-4**, **10**, **11** et **12** (Figure III. 21), n'ont pas été performant pour un couplage phosphoramidite dans les protocoles automatisés standards. Les effets stériques, électroniques ou de solubilité jouent certainement un rôle dans ces faibles rendements, mais des études supplémentaires restent nécessaires à de telles affirmations.

Par ailleurs, un nouveau support solide universel (Unysupport **3**, Figure III. 6) a été utilisé dans ces travaux. Ce dernier permet de s'affranchir d'une masse superflue en bout de chaîne (généralement une nucléobase de l'ADN), auparavant nécessaire pour labeliser le dernier bloc en spectrométrie de masse.[27] Cependant, l'application de ce support aux macromolécules numériques n'a dans un premier temps pas permis d'obtenir de séquences monodisperses, en cause d'une impureté majeure identifiée à + 293 Da. Ce défaut de masse a pu être attribué à la mauvaise déphosphorylation du support utilisé lors du processus de clivage. Après de nombreuses optimisations (températures, temps, solutions de clivage, purifications), une température élevée pendant le clivage a semblé efficace pour s'affranchir de cette impureté. Cependant, cette solution n'est pas applicable aux polymères développés dans cette thèse à cause des alcoxyamines fragiles incorporées dans la séquence. Finalement, un compromis a été trouvé en utilisant une température à 50°C pendant 20h.

Par ailleurs, après avoir constaté une efficacité de couplage affaiblie lorsque les monomères utilisés deviennent vieillissants (Figure III. 22), de nouveaux polymères ont été construit en respectant strictement les conditions anhydres (préparation le jour-même des solutions d'unités codantes et des marqueurs clivables). L'application de ces paramètres a permis, en incorporant les ***TISP3*** les plus performants, d'obtenir des polymères informationnels monodisperses. De plus, l'impureté liée au mauvais clivage du support n'a quasiment plus été observable dans les spectres de masse. L'élimination totale de l'eau dans les monomères phosphoramidites s'est donc révélée être extrêmement importante pour la construction de tels polymères.

Finalement, un nouveau design ***TISP4*** a été conçu avec l'objectif d'optimiser l'efficacité de couplage de marqueurs encombrés, tel que le di-*tert*butyle (***TISP3*-10**). Pour ce faire, la chaîne propyle de l'amino-TEMPO a été substituée par une chaîne héxyle plus longue (Schéma III. 3). Cependant, les résultats obtenus après analyse du monomère ***TISP4***, ont démontré des impuretés majeures correspondant au groupe phosphoramidite hydrolysé et/ou fonctionnalisé par la TEA. Ces défauts proviendraient certainement des conditions de colonnes utilisées, mais des tests supplémentaires restent encore nécessaires pour étudier ces hypothèses.

Ces derniers résultats ont notamment permis une discussion quant à la pertinence du design étudié dans cette thèse. En outre, l'objectif de ces travaux a été de diminuer le nombre de monomères non informationnels utilisés dans le polymère, mais aussi de réaliser une économie d'atome et de masse moléculaire. En comparant la stratégie ***RISC2* + marqueur de masse** à la stratégie **TISP3**, il est clair que cette dernière reste la plus

avantageuse (Figure III. 25). Cependant, dans le cas du ***TISP4***, l'ancien design reste parfois plus avantageux en fonction des substituants choisis.

Dans l'ensemble, les macromonomères ***TISP3*** élaborés dans ce chapitre établissent aujourd'hui le meilleur choix pour incorporer un marqueur de masse clivable dans une séquence polymère. Alors que certains espaceurs n'ont pas démontré de bons résultats, la chimiothèque obtenue n'est bien sûr pas limitée aux dérivés benzyles choisis, et une plus grande variété d'étiquette de masse peut être envisagée pour l'avenir. Effectivement, la chimie présente une large diversité de possibilité, tel que les différentes options vues au début du Chapitre 2, avec des couplages catalysés au palladium ou encore des réactions click.

Conclusion Générale & Perspectives

Les travaux effectués au cours de cette thèse ont permis d'élaborer de nouveaux marqueurs d'octets clivables pour la synthèse de poly(phosphodiesters) synthétiques numériques.

De tels polymères sont composés de monomères synthétiques codants pour 0 ou 1, avec des masses molaires distinctes, permettant de retrouver l'information encodée par spectrométrie de masse (Figure C&P. 1A). Leur synthèse peut être effectuée par la chimie de phosphoramidite automatisée, en respectant parfaitement l'enchainement des unités codantes. Afin de faciliter l'analyse de tels polymères, la flexibilité des macromolécules synthétiques permet d'étudier de nouveaux designs en adaptant leur structure à l'appareil analytique utilisé. En outre, il est possible d'insérer des liaisons fragiles alcoxyamines permettant de sélectivement fragmenter l'intersection des segments codés, et des marqueurs de masse permettant de retrouver la signature de chaque fragment sur le spectre MS. Cependant, l'ajout de ces monomères non informationnels réduit la densité d'information, et augmente la longueur de la chaîne polymère. De plus, ils impliquent des étapes supplémentaires de couplage, ce qui diminue le rendement global et augmente le nombre d'atomes de la macromolécule numérique. Pour ces raisons, l'objectif de cette thèse a été d'élaborer un nouvel espaceur clivable pour limiter le nombre de blocs constitutifs non codants. Le nouveau design ***TISP3*** décrit dans cette thèse permet efficacement de répondre à ce défi. Effectivement, en combinant une liaison alcoxyamine fragile jouant le rôle d'espaceur clivable, et un groupement chimique variable jouant le rôle de marqueur de masse, ce macromonomère ***TISP3*** permet d'économiser, par rapport au design ***RISC2*** précédemment décrit, l'ajout d'un monomère par segment codant.

Afin de développer ce nouveau design, plusieurs approches ont été étudiées. Dans le **Chapitre 2**, une première stratégie a été d'utiliser un alcyne (***TISP***) synthétisé de manière convergente grâce à une étape clef d'ATRA impliquant un amino-TEMPO modifié (**T1**-**T10**). Son schéma de synthèse est efficace et a permis d'obtenir le marqueur clivable avec de bons rendements. De plus, son incorporation en chaîne polymère a été démontrée. Cependant, la liaison alcyne compromet un séquençage pseudo-MS^3 optimal à cause des radicaux formés en CID. Particulièrement, un défaut de masse de +/- 1 Da lié à une abstraction d'hydrogène a été observé sur les signaux MS. Afin de pallier ce problème, la réduction de l'alcyne a été étudiée. Malgré les diverses tentatives d'hydrogénation, la présence de l'alcool en α de la liaison insaturée a empêché d'obtenir une réduction quantitative. Ceci est principalement dû à la formation d'un sous-produit correspondant à une déshydratation du composé désiré. Une stratégie de protection/déprotection impliquant le TBDMS a alors été testée, et a finalement permis d'obtenir le macromonomère saturé ***TISP2*** souhaité. Cependant, cette approche présente des défauts puisqu'elle utilise des étapes supplémentaires, et il est difficile d'obtenir le marqueur clivable pure, avec de bons rendements. Pour ces raisons, une nouvelle stratégie se basant sur le design ***RISC2*** préalablement décrite a été considérée pour construire un nouveau macromonomère ***TISP3***.

Le **Chapitre 3** a abordé cette nouvelle approche en impliquant un intermédiaire du ***RISC2*** et un amino-TEMPO modifié lors d'une ATRA, pour former le marqueur clivable ***TISP3*** désiré avec de bons rendements. Après une incorporation en chaîne polymère

concluante, et un séquençage pseudo-MS^3 aboutit, une chimiothèque de douze marqueurs clivables a pu être réalisée. L'efficacité de couplage de chacun de ces monomères a été testée dans des conditions standard de chimie de phosphoramidite automatisée (Figure C&P. 1B). En général, les macromonomères ***TISP3*** ont présenté de bons résultats de couplages. Cependant, certains dérivés benzyles comme le di-*tert*benzyle (***TISP3*-10**) et le mono-CF_3 (***TISP3*-4)** n'ont pas été incorporé efficacement en chaîne polymère. Ceci peut s'expliquer par un problème de solubilité, ou de gêne stérique. Pour tenter de résoudre ce problème, un nouveau design ***TISP4***, possédant le groupe phosphoramidite réactif plus éloigné du groupe fonctionnel benzyle, a été élaboré. Toutefois, ce dernier n'a que très peu permis d'optimiser l'efficacité de couplage du dérivé di-*tert*benzyle. Ceci illustre l'importance du choix du groupe variable pour le macromonomère ***TISP3*** (ou ***TISP4***).

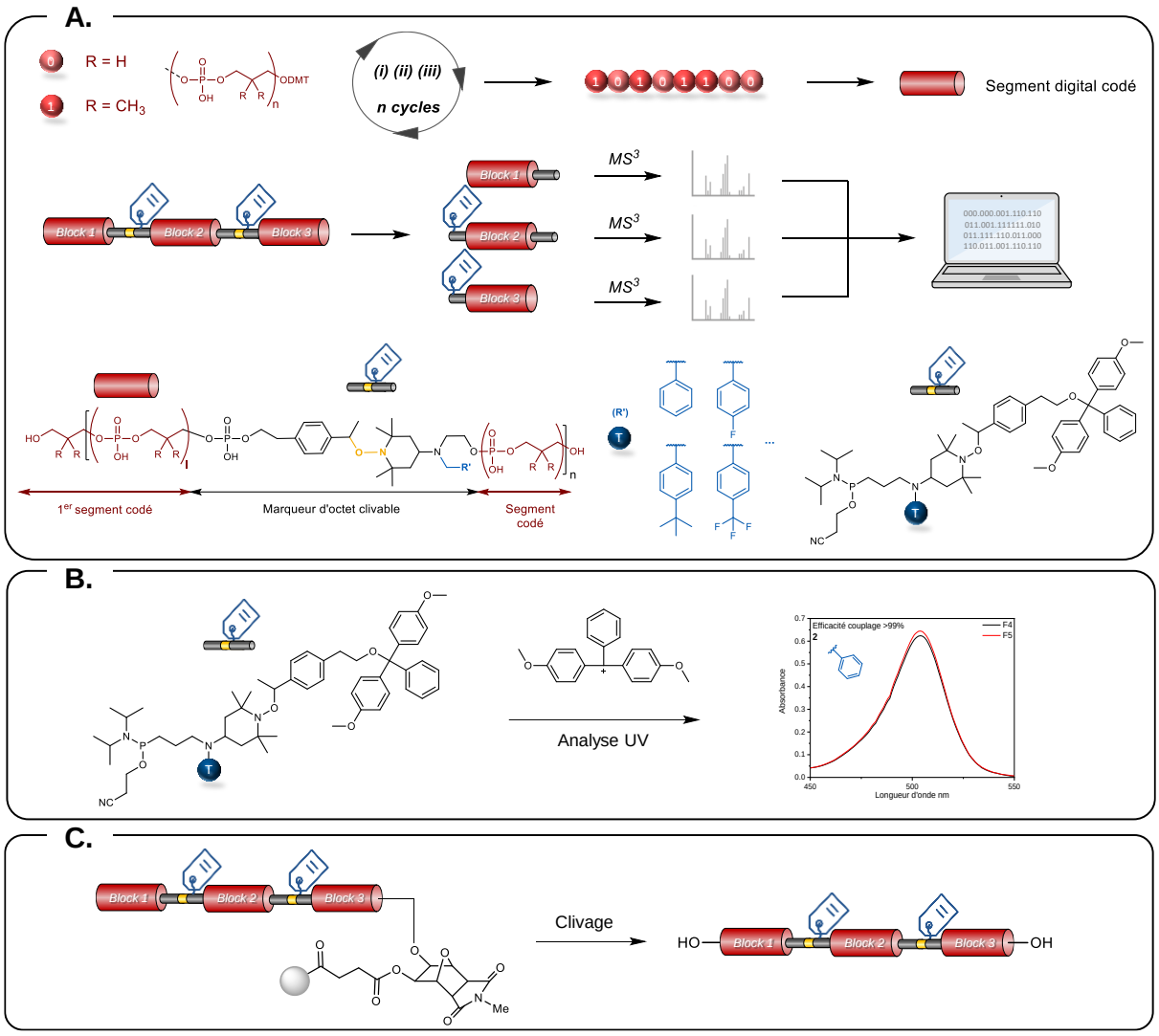


Figure C&P. 1 : Principaux faits marquants de cette thèse.
***A.** Concept général décrit au cours de cette thèse. **B.** Étude de l'efficacité de couplage des macromonomères **TISP3**. **C.** Utilisation d'un nouveau support solide universel.*

Finalement, le **Chapitre 3**, a décrit l'utilisation d'un nouveau support solide pour la préparation automatisée de poly(phosphodiesters) synthétiques. Les supports

traditionnellement employés dans notre groupe étaient des CPG fonctionnalisés de bases nucléiques A, T, G ou C. Ces deniers sont utiles lorsque la stratégie ***RISC2*** + **marqueur de masse** est employée pour labeliser le dernier segment d'information des polymères numériques. Cependant, la nouvelle approche ***TISP3*** ne nécessite plus de marqueurs indépendants. Par conséquent, l'utilisation d'un nouveau support solide universel (*Glen Unysupport*^{TM}) permet de s'affranchir du groupement nucléique, et de terminer en bout de chaîne avec une fonction alcool, sans groupes fonctionnels superflus (Figure C&P. 1C). Pour ces raisons, une série de polymères numériques comportant 3 blocs (3 octets), impliquant cette nouvelle résine et le ***TISP3***, a été réalisée. Enfin, diverses optimisations, notamment pour éliminer un sous-produit lié à un mauvais clivage du support, ont permis d'identifier un protocole performant pour obtenir des poly(phosphodiesters) numériques uniformes incorporant le nouveau macromonomère ***TISP3***.

Les travaux décrits dans cette thèse démontrent les possibilités infinies qu'offrent la chimie synthétique dans la construction de polymères numériques. Comparativement à l'ADN, les macromolécules non naturelles semblent posséder une immense marge d'optimisation pour la synthèse et la lecture de polymères informationnels. Les discussions faites dans le (Chapitre 1, 3.2) sur les différentes chimies utilisées ces 5-10 dernières années montrent bien la progression obtenue dans ce domaine, particulièrement sur un laps de temps très court. Par exemple, la chimie de phosphoramidite utilisée pour la synthèse d'oligonucléotides a été optimisée pendant plusieurs décennies, avant de pouvoir être automatisée et d'obtenir des séquences parfaitement uniformes. D'un autre côté, sur l'espace de quelques années, plusieurs méthodes de préparations de polymères informationnels synthétiques ont d'ores et déjà pu être automatisées (poly(phosphodiesters), polycarbamates, etc.). Cependant, ces dernières nécessitent encore du développement afin d'espérer un jour pouvoir écrire, stocker et lire de l'information comme il est possible de le faire aujourd'hui sur les ordinateurs.

Pour ce faire, diverses actions doivent être mises en œuvre pour répondre aux limites, aussi bien synthétiques, qu'analytiques. En ce qui concerne le premier point, les deux problèmes principaux sont la vitesse d'écriture, et la longueur maximale de la chaîne. En effet, il est impératif de trouver un moyen de synthèse rapide et efficace pour aspirer concurrencer les outils informatiques actuels. Alors que la chimie de phosphoramidite automatisée nécessite environ 5 minutes pour fabriquer 1 bit (1 cycle itératif), l'objectif d'écriture d'1 bit/seconde permettrait de se rapprocher de ce qu'il est possible de faire aujourd'hui sur les claviers numériques. Pour réaliser une telle prouesse à l'échelle moléculaire, de nouvelles réactions chimiques doivent être explorées. Par exemple, des réactions plus rapides, telles que les réactions photochimiques ou click, peuvent être développer pour la construction rapide, contrôlée, et automatisée de polymère informationnels.

Autrement, les longueurs de chaînes maximales atteintes nécessitent également de l'amélioration. En ce qui concerne la construction d'oligonucléotides par chimie de phosphoramidite automatisée, de nombreuses études évoquent des chaînes polymères allant de 100 à 200 nucléotides.[8,56,57,187,241–244] Mais il est difficile de trouver dans la

littérature la réelle limite d'un tel outil de synthèse. Dans le cas des poly(phosphodiesters) synthétiques numériques, la longueur maximale obtenue a été publiée en 2021 avec 124 unités, codant le portrait d'Antoine de Lavoisier (Figure I. 34).[29] Toutefois, cette limite n'a également jamais été explorée. Afin de repousser de tels records, des tests peuvent être effectués en laboratoire sur les robots synthétiseurs d'ADN, en construisant différentes chaînes polymères de longueurs différentes, en utilisant les monomères phosphoramidites synthétiques **0** et/ou **1** décrits précédemment. La taille de la chaîne serait alors croissante, jusqu'à ne plus être capable de l'identifier par MS ou RMN. En outre, cette dernière technique est un bon outil pour vérifier l'uniformité d'une séquence par DOSY, ou encore le nombre d'unité par intégration des signaux liés au phosphore. De telles études ont été réalisées au cours de cette thèse, et les résultats préliminaires obtenus sont encourageants. Une fois la limite atteinte, celle-ci peut être étendue à l'aide de plusieurs outils. Par exemple, les temps de couplages et les concentrations des monomères peuvent être augmentées, mais ceci n'améliorerait que très peu la longueur maximale de chaîne qui serait limitée par un rendement de synthèse tendant vers 0%. D'autres moyens existent pour trouver une solution à ce défi, notamment en s'inspirant de l'ADN. En construisant des polymères biohybrides incorporant des séquences synthétiques binaires et des séquences naturelles nucléotidiques, des systèmes supramoléculaires codés peuvent être construits. Ces derniers profiteraient de l'autoassemblage de brins complémentaires d'ADN pour étendre la chaîne polymère au-delà des limites permises par la synthèse phosphoramidite classique. Une preuve de concept a récemment été démontrée par notre groupe avec des segments synthétiques homogènes (Figure C&P. 2).[385]

Toutefois, l'optimisation de l'écriture d'information n'est pas le seul levier à actionner pour progresser dans le domaine du stockage numérique à l'échelle moléculaire. La lecture d'une grande quantité d'information de manière rapide et fiable est essentielle pour une telle application. Aujourd'hui, la MS/MS est l'outil analytique le plus utilisé pour décoder l'information contenue dans les polymères numériques. Cependant, l'utilisation de cette technique représente pour le moment un frein puisqu'elle ne permet pas une lecture en temps réel, et nécessite plusieurs expériences d'analyse.

Pour répondre à ces enjeux, le projet ANR impliquant les travaux de cette thèse ont également pour but d'employer l'IMS (introduit précédemment Figure I. 30). Des séquences polymères de 3 à 5 blocs incorporant les macromonomères ***TISP3*** ont été synthétisées pour tester cet outil analytique, et des résultats encourageants ont été obtenus. Les marqueurs insérés dans les différents octets permettent une séparation définie des fragments formés en source, avant de pouvoir à postériori dissocier les liaisons phosphates et retrouver l'information encodée dans chacun des blocs. Dans ce cas, l'ordre des marqueurs doit être choisi en fonction de leur comportement en IMS. De ce fait, des modèles théoriques basés sur les sections efficaces de collisions ioniques (CCS) de chaque marqueur ont été calculés.

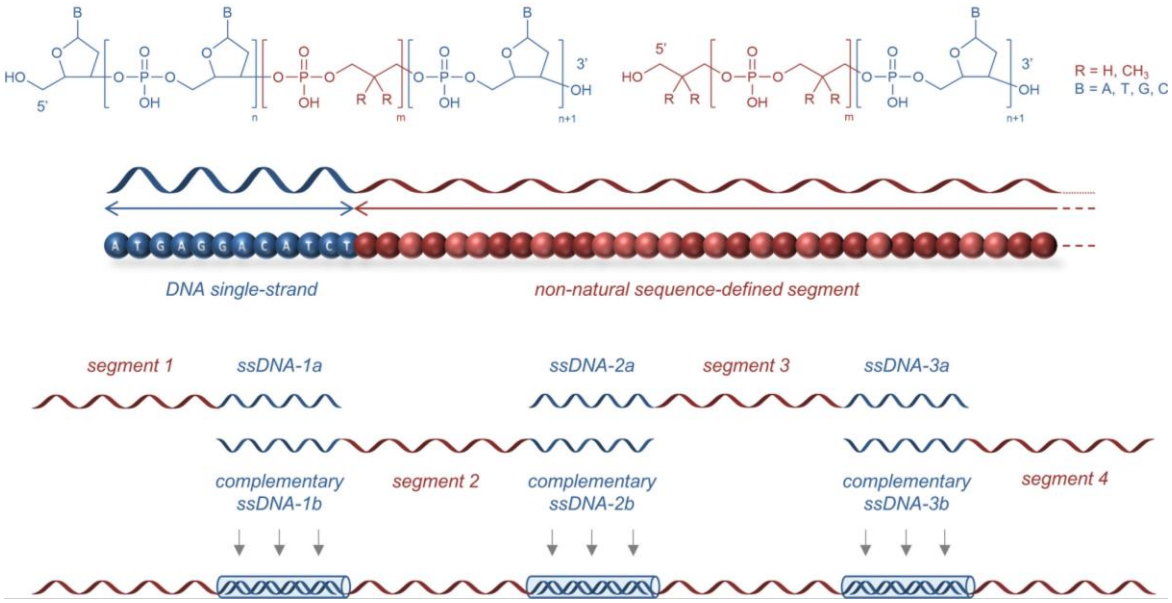


Figure C&P. 2 : Concept général étudié pour la synthèse de poly(phosphodiesters) biohybrides utilisés pour l'auto-assemblage guidé par l'ADN.
Reproduit avec permission de Mondal, T. et al. Large Sequence-Defined Supramolecules Obtained by the DNA-Guided Assembly of Biohybrid Poly(phosphodiester)s. Macromolecules 54, 3423–3429 (2021).[385] Copyright 2021 American Chemical Society.

Toutefois, une des contraintes observées dans le design ***TISP3*** est le nombre de marqueurs utilisable possible dans une même chaîne polymère (la partie 2.4 a présenté ces limites), ce qui restreint le nombre de blocs codants. Une des pistes à explorer pour le futur de ce projet serait l'utilisation d'autres outils de chimie de synthèse pour varier la structure du ***TISP3***. Par exemple, la chimie click CUAAC est utile pour modifier facilement un groupe fonctionnel en y incorporant un alcyne. Des couplages catalysés au palladium sont également intéressant pour apporter de la diversité efficacement. Particulièrement le Suzuki-Miyaura qui pourrait servir à varier le nombre de groupes phényles attachés au ***TISP3***. Cela permettrait notamment d'obtenir des structures possédant des encombrements stériques très différents, pouvant servir à une meilleure séparation en IMS.

Pourtant, d'autres techniques de lectures doivent être considérées pour les polymères synthétiques numériques, particulièrement pour une lecture en temps réel, bit par bit. Le séquençage par nanopores pourrait répondre à ce défi. La preuve de concept décrite précédemment (partie 3.3.2.2) n'a démontré la lecture de seulement quelques bits. En adaptant le design de la structure polymère aux pores de la protéine, une plus grande quantité d'information pourrait être analysée. À cet effet, des travaux préliminaires ont récemment été réalisé dans notre groupe, notamment avec des bouts de chaînes fonctionnalisé de biotine, possédant une affinité particulière avec l'entrée des nanopores de certaines protéines. Aussi, les monomères codants peuvent bien sûr être modifiés pour optimiser la vitesse de translocation à l'intérieur des pores, et permettre une meilleure lecture.

Dans l'ensemble, les recherches menées tout au long de cette thèse, ainsi que les exemples tirés de la littérature, mettent en évidence le fort potentiel de l'utilisation des

polymères synthétiques pour le stockage de données numériques. Comparativement aux polymères biologiques, ces macromolécules peuvent être modifiées à souhait et adaptées aux différents outils analytiques à disposition aujourd'hui. Les poly(phosphodiesters) sont particulièrement intéressants par leur méthode de synthèse robuste et rapide. Même si des optimisations de synthèse et de lecture restent à faire, ils sont une solution prometteuse pour le stockage numérique de demain.

Partie Expérimentale

Tables des matières pour la Partie Expérimentale

Liste des Figures pour la Partie Expérimentale 180

1. Matériels et réactifs 185

1.1. Réactifs pour la synthèse 185

1.2. Instruments et méthodes 186

1.2.1. Résonance magnétique nucléaire (RMN). 186

1.2.2. Spectroscopie UV. 186

1.2.3. Spectrométrie de masse (MS). 186

1.2.4. Chromatographie en phase liquide à haute performance (HPLC) 186

1.2.5. Résonnance paramagnétique électronique (RPE) 187

1.2.6. Langage ASCII utilisé 187

1.2.7. Chimie automatisée sur support solide ABI3900 et Expedite 8909 188

2. Procédures expérimentales 191

2.1. Synthèse des intermédiaires Amino-TEMPO T1-T11 191

2.2. Synthèse TISP et TISP2 194

2.3. Synthèse TISP3 201

2.4. Synthèse TISP4 210

3. Spectres RMN 212

3.1. Intermédiaires TISP et TISP2 212

3.1.1. RMN ^{31}P TISP et TISP2 225

3.2. TISP3 226

3.2.1. RMN ^{1}H et ^{13}C 226

3.2.2. RMN ^{31}P 248

3.2.3. RMN ^{19}F 249

3.3. TISP4 250

3.3.1. RMN ^{1}H et ^{13}C 250

3.3.2. RMN ^{31}P 253

4. Spectres MS 254

4.1. Amino-TEMPO T1-T13 254

4.2. $PdCl_2(PPh_3)_2$ 261

4.3. TISP et TISP2 261

4.4. TISP3 265
4.5. TISP4 270
4.5.1. T12-T13 270
4.5.2. f12 et Espaceur TISP4 271
4.6. Données complémentaires sur les polymères 273
4.6.1. TISP (Chapitre 2) 273
4.6.2. TISP3 (Chapitre 3) 277
5. Spectres UV (efficacité de couplage) 301
6. Chromatogrammes HPLC 303

Liste des Figures pour la Partie Expérimentale

Figure PE.1. 1 : Protocole standard1µmol du robot Expedite 8909 pour la position A. 189
Figure PE.1. 2 : Protocole standard 1µmol du robot Expedite 8909 pour la position G. 189
Figure PE.1. 3 : Protocole standard des positions 5-6-7 utilisé pour l'étude d'efficacité de couplage des Espaceurs **TISP3** et **TISP4**. 190
Figure PE.1. 4 : Protocole modifié pour un temps de couplage rallongé (jaune). 190
Figure PE.1. 5 : Protocole standard 1µmol utilisé pour la synthèse de polymère sur le synthétiseur ABI3900. 191

Figure PE.3. 1 : Spectre RMN ^{1}H du composé **a1**. 212
Figure PE.3. 2 : Spectre RMN ^{13}C du composé **a1.** 212
Figure PE.3. 3 : Spectre RMN ^{1}H du composé **b1.** 213
Figure PE.3. 4 : Spectre RMN ^{13}C du composé **b1.** 213
Figure PE.3. 5 : Spectre RMN ^{1}H du composé **c1**. 214
Figure PE.3. 6 : Spectre RMN ^{13}C du composé **c1**. 214
Figure PE.3. 7 : Spectre RMN ^{1}H du composé **d1**. 215
Figure PE.3. 8 : Spectre RMN ^{13}C du composé **d1**. 215
Figure PE.3. 9 : Spectre RMN ^{1}H du composé **e1**. 216
Figure PE.3. 10 : Spectre RMN ^{13}C du composé **e1**. 216
Figure PE.3. 11 : Spectre RMN ^{1}H du composé **TISP**. 217
Figure PE.3. 12 : Spectre RMN ^{13}C du composé **TISP**. 217
Figure PE.3. 13 : Spectre RMN ^{1}H du composé **e1-F**. 218
Figure PE.3. 14 : Spectre RMN ^{1}H du composé **TISP-F**. 218
Figure PE.3. 15 : Spectre RMN ^{13}C du composé **TISP-F**. 219
Figure PE.3. 16 : Spectre RMN ^{1}H du composé **a2**. 219
Figure PE.3. 17 : Spectre RMN ^{13}C du composé **a2**. 219
Figure PE.3. 18 : Spectre RMN ^{1}H du composé **b2**. 220
Figure PE.3. 19 : Spectre RMN ^{13}C du composé **b2**. 220
Figure PE.3. 20 : Spectre RMN ^{1}H du composé **c2**. *En gris, sous-produit de la réaction. 220
Figure PE.3. 21 : Spectre RMN ^{13}C du composé **c2**. 221
Figure PE.3. 22 : Spectre RMN ^{1}H du composé **d2**. 221
Figure PE.3. 23 : Spectre RMN ^{13}C du composé **d2**. Brut de la réduction alcyne à partir de **c1**, montrant l'impureté majeure identifiée en gris*. 221
Figure PE.3. 24 : Spectre RMN ^{1}H du sous-produit **SP** de la réaction menant à l'intermédiaire **d2**. 222
Figure PE.3. 25 : Spectre RMN ^{1}H du composé **e2**. 222
Figure PE.3. 26 : Spectre RMN ^{13}C du composé **e2**. 223
Figure PE.3. 27 : Spectre RMN ^{1}H du composé **g2**. 223
Figure PE.3. 28 : Spectre RMN ^{13}C du composé **g2**. 224

Figure PE.3. 29 : Spectre RMN ^{1}H du composé **TISP2**. 224
Figure PE.3. 30 : Spectre RMN ^{13}C du composé **TISP2**. 225
Figure PE.3. 31 : Spectres RMN ^{31}P des composés **TISP**, **TISP-F** et **TISP2**. 225
Figure PE.3. 32 : Spectre RMN ^{1}H du composé **f2**. 226
Figure PE.3. 33 : Spectre RMN ^{13}C du composé **f2**. 226
Figure PE.3. 34 : Spectre RMN ^{1}H de l'Espaceur **TISP3-2**. 227
Figure PE.3. 35 : Spectre RMN ^{13}C de l'Espaceur **TISP3-2**. 227
Figure PE.3. 36 : Spectre RMN ^{1}H du composé **f3**. 228
Figure PE.3. 37 : Spectre RMN ^{13}C du composé **f3**. 228
Figure PE.3. 38 : Spectre RMN ^{1}H de l'Espaceur **TISP3-3**. 229
Figure PE.3. 39 : Spectre RMN ^{13}C de l'Espaceur **TISP3-3**. 229
Figure PE.3. 40 : Spectre RMN ^{1}H du composé **f4**. 230
Figure PE.3. 41 : Spectre RMN ^{13}C du composé **f4**. 230
Figure PE.3. 42 : Spectre RMN ^{1}H de l'Espaceur **TISP3-4**. 231
Figure PE.3. 43 : Spectre RMN ^{13}C de l'Espaceur **TISP3-4**. 231
Figure PE.3. 44 : Spectre RMN ^{1}H du composé **f5**. 232
Figure PE.3. 45 : Spectre RMN ^{13}C du composé **f5**. 232
Figure PE.3. 46 : Spectre RMN ^{1}H de l'Espaceur **TISP3-5**. 233
Figure PE.3. 47 : Spectre RMN ^{13}C de l'Espaceur **TISP3-5**. 233
Figure PE.3. 48 : Spectre RMN ^{1}H du composé **f6**. 234
Figure PE.3. 49 : Spectre RMN ^{13}C du composé **f6**. 234
Figure PE.3. 50 : Spectre RMN ^{1}H de l'Espaceur **TISP3-6**. 235
Figure PE.3. 51 : Spectre RMN ^{13}C de l'Espaceur **TISP3-6**. 235
Figure PE.3. 52 : Spectre RMN ^{1}H du composé **f7**. 236
Figure PE.3. 53 : Spectre RMN ^{13}C du composé **f7**. 236
Figure PE.3. 54 : Spectre RMN ^{1}H de l'Espaceur **TISP3-7**. 237
Figure PE.3. 55 : Spectre RMN ^{13}C de l'Espaceur **TISP3-7**. 237
Figure PE.3. 56 : Spectre RMN ^{1}H du composé **f8**. 238
Figure PE.3. 57 : Spectre RMN ^{13}C du composé **f8**. 238
Figure PE.3. 58 : Spectre RMN ^{1}H de l'Espaceur **TISP3-8**. 239
Figure PE.3. 59 : Spectre RMN ^{13}C de l'Espaceur **TISP3-8**. 239
Figure PE.3. 60 : Spectre RMN ^{1}H du composé **f9**. 240
Figure PE.3. 61 : Spectre RMN ^{13}C du composé **f9**. 240
Figure PE.3. 62 : Spectre RMN ^{1}H de l'Espaceur **TISP3-9**. 241
Figure PE.3. 63 : Spectre RMN ^{13}C de l'Espaceur **TISP3-9**. 241
Figure PE.3. 64 : Spectre RMN ^{1}H du composé **f10**. 242
Figure PE.3. 65 : Spectre RMN ^{13}C du composé **f10**. 242
Figure PE.3. 66 : Spectre RMN ^{1}H de l'Espaceur **TISP3-10**. 243
Figure PE.3. 67 : Spectre RMN ^{13}C de l'Espaceur **TISP3-10**. 243
Figure PE.3. 68 : Spectre RMN ^{1}H du composé **f11**. 244
Figure PE.3. 69 : Spectre RMN ^{13}C du composé **f11**. 244
Figure PE.3. 70 : Spectre RMN ^{1}H de l'Espaceur **TISP3-11**. 245
Figure PE.3. 71 : Spectre RMN ^{13}C de l'Espaceur **TISP3-11**. 245
Figure PE.3. 72 : Spectre RMN ^{1}H du composé **f12**. 246

Figure PE.3. 73 : Spectre RMN ^{13}C du composé **f12**. 246
Figure PE.3. 74 : Spectre RMN ^{1}H de l'Espaceur **TISP3-12**. 247
Figure PE.3. 75 : Spectre RMN ^{13}C de l'Espaceur **TISP3-12**. 247
Figure PE.3. 76 : Spectres RMN ^{31}P des Espaceurs **TISP3-2-10**. 248
Figure PE.3. 77 : Spectres RMN ^{31}P des Espaceurs **TISP3-11-12**. 249
Figure PE.3. 78 : Spectres RMN ^{19}F des Espaceurs **TISP3-4**, **TISP3-8** et **TISP3-11**. . 249
Figure PE.3. 79 : Spectre RMN ^{1}H du composé **f13**. 250
Figure PE.3. 80 : Spectre RMN ^{13}C du composé **f13**. 250
Figure PE.3. 81 : Spectre RMN ^{1}H de l'Espaceur **TISP4**. 251
Figure PE.3. 82 : Spectre RMN ^{13}C de l'Espaceur **TISP4**. 251
Figure PE.3. 83 : Spectre RMN ^{1}H du **sous-produit phosphoramidite** identifié lors de la synthèse de **TISP4**. 252
Figure PE.3. 84 : Spectre RMN ^{13}C du **sous-produit phosphoramidite** identifié lors de la synthèse de **TISP4**. 252
Figure PE.3. 85 : Spectres RMN ^{31}P du composé **TISP4** et de son sous-produit. 253

Figure PE.4. 1 : Spectres MS du composé **T1**. 254
Figure PE.4. 2 : Spectres MS du composé **T2**. 255
Figure PE.4. 3 : Spectres MS du composé **T3.** 255
Figure PE.4. 4 : Spectres MS du composé **T4**. 256
Figure PE.4. 5 : Spectres MS du composé **T5**. 256
Figure PE.4. 6 : Spectres MS du composé **T6**. 257
Figure PE.4. 7 : Spectres MS du composé **T7**. 257
Figure PE.4. 8 : Spectres MS du composé **T8**. 258
Figure PE.4. 9 : Spectres MS du composé **T9**. 259
Figure PE.4. 10 : Spectres MS du composé **T10**. 260
Figure PE.4. 11 : Spectres MS du composé **T11**. 260
Figure PE.4. 12 : Spectres MS du catalyseur $PdCl_2(PPh_3)_2$ utilisé pour la synthèse des intermédiaires **c1** et **b2**. 261
Figure PE.4. 13 : Spectres MS de l'Espaceur **TISP**. 262
Figure PE.4. 14 : Spectres MS de l'Espaceur **TISP-F**. 262
Figure PE.4. 15 : Spectres MS de l'Espaceur **TISP2**. 263
Figure PE.4. 16 : Spectres ESI-MS/MS des impuretés observées dans la synthèse de l'Espaceur **TISP2**. 263
Figure PE.4. 17 : Spectres MS de l'Espaceur **TISP3-2**. 265
Figure PE.4. 18 : Spectres MS de l'Espaceur **TISP3-3**. 265
Figure PE.4. 19 : Spectres MS de l'Espaceur **TISP3-4**. 266
Figure PE.4. 20 : Spectres MS de l'Espaceur **TISP3-5**. 266
Figure PE.4. 21 : Spectres MS de l'Espaceur **TISP3-6**. 267
Figure PE.4. 22 : Spectres MS de l'Espaceur **TISP3-7**. 267
Figure PE.4. 23 : Spectres MS de l'Espaceur **TISP3-8**. 268
Figure PE.4. 24 : Spectres MS de l'Espaceur **TISP3-9**. 268
Figure PE.4. 25 : Spectres MS de l'Espaceur **TISP3-10**. 269

Figure PE.4. 26 : Spectres MS de l'Espaceur **TISP3-11**. 269
Figure PE.4. 27 : Spectres MS de l'Espaceur **TISP3-12**. 270
Figure PE.4. 28 : Spectres MS du composé **T12**. 270
Figure PE.4. 29 : Spectres MS du composé **T13**. 271
Figure PE.4. 30 : Spectres MS du composé **f12**. 271
Figure PE.4. 31 : Spectres MS de l'Espaceur **TISP4**. 272
Figure PE.4. 32 : Spectres MS de **P1**. 273
Figure PE.4. 33 : Spectres MS/MS des impuretés observées dans **P1** (Figure PE.4. 32). 274
Figure PE.4. 34 : Spectres MS/MS du polymère ciblé et des impuretés observées dans l'échantillons **P2**. 275
Figure PE.4. 35 : Spectres MS/MS des impuretés observées dans **P3**. 276
Figure PE.4. 36 : Spectres MS/MS du polymère ciblé et des impuretés observées dans l'échantillon **P7**. 277
Figure PE.4. 37 : Spectres MS du polymère ciblé des impuretés observées dans l'échantillon **P8**. 278
Figure PE.4. 38 : Spectres MS du polymère ciblé et de l'impureté X observée dans l'échantillon **P9**. 279
Figure PE.4. 39 : Spectres MS de l'échantillon **P10**. 280
Figure PE.4. 40 : Spectres MS/MS des impuretés observées dans l'échantillon **P10** (Figure PE.4. 39). 281
Figure PE.4. 41 : Spectres MS de l'échantillon **P11**. 282
Figure PE.4. 42 : Spectres MS/MS des impuretés observées dans l'échantillon **P11** (Figure PE.4. 41). 283
Figure PE.4. 43 : Spectres MS/MS des différentes espèces observées dans les échantillons **P12** et **P13**. 284
Figure PE.4. 44 : Spectres ESI-MS en mode négatif des échantillons **P15**, **P16** et **P17**. Les pics annotés en gris (annotation non exhaustive) correspondent à des fragments formés en source. 285
Figure PE.4. 45 : Spectres ESI(–)-MS des échantillons **P18** et **P19**. Les blocs libérés en source sont annotés en noir et les pics annotés en gris sont des fragments de très faible état de charge. 286
Figure PE.4. 46 : Spectres ESI(–)-MS de l'échantillon **P21**. Les blocs libérés en source sont annotés en gris et les pics annotés en noir sont des fragments de très faible état de charge. 287
Figure PE.4. 47 : Spectres ESI(–)-MS de l'échantillon **P24**. De nombreux échanges H/Na (annotés par un astérisque) sont observés. L'échantillon présente trop de sel de TCA pour être ionisé correctement. 287
Figure PE.4. 48 : Spectres MS de l'échantillon **P25**. 288
Figure PE.4. 49 : Spectres ESI(–)-MS de l'échantillon **P26**. 289
Figure PE.4. 50 : Spectre MS/MS de l'ion $[P - 6H]^{6-}$ à m/z 780.7 dans l'échantillon **P28**. 289
Figure PE.4. 51 : Spectre MS/MS de l'ion $[P - 6H]^{6-}$ à m/z 780.7 dans l'échantillon **P29**. 290

Figure PE.4. 52 : Spectre MS/MS de l'ion $[P - 6H]^{6-}$ à m/z 799.2 dans l'échantillon **P31**. 290
Figure PE.4. 53 : Spectres MS de l'échantillon **P33**..... 291
Figure PE.4. 54 : Spectres MS de l'échantillon **P34**..... 292
Figure PE.4. 55 : Séquençage pseudo-MS^3 du polymère ciblé dans l'échantillon **P34** (Figure PE.4. 54). 293
Figure PE.4. 56 : Spectres MS de l'échantillon **P35**..... 294
Figure PE.4. 57 : Spectres MS de l'échantillon **P36**..... 295
Figure PE.4. 58 : Spectre ESI(–)-MS de l'échantillon **P37**. 296
Figure PE.4. 59 : Spectre ESI(–)-MS de l'échantillon **P38**. 296
Figure PE.4. 60 : Spectre ESI(–)-MS de l'échantillon **P39**. 297
Figure PE.4. 61 : Spectre ESI(–)-MS de l'échantillon **P40**. 297
Figure PE.4. 62 : Spectres MS de l'échantillon **P41**..... 298
Figure PE.4. 63 : Spectres MS de l'échantillon **P42**..... 299
Figure PE.4. 64 : Spectres MS de l'échantillon **P43**..... 300

Figure.PE.5. 1 : Spectres UV enregistrés dans du CH_2Cl_2 pour les fractions de DMT déprotégées lors de la synthèse d'oligomères incorporant **TISP3**. 301
Figure.PE.5. 2 : Spectres UV enregistrés dans du CH_2Cl_2 pour les fractions de DMT déprotégées lors de la synthèse d'oligomères **TISP3**. 302

Figure PE.6. 1 : Spectres HPLC des polymères **P28** et **P29** à différentes températures de clivage. 303
Figure PE.6. 2 : HPLC des polymères **P33**, **P34**, **P35**, **P36**, **P38** et **P42**. 304

1. Matériels et réactifs

1.1. Réactifs pour la synthèse

3-Amino-1-propanol (>99,0%, TCI), 6-Amino-1-hexanol (>97.0%, TCI), 4-oxo-2,2,6,6-tétraméthylpipéridine 1-oxyde (>95,0%, TCI), sodium triacétoxyborohydrure (>97,0%, Sigma Aldrich), bromure de benzyle (98,0%, Sigma Aldrich), bromure de 4-fluorobenzyle (>97%, TCI), bromure de 4-(trifluorométhyl)benzyle (>98%, TCI), 4-bromométhylbiphényle (>98%, TCI), 2-(bromométhyl)naphthalène (96%, Alfa Aesar), bromure de 2,3,4,5,6-pentafluorobenzyl (99%, Sigma Aldrich), 3-(bromométhyl)pyrène (95%, ABCR), 3,5-bis(trifluorométhyl)benzyle bromure (>96%, TCI), bromure de 4-tert-butylbenzyle (97%, Sigma Aldrich), bromure de 3,5-di-tert-butylbenzyle (>97%, Sigma Aldrich), Cu0 (poudre, 99,5%, Alfa Aesar), CuBr (I) (poudre, 98%, Sigma Aldrich), N,N,N',N",N"-pentaméthyldiéthylènetriamine (PMDETA, 99%, Sigma Aldrich), O-2-cyanoéthyl-N,N-diisopropylchlorophosphoramidite (97%, ABCR), borohydrure de sodium (NaBH4, 99%, Sigma Aldrich), alcool propargylique (2-propyn-1-ol, 99%, Sigma Aldrich), CuI (I) (poudre, 98%, Sigma Aldrich), chlorure de diméthoxytrityle (DMT-Cl, 97%, Atarax/ChemGenes), platine sur charbon (Pt/C, charge de 10 %, Sigma Aldrich), carbonate de potassium anhydre (K2CO3, 99,5%, Fisher Chemical), sulfate de sodium anhydre (Na2SO4, 99,5%, Fisher Chemical), bicarbonate de sodium (NaHCO3, 99%, Fisher Chemical), 1,2-dichloroéthane anhydre (>99,0%, Sigma Aldrich), acétonitrile anhydre (ACN, 99,8%, Sigma Aldrich), dichlorométhane anhydre (DCM, >99,8%, stabilisé avec 50 ppm d'amylène, Sigma Aldrich), tétrahydrofurane anhydre (THF, >99,9%, exempt d'inhibiteurs, Sigma Aldrich), n-hexane (95%, Carlo Erba), acétate d'éthyle (EtOAc, 99,8%, Carlo Erba), éthanol (EtOH, 99,9%, Carlo Erba), DCM (99,8%, stabilisé avec de l'éthanol ou de l'amylène, Carlo Erba), cyclohexane (99,8%, Carlo Erba), éther diéthylique (Et_2O, 99%, pur-stabilisé avec 250-400 ppm de 2,6-di-tert-butyl-4-méthylphénol (BHT), Carlo Erba), THF (99,5%, pur-stabilisé avec 250-400 ppm de BHT, Carlo Erba), pyridine anhydre (99,8% en Sure/Seal™, Sigma Aldrich), N,N-diisopropyléthylamine (DIPEA, 99%, TCI), triéthylamine (Et_3N, >97%, ACROS Organics), acide acétique glaciale (AcOH, 99,7%, Fisher Chemical), HCl concentré (37% dans l'eau), gel de silice pour colonne de chromatographie (de haute pureté, taille de pore 60 Å, 230 - 400 mesh, SiO2, Fluka), chlorure de sodium (pour solution saline, Cedo). Tous ces solvants ont été utilisés tels quels. Le $PdCl_2(PPh_3)_2$ a été synthétisé en adaptant une procédure déjà décrite.[386] Pour la synthèse automatisée de phosphoramidites : acétonitrile anhydre (diluant de phosphoramidite et lavages à sec, ChemGenes), acétonitrile (≥99,9%, lavages, Roth), réactif d'activation (0,25 M de 5-(éthylthio)-1H-tétrazole dans le MeCN, ChemGenes), Cap A (anhydride acétique/pyridine/THF, ChemGenes), Cap B (10 % de N-méthylimidazole dans le THF, ChemGenes), réactif de déparage du DMT (3% TCA dans le DCM, Roth), support Glen UnySupport^{tm} PS (200 nmol en cartouche, Glen Research), dT-CPG 1000 (1 µmol en cartouche, taille de pore de 1000 Å, Glen Research), cartouche de purification d'ADN Glen-Pak (de 10 nmol à 1,0 µmol, Glen Research) et solution d'oxydation (0,02 M d'iode/pyridine/H_2O/THF, ChemGenes) ont été utilisés tels quels. Les monomères de phosphoramidites **0** et **1** ont été préparés suivant une procédure précédemment

décrite.[296] Tous les monomères de phosphoramidites ont été conservés au congélateur à -18°C.

1.2. Instruments et méthodes

1.2.1. Résonance magnétique nucléaire (RMN).

Les données RMN ont été enregistrées sur un spectromètre Bruker Avance 400 équipé d'un aimant Ultrashield. Les déplacements chimiques (δ) sont rapportés en parties par million (ppm) par rapport au signal résiduel du solvant (RMN 1H, $CDCl_3$: δ = 7,26 ppm ; RMN ^{13}C, $CDCl_3$: δ = 77,16 ppm ; RMN ^{1}H, $(CD_3)_2CO$: δ = 2,05 ppm ; RMN 13C, $(CD_3)_2CO$: δ = 29,84 ppm). Les spectres RMN ^{1}H ont été enregistrés à 400,13 MHz, les spectres RMN ^{13}C à 100,62 MHz, les spectres RMN ^{31}P à 161,96 MHz et les spectres RMN ^{19}F à 376 MHz. Les RMN ^{13}C, ^{19}F et ^{31}P ont été réalisées en mode découplé. Les attributions des signaux RMN ont été réalisées grâce à des expériences préliminaires en 2D non présentées dans ce manuscrit (COSY et HSQC). Les signaux issus d'échanges de protons avec les solvants deutérés utilisés ont été notés *$CDCl_3$* et *$(CD_3)_2CO$* sur les différents spectres.

1.2.2. Spectroscopie UV.

Les données UV ont été enregistrées sur un spectrophotomètre Jasco V-670. Les fractions de DMT déprotégées ont été collectées dans une fiole volumétrique de 5 ml, complétées à 5 ml avec du dichlorométhane. Un volume de 250 µL a été transféré dans une autre fiole volumétrique de 5 ml et complété à 5 ml avec une solution d'acide trichloroacétique à 3% dans du dichlorométhane (V/V). L'absorbance a été mesurée à 504 nm en utilisant du dichlorométhane comme solvant de référence.[25,381]

1.2.3. Spectrométrie de masse (MS).

Les données MS présentées dans ce rapport ont été effectuées à Marseille à l'Institut de Chimie Radicalaire (ICR) dans l'équipe du Professeur Laurence Charles. Tous les échantillons ont été dissous dans une solution H_2O/ACN (50:50, v/v) contenant 0,1% d'acide pourmique, dilués davantage dans une solution méthanolique d'acétate d'ammonium (3 mM) et introduits à un débit de 10 µL/min dans la source d'ionisation par électrospray (ESI) fonctionnant en mode négatif (tension capillaire : -2,27 kV) sous un flux de gaz de désolvatation (N_2) de 100 L/h chauffé à 35°C. La tension de cône était de -10 V ou -20 V pour les expériences MS et MS^2, tandis qu'elle était ajustée entre -30 V et -50 V pour induire une fragmentation en source dans les expériences MS^3. La dissociation induite par collision (CID) a été effectuée dans le dispositif de piège à ions en utilisant de l'argon comme gaz de collision après la sélection de l'ion précurseur primaire (en MS^2) ou secondaire (en pseudo-MS^3) dans le spectromètre de masse à quadrupôle. Les analyses des données ont été réalisées à l'aide du programme MassLynx 4.1 fourni par Waters.

1.2.4. Chromatographie en phase liquide à haute performance (HPLC)

La HPLC à échange d'anions a été utilisée pour vérifier la pureté des polymères sur un système HPLC analytique 1220 Infinity II d'Agilent Technologies. L'appareil était équipé d'une colonne pour les oligonucléotides PA100 Thermo Scientific™ Dionex™ DNAPac™ (pores de 13 µm, 4 x 250 mm) et d'un détecteur UV (260 et 280 nm). La phase A a été

préparée avec 10% d'acétonitrile et 20% de NH_3 à 2 M dans de l'eau MilliQ, et la phase B avec du NaCl à 2,5 M dans de l'eau MilliQ. L'élution a été réalisée en appliquant un gradient linéaire de la phase B (de 5% à 30%) sur 25 minutes à un débit de 1 mL/min. Les données HPLC ont été traitées et normalisées à l'aide d'Origin 2018.

1.2.5. Résonnance paramagnétique électronique (RPE)

Les expériences RPE ont été réalisées à l'Institut de Chimie Radicalaire de Marseille par Jean-Louis CLEMENT. Le facteur-g de Landé et les constantes de couplage (en mT) ont été obtenus à partir d'expériences réalisées sur un spectromètre Bruker Elexsys E500 équipé d'une cavité RPE Bruker SHQ. Les échantillons de nitroxide **T1** à **T10** ont été dilués dans du tert-butylbenzène (600 µL à une concentration de 10-4 M). Des tubes RMN en verre ont été introduits dans la cavité RPE et les spectres ont été enregistrés (temps de balayage 40 s, largeur de balayage 100 G, puissance 10 mW, amplitude de modulation 1,0 G, temps de conversion 40 ms, fréquence de modulation 100 kHz). Le spectre RPE du (2,2,6,6-tétraméthylpipéridin-1-yl)oxyl (TEMPO) a été enregistré avant les échantillons de nitroxides et correspond à la littérature avec un facteur g de 2.00608.

1.2.6. Langage ASCII utilisé

Pour toutes les synthèses effectuées dans cette thèse, le langage ASCII a été utilisé pour encoder de l'information numérique à l'échelle moléculaire. Ci-dessous, un Tableau représentant les détails de cet alphabet.

Caractère	Code binaire	Caractère	Code binaire
A	01000001	**a**	01100001
B	01000010	**b**	01100010
C	01000011	**c**	01100011
D	01000100	**d**	01100100
E	01000101	**e**	01100101
F	01000110	**f**	01100110
G	01000111	**g**	01100111
H	01001000	**h**	01101000
I	01001001	**i**	01101001
J	01001010	**j**	01101010
K	01001011	**k**	01101011
L	01001100	**l**	01101100
M	01001101	**m**	01101101
N	01001110	**n**	01101110
O	01001111	**o**	01101111
P	01010000	**p**	01110000
Q	01010001	**q**	01110001
R	01010010	**r**	01110010
S	01010011	**s**	01110011
T	01010100	**t**	01110100
U	01010101	**u**	01110101
V	01010110	**v**	01110110
W	01010111	**w**	01110111

X	01011000	**x**	01111000
Y	01011001	**y**	01111001
Z	01011010	**z**	01111010

Tableau PE.1. 1 : Langage ASCII utilisé au cours de cette thèse.

1.2.7. Chimie automatisée sur support solide ABI3900 et Expedite 8909

Les polymères synthétisés pour étudier l'efficacité de couplage des marqueurs clivables ont été réalisés sur l'Expedite 8909. Tous les autres polymères montrés dans cette thèse et dans cette partie expérimentale ont été synthétisés sur l'ABI3900. Pour chacun des deux robots synthétiseurs, les expériences se sont effectuées sous argon dans des conditions rigoureusement sèches à l'aide de la méthode standard automatisée de phosphoramidite en phase solide. La méthode typique comporte quatre étapes : (i) déprotection, (ii) couplage, (iii) oxydation et (iv) piégeage, comme décrit précédemment.[25] De manière générale, les monomères phosphoramidites ont été dissous dans de l'acétonitrile anhydre sous argon (solution à 55 mM pour ABI3900 et 0.1M pour Expedite 8909). Les monomères ont été placés dans le synthétiseur avec tous les réactifs nécessaires (réactif d'activation, Cap A, Cap B, réactif de retrait de DMT) et conditionnés deux fois. Le support solide (polystyrène Glen UnySupport^{tm} 200 nmol, ou dT CPG 1000 A° 1 µmol) a été placé dans le synthétiseur et les séquences ont été réalisées en mode DMT-On. Une fois la synthèse terminée, la colonne de support solide a été retirée du synthétiseur et la séquence protégée par DMT a été libérée du support. Une solution d'ammoniac à 28 % et de méthylamine (1/1, v/v) a été utilisée pour libérer les polymères protégés par le DMT du support solide à température ambiante pendant 10 minutes. La solution résultante a été scellée dans un flacon de 2 mL et laissée dans une étuve pendant la nuit à 50 °C (pour effectuer l'étape de déphosphorylation et récupérer le polymère souhaité, dans le cas de dT CPG, le polymère a été clivé pendant 45 min à température ambiante avec AMA et directement purifié). Le mélange a ensuite été purifié sur une cartouche de purification d'ADN Glen-Pak en suivant la procédure classique pour les oligos en mode DMT-ON (décrit Chapitre 2). La solution de polymère isolée a été lyophilisée et obtenue sous forme de poudre blanche avec des rendements modérés (≈ 40% pour une séquence de 27 monomères). Les rendements obtenus ne représentent, dans la plupart des cas, pas la réalité à la suite des résultats polydisperses observés, ou de la quantité de sel encore non négligeable dans l'échantillon.

1.2.7.1 Protocoles utilisés pour l'Expedite 8909

Les actions décrites ci-dessous sont exécutées par le synthétiseur Expedite 8909 et représentent un cycle complet du protocole standard (ou modifié) pour l'ajout d'un monomère. Les commentaires pour chaque ligne sont indiqués entre guillemets ou après un astérisque. Le premier numéro de chaque ligne correspond à la position du réactif (ou à la commande pour la détection UV), le deuxième indique le nombre de pulsations nécessaires pour introduire le réactif dans la colonne, et le troisième indique la durée pendant laquelle ces pulsations doivent être effectuées. Lorsqu'un "0" est indiqué dans cette colonne, la pompe fonctionne à sa vitesse maximale.

Protocole position A

```
/* ------------------------------------------------------------------------------------ */
/*       Function              Mode  Amount Time(sec)          Description               */
/*                                    /Arg1   /Arg2                                     */
/* ------------------------------------------------------------------------------------ */
$Deblocking
 144 /*Index Fract. Coll.  */ NA         1     0  "Event out ON"
   0 /*Default             */ WAIT       0   1.5  "Wait"
 141 /*Trityl Mon. On/Off  */ NA         1     1  "START data collection"
  16 /*Dblk                */ PULSE     10     0  "Dblk to column"
  16 /*Dblk                */ PULSE     50    49  "Deblock"
  38 /*Diverted Wsh A      */ PULSE     40     0  "Flush system with Wsh A"
 141 /*Trityl Mon. On/Off  */ NA         0     1  "STOP data collection"
  38 /*Diverted Wsh A      */ PULSE     40     0  "Flush system with Wsh A"
 144 /*Index Fract. Coll.  */ NA         2     0  "Event out OFF"
$Coupling
   1 /*Wsh                 */ PULSE      5     0  "Flush system with Wsh"
   2 /*Act                 */ PULSE      5     0  "Flush system with Act"
  18 /*A + Act             */ PULSE      5     0  "Monomer + Act to column"
  18 /*A + Act             */ PULSE      2    16  "Couple monomer"
   2 /*Act                 */ PULSE      3    24  "Couple monomer"
   1 /*Wsh                 */ PULSE      7    56  "Couple monomer"
   1 /*Wsh                 */ PULSE      8     0  "Flush system with Wsh"
$Capping
  12 /*Wsh A               */ PULSE     20     0  "Flush system with Wsh A"
  13 /*Caps                */ PULSE      8     0  "Caps to column"
  12 /*Wsh A               */ PULSE      6    15  "Cap"
  12 /*Wsh A               */ PULSE     14     0  "Flush system with Wsh A"
$Oxidizing
  15 /*Ox                  */ PULSE     15     0  "Ox to column"
  12 /*Wsh A               */ PULSE     15     0  "Flush system with Wsh A"
$Capping
  13 /*Caps                */ PULSE      7     0  "Caps to column"
  12 /*Wsh A               */ PULSE     30     0  "End of cycle wash"
```

Figure PE.1. 1 : Protocole standard1µmol du robot Expedite 8909 pour la position A.

Protocole position G

```
/* ------------------------------------------------------------------------------------ */
/*       Function              Mode  Amount Time(sec)          Description               */
/*                                    /Arg1   /Arg2                                     */
/* ------------------------------------------------------------------------------------ */
$Deblocking
 144 /*Index Fract. Coll.  */ NA         1     0  "Event out ON"
   0 /*Default             */ WAIT       0   1.5  "Wait"
 141 /*Trityl Mon. On/Off  */ NA         1     1  "START data collection"
  16 /*Dblk                */ PULSE     10     0  "Dblk to column"
  16 /*Dblk                */ PULSE     50    49  "Deblock"
  38 /*Diverted Wsh A      */ PULSE     40     0  "Flush system with Wsh A"
 141 /*Trityl Mon. On/Off  */ NA         0     1  "STOP data collection"
  38 /*Diverted Wsh A      */ PULSE     40     0  "Flush system with Wsh A"
 144 /*Index Fract. Coll.  */ NA         2     0  "Event out OFF"
$Coupling
   1 /*Wsh                 */ PULSE      5     0  "Flush system with Wsh"
   2 /*Act                 */ PULSE      5     0  "Flush system with Act"
  20 /*G + Act             */ PULSE      6     0  "Monomer + Act to column"
  20 /*G + Act             */ PULSE      1     8  "Couple monomer"
   2 /*Act                 */ PULSE      4    32  "Couple monomer"
   1 /*Wsh                 */ PULSE      7    56  "Couple monomer"
   1 /*Wsh                 */ PULSE      8     0  "Flush system with Wsh"
$Capping
  12 /*Wsh A               */ PULSE     20     0  "Flush system with Wsh A"
  13 /*Caps                */ PULSE      8     0  "Caps to column"
  12 /*Wsh A               */ PULSE      6    15  "Cap"
  12 /*Wsh A               */ PULSE     14     0  "Flush system with Wsh A"
$Oxidizing
  15 /*Ox                  */ PULSE     15     0  "Ox to column"
  12 /*Wsh A               */ PULSE     15     0  "Flush system with Wsh A"
$Capping
  13 /*Caps                */ PULSE      7     0  "Caps to column"
  12 /*Wsh A               */ PULSE     30     0  "End of cycle wash"
```

Figure PE.1. 2 : Protocole standard 1µmol du robot Expedite 8909 pour la position G.

Protocole positions 5-6-7

```
/* ------------------------------------------------------------------------------------ */
/*       Function              Mode  Amount Time(sec)          Description               */
/*                                   /Arg1   /Arg2                                       */
/* ------------------------------------------------------------------------------------ */
$Deblocking
 144 /*Index Fract. Coll.  */ NA         1     0  "Event out ON"
   0 /*Default             */ WAIT       0   1.5  "Wait"
 141 /*Trityl Mon. On/Off  */ NA         1     1  "START data collection"
  16 /*Dblk                */ PULSE     10     0  "Dblk to column"
  16 /*Dblk                */ PULSE     50    49  "Deblock"
  38 /*Diverted Wsh A      */ PULSE     40     0  "Flush system with Wsh A"
 141 /*Trityl Mon. On/Off  */ NA         0     1  "STOP data collection"
  38 /*Diverted Wsh A      */ PULSE     40     0  "Flush system with Wsh A"
 144 /*Index Fract. Coll.  */ NA         2     0  "Event out OFF"
$Coupling
   1 /*Wsh                 */ PULSE      5     0  "Flush system with Wsh"
   2 /*Act                 */ PULSE      5     0  "Flush system with Act"
  22 /*5 + Act             */ PULSE      6     0  "Monomer + Act to column"
  22 /*5 + Act             */ PULSE      1     8  "Couple monomer"
   2 /*Act                 */ PULSE      4    32  "Couple monomer"
   1 /*Wsh                 */ PULSE      7    56  "Couple monomer"
   1 /*Wsh                 */ PULSE      8     0  "Flush system with Wsh"
$Capping
  12 /*Wsh A               */ PULSE     20     0  "Flush system with Wsh A"
  13 /*Caps                */ PULSE      8     0  "Caps to column"
  12 /*Wsh A               */ PULSE      6    15  "Cap"
  12 /*Wsh A               */ PULSE     14     0  "Flush system with Wsh A"
$Oxidizing
  15 /*Ox                  */ PULSE     15     0  "Ox to column"
  12 /*Wsh A               */ PULSE     60     0  "Flush system with Wsh A"
$Capping
  13 /*Caps                */ PULSE      7     0  "Caps to column"
  12 /*Wsh A               */ PULSE      6    10  "Cap"
  12 /*Wsh A               */ PULSE     30     0  "End of cycle wash"
```

*Figure PE.1. 3 : Protocole standard des positions 5-6-7 utilisé pour l'étude d'efficacité de couplage des Espaceurs **TISP3** et **TISP4**.*
En vert, une quantité de lavage augmentée pour ne pas récupérer le réactif d'oxydation lors de la collection des fractions de déprotection du DMT.

Protocole position G – Long couplage

```
/* ------------------------------------------------------------------------------------ */
/*       Function              Mode  Amount Time (sec)          Description               */
/*                                   /Arg1   /Arg2                                       */
/* ------------------------------------------------------------------------------------ */
$Deblocking
 144 /*Index Fract. Coll.  */ NA         1     0  "Event out ON"
   0 /*Default             */ WAIT       0   1.5  "Wait"
 141 /*Trityl Mon. On/Off  */ NA         1     1  "START data collection"
  16 /*Dblk                */ PULSE     10     0  "Dblk to column"
  16 /*Dblk                */ PULSE     50    49  "Deblock"
  38 /*Diverted Wsh A      */ PULSE     80     0  "Flush system with Wsh A"
 141 /*Trityl Mon. On/Off  */ NA         0     1  "STOP data collection"
  38 /*Diverted Wsh A      */ PULSE     80     0  "Flush system with Wsh A"
 144 /*Index Fract. Coll.  */ NA         2     0  "Event out OFF"
$Coupling
   1 /*Wsh                 */ PULSE     15     0  "Flush system with Wsh"
   2 /*Act                 */ PULSE      5     0  "Flush system with Act"
  20 /*G + Act             */ PULSE      6     0  "Monomer + Act to column"
  20 /*G + Act             */ PULSE      2    75  "Couple monomer"
   2 /*Act                 */ PULSE      3    75  "Couple monomer"
   1 /*Wsh                 */ PULSE      7   150  "Couple monomer"
   1 /*Wsh                 */ PULSE      8     0  "Flush system with Wsh"
$Capping
  12 /*Wsh A               */ PULSE     20     0  "Flush system with Wsh A"
  13 /*Caps                */ PULSE      8     0  "Caps to column"
  12 /*Wsh A               */ PULSE      6    90  "Cap"
  12 /*Wsh A               */ PULSE     14     0  "Flush system with Wsh A"
$Oxidizing
  15 /*Ox                  */ PULSE     15     0  "Ox to column"
  12 /*Wsh A               */ PULSE     60     0  "Flush system with Wsh A"
$Capping
  13 /*Caps                */ PULSE      7     0  "Caps to column"
  12 /*Wsh A               */ PULSE      6    90  "Cap"
  12 /*Wsh A               */ PULSE     30     0  "End of cycle wash"
```

Figure PE.1. 4 : Protocole modifié pour un temps de couplage rallongé (jaune).

1.2.7.2 Protocoles utilisés pour l'ABI3900.

PREPROCESSING	Purges	# Iterations	Wait sec.	Command Code	Param1 Chemical	Param2 Volume
PREWASH1		1	0	DISP	ACN	250
	LONG_PURGE					
START_LOOP						
DETRITYLATION		3	0	DISP	DEBLOCK	180
	LONG_PURGE					
TCA_WASH		1	0	DISP	ACN	250
	LONG_PURGE					
COUPLING		3	0	DISP	ACTIVATOR	115
				DISP	AMIDITE	75
	LONG_PURGE					
CAPPING		1	0	DISP	CAPB	80
				DISP	CAPA	80
	LONG_PURGE					
OXIDATION		1	0	DISP	OXIDIZER	150
	LONG_PURGE					
OX_WASH		1	0	DISP	ACN	250
	LONG_PURGE					
CAPPING		1	0	DISP	CAPB	80
				DISP	CAPA	80
	LONG_PURGE					
TRITYLOFF		3	0	DISP	DEBLOCK	180
	LONG_PURGE					
FINAL_FLUSH		4	0	DISP	ACN	250
	LONG_PURGE					
DRY_SUPPORT		1	0	DISP	ACN	180
	DRY_BEADS					
END_PROGRAM						

Figure PE.1. 5 : Protocole standard 1µmol utilisé pour la synthèse de polymère sur le synthétiseur ABI3900.
Chaque ligne décrit les commandes suivies par le robot, avec un nombre précis d'itération, de volume, et de temps de couplage.

2. Procédures expérimentales

2.1. Synthèse des intermédiaires Amino-TEMPO T1-T11

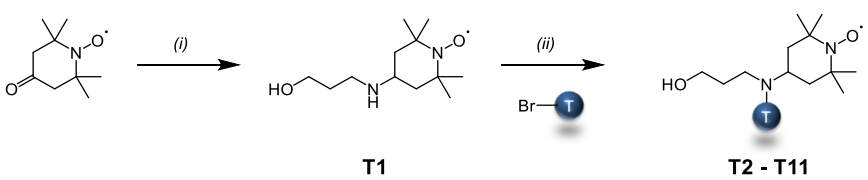
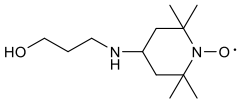
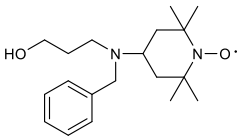


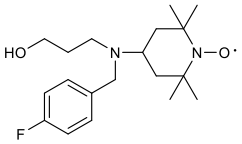
*Schéma PE.II. 1 : Stratégie de synthèse des amino-TEMPO modifiés **T2**-**T11**.*
i) $Na(CH_3COO)_3BH$, CH_3COOH, amino propanol, 1,2-dichloroéthane, t.a., 17h; ii) Acétonitrile, K_2CO_3, 83°C, 17h.



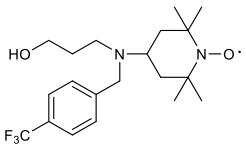
T1. La synthèse de **T1** a été adaptée à partir d'une procédure précédemment décrite.[364] Un ballon rond de 250 mL chargé de 3-amino-1-propanol (2,25 g, 30 mmol, 1 éq), de 4-oxo-2,2,6,6-tétraméthylpipéridine-1-oxyl (5,10 g, 30 mmol, 1 éq) et de 1,2-dichloroéthane anhydre (105 mL) a été placé sous une atmosphère d'argon. Du triacétoxyborohydrure de sodium (8,90 g, 42 mmol, 1,4 éq) et de l'AcOH (1,71 mL, 30 mmol, 1 éq) ont ensuite été ajoutés. Le mélange a été agité à température ambiante pendant 24 heures. Le mélange de réaction a directement été purifié par chromatographie sur colonne sur gel de silice (EtOAc/EtOH/Et_3N : 60/20/2). Le composé final a été dissous dans du toluène et évaporé pour complètement éliminer l'AcOH, pour obtenir 5,86 g de **T1** sous forme de solide orange avec un rendement de 85%. HRMS m/z : $[M+H]^+$ calculé pour $C_{12}H_{26}N_2O_2^{\bullet+}$ 230,1989 ; trouvé 230,1984. Facteur g de RPE = 2,00605, aN = 1,554 mT.



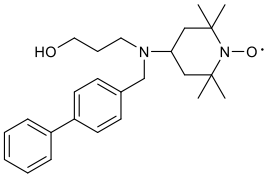
T2. Un ballon rond de 250 mL équipé d'un réfrigérant a été chargé de bromure de benzyle (2,30 g, 13,4 mmol, 1,5 éq.) et d'ACN anhydre (50 mL) sous une atmosphère d'argon. Ensuite, **T1** (2,06 g, 89,8 mmol, 1,0 éq.) dissous dans de l'ACN anhydre (13 mL) a été ajouté, suivi de K_2CO_3 (3,72 g, 26,9 mmol, 3,0 éq.). La réaction a été agité à reflux (82°C) pendant une nuit. Après 20 heures, le mélange de réaction a été lavé avec de la saumure (30 mL) et de l'eau (10 mL), puis extrait avec de l'EtOAc (3 x 30 mL). Les différentes fractions organiques combinées ont été séchées avec Na_2SO_4, filtrées et le solvant a été évaporé sous pression réduite. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (n-hexane / EtOAc : 8/2 à 1/1 avec 1% d'Et_3N) pour obtenir 1,59 g de **T2** sous forme de solide orange avec un rendement de 56%. HRMS m/z : $[M+H]^+$ calculé pour $C_{19}H_{32}N_2O_2^{\bullet+}$ 320,2458 ; trouvé 320,2462. Facteur g de RPE = 2,00610, aN = 1,528 mT.



T3. Le protocole est le même que pour **T2** en utilisant du bromure de 4-fluorobenzyle (2,59 g, 13,7 mmol, 1,5 éq.) au lieu du bromure de benzyle. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (n-hexane / EtOAc : 9/1 à 1/1 avec 1% d'Et_3N) pour obtenir 1,72 g de **T3** sous forme d'huile orange avec un rendement de 56%. HRMS m/z : $[M+H]^+$ calculé pour $C_{19}H_{31}N_2O_2F^{\bullet+}$ 338,2364 ; trouvé 338,2365. Facteur g de RPE = 2,00610, aN = 1,548 mT.

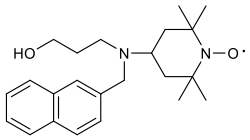


T4. Le protocole est le même que pour **T2** en utilisant du bromure de 4-(trifluorométhyl)benzyle (1,88 g, 7,85 mmol, 1,5 éq.) au lieu du bromure de benzyle. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (n-hexane / EtOAc : 8/2 à 1/1 avec 1% d'Et_3N) pour obtenir 0,74 g de **T4** sous forme d'huile orange avec un rendement de 37%. HRMS m/z : $[M+H]^+$ calculé pour $C_{20}H_{31}N_2O_2F_3^{\bullet+}$ 388,2332 ; trouvé 388,2332. Facteur g de RPE = 2,00612, aN = 1,533 mT.

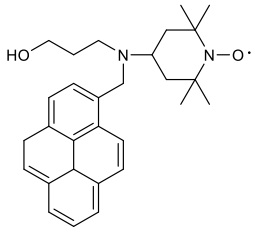


T5. Le protocole est le même que pour **T2** en utilisant du 4-bromométhylbiphényle (1,88 g, 7,59 mmol, 1,5 éq.) au lieu du bromure de benzyle. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (n-hexane / EtOAc :

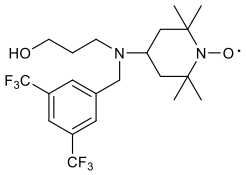
8/2 à 1/1 avec 1% d'Et_3N) pour obtenir 1,38 g de **T5** sous forme d'huile orange avec un rendement de 69%. HRMS m/z : $[M+H]^+$ calculé pour $C_{25}H_{36}N_2O_2^{\bullet+}$ 396,2771 ; trouvé 396,2768. Facteur g de RPE = 2,00614, aN = 1,540 mT.



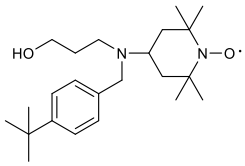
T6. Le protocole est le même que pour **T2** en utilisant du 2-(bromométhyl)naphthalène (3,16 g, 13,7 mmol, 1,5 éq.) au lieu du bromure de benzyle. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (n-hexane / EtOAc : 8/2 à 2/8 avec 1% d'Et_3N) pour obtenir 1,78 g de **T6** sous forme d'huile orange avec un rendement de 53%. HRMS m/z : $[M+H]^+$ calculé pour $C_{23}H_{34}N_2O_2^{\bullet+}$ 370,2615 ; trouvé 370,2623. Facteur g de RPE = 2,00613, aN = 1,532 mT.



T7. Le protocole est le même que pour **T2** en utilisant du 3-(bromométhyl)pyrène (4,26 g, 13,7 mmol, 1,5 éq.) au lieu du bromure de benzyle. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (n-hexane / EtOAc : 85/15 à EtOAc 100% avec 1% d'Et_3N) pour obtenir 1,96 g de **T7** sous forme d'huile orange avec un rendement de 48%. HRMS m/z : $[M+H]^+$ calculé pour $C_{29}H_{36}N_2O_2^{\bullet+}$ 444,2771 ; trouvé 444,2787. Facteur g de RPE = 2,00609, aN = 1,551 mT.

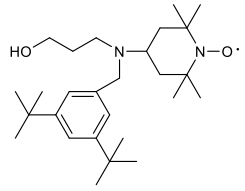


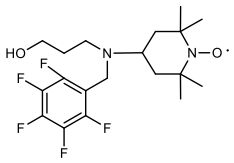
T8. Même protocole que **T2** en utilisant du 5-bis(trifluorométhyl)benzyle (4,21 g, 13,7 mmol, 1,5 éq.) au lieu du bromure de benzyle. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (n-hexane / EtOAc : 9/1 à 1/1 avec 1% d'Et_3N) pour obtenir 2,06 g de **T8** sous forme d'huile orange avec un rendement de 49%. HRMS m/z : $[M+H]^+$ calculé pour $C_{21}H_{30}N_2O_2F_6^{\bullet+}$ 456,2206 ; trouvé 456,2211. Facteur g de RPE = 2,00606, aN = 1,541 mT.



T9. Même protocole que **T2** en utilisant du 4-*tert*-butylbenzyl (3,12 g, 13,7 mmol, 1,5 éq.) au lieu du bromure de benzyle. Le produit brut a été purifié par chromatographie sur colonne sur gel de silice (n-hexane / EtOAc : 9/1 à 8/2 avec 1% d'Et_3N) pour obtenir 2,76 g de **T9** sous forme d'huile orange avec un rendement de 80%. HRMS m/z : $[M+H]^+$ calculé pour $C_{23}H_{40}N_2O_2^{\bullet+}$ 376,3084 ; trouvé 376,3089. Facteur g de RPE = 2,00609, aN = 1,546 mT.

T10 : Le protocole est le même que pour **T2** en utilisant du 3,5-di-tert-butylbenzyle (2,22 g, 7,85 mmol, 1,5 éq.) au lieu du bromure de benzyle. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (n-hexane / EtOAc : 9/1 à 8/2 avec 1% d'Et_3N) pour obtenir 1,84 g de **T10** sous forme d'huile orange avec un rendement de 82%. HRMS m/z : $[M+H]^+$ calculé pour $C_{27}H_{48}N_2O_2^{\bullet+}$ 432,3710 ; trouvé 432,3710. Facteur g de RPE = 2,00611, aN = 1,546 mT.





T11 : Le protocole est le même que pour **T2** en utilisant du 2,3,4,5,6-pentafluorobenzyl (2.22 g, 8.50 mmol, 1,5 éq.) au lieu du bromure de benzyle. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (cyclohexane / EtOAc : 9/1 à 8/2 avec 1% d'Et_3N) pour obtenir 1,68 g de **T11** sous forme d'huile orange avec un rendement de 72%. HRMS m/z : $[M+H]^+$ calculé pour $C_{19}H_{27}N_2O_2F_5{}^{\bullet+}$ 410.1987 ; trouvé 410.1994.

2.2. Synthèse TISP et TISP2

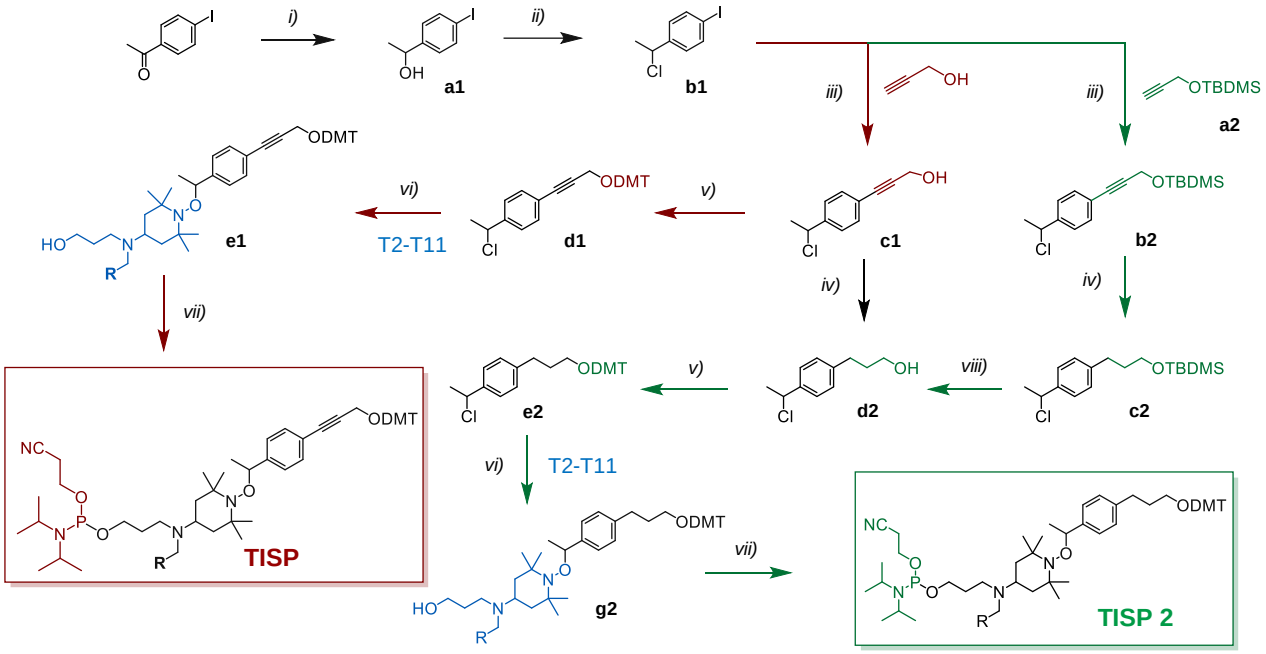
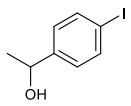
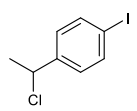


*Schéma PE.II. 2 : Stratégies **TISP** et **TISP2** pour la préparation de marqueurs d'octets clivables. i) $NaBH_4$, EtOH/THF, 0°C à t.a., 3h; (ii) HCl 37%, THF, 0°C à t.a., 4h; (iii) $Pd(PPh_3)_2$, CuI, Et_3N, t.a., 17h. iv) Pd/C ou Pt/C, H_2, THF anhydre, 1h; (v) Chlorure de 4,4-diméthoxytrityle (DMT-Cl), Pyridine anhydre, THF anhydre, 0°C à t.a., 1h; (vi) CuBr, Cu(0), PMDETA, THF, t.a ., 17h; (vii) 2-Cyanoéthyle N,N-diisopropylechlorophosphoramidite, DIPEA, DCM anhydre, 0°C à t.a., 1h. viii) Fluorure d'ammonium tétra-n-butyle (TBAF), THF, 0°C à t.a., 3h..*

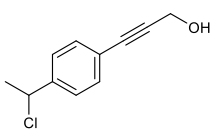


a1. Dans un ballon rond de 250 mL sous atmosphère d'argon, du $NaBH_4$ (1,85 g, 43,77 mmol) a été ajouté en 3 fois (3 x 1 minutes) à une solution de 4-iodoacétophénone (8 g, 32,51 mmol) solubilisée dans un mélange d'éthanol (35 mL) et de THF (35 mL) à 0°C. Après agitation pendant 30 minutes, le bain de glace a été retiré, et la réaction a été poursuivie pendant 2 heures supplémentaires à température ambiante. Après vérification de la réduction par chromatographie sur couche mince (CCM) (EtOAc/Cyclohexane ; 3/7 ; Rf = 0,5), le mélange résultant a été neutralisé par l'ajout de 2M HCl (37%) à 0°C (60 mL). Le produit désiré a ensuite été lavé à l'aide de saumure (50 mL) et extrait à l'aide d'EtOAc (4x60 mL). Les phases organiques ont été combinées, lavées avec de l'eau (2x 50 mL), séchées avec Na_2SO_4, filtrées, et concentrées sous pression réduite pour obtenir une huile orange qui a été utilisée pour la réaction suivante sans purification supplémentaire (8 g, rendement de 100%). HRMS m/z : $[M+Na]^+$ calculé pour $C_8H_9OINa^+$ 270,9590 ; trouvé, 270,9589. RMN ^1H (400 MHz, Chloroforme-*d*), δ (ppm) 7,72 - 7,63 (m, 2H), 7,16 - 7,08 (m, 2H), 4,85 (q, J = 6,5 Hz, 1H),

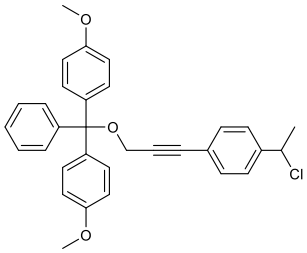
1,46 (d, J = 6,5 Hz, 3H). RMN ^{13}C (101 MHz, Chloroforme-*d*), δ (ppm) : 145,80, 137,83, 127,80, 93,08, 70,08, 25,57.



b1. Dans un ballon rond de 250 mL sous atmosphère d'argon, de l'acide chlorhydrique concentré (45 mL) a été ajouté lentement à une solution de **a1** (8 g, 32,25 mmol) dans du THF (15 mL) à 0°C. Après agitation pendant 45 minutes, le bain de glace a été retiré, et la réaction a été poursuivie toute la nuit à température ambiante. Après vérification de la complétion de la chloration par CCM, comme pour la synthèse de **a1** (Rf = 0,8), le produit désiré a été lavé en utilisant un mélange de saumure et d'eau (80 mL ; 1/1) et extrait à l'aide d'EtOAc (3x 60 mL). Les phases organiques ont été combinées et neutralisées avec du bicarbonate de sodium concentré (1x 80 mL) et de l'eau (2x 50 mL). Le mélange jaunâtre a ensuite été séché avec Na_2SO_4, filtré, et concentré sous pression réduite pour obtenir une huile rouge qui a été utilisée pour la réaction suivante sans purification supplémentaire (8,4 g, rendement de 98%).^{1}H RMN (400 MHz, Chloroforme-*d*), δ (ppm): 7.73 – 7.65 (m, 2H), 7.21 – 7.13 (m, 2H), 5.02 (q, J = 6.8 Hz, 1H), 1.82 (d, J = 6.8 Hz, 3H). ^{13}C RMN (101 MHz, Chloroforme-*d*), δ (ppm) 142.87, 138.13, 128.86, 94.23, 58.32, 26.81. (HRMS non possible pour cause d'un composé trop volatile).

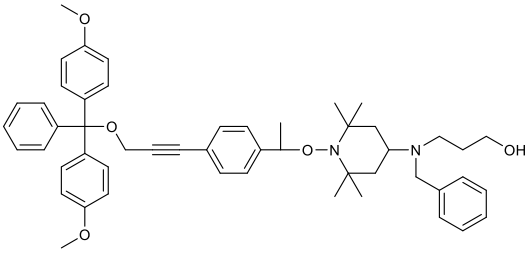


c1. Dans un ballon rond de 250 mL sous atmosphère d'argon, **b1** (8,3 g, 31,14 mmol) a été mélangé avec $PdCl_2(PPh_3)_2$ (0,65 g, 0,93 mmol) et CuI (0,119 g, 0,62 mmol). La vaisselle a été évacuée 3 fois sous atmosphère d'argon avant d'ajouter de l'Et_3N (50 mL). Pendant que la réaction était agitée sous atmosphère d'argon, de l'alcool propargylique (1,92 g, 34,25 mmol) a été ajouté au mélange avec une seringue. La réaction a ensuite été agitée à température ambiante et était complète après 17h, comme en témoigne le suivi par CCM (EtOAc/cyclohexane/Et_3N ; 30/68/2 ; Rf = 0,4). Le mélange réactionnel orange a ensuite été filtré à travers de la célite (545) et rincé à l'aide d'EtOAc. Le filtrat rouge a été concentré sous pression réduite avant d'être purifié sur gel de silice (EtOAc/cyclohexane/Et_3N ; 10/88/2 à 30/68/2) pour obtenir une huile sombre rouge (5,5 g, rendement de 91%).^{1}H RMN (400 MHz, Chloroforme-*d*), δ (ppm): 7.46 – 7.39 (m, 2H), 7.41 – 7.33 (m, 2H), 5.07 (q, J = 6.8 Hz, 1H), 4.50 (s, 2H), 1.83 (d, J = 6.8 Hz, 3H). ^{13}C RMN (101 MHz, Chloroforme-*d*), δ (ppm) 143.18, 132.07, 126.68, 122.61, 87.83, 85.37, 58.30, 51.80, 26.47. (HRMS non possible pour cause d'un composé trop volatile). $PdCl_2(PPh_3)_2$ a été synthétisé en adaptant une procédure déjà décrite,[386] et son spectre de masse est montré en partie 4.2, Figure PE.4. 12.

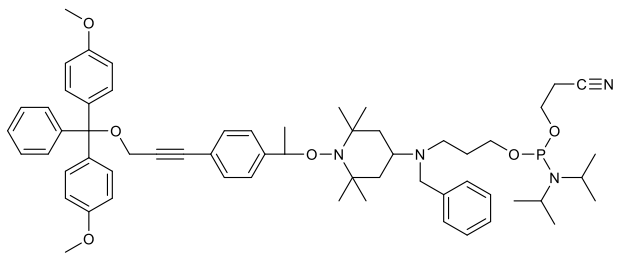


d1. c1 (2 g, 10,27 mmol) a été dissous dans du THF anhydre (30 mL) à 0°C, puis ajouté goutte à goutte à une solution de DMT-Cl (4,18 g, 12,32 mmol) dans de la pyridine anhydre (1,1 mL, 13,36 mmol) et du THF anhydre (10 mL) sous atmosphère d'argon. Le mélange a été agité toute la nuit à température ambiante. Après vérification de la complétion de la protection par CCM (EtOAc/Cyclohexane/Et_3N ; 10/89/1), la réaction a été stoppée par l'addition d'eau (100 mL) et le produit désiré a été extrait avec de l'EtOAc (3x 50 mL). Les phases organiques ont été combinées, séchées avec du Na_2SO_4, filtrées, puis

concentrées sous pression réduite pour obtenir une huile orange. Le produit brut a été purifié sur gel de silice (EtOAc/Cyclohexane/Et_3N ; 5/94/1 à 25/64/1) pour obtenir une poudre légèrement orangée (3,8 g, rendement de 74%). ^{1}H RMN (300 MHz, Acétone-*d_6*), δ (ppm): 7.55 – 7.42 (m, 2H), 7.45 – 7.24 (m, 11H), 6.96 – 6.83 (m, 4H), 5.27 (q, J = 6.8 Hz, 1H), 4.03 (s, 2H), 3.79 (s, 6H), 1.82 (d, J = 6.8 Hz, 3H). ^{13}C RMN (101 MHz, Acétone-*d_6*), δ (ppm) 159.51, 145.63, 143.95, 136.20, 132.46, 132.43, 130.68, 128.59, 128.50, 127.48, 123.41, 113.81, 87.57, 87.36, 84.96, 58.86, 55.29, 53.65, 26.35**.** (HRMS non possible pour cause d'un composé trop volatile).

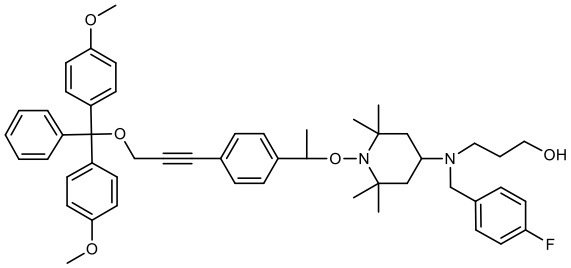


e1. Dans un ballon rond de 250 mL préalablement séché, une solution dégazée de **d1** (0,37 g, 0,73 mmol, 1,0 équivalent) et **T2** (0,35 g, 1,1 mmol, 1,5 équivalents) dans du THF anhydre (10 mL) a été ajoutée goutte à goutte à une solution dégazée de CuBr (0,69 g, 4,82 mmol, 1,5 équivalents), Cu(0) (0,31 g, 4,82 mmol, 1,5 équivalents) et de N,N,N',N",N"-pentaméthyldiéthylènetriamine (PMDETA) (2,01 mL, 9,64 mmol, 3,0 équivalents) dans du THF anhydre (10 mL) sous atmosphère d'argon. Le mélange réactionnel a été agité toute la nuit à température ambiante et suivi par CCM (Cyclohexane/EtOAc/Et_3N ; 6/4/0,2 ; Rf = 0,3). Après 24 heures, le réactif de départ **d1** avait presque entièrement réagi, donc la réaction a été stoppée, et le produit brut a été filtré directement sur de la Célite 545, lavé avec du THF, puis évaporé sous pression réduite. Le résidu vert a été lavé avec un mélange d'eau/saumure (1:1) et extrait avec de l'éther diéthylique (Et_2O). Les phases organiques combinées ont été séchées avec du Na_2SO_4, filtrées, évaporées sous pression réduite, et purifiées sur gel de silice (Cyclohexane/EtOAc/Et_3N ; 9/1/0,2 à 6/4/0,2) pour obtenir 0,49 g d'**e1** sous forme de poudre cristalline blanche, avec un rendement de 86%. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{51}H_{61}N_2O_5^+$ 781.4575; trouvé, 781.4575. ^{1}H RMN (400 MHz, Acétone-*d6*) δ 7.55 – 7.48 (m, 2H), 7.44 – 7.18 (m, 16H), 6.96 – 6.87 (m, 4H), 4.82 (q, J = 6.6 Hz, 1H), 4.01 (s, 2H), 3.79 (s, 6H), 3.63 (s, 2H), 3.57 (t, J = 5.9 Hz, 2H), 3.04 – 2.87 (m, 1H), 2.66 – 2.58 (m, 2H), 1.71 – 1.43 (m, 6H), 1.45 (d, J = 6.7 Hz, 3H), 1.33 (s, 3H), 1.08 (s, 3H), 0.94 (s, 3H), 0.66 (s, 3H). ^{13}C RMN (101 MHz, Acétone-*d_6*), δ (ppm) 159.36, 146.50, 145.49, 141.18, 136.07, 131.91, 130.50, 129.05, 128.57, 128.43, 128.33, 127.30, 127.27, 127.17, 121.98, 113.64, 87.34, 86.28, 85.27, 83.27, 61.71, 60.70, 60.38, 55.13, 54.84, 53.50, 47.98, 41.46, 34.64, 23.07, 20.82.

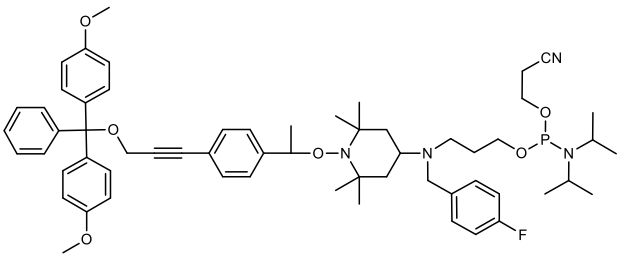


Espaceur *TISP*. Une solution d'O-2-cyanoéthyl-N,N-diisopropylchlorophosphoramidite (0,22 g, 0,93 mmol, 1,5 équivalents) dans du DCM anhydre (10 mL) a été refroidie à 0°C et ajoutée goutte à goutte à une solution d'**e1** (0,49 g, 0,62 mmol, 1,0 équivalent) et de DIPEA (2,3 mL, 13,2 mmol, 5,0 équivalents) dans du DCM anhydre (10 mL) également à 0°C sous atmosphère d'argon. Le bain de glace a été retiré directement après l'addition. Après 2 heures d'agitation à

température ambiante, la réaction a été arrêtée, et le mélange a été évaporé sous pression réduite pour obtenir une huile jaune contenant de la DIPEA. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (Cyclohexane/EtOAc/DIPEA ; 95/5/2 à 65/35/2) pour obtenir 0,29 g de **TISP** sous forme de poudre cristalline, avec un rendement de 48%. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{60}H_{78}N_4O_6P^+$ 981.5654; trouvé, 981.5683. ^{1}H RMN (400 MHz, Acétone-*d6*) δ 7.55 – 7.48 (m, 2H), 7.40 – 7.20 (m, 16H), 6.96 – 6.87 (m, 4H), 4.82 (q, J = 6.6 Hz, 1H), 4.00 (s, 2H), 3.79 (s, 6H), 3.84 – 3.70 (m, 2H), 3.69 – 3.58 (m, 2H), 3.63 (s, 2H), 2.91 (m, 1H), 2.71 (t, J = 6.1, 2H), 2.60 (m, 2H), 1.81 – 1.47 (m, 6H), 1.45 (d, J = 6.7 Hz, 3H), 1.33 (s, 3H), 1.18 (dd, J = 6.8, 3.1 Hz, 12H), 1.10 (s, 3H), 0.96 (s, 3H), 0.66 (s, 3H). ^{13}C RMN (101 MHz, Acétone-d_6), δ (ppm) 160.14, 147.29, 146.28, 142.57, 136.85, 132.68, 131.28, 129.60, 129.27, 129.21, 129.10, 128.07, 127.77, 122.75, 119.34, 114.41, 88.12, 87.05, 86.06, 84.01, 62.85, 61.49, 61.17, 59.77, 55.89, 55.78, 54.27, 50.47, 47.50, 44.10, 35.47, 25.34, 23.82, 21.65, 21.14. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 146.76.

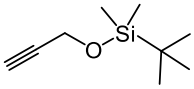


e1-F. La même procédure que pour la synthèse de **e1** mais **T3** au lieu de **T2**. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (Cyclohexane/EtOAc/TEA ; 85/15/2 à 50/50/2) pour obtenir 0,94 g de **e1-F** sous forme de poudre cristalline, avec un rendement de 70 %. ^{1}H NMR (400 MHz, Acétone-d_6) δ 7.56 – 7.49 (m, 2H), 7.45 – 7.17 (m, 13H), 7.12 – 7.03 (m, 2H), 6.99 – 6.89 (m, 4H), 4.83 (q, *J* = 6.7 Hz, 1H), 4.02 (s, 2H), 3.80 (s, 6H), 3.62 (s, 2H), 3.57 (t, *J* = 5.9 Hz, 2H), 3.03 – 2.86 (m, 1H), 2.61 (td, *J* = 6.6, 1.3 Hz, 2H), 1.71 – 1.39 (m, 6H), 1.46 (d, *J* = 6.7 Hz, 3H), 1.34 (s, 3H), 1.11 (s, 3H), 0.96 (s, 3H), 0.67 (s, 3H). L'analyse RMN ^{13}C effectuée pour cet intermédiaire n'est pas montrée car le spectre obtenu ne montre pas tous les signaux attendus en raison de la concentration trop faible de l'échantillon. Cependant, l'analyse complète du ***TISP*-F** ci-après permet de confirmer avec certitude la bonne obtention de l'intermédiaire **e1-F.**

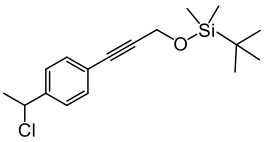


TISP*-F**. La même procédure que pour la synthèse de ***TISP mais avec **e1-F** au lieu de **e1.** Le produit brut a été purifié par chromatographie sur colonne de gel de silice (Cyclohexane/EtOAc/DIPEA ; 95/5/2 à 65/35/2) pour obtenir 0,5 g de ***TISP*-F** sous forme de poudre cristalline, avec un rendement de 43%. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{60}H_{76}N_4O_6PF^+$ 999.5559; trouvé, 999.5573. ^{1}H NMR (400 MHz, Acétone-d_6) δ 7.55 – 7.43 (m, 2H), 7.45 – 7.18 (m, 14H), 7.10 – 7.02 (m, 2H), 6.95 – 6.87 (m, 4H), 4.82 (q, *J* = 6.5 Hz, 1H), 4.01 (s, 2H), 3.79 (s, 6H), 3.89 – 3.64 (m, 2H), 3.67 – 3.55 (m, 4H), 2.93 – 2.82 (m, 1H), 2.72 (td, *J* = 6.0, 2.6 Hz, 2H), 2.59 (tt, *J* = 6.7, 3.3 Hz, 2H), 1.76 – 1.36 (m, 6H), 1.45 (d, *J* = 6.6 Hz, 3H), 1.33 (s, 3H), 1.18 (dd, *J* = 6.8, 5.2 Hz, 12H), 1.12 (s, 3H), 0.97 (s, 3H), 0.66 (s, 3H). ^{13}C RMN (101 MHz, Acétone) δ 161.32, 159.73, 146.86, 145.87, 138.18, 136.45, 132.29,

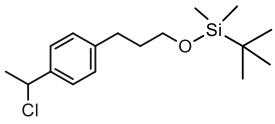
130.88, 130.80, 128.80, 128.70, 127.66, 122.36, 118.94, 115.57, 115.36, 114.02, 87.72, 86.67, 85.67, 83.62, 62.46, 61.08, 60.76, 59.16, 55.50, 54.53, 53.88, 50.11, 47.11, 43.71, 34.98, 25.02, 23.43, 21.27, 20.74. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 146.80.



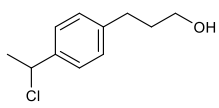
a2. Adapté de la littérature.[387] Dans un ballon rond de 500 mL, 2-propyn-1-ol (4,0 mL, 69,4 mmol, 1,0 équivalent) a été solubilisé dans du DCM anhydre (100mL) et refroidit à 0°C. TBDMSCl (10,45 g, 69,4 mmol, 1,0 équivalent) et imidazole (9.44 g, 138,71 mmol, 2,0 équivalents) ont été successivement ajoutés à la solution. Le bain de glace a directement été retiré et la réaction a été agitée 1h30 à température ambiante. La solution a ensuite été lavée avec de la saumure (50 mL), extrait avec du DCM trois fois. Les phases organiques combinées ont été lavée avec une solution de NH_4Cl (2 fois 50 mL), séchées avec Na_2SO_4 et concentrées sous pression réduite pour obtenir une huile transparente (rendement >99%). Le composé étant hautement volatile, le DCM n'a pas pu être totalement éliminé. Le produit brut a été utilisé sans purification supplémentaire pour synthétiser l'intermédiaire **b2**. ^{1}H RMN (400 MHz, Chloroforme-*d*) δ 4.31 (d, *J* = 2.4 Hz, 2H), 2.39 (t, *J* = 2.4 Hz, 1H), 0.91 (s, 9H), 0.13 (s, 6H). ^{13}C NMR (101 MHz, Acétone) δ 83.35, 74.40, 51.90, 26.09, 18.71, -5.09.



b2. b2 a été synthétisé avec la même procédure que pour **c1**, mais avec **a2** au lieu de l'alcool propargylique. Il a été purifié par chromatographie sur colonne de gel de silice (Cyclohexane /EtOAc/Et_3N ; 99/01/0,2 à 85/15/0,2) pour obtenir **b2** sous forme d'huile jaunâtre, avec un rendement de 61%.^{1}H RMN (400 MHz, Acétone-*d6*), δ (ppm) 7.53 – 7.41 (m, 4H), 5.27 (q, J = 6.8 Hz, 1H), 4.58 (s, 2H), 1.82 (d, J = 6.8 Hz, 3H), 0.93 (s, 9H), 0.17 (s, 6H). ^{13}C RMN (101 MHz, Acétone) δ 144.27, 132.47, 127.74, 123.68, 89.58, 84.78, 59.06, 52.57, 26.55, 26.16, 18.79, -4.91.

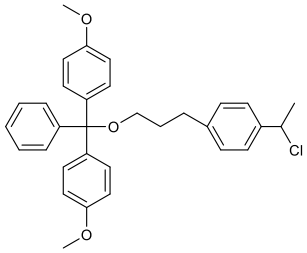


c2. Dans un ballon rond de 100 mL préalablement séché sous atmosphère d'argon, **b2** (1.15 g, 3.71 mmol, 1,0 équivalent) a été solubilisé dans du THF anhydre (30 mL) à température ambiante. Un autre ballon rond de 100 mL préalablement séché, équipé d'un ballon d'hydrogène (H_2), a été chargé avec du Pt/C (0,145 g, 0,74 mmol, 0,2 équivalent) et purgé 3 fois avec de l'argon, puis de l'hydrogène (H_2). Ensuite, la solution de **b2** a été ajoutée au ballon contenant Pt/C, toujours sous atmosphère d'hydrogène (H_2), et a été agitée à température ambiante pendant 24 heures. Après cela, la RMN du produit brut a montré une conversion complète du réactif de départ **b2**. La réaction a été stoppée en purgeant l'atmosphère avec de l'argon. Le mélange a ensuite été filtré à travers de la Célite 545, lavé avec du THF, puis évaporé sous pression réduite pour obtenir une huile jaune avec 100% de rendement. Le produit brut a été identifié comme étant pure en RMN et a donc été utilisé sans purification supplémentaire pour la déprotection donnant **d2.** ^{1}H NMR (400 MHz, Chloroforme-*d*) δ 7.32 – 7.28 (m, 2H), 7.20 – 7.16 (m, 2H), 5.09 (q, *J* = 6.9 Hz, 1H), 3.63 (td, *J* = 6.3, 1.1 Hz, 2H), 2.71 – 2.52 (m, 2H), 1.84 (dd, *J* = 6.8, 1.2 Hz, 3H), 1.82 (ddd, *J* = 11.4, 5.4, 2.7 Hz, 2H, 0.98 – 0.87 (m, 9H), 0.10 – 0.02 (m, 6H).^{13}C NMR (101 MHz, Chloroforme-*d*) δ 128.54, 128.51, 126.41, 126.22, 62.04, 58.58, 33.91, 26.27, 25.78, 24.29, -5.44.

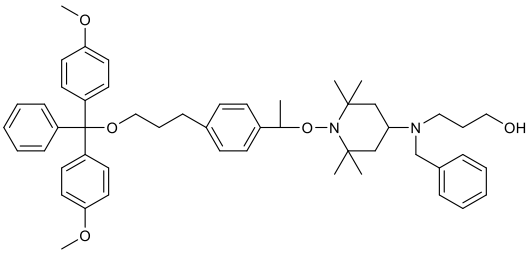


d2. À partir de **c1** : Dans un ballon rond de 100 mL préalablement séché sous atmosphère d'argon, **c1** (0,82 g, 4,21 mmol, 1,0 équivalent) a été solubilisé dans du THF anhydre (40 mL) à température ambiante. Un autre ballon rond de 100 mL préalablement séché, équipé d'un ballon d'hydrogène (H_2), a été chargé avec du Pt/C (0,164 g, 0,84 mmol, 0,2 équivalent) et purgé 3 fois avec de l'argon, puis de l'hydrogène (H_2). Ensuite, la solution de **c1** a été ajoutée au ballon contenant Pt/C, toujours sous atmosphère d'hydrogène (H_2), et a été agitée à température ambiante pendant 24 heures. Après cela, la RMN du produit brut a montré une conversion complète du réactif de départ **c1**. La réaction a été stoppée en purgeant l'atmosphère avec de l'argon. Le mélange a ensuite été filtré à travers de la Célite 545, lavé avec du THF, puis évaporé sous pression réduite pour obtenir une huile jaune. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (Cyclohexane/EtOAc ; 95/5 à 7/3) pour obtenir 0,22 g de **d2** sous forme d'une huile jaunâtre, avec un rendement de 27%. Un sous-produit majeur **SP** (1-(1-chloroéthyl)-4-propylbenzène) correspondant à une perte de H_2O du produit désiré a également été identifié, avec un rendement de 60%. La Figure PE.3. 24 montre le spectre RMN de ce dernier. ^{1}H RMN (400 MHz, Chloroforme-*d*) δ 7.33 – 7.20 (m, 2H), 7.20 – 7.08 (m, 2H), 4.88 (q, *J* = 6.5 Hz, 1H), 2.58 (dd, *J* = 8.6, 6.7 Hz, 2H), 1.70 – 1.57 (m, 2H), 1.49 (d, *J* = 6.4 Hz, 3H), 0.94 (t, *J* = 7.3 Hz, 3H).

d2. À partir de **c2** : dans un ballon rond de 100 mL préalablement séché sous atmosphère d'argon, **c2** (0,93g, 2.97 mmol, 1,0 équivalent) a été solubilisé dans du THF (25 mL) à 0°C. TBAF (1M dans THF, 4,0 mL, 4,00 mmol, 1.35 équivalent) a été ajouté goute à goute à la solution contenant **c2**. Après 15min, le bain à glace a été retiré et la réaction agitée pendant 2h30. Après vérification par CCM, la réaction était complète. La solution résultante a été lavée avec NH_4Cl (70 mL), $NaHCO_3$ (70 mL) puis de la saumure (70 mL). Les phases organiques ont été récupérées avec EtOAC, filtrées avec Na_2SO_4 puis concentrées sous pression réduite. Le brut a été purifié par chromatographie sur colonne de gel de silice (Cyclohexane/EtOAc/TEA ; 94/5/1 à 69/30/1) donnant une huile jaune avec 76% de rendement. ^{1}H RMN (400 MHz, Chloroforme-*d*), δ (ppm) 7.42 – 7.27 (m, 2H), 7.26 – 7.10 (m, 2H), 5.09 (q, J = 6.9 Hz, 1H), 3.69 (td, J = 6.4, 2.3 Hz, 2H), 2.77 – 2.66 (m, 2H), 2.01 – 1.81 (m, 3H), 1.85 (d, J = 6.8 Hz, 2H). ^{13}C RMN (101 MHz, Chloroforme-*d*) δ (ppm) 142.13, 140.52, 128.80, 126.70, 62.32, 58.86, 34.20, 31.86, 26.55.

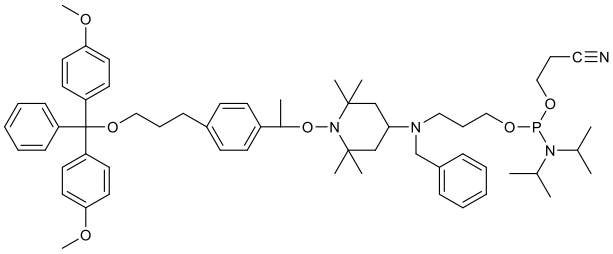


e2. Ce composé a été synthétisé en suivant la même procédure que **d1**, mais en utilisant **d2** au lieu de **c1**. Il a été purifié par chromatographie sur colonne de gel de silice (*n*-hexane/EtOAc/Et_3N ; 99/01/0,2 à 85/15/0,2) pour obtenir **e2** sous forme de poudre cristalline blanche, avec un rendement de 61%.^{1}H RMN (400 MHz, Chlorome-*d*), δ (ppm) 7.48 – 7.41 (m, 2H), 7.37 – 7.25 (m, 8H), 7.25 – 7.16 (m, 1H), 7.12 (m, 2H), 6.86 – 6.78 (m, 4H), 5.08 (q, J = 6.7 Hz, 1H), 3.79 (s, 6H), 3.09 (td, J = 6.3, 1.9 Hz, 2H), 2.72 (t, J = 7.5 Hz, 2H), 1.91 (m, 2H), 1.84 (d, J = 6.7 Hz, 3H). ^{13}C RMN (101 MHz, Chloroforme-*d*) δ (ppm) 158.77, 145.76, 142.85, 140.58, 137.02, 130.46, 129.13, 128.62, 128.16, 127.05, 126.84, 113.42, 86.17, 62.84, 59.28, 55.64, 32.69, 32.06, 26.90.



g2. Le composé a été synthétisé en suivant la même procédure que **e1**, mais en utilisant **e2** au lieu de **d1**. Il a été purifié par chromatographie sur colonne de gel de silice (*n*-hexane/EtOAc/Et_3N ; 9/1/0,2 à 7/3/0,2) pour obtenir **g2** sous forme de poudre cristalline blanche, avec un rendement de 54%. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{51}H_{65}N_2O_5^+$ 785.4888; trouvé, 785.4883. ^{1}H RMN (400 MHz, Acétone-*d6*), δ (ppm) 7.52 – 7.45 (m, 2H), 7.38 – 7.29 (m, 11H), 7.21 – 7.16 (m, 2H), 7.11 (d, J = 8.0 Hz, 2H), 6.93 – 6.84 (m, 4H), 4.76 (q, J = 6.6 Hz, 1H), 3.80 (s, 6H), 3.64 (s, 2H), 3.58 (t, J = 5.9 Hz, 2H), 3.07 (t, J = 6.3 Hz, 2H), 2.94 (t, J = 12.2 Hz, 1H), 2.72 (t, J = 7.6 Hz, 2H), 2.63 (t, J = 6.6 Hz, 2H), 1.72 – 1.38 (m, 6H), 1.43 (d, J = 6.6 Hz, 3H), 1.33 (s, 3H), 1.09 (s, 3H), 0.93 (s, 3H), 0.61 (s, 3H). ^{13}C RMN (101 MHz, Chloroforme-*d*) δ (ppm) 159.54, 146.68, 143.71, 141.66, 141.61, 137.48, 130.95, 129.57, 129.09, 129.08, 129.05, 128.63, 127.70, 127.59, 127.50, 113.91, 86.64, 83.77, 63.16, 62.37, 61.13, 60.85, 55.63, 55.36, 48.60, 42.01, 35.02, 23.66, 21.44.

Espaceur *TISP2*. Ce composé a été synthétisé en suivant la même procédure que pour ***TISP***, mais en utilisant **g2** au lieu d'**e1**. Il a été purifié par chromatographie sur colonne de gel de silice (*n*-hexane/EtOAc/DIPEA ; 95/5/2 à 65/35/2) pour obtenir ***TISP2*** sous forme de poudre cristalline blanche, avec un rendement de 36%. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{60}H_{82}N_4O_6P^+$ 985.5967; trouvé, 985.5971. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.51 – 7.43 (m, 2H), 7.39 – 7.25 (m, 10H), 7.25 – 7.15 (m, 4H), 7.10 (d, J = 8.0 Hz, 2H), 6.91 – 6.82 (m, 4H), 4.75 (q, J = 6.6 Hz, 1H), 3.78 (s, 6H), 3.84 – 3.68 (m, 2H), 3.68 – 3.57 (m, 2H), 3.63 (s, 2H), 3.07 (t, J = 6.3 Hz, 2H), 2.96 – 2.87 (m, 1H), 2.71 (m, 4H), 2.60 (m, 2H), 1.95 – 1.84 (m, 2H), 1.77 – 1.26 (m, 6H), 1.43 (d, J = 6.6 Hz, 3H), 1.32 (s, 3H), 1.25 – 1.14 (m, 12H), 1.09 (s, 3H), 0.94 (s, 3H), 0.61 (s, 3H). ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 159.41, 146.54, 143.58, 142.14, 141.52, 137.33, 137.30, 130.79, 129.16, 128.90, 128.84, 128.46, 127.45, 127.34, 118.87, 113.74, 86.49, 83.55, 63.00, 62.28, 60.97, 59.33, 55.44, 55.37, 50.10, 47.13, 43.68, 39.13, 35.11, 32.56, 25.04, 23.45, 21.31, 20.71. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 146.78.



2.3. Synthèse TISP3

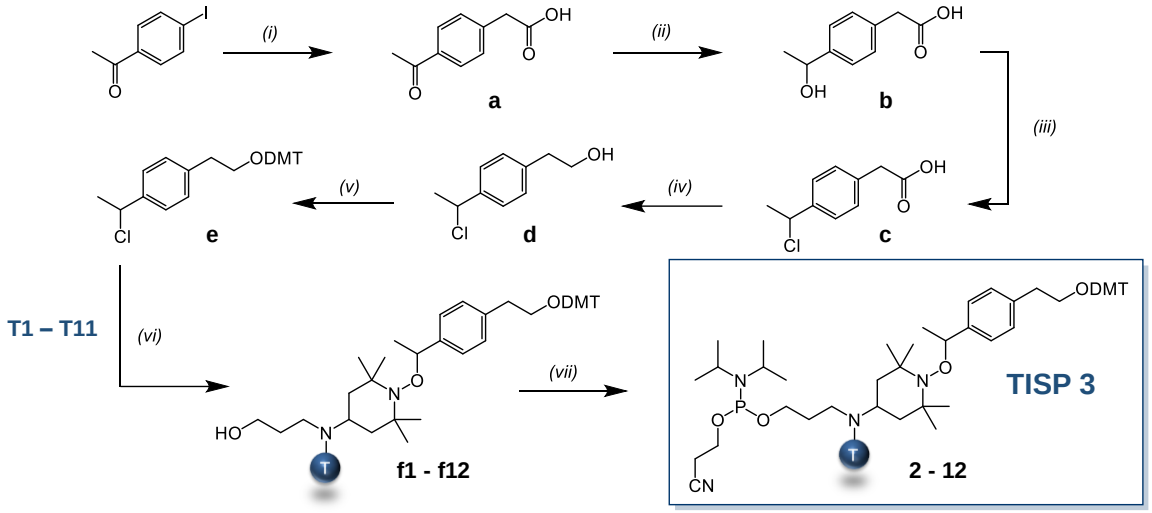
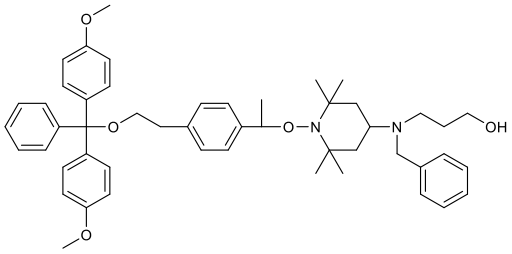


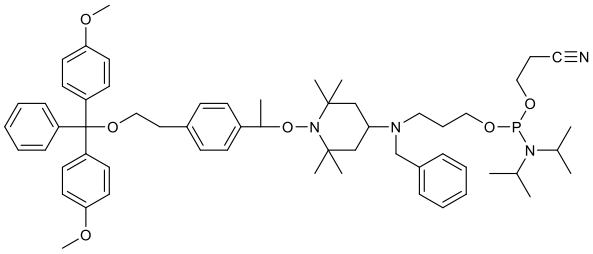
*Schéma PE.II. 3 : Approche choisie pour la synthèse de nouveaux marqueurs d'octets clivables **TISP3**.*
i) $PdCl_2$, AgOAc, NaOAc, AcOH, 130°C; (ii) $NaBH_4$, EtOH/THF, 0°C à t.a., 3h; (iii) HCl 37%, THF, 0°C à t.a., 4h; (iv) BH_3·SMe_2, THF anhydre, t.a., 17h; (v) Chlorure de diméthoxytrityle (DMT-Cl), Pyridine anhydre, THF anhydre, 0°C à t.a., 1h; (vi) CuBr, Cu(0), PMDETA, THF anhydre, t.a ., 17h; (vii) 2-Cyanoéthyle N,N-diisopropylechlorophosphoramidite, DIPEA, DCM anhydre, 0°C à t.a., 1h.

L'intermédiaire **a** a été synthétisé selon une procédure décrite dans la littérature[363] et la synthèse des intermédiaires **b**, **c**, **d** et **e** ont été synthétisés selon notre article précédent.[355] Les procédures et caractérisations des intermédiaires **a**, **b**, **c**, **d** et **e** ne sont par conséquent pas montrées ici, mais seulement les intermédiaires **f2** - **f12** et **TISP3** - **2-12.**

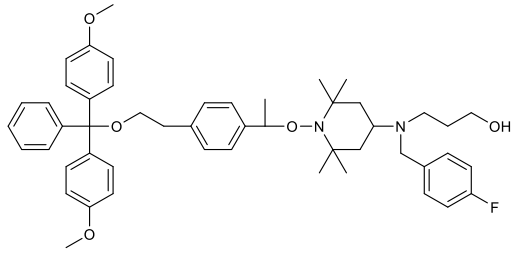


f2. Dans un ballon rond de 250 mL séché, une solution dégazée de **e** (1,565 g, 3,21 mmol, 1,0 éq) et **T2** (1,54 g, 4,82 mmol, 1,5 éq) dans du THF anhydre (33 mL) a été ajoutée goutte à goutte à une solution dégazée de CuBr (0,69 g, 4,82 mmol, 1,5 éq), Cu0 (0,31 g, 4,82 mmol, 1,5 éq) et PMDETA (2,01 mL, 9,64 mmol, 3,0 éq) dans du THF anhydre (33 mL) sous atmosphère d'argon. Le mélange réactionnel a été agité toute la nuit à température ambiante. Après 17h, et vérification de la complétion de l'ATRA par CCM (n-hexane/EtOAc/Et3N ; 5/5/0,2), **e** n'était pas complètement consommé, alors la réaction a été poursuivie pendant 24 heures supplémentaires à température ambiante sous atmosphère d'argon. Après 41 heures, le réactif de départ **e** avait presque entièrement réagi. La réaction a été arrêtée et le brut a directement été filtré sur de la Célite 545, lavé avec du THF, et évaporé sous atmosphère réduite. Le résidu vert a été lavé grâce à un mélange d'eau/saumure (1:1) et extrait avec de l'éther diéthylique (Et_2O). Les phases organiques combinées ont été séchées avec du Na_2SO_4 anhydre, filtrées, évaporées sous vide, et purifiées sur colonne de gel de silice (n-hexane/EtOAc/Et_3N ; 9/1/0,2 à 7/3/0,2) pour obtenir 2,08 g de **f2** sous forme d'une poudre cristalline blanche, avec un rendement de 82 %. HRMS *m/z*: $[M+H]^+$ calculée pour $C_{50}H_{63}N_2O_5^+$ 771.4732; trouvé, 771.4719. ^1H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.41

– 7.20 (m, 18H), 6.88 – 6.80 (m, 4H), 4.81 (q, *J* = 6.6 Hz, 1H), 3.78 (s, 6H), 3.63 (s, 2H), 3.58 (t, *J* = 5.9 Hz, 2H), 3.26 (t, *J* = 6.7 Hz, 2H), 2.99 – 2.86 (m, 1H), 2.87 (t, *J* = 6.6 Hz, 2H), 2.62 (td, *J* = 6.5, 1.9 Hz, 2H), 1.71 – 1.38 (m, 6H), 1.46 (d, *J* = 6.7 Hz, 3H), 1.34 (s, 3H), 1.09 (s, 3H), 0.94 (s, 3H), 0.69 (s, 3H); ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 159.73, 146.73, 144.46, 141.86, 139.47, 137.52, 131.16, 130.09, 129.77, 129.29, 129.26, 128.80, 127.89, 127.72, 127.70, 114.09, 86.9, 84.00, 65.76, 62.48, 61.28, 61.08, 55.86, 55.55, 49.77, 48.73, 42.18, 37.34, 35.40, 31.88, 24.02, 21.61.

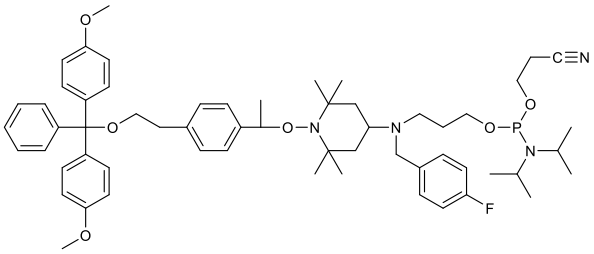


Espaceur *TISP3*-2. Une solution d'O-2-cyanoéthyl-N,N-diisopropylchlorophosphoramidite (0,94 g, 3,96 mmol, 1,5 éq) dans du DCM anhydre (40 mL) a été refroidie à 0°C, puis ajoutée goutte à goutte à une solution de **f2** (2,08 g, 2,64 mmol, 1,0 éq) et de diisopropyléthylamine (DIPEA) (2,3 mL, 13,2 mmol, 5,0 éq) dans du DCM anhydre (40 mL), également à 0°C, sous atmosphère d'argon. Le bain de glace a été retiré directement après l'ajout. Après 2 heures d'agitation à température ambiante, la réaction a été arrêtée et le mélange a été évaporé sous pression réduite pour obtenir une huile jaune, qui a ensuite directement été purifiée par chromatographie sur colonne de gel de silice (*n*-hexane/EtOAc/DIPEA ; 95/5/2 à 65/35/2) pour obtenir 1,12 g de ***TISP3*-2** sous forme de poudre cristalline blanche, avec un rendement de 45%. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{59}H_{80}N_4O_6P^+$ 971.5810; trouvé, 971.5760. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.44 – 7.13 (m, 18H), 6.88 – 6.79 (m, 4H), 4.81 (q, *J* = 6.6 Hz, 1H), 3.77 (s, 6H), 3.86 – 3.67 (m, 2H), 3.63 (s, 2H), 3.73 – 3.57 (m, 2H), 3.26 (t, *J* = 6.6 Hz, 2H), 2.93 – 2.88 (m, 1H), 2.86 (t, *J* = 6.7 Hz, 2H), 2.86 – 2.77 (m, 2H), 2.70 – 2.66 (m, 2H), 2.65 – 2.55 (m, 2H), 1.77 – 1.37 (m, 6H), 1.46 (d, *J* = 6.6 Hz, 3H), 1.33 (s, 3H), 1.20 – 1.16 (m, 12H), 1.10 (s, 3H), 0.96 (s, 3H), 0.69 (s, 3H); ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 159.31, 146.31, 144.06, 142.06, 139.02, 137.10, 130.74, 129.68, 129.30, 129.11, 128.84, 128.80, 128.39, 127.31, 127.29, 118.80, 113.68, 86.53, 83.54, 65.35, 62.39, 60.86, 60.68, 59.14, 55.47, 55.30, 50.00, 47.06, 43.60, 42.09, 36.94, 35.03, 24.98, 23.62, 21.23, 20.80. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 146.83.

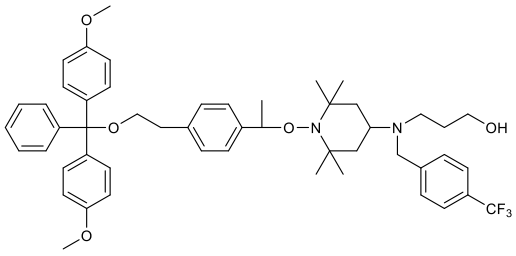


f3. Même procédure que pour **f2** mais en utilisant **T3** au lieu de **T2** et agité pour 24h. Le composé **f3** était obtenu sous forme de poudre blanche avec 75% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{50}H_{62}N_2O_5F^+$ 789.4637; trouvé, 789.4643. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.36 – 7.32 (m, 4H), 7.30 – 7.12 (m, 11H), 7.09 – 7.00 (m, 2H), 6.86 – 6.78 (m, 4H), 4.78 (q, *J* = 6.6 Hz, 1H), 3.76 (s, 6H, *d*), 3.60 (s, 2H), 3.54 (t, *J* = 6.0 Hz, 2H), 3.24 (t, *J* = 6.7 Hz, 2H), 2.91 (m, 1H), 2.85 (t, *J* = 6.7 Hz, 2H), 2.65 – 2.55 (m, 2H), 1.65 – 1.45 (m, 6H), 1.44 (d, *J* = 6.7 Hz, 3H), 1.32 (s, 3H), 1.08 (s, 3H), 0.94 (s, 3H), 0.67 (s, 3H); ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 163.63, 161.22, 159.22, 146.25, 143.96, 138.98, 137.39, 137.05, 131.01, 130.69, 129.64, 128.79, 128.35,

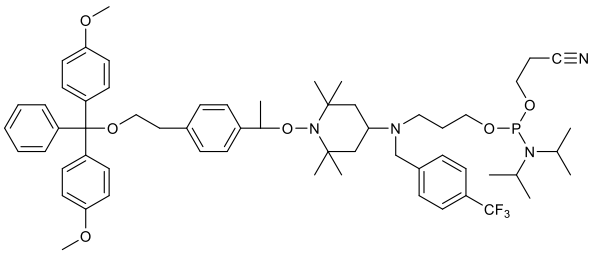
127.24, 115.58, 113.63, 86.50, 83.56, 65.27, 61.94, 60.82, 60.63, 55.39, 54.25, 49.42, 48.23, 41.74, 36.92, 34.99, 23.66, 21.25. ^{19}F RMN (376 MHz, Acétone-*d6*) δ (ppm) -117.97.



Espaceur *TISP3*-3. Même procédure que pour ***TISP3*-2** mais en utilisant **f3** au lieu de **f2**. Le composé ***TISP3*-3** était obtenu sous forme de poudre blanche avec 40% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{59}H_{79}N_4O_6PF^+$ 989.5716; trouvé, 989.5717. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.43 – 7.34 (m, 4H), 7.31 – 7.15 (m, 11H), 7.10 – 7.00 (m, 2H), 6.88 – 6.79 (m, 4H), 4.80 (q, *J* = 6.7 Hz, 1H), 3.87 – 3.72 (m, 2H), 3.77 (s, 6H), 3.75 – 3.56 (m, 2H), 3.62 (s, 3H), 3.26 (t, *J* = 6.7 Hz, 2H), 3.01 (p, *J* = 6.6 Hz, 1H), 2.93 – 2.78 (m, 4H), 2.78 – 2.68 (m, 2H), 2.66 – 2.53 (m, 2H), 1.78 – 1.37 (m, 6H), 1.46 (d, *J* = 6.6 Hz, 3H), 1.33 (s, 3H), 1.18 (ddd, *J* = 6.6, 5.6, 1.0 Hz, 12H), 1.11 (s, 3H), 0.98 (s, 3H), 0.69 (s, 3H). ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 161.71, 159.83, 146.83, 144.57, 139.56, 138.64, 137.63, 131.27, 131.23, 130.14, 129.33, 128.84, 127.79, 119.33, 115.95, 114.12, 87.05, 84.05, 65.83, 62.86, 61.37, 61.18, 59.56, 55.83, 54.91, 50.54, 49.09, 47.53, 44.09, 42.90, 37.38, 35.51, 31.76, 25.41, 24.04, 21.68, 21.13. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 146.81. ^{19}F RMN (376 MHz, Acétone-*d6*) δ (ppm) -118.12.

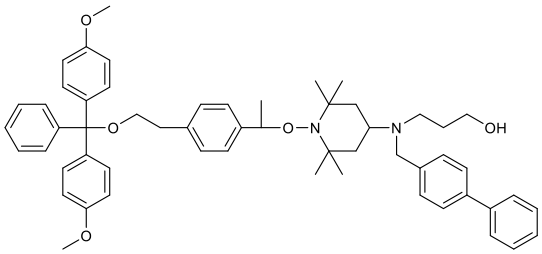


f4. Même procédure que pour **f2** mais en utilisant **T4** au lieu de **T2** et agité pour 72h. Le composé **f4** était obtenu sous forme de poudre blanche avec 89% de rendement (comparé à **f2** où le temps de réaction était de 41h). HRMS *m/z*: $[M+H]^+$ calculé pour $C_{51}H_{62}N_2O_5F_3^+$ 839.4605; trouvé, 839.4613. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.67 – 7.54 (m, 4H), 7.40 – 7.32 (m, 2H), 7.29 – 7.13 (m, 11H), 6.86 – 6.78 (m, 4H), 4.79 (q, J = 6.6 Hz, 1H), 3.75 (s, 6H), 3.72 (s, 2H), 3.54 (t, J = 6.0 Hz, 2H), 3.24 (t, J = 6.7 Hz, 2H), 2.96 – 2.88 (m, 1H), 2.85 (t, J = 6.6 Hz, 2H), 2.61 (m, 2H), 1.69 – 1.38 (m, 6H), 1.44 (d, J = 6.6 Hz, 3H), 1.32 (s, 3H), 1.10 (s, 3H), 0.96 (s, 3H), 0.67 (s, 3H). ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 159.06, 146.78, 146.06, 143.77, 138.82, 136.86, 130.46, 130.45, 129.38, 128.56, 128.07, 127.01, 126.96, 125.43, 125.39, 113.35, 86.27, 83.34, 65.05, 61.09, 60.59, 60.38, 55.05, 54.34, 49.88, 41.72, 36.60, 34.69, 23.28, 20.88. ^{19}F RMN (376 MHz, Acétone-*d6*) δ (ppm) -62.71.

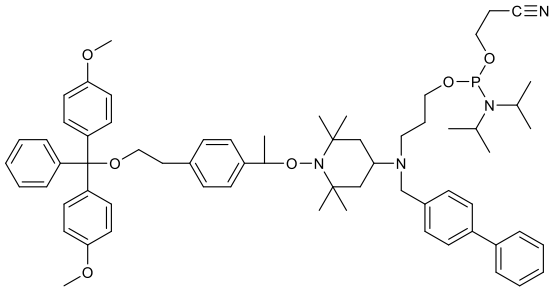


Espaceur *TISP3*-4. Même procédure que pour ***TISP3*-2** mais en utilisant **f4** au lieu de **f2**. Le composé ***TISP3*-4** était obtenu sous forme de poudre blanche avec 46% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{60}H_{79}N_4O_6PF_3^+$ 1039.5684; trouvé, 1039.5694. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.62 (q, J = 8.3 Hz, 4H), 7.41 – 7.33 (m, 2H), 7.30 – 7.23

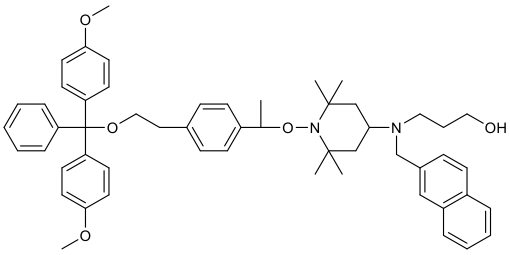
(m, 8H), 7.22 – 7.14 (m, 3H), 6.86 – 6.79 (m, 4H), 4.80 (q, J = 6.6 Hz, 1H), 3.88 – 3.73 (m, 2H), 3.77 (s, 6H), 3.73 (s, 2H), 3.71 – 3.56 (m, 2H), 3.25 (t, J = 6.6 Hz, 2H), 2.97 – 2.88 (m, 1H), 2.86 (t, J = 6.7 Hz, 2H), 2.76 – 2.68 (m, 2H), 2.63 (m, 2H), 1.81 – 1.40 (m, 6H), 1.46 (d, J = 6.6 Hz, 3H), 1.33 (s, 3H), 1.18 (d, J = 1.0 Hz, 3H), 1.19 – 1.10 (m, 12H), 0.98 (s, 3H), 0.69 (s, 3H).^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 159.31, 147.34, 146.33, 144.04, 139.06, 137.13, 130.74, 129.69, 129.48, 128.85, 128.37, 127.30, 127.26, 125.70, 125.66, 118.78, 113.66, 86.58, 83.62, 65.33, 62.18, 60.88, 60.70, 59.03, 55.38, 54.78, 50.62, 48.75, 43.63, 36.97, 35.12, 24.99, 23.67, 21.34, 20.67. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 146.92. ^{19}F RMN (376 MHz, Acétone-*d6*) δ (ppm) -62.42.



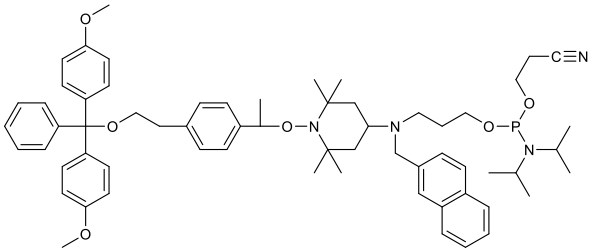
f5. Même procédure que pour **f2** mais en utilisant **T5** au lieu de **T2** et agité pour 72h. Le composé **f5** était obtenu sous forme de poudre blanche avec 75% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{56}H_{67}N_2O_5^+$ 847.5045; trouvé, 847.5029. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.69 – 7.55 (m, 4H), 7.44 (m, 4H), 7.40 – 7.30 (m, 3H), 7.29 – 7.18 (m, 11H), 6.86 – 6.79 (m, 4H), 4.82 (q, J = 6.7 Hz, 1H), 3.76 (s, 6H), 3.67 (s, 2H), 3.59 (t, J = 5.9 Hz, 2H), 3.25 (t, J = 6.6 Hz, 2H), 2.98 (m, 1H), 2.86 (t, J = 6.6 Hz, 2H), 2.65 (t, J = 6.6 Hz, 2H), 1.75 – 1.39 (m, 6H), 1.46 (d, J = 6.7 Hz, 3H), 1.34 (s, 3H), 1.11 (s, 3H), 0.97 (s, 3H), 0.70 (s, 3H). ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 159.37, 146.38, 144.11, 141.56, 140.85, 140.16, 139.11, 137.19, 130.80, 130.78, 129.93, 129.72, 129.64, 128.90, 128.42, 127.52, 127.40, 127.34, 127.31, 113.71, 86.62, 83.66, 65.39, 62.11, 60.96, 60.77, 55.42, 54.81, 48.49, 41.90, 36.97, 35.08, 23.69, 21.30.



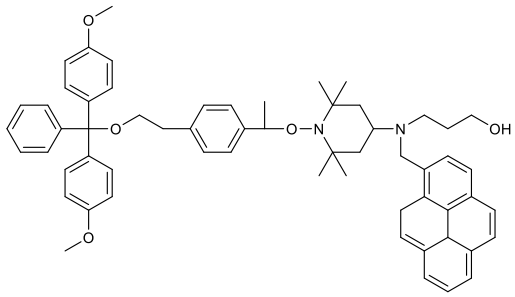
Espaceur *TISP3*-5. Même procédure que pour ***TISP3*-2** mais en utilisant **f5** au lieu de **f2**. Le composé ***TISP3*-5** était obtenu sous forme de poudre blanche avec 34% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{65}H_{84}N_4O_6P^+$ 1047.6123; trouvé, 1047.6107. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.69 – 7.63 (m, 2H), 7.63 – 7.57 (m, 2H), 7.49 – 7.41 (m, 4H), 7.40 – 7.29 (m, 3H), 7.29 – 7.17 (m, 11H), 6.86 – 6.80 (m, 4H), 4.81 (q, J = 6.6 Hz, 1H), 3.85 – 3.70 (m, 2H), 3.76 (s, 6H), 3.68 (s, 2H), 3.70 – 3.57 (m, 2H), 3.25 (t, J = 6.7 Hz, 2H), 3.01 – 2.91 (m, 1H), 2.89 – 2.78 (m, 4H), 2.75 – 2.65 (m, 2H), 2.66 – 2.52 (m, 2H), 1.80 – 1.39 (m, 6H), 1.46 (d, J = 6.6 Hz, 3H), 1.34 (s, 3H), 1.17 (m, 12H), 1.13 (s, 3H), 0.99 (s, 3H), 0.70 (s, 3H). ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 159.63, 146.63, 144.39, 141.91, 141.76, 140.25, 139.36, 137.42, 131.03, 129.95, 129.92, 129.86, 129.14, 128.65, 128.13, 127.74, 127.59, 127.56, 127.54, 119.14, 113.93, 86.85, 83.86, 65.64, 62.70, 61.20, 61.01, 59.38, 55.64, 55.19, 50.51, 47.51, 43.90, 37.18, 35.32, 25.23, 23.86, 21.54, 21.00. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 146.80.



f6. Même procédure que pour **f2** mais en utilisant **T6** au lieu de **T2** et agité pour 72h. Le composé **f6** était obtenu sous forme de poudre blanche avec 69% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{54}H_{65}N_2O_5^+$ 821.4888; trouvé, 821.4895. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.90 – 7.79 (m, 4H), 7.57 – 7.41 (m, 3H), 7.41 – 7.33 (m, 2H), 7.27 – 7.14 (m, 11H), 6.87 – 6.78 (m, 4H), 4.80 (q, J = 6.6 Hz, 1H), 3.79 (s, 2H), 3.76 (s, 6H), 3.57 (t, J = 5.9 Hz, 2H), 3.25 (t, J = 6.6 Hz, 2H), 3.00 (m, 1H), 2.86 (m, 2H), 2.67 (td, J = 6.6, 1.9 Hz, 2H), 1.82 – 1.40 (m, 6H), 1.45 (d, J = 6.7 Hz, 3H), 1.34 (s, 3H), 1.07 (s, 3H), 0.93 (s, 3H), 0.69 (s, 3H). ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 159.45, 146.45, 144.18, 139.32, 139.19, 137.25, 134.39, 133.71, 130.85, 129.76, 128.95, 128.61, 128.46, 128.41, 127.87, 127.85, 127.40, 127.36, 126.78, 126.30, 113.74, 86.66, 83.70, 65.44, 61.99, 61.01, 60.81, 55.44, 55.38, 48.42, 41.96, 36.98, 35.08, 23.66, 21.25.

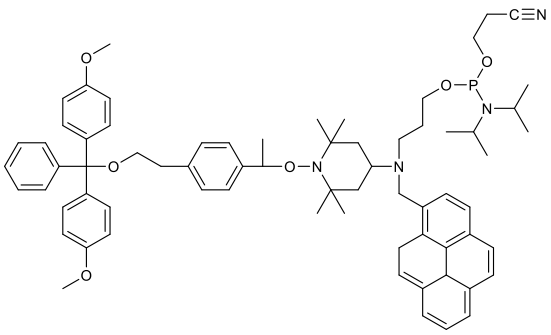


Espaceur *TISP3*-6. Même procédure que pour ***TISP3*-2** mais en utilisant **f6** au lieu de **f2**. Le composé ***TISP3*-6** était obtenu sous forme de poudre blanche avec 53% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{63}H_{82}N_4O_6P^+$ 1021.5967; trouvé, 1021.5964. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.89 – 7.79 (m, 4H), 7.65 – 7.34 (m, 5H), 7.30 – 7.11 (m, 11H), 6.87 – 6.78 (m, 4H), 4.81 (q, *J* = 6.6 Hz, 1H), 3.79 (s, 2H), 3.76 (s, 6H), 3.80 – 3.53 (m, 4H), 3.25 (t, *J* = 6.6 Hz, 2H), 3.07 – 2.92 (m, 2H), 2.85 (t, *J* = 6.7 Hz, 3H), 2.72 – 2.57 (m, 4H), 1.81 – 1.47 (m, 6H), 1.45 (d, *J* = 6.7 Hz, 3H), 1.34 (s, 3H), 1.14 (m, 12H), 1.10 (s, 3H), 0.96 (s, 3H), 0.70 (s, 3H). ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 158.58, 145.61, 143.35, 139.04, 138.30, 136.41, 133.58, 132.87, 130.03, 128.97, 128.14, 127.73, 127.67, 127.62, 127.06, 126.68, 126.57, 125.90, 125.39, 118.06, 112.96, 85.86, 82.87, 64.63, 61.64, 60.20, 60.03, 58.32, 54.70, 49.55, 47.93, 46.55, 42.88, 36.26, 34.46, 24.22, 23.01, 20.66, 19.94. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 146.61.

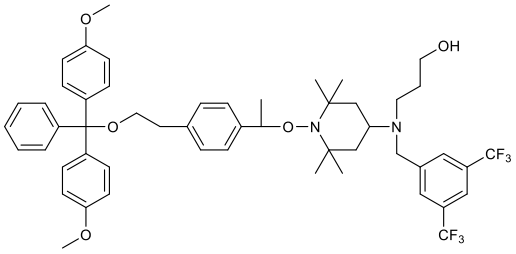


f7. Même procédure que pour **f2** mais en utilisant **T7** au lieu de **T2** et agité pour 72h. It était obtenu sous forme de poudre blanche avec 71% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{60}H_{67}N_2O_5^+$ 895.5045; trouvé, 895.5073. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 8.62 (d, J = 9.3 Hz, 1H), 8.28 – 8.18 (m, 3H), 8.21 – 8.07 (m, 4H), 8.03 (t, J = 7.6 Hz, 1H), 7.40 – 7.32 (m, 2H), 7.29 – 7.12 (m, 11H), 6.86 – 6.77 (m, 4H), 4.79 (q, J = 6.6 Hz, 1H), 4.37 (s, 2H), 3.74 (s, 6H), 3.52 (t, J = 6.1 Hz, 2H), 3.24 (t, J = 6.6 Hz, 2H), 3.06 – 2.96 (m, 1H), 2.84 (t, J = 6.7 Hz, 2H), 2.74 (td, J = 6.7, 2.1 Hz, 2H), 1.80 – 1.70 (m, 1H), 1.74 – 1.63 (m, 2H), 1.86 – 1.52 (m, 6H), 1.43 (d, J = 6.7 Hz, 3H), 1.33 (s, 3H), 0.99 (s,

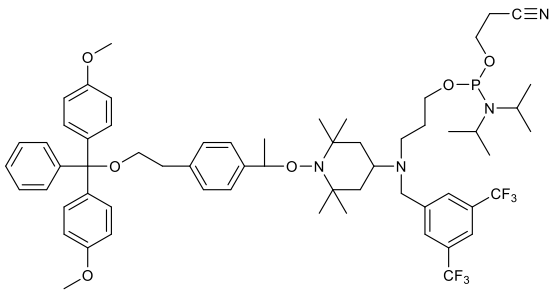
3H), 0.85 (s, 3H), 0.69 (s, 3H). ^{13}C RMN (400 MHz, Acétone-*d6*) δ (ppm) 159.43, 146.43, 144.16, 139.16, 137.24, 137.22, 135.07, 132.28, 131.85, 131.40, 130.84, 130.82, 130.39, 129.74, 128.94, 128.45, 128.40, 127.77, 127.38, 127.34, 126.85, 125.83, 125.63, 125.53, 125.37, 124.86, 113.73, 86.65, 83.68, 65.42, 61.52, 61.01, 60.81, 55.43, 53.71, 48.18, 42.02, 36.97, 35.11, 23.66, 21.25.



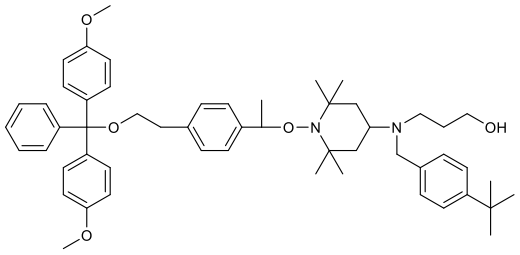
Espaceur *TISP3*-7. Même procédure que pour ***TISP3*-2** mais en utilisant **f7** au lieu de **f2**. Le composé ***TISP3*-7** était obtenu sous forme de poudre blanche avec 61% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{69}H_{84}N_4O_6P^+$ 1095.6123; trouvé, 1095.6128. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 8.65 (d, J = 9.3 Hz, 1H), 8.32 – 8.01 (m, 8H), 7.41 – 7.34 (m, 2H), 7.30 – 7.22 (m, 8H), 7.21 – 7.17 (m, 3H), 6.87 – 6.79 (m, 4H), 4.81 (q, J = 6.6 Hz, 1H), 4.39 (s, 2H), 3.76 (s, 6H), 3.72 – 3.45 (m, 4H), 3.25 (t, J = 6.7 Hz, 2H), 3.07 – 2.97 (m, 1H), 2.88 – 2.83 (m, 3H), 2.78 – 2.70 (m, 2H), 2.62 – 2.55 (m, 2H), 1.93 – 1.52 (m, 6H), 1.45 (d, J = 6.7 Hz, 3H), 1.36 (s, 3H), 1.23 – 1.01 (m, 12H), 1.03 (s, 3H), 0.89 (s, 3H), 0.71 (s, 3H). ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 159.44, 146.44, 144.17, 139.17, 137.24, 135.28, 132.29, 131.89, 131.42, 130.84, 130.42, 129.75, 128.95, 128.91, 128.45, 128.42, 127.72, 127.40, 127.35, 126.84, 125.83, 125.80, 125.67, 125.55, 125.38, 124.99, 118.91, 113.74, 86.66, 83.66, 65.44, 62.37, 61.03, 60.84, 59.08, 55.44, 54.07, 50.67, 47.34, 43.64, 36.99, 35.16, 24.77, 23.64, 21.29, 20.66. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 146.62.



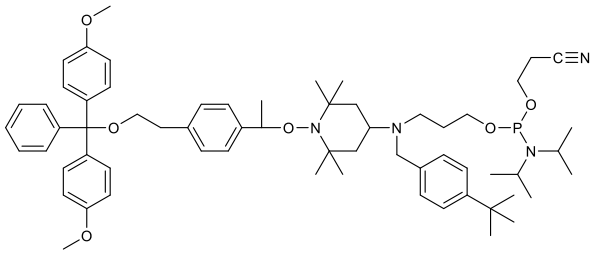
f8. Même procédure que pour **f2** mais en utilisant **T8** au lieu de **T2** et agité pour 72h. Le composé **f8** était obtenu sous forme de poudre blanche avec 84% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{52}H_{61}N_2O_5F_6^+$ 907.4479; trouvé, 907.4504. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 8.02 (s, 2H), 7.89 (s, 1H), 7.41 – 7.33 (m, 2H), 7.29 – 7.19 (m, 11H), 6.87 – 6.79 (m, 4H), 4.80 (q, *J* = 6.7 Hz, 1H), 3.87 (s, 2H), 3.77 (s, 3H), 3.55 (t, *J* = 6.1 Hz, 2H), 3.25 (t, *J* = 6.6 Hz, 2H), 3.01 – 2.88 (m, 1H), 2.86 (t, *J* = 6.6 Hz, 2H), 2.69 – 2.60 (m, 2H), 1.73 – 1.39 (m, 6H), 1.46 (d, *J* = 6.6 Hz, 3H), 1.33 (s, 3H), 1.13 (s, 3H), 0.98 (s, .3H), 0.69 (s, 3H); ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 159.14, 146.18, 145.91, 143.85, 138.93, 136.97, 131.36, 130.58, 129.53, 128.70, 128.21, 127.13, 123.00, 120.92, 113.50, 86.42, 83.51, 65.17, 60.82, 60.71, 60.30, 55.19, 53.93, 50.29, 41.91, 36.80, 34.86, 23.50, 21.07. ^{19}F RMN (376 MHz, Acétone-*d6*) δ (ppm) -63.28.



Espaceur *TISP3*-8. Même procédure que pour ***TISP3*-2** mais en utilisant **f8** au lieu de **f2**. Le composé ***TISP3*-8** était obtenu sous forme de poudre blanche avec 27% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{61}H_{78}N_4O_6PF_6^+$ 1107.5558; trouvé, 1107.5574. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 8.02 (s, 2H), 7.89 (s, 1H), 7.41 – 7.33 (m, 2H), 7.28 – 7.18 (m, 11H), 6.90 – 6.75 (m, 4H), 4.80 (q, *J* = 6.6 Hz, 1H), 3.89 (s, 2H), 3.76 (s, 6H), 3.77 – 3.52 (m, 4H), 3.25 (t, *J* = 6.7 Hz, 2H), 2.92 (m, 1H), 3.07 – 2.78 (m, 2H), 2.86 (t, *J* = 6.7 Hz, 2H), 2.74 – 2.59 (m, 4H), 1.78 – 1.40 (m, 6H), 1.46 (d, *J* = 6.6 Hz, 3H), 1.34 (s, 3H), 1.21 – 1.09 (m, 12H), 1.13 (s, 3H), 0.97 (s, 3H), 0.69 (s, 3H). ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 159.37, 146.39, 144.07, 139.13, 137.18, 130.79, 129.73, 129.36, 128.90, 128.41, 127.34, 127.29, 125.93, 123.23, 121.16, 118.79, 113.70, 86.62, 83.69, 65.38, 62.30, 60.91, 60.72, 59.11, 55.40, 54.30, 50.88, 47.75, 43.54, 36.99, 35.08, 24.87, 23.67, 21.28, 20.74. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 147.03. ^{19}F RMN (376 MHz, Acétone-*d6*) δ (ppm) -63.23.

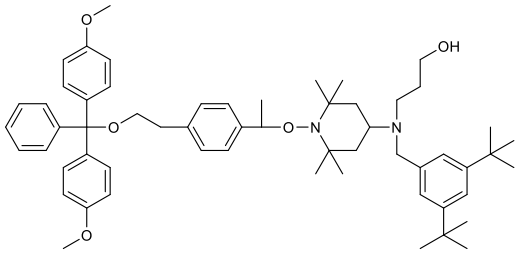


f9. Même procédure que pour **f2** mais en utilisant **T9** au lieu de **T2** et agité pour 42h. Le composé **f9** était obtenu sous forme de poudre blanche avec 65% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{54}H_{71}N_2O_5^+$ 827.5358; trouvé, 827.5357. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.41 – 7.32 (m, 4H), 7.31 – 7.07 (m, 13H), 6.87 – 6.80 (m, 4H), 4.80 (q, *J* = 6.7 Hz, 1H), 3.76 (s, 6H), 3.59 (s, 2H), 3.56 (m, 2H), 3.26 (t, *J* = 6.6 Hz, 2H), 2.99 – 2.87 (m, 1H), 2.86 (t, *J* = 6.6 Hz, 2H), 2.65 – 2.58 (m, 2H), 1.70 – 1.40 (m, 6H), 1.46 (d, *J* = 6.7 Hz, 3H), 1.33 (s, 3H), 1.09 (s, 3H), 0.95 (s, 3H), 0.69 (s, 3H); ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 158.59, 149.40, 145.60, 143.33, 138.32, 137.50, 136.39, 130.03, 130.01, 128.97, 128.38, 128.13, 127.68, 126.58, 124.94, 112.96, 85.82, 82.86, 64.63, 61.54, 60.15, 59.97, 54.75, 53.96, 48.53, 41.06, 36.23, 34.30, 34.11, 30.96, 22.95, 20.51.

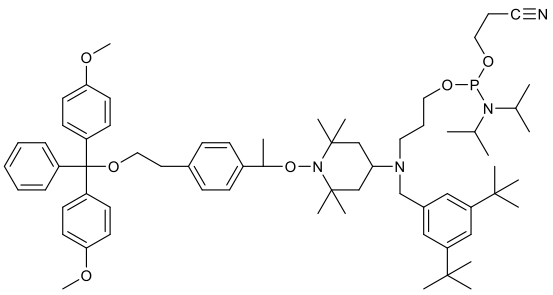


Espaceur *TISP3*-9. Même procédure que pour ***TISP3*-2** mais en utilisant **f9** au lieu de **f2**. Le composé ***TISP3*-9** était obtenu sous forme de poudre blanche avec 44% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{63}H_{88}N_4O_6P^+$ 1027.6436; trouvé, 1027.6412. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.41 – 7.29 (m, 4H), 7.29 – 7.06 (m, 13H), 6.86 – 6.80 (m, 4H), 4.80 (q, *J* = 6.6 Hz, 1H), 3.84 – 3.75 (m, 2H), 3.76 (s, 6H), 3.63 (m, 2H), 3.59 (s, 2H), 3.25 (t, *J* = 6.6 Hz, 2H), 3.01 (m, 1H), 2.97 – 2.77 (m, 2H), 2.86 (t, *J* = 6.6 Hz, 2H), 2.70 (m, 2H), 2.59 (m, 2H), 1.76 – 1.40 (m, 6H), 1.46 (d, *J* = 6.6 Hz, 3H), 1.33 (s, 3H), 1.21 – 1.16 (m, 12H), 1.11 (s, 3H), 0.98 (s, 3H), 0.69 (s, 3H); ^{13}C RMN (101 MHz,

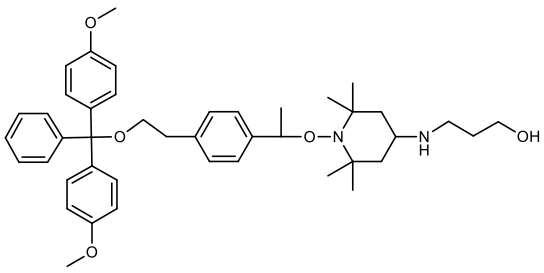
Acétone-*d6*) δ (ppm) 158.58, 149.13, 145.59, 143.35, 138.29, 138.20, 136.39, 130.01, 128.95, 128.26, 128.12, 128.07, 127.66, 126.56, 124.83, 118.05, 112.95, 85.83, 82.82, 64.63, 61.75, 60.14, 59.98, 58.60, 54.72, 54.12, 49.27, 46.44, 42.88, 36.23, 34.40, 34.09, 30.99, 24.09, 22.95, 20.55, 20.04. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 146.81.



f10. Même procédure que pour **f2** mais en utilisant **T10** au lieu de **T2** et agité pour 42h. Le composé **f10** était obtenu sous forme de poudre blanche 41% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{58}H_{79}N_2O_5^+$ 883.5984; trouvé, 883.5991. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.40 – 7.30 (m, 3H), 7.28 – 7.13 (m, 14H), 6.86 – 6.77 (m, 4H), 4.78 (q, J = 6.6 Hz, 1H), 3.75 (s, 6H), 3.67 – 3.56 (m, 4H), 3.24 (t, J = 6.7 Hz, 2H), 2.93 – 2.87 (m, 1H), 2.84 (t, J = 6.8 Hz, 2H), 2.63 (td, J = 6.2, 2.6 Hz, 2H), 1.73 – 1.40 (m, 6H), 1.43 (d, J = 6.6 Hz, 3H), 1.31 (s, 3H), 1.29 (s, 18H), 1.01 (s, 3H), 0.87 (s, 3H), 0.66 (s, 3H).; ^{13}C RMN (400 MHz, Acétone-*d6*) δ (ppm) 159.05, 150.81, 146.06, 143.79, 139.79, 138.78, 136.84, 130.45, 129.37, 128.55, 128.06, 126.97, 126.95, 123.29, 120.86, 113.35, 86.26, 83.30, 65.05, 62.35, 60.61, 60.43, 55.44, 55.05, 48.49, 41.29, 36.60, 34.93, 34.68, 31.44, 23.30, 20.92.

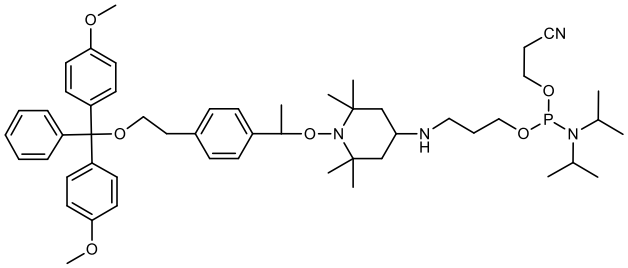


Espaceur *TISP3*-10. Même procédure que pour ***TISP3*-2** mais en utilisant **f10** au lieu de **f2**. Le composé ***TISP3*-10** était obtenu sous forme de poudre blanche avec 24% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{67}H_{96}N_4O_6P^+$ 1083.7062; trouvé, 1083.7064. ^{1}H RMN (400 MHz, Acétone-*d6*) δ (ppm) 7.42 – 7.30 (m, 3H), 7.30 – 7.14 (m, 13H), 6.86 – 6.79 (m, 4H), 4.80 (q, J = 6.6 Hz, 1H), 3.87 – 3.75 (m, 2H), 3.75 (s, 6H), 3.78 – 3.53 (m, 2H), 3.65 (s, 2H), 3.26 (t, J = 6.6 Hz, 2H), 3.02 – 3.05 (m, 1H), 2.94 – 2.80 (m, 4H), 2.68 (q, J = 6.2 Hz, 2H), 2.61 (p, J = 6.1 Hz, 2H), 1.81 – 1.38 (m, 6H), 1.46 (d, J = 6.6 Hz, 3H), 1.33 (s, 3H), 1.33 (s, 18H), 1.20 (dd, J = 6.8, 1.2 Hz, 12H), 1.06 (s, 3H), 0.92 (s, 3H), 0.70 (s, 3H). ^{13}C RMN (101 MHz, Acétone-*d6*) δ (ppm) 159.33, 150.93, 146.35, 144.11, 140.87, 139.02, 137.13, 130.75, 129.68, 128.86, 128.38, 127.27, 123.29, 120.92, 118.78, 113.67, 86.58, 83.55, 65.37, 62.44, 60.92, 60.78, 59.36, 56.07, 55.40, 49.53, 46.57, 43.65, 36.96, 35.25, 35.14, 31.88, 24.84, 23.69, 21.35, 20.71. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 146.96.

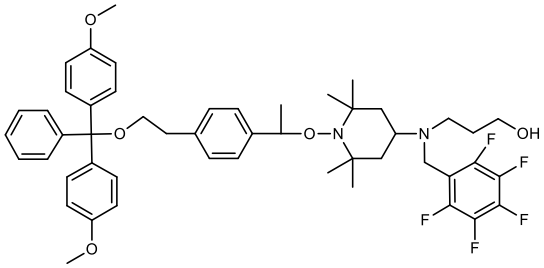


f11. Même procédure que pour **f2** mais en utilisant **T1** au lieu de **T2**. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (EtOAc/MeOH/TEA ; 80/20/2 à 50/50/2) et le composé **f11** a été obtenu sous forme de poudre blanche avec 70% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{43}H_{57}N_2O_5^+$ 681.4262 ; trouvé, 681.4279. ^{1}H RMN (400 MHz, Acétone-d_6) δ 7.42 – 7.35 (m, 2H), 7.29 – 7.20 (m, 11H), 6.88 – 6.81 (m, 4H), 4.82 (q, *J* = 6.6 Hz, 1H), 3.78 (s, 6H),

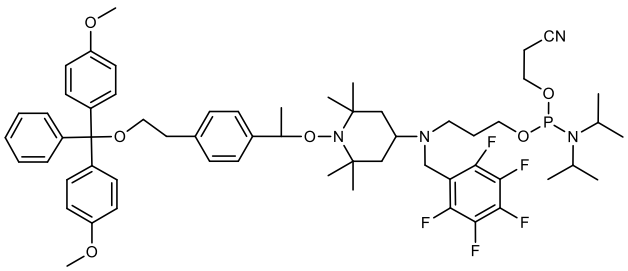
3.63 (t, *J* = 5.8 Hz, 2H), 3.27 (t, *J* = 6.7 Hz, 2H), 2.88 (t, *J* = 6.6 Hz, 3H), 2.77 (t, *J* = 6.2 Hz, 2H), 1.48 (d, *J* = 6.6 Hz, 3H), 1.70 – 1.09 (m, 6H), 1.33 (s, 3H), 1.23 (s, 3H), 1.08 (s, 3H), 0.69 (s, 3H). ^{13}C RMN (101 MHz, Acétone) δ 159.40, 146.41, 144.18, 139.12, 137.20, 130.81, 129.73, 128.91, 128.43, 127.35, 127.32, 113.72, 86.63, 83.66, 65.43, 62.65, 60.53, 60.36, 55.43, 48.87, 47.43, 46.20, 36.97, 34.99, 33.11, 23.71, 21.51.



Espaceur *TISP3*-11. Même procédure que pour ***TISP3*-2** mais en utilisant **f11** au lieu de **f2**. Le produit brut a été purifié par chromatographie sur gel de silice (EtOAc/MeOH/TEA ; 99/1/2 à 90/10/2) et le composé ***TISP3*-11** était obtenu sous forme de poudre blanche avec 10% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{52}H_{74}N_4O_6P^+$ 881.5341; trouvé, 881.5347. ^{1}H RMN (400 MHz, Acétone-d_6) δ 7.42 – 7.34 (m, 2H), 7.28 – 7.18 (m, 11H), 6.86 – 6.79 (m, 4H), 4.82 (q, *J* = 6.6 Hz, 1H), 4.03 – 3.77 (m, 2H), 3.76 (s, 6H), 3.77 – 3.60 (m, 2H), 3.26 (t, *J* = 6.6 Hz, 2H), 2.86 (t, *J* = 6.6 Hz, 2H), 2.91 – 2.58 (m, 6H), 1.86 – 1.60 (m, 2H), 1.47 (dd, *J* = 6.7, 2.0 Hz, 3H), 1.32 (s, 3H), 1.37 – 1.11 (m, 4H), 1.22 (s, 3H), 1.23 – 1.15 (m, 12H), 1.09 (s, 3H), 0.69 (s, 3H). ^{13}C RMN (101 MHz, Acétone) δ 159.21, 146.22, 144.03, 138.92, 137.01, 130.61, 129.54, 128.72, 128.24, 127.14, 118.77, 113.52, 86.44, 83.46, 65.24, 62.48, 60.34, 59.02, 55.24, 48.58, 45.85, 44.05, 43.51, 36.78, 34.82, 24.72, 23.55, 21.35, 20.52. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 146.79.



f12. Même procédure que pour **f2** mais en utilisant **T11** au lieu de **T2**. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (*n*-hexane/EtOAc/TEA ; 80/20/2 à 60/40/2) et le composé **f12** a été obtenu sous forme de poudre blanche avec 72 % de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{50}H_{58}N_2O_5F_5^+$ 861.4260 ; trouvé, 861.4260. ^{1}H RMN (400 MHz, Acétone-d_6) δ 7.42 – 7.34 (m, 2H), 7.31 – 7.21 (m, 8H), 7.24 – 7.14 (m, 3H), 6.87 – 6.79 (m, 4H), 4.82 (q, *J* = 6.6 Hz, 1H), 3.76 (s, 6H), 3.76 (s, 2H), 3.47 (t, *J* = 4.8 Hz, 2H), 3.26 (t, *J* = 6.7 Hz, 2H), 2.94 (dt, *J* = 11.3, 7.0 Hz, 1H), 2.87 (d, *J* = 6.6 Hz, 1H), 2.62 (t, *J* = 6.7 Hz, 2H), 1.65 – 1.55 (m, 4H), 1.48 (d, *J* = 8.5 Hz, 2H), 1.47 (d, *J* = 6.6 Hz, 3H), 1.35 (s, 3H), 1.18 (s, 3H), 1.04 (s, 3H), 0.70 (s, 3H). ^{13}C RMN (101 MHz, Acétone) δ 159.64, 146.64, 144.34, 139.42, 137.44, 137.43, 131.04, 131.03, 129.97, 129.14, 128.65, 127.61, 127.55, 113.93, 86.86, 83.94, 65.64, 61.26, 61.13, 60.92, 55.63, 51.04, 47.82, 42.18, 37.19, 35.28, 23.85, 21.43.



Espaceur *TISP3*-12. Même procédure que pour ***TISP3*-2** mais en utilisant **f12** au lieu de **f2**. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (*n*-hexane/EtOAc/TEA ; 90/10/2 à 60/40/2) et le composé ***TISP3*-12** était obtenu sous forme de poudre blanche avec 10% de rendement (le réactif de départ ne semblait pas avoir totalement été consommé après 2 h de réaction, et un sous-produit correspondant au phosphite a pu être observé). HRMS *m/z*: $[M+H]^+$ calculé pour $C_{52}H_{74}N_4O_6F_5P^+$ 1061.5339; trouvé, 1061.5340. ^{1}H NMR (400 MHz, Acétone-d_6) δ 7.38 – 7.31 (m, 2H), 7.27 – 7.10 (m, 11H), 6.83 – 6.75 (m, 4H), 4.79 (q, *J* = 6.6 Hz, 1H), 3.86 – 3.67 (m, 2H), 3.74 (dd, *J* = 2.8, 1.4 Hz, 2H), 3.72 (s, 6H), 3.58 (ddd, *J* = 13.7, 6.7, 3.2 Hz, 2H), 3.23 (t, *J* = 6.6 Hz, 2H), 2.83 (t, *J* = 6.6 Hz, 2H), 2.63 – 2.54 (m, 2H), 1.71 – 1.38 (m, 6H), 1.44 (d, *J* = 6.6 Hz, 3H), 1.32 (s, 3H), 1.20 – 1.13 (m, 12H), 1.12 (s, 3H), 1.03 – 0.92 (s, 3H), 0.67 (s, 3H).^{13}C NMR (101 MHz, Acétone) δ 159.39, 146.40, 144.08, 139.17, 137.20, 130.80, 129.74, 128.91, 128.43, 127.38, 127.35, 127.32, 118.85, 113.71, 86.64, 83.67, 65.41, 62.06, 60.90, 60.71, 59.13, 55.42, 50.89, 46.57, 43.66, 43.53, 36.98, 35.11, 24.86, 23.63, 21.29, 20.68. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 147.02.

2.4. Synthèse TISP4

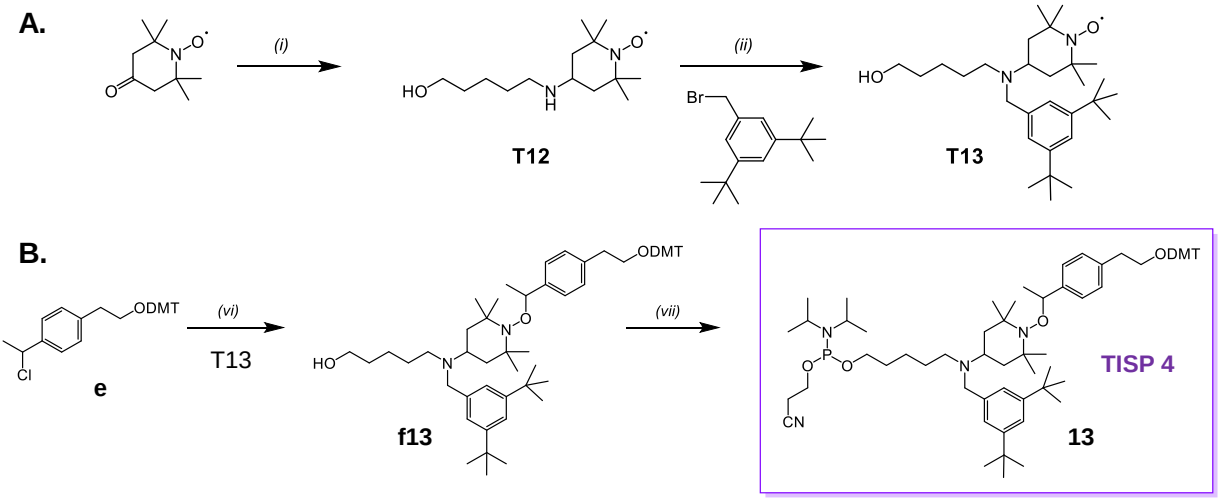
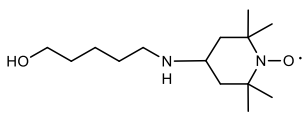
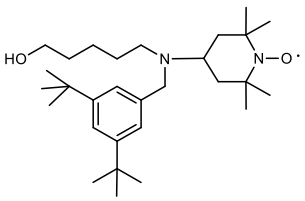


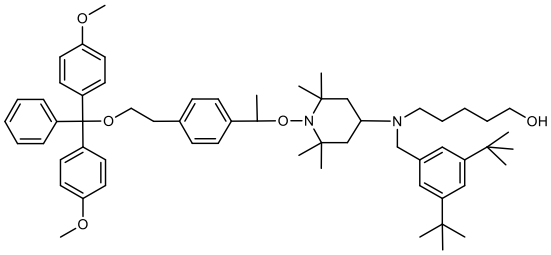
*Schéma PE.II. 4 : Stratégie de synthèse de **A. T13** et **B**. de l'Espaceur **TISP4**.*
i) $Na(CH_3COO)_3BH$, CH3COOH, amino propanol, 1,2-dichloroéthane, t.a., 17h; ii) Acétonitrile, K_2CO_3, 83°C, 17h ; (vi) CuBr, Cu(0), PMDETA, THF, t.a ., 17h; (vii) 2-Cyanoéthyle N,N-diisopropylechlorophosphoramidite, DIPEA, DCM anhydre, 0°C à t.a., 1h.



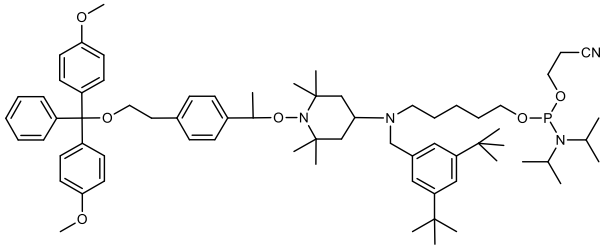
T12. Le protocole est le même que pour **T1** mais avec amino-1-hexanol au lieu de amino-1-propanol. **T12** a été obtenu sous forme de solide orange avec un rendement de 83%. HRMS m/z : $[M+H]^+$ calculé pour $C_{15}H_{32}N_2O_2^{\bullet+}$ 272.2458 ; trouvé 272.2457.



T13. Le protocole est le même que pour **T10** mais en utilisant **T12** au lieu de **T1**. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (cyclohexane/EtOAc : 9/1 à 1/1 avec 1% d'Et_3N) pour obtenir 1,61 g de **T13** sous forme d'huile orange avec un rendement de 93%. HRMS m/z : $[M+H]^+$ calculé pour $C_{30}H_{64}N_2O_2^{\bullet+}$ 474.4180 ; trouvé 474.4181.



f13. Même procédure que pour **f2** mais en utilisant **T13** au lieu de **T2**. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (*n*-heptane/EtOAc/TEA : 95/5/2 à 50/50/2) pour obtenir 0.6 g de **f13** sous forme de poudre blanche avec 53% de rendement. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{58}H_{79}N_2O_5^+$ 883.5984; trouvé, 883.5991. ^{1}H NMR (500 MHz, Acétone) δ 7.37 (d, *J* = 7.6 Hz, 2H), 7.31 (d, *J* = 1.9 Hz, 1H), 7.26 – 7.16 (m, 13H), 6.85 – 6.80 (m, 4H), 4.79 (q, *J* = 6.6 Hz, 1H), 3.76 (s, 6H), 3.50 (q, *J* = 6.6 Hz, 2H), 3.36 (t, *J* = 5.4 Hz, 1H), 3.25 (t, *J* = 6.7 Hz, 2H), 2.86 (d, *J* = 6.3 Hz, 2H), 2.45 (td, *J* = 6.8, 2.6 Hz, 2H), 1.62 – 1.34 (m, 15H), 1.30 (d, *J* = 1.4 Hz, 21H), 1.05 (s, 3H), 0.91 (s, 3H), 0.67 (s, 3H). ^{13}C NMR (126 MHz, Acétone) δ 159.75, 151.25, 146.76, 144.55, 141.47, 139.46, 137.54, 131.15, 130.06, 129.25, 128.76, 127.66, 123.70, 121.24, 114.04, 86.96, 83.97, 65.76, 62.78, 61.33, 61.16, 56.05, 55.75, 50.12, 42.53, 37.28, 35.60, 35.42, 34.20, 32.89, 32.16, 28.20, 26.96, 24.03, 21.66.



Espaceur *TISP4*. Même procédure que pour **Espaceur *TISP3*-2** mais en utilisant **f13** au lieu de **f2**. Le produit brut a été purifié par chromatographie sur colonne de gel de silice (*n*-heptane/EtOAc/TEA ; 95/5/1 à 70/30/1) et le composé **TISP4** était obtenu sous forme de poudre blanche avec 20 % de rendement. Une fraction impure correspondant au phosphite a pu être isolée (Figure PE.3. 83). Afin d'améliorer les rendements obtenus, du *n*-hexane et de la DIPEA seraient adéquat pour une meilleure purification. HRMS *m/z*: $[M+H]^+$ calculé pour $C_{70}H_{102}N_4O_6P^+$ 1125.7532 ; trouvé, 1125.7517. ^{1}H NMR (500 MHz, Acétone) δ 7.40 – 7.34 (m, 2H), 7.31 (t, *J* = 1.9 Hz, 1H), 7.25 (s, 2H), 7.27 – 7.14 (m, 11H), 6.86 – 6.79 (m, 4H), 4.79 (q, *J* = 6.6 Hz, 1H), 3.89 – 3.74 (m, 2H), 3.76 (s, 6H), 3.70 – 3.56 (m, 3H), 3.60 (s, 2H), 3.25 (t, *J* = 6.7 Hz, 2H), 2.89 – 2.80 (m, 4H), 2.72 (t, *J* = 6.1 Hz, 2H), 2.46 (td, *J* = 6.7, 2.6 Hz, 2H), 1.64 – 1.25 (m, 14H), 1.44 (d, *J* = 6.7 Hz, 3H), 1.32 (s, 3H), 1.30 (d, *J* = 2.0 Hz, 18H), 1.18 (dd, *J* = 6.8, 5.4 Hz, 12H), 1.06 (s, 3H), 0.91 (s, 3H), 0.68 (s, 3H). ^{13}C NMR (126 MHz, Acétone) δ 159.28, 150.78, 146.29, 144.08, 141.00, 138.99, 137.08, 130.68, 130.68, 129.59, 128.78, 128.29, 127.20, 127.19, 123.24, 120.78, 118.81, 113.58, 86.49, 83.50, 65.29, 64.01, 60.86, 60.70, 59.23, 55.62, 55.28, 49.68, 43.53, 42.09, 36.82, 35.13, 34.94, 31.91, 31.72, 27.52, 26.49, 24.80, 23.57, 21.21, 20.57. ^{31}P RMN (162 MHz, Acétone-*d6*) δ (ppm) 146.96.

3. Spectres RMN

3.1. Intermédiaires TISP et TISP2

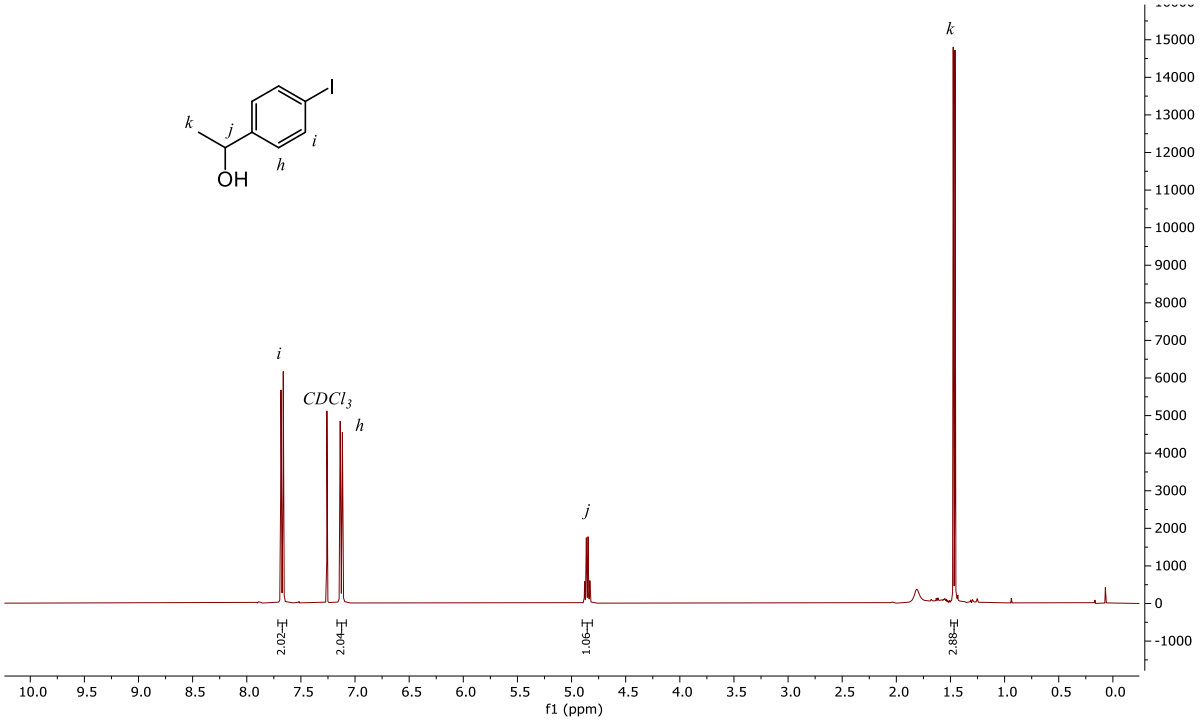


*Figure PE.3. 1 : Spectre RMN ^{1}H du composé **a1**.*

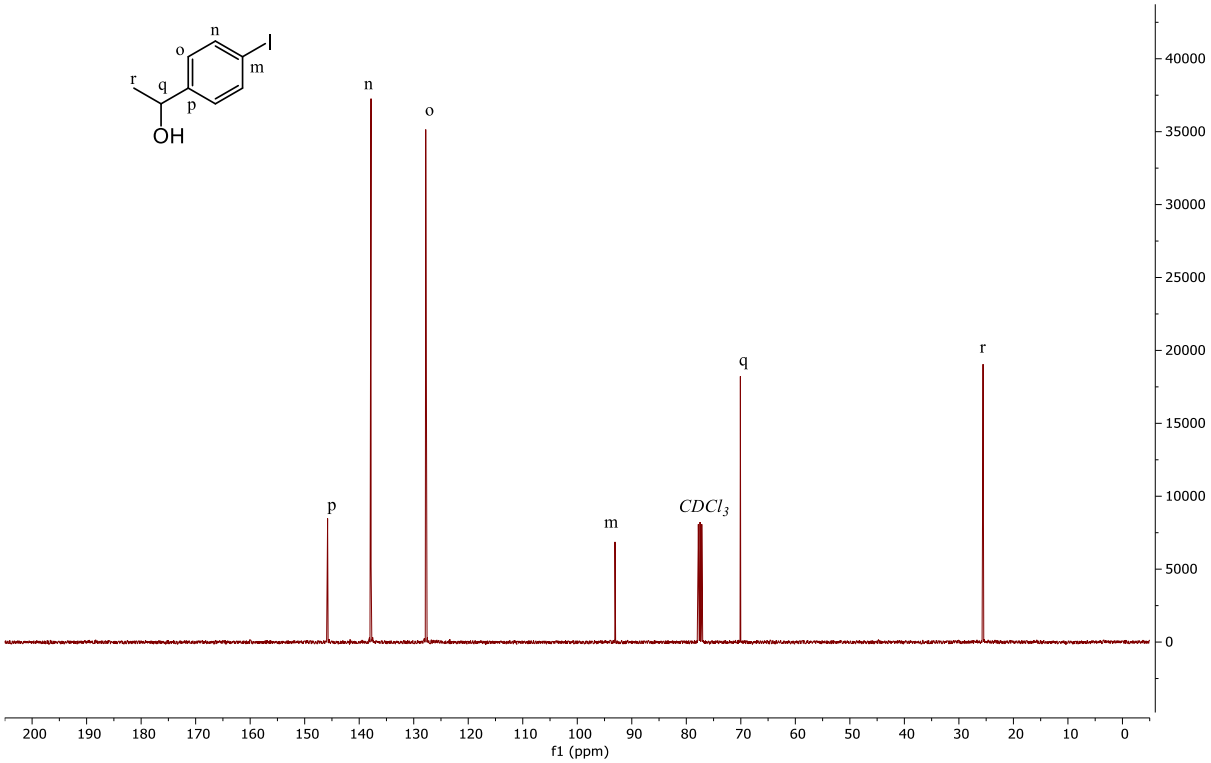


*Figure PE.3. 2 : Spectre RMN ^{13}C du composé **a1.***

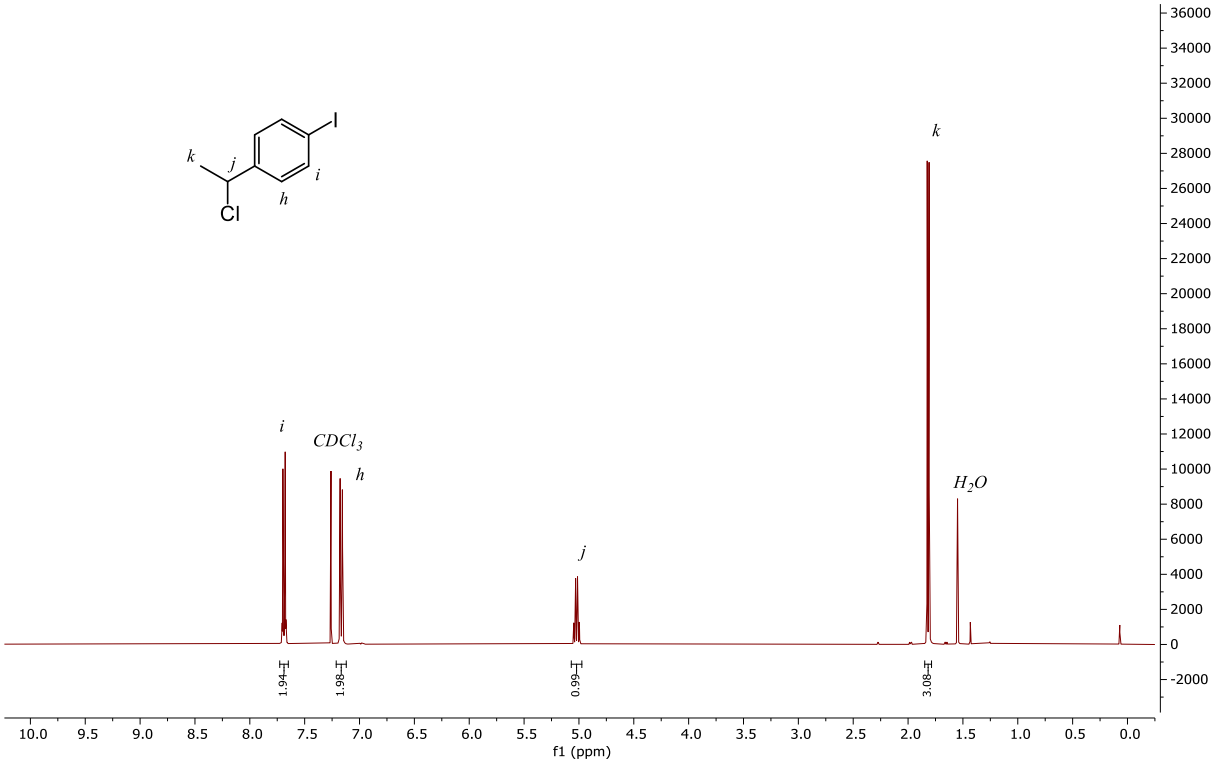


*Figure PE.3. 3 : Spectre RMN ^{1}H du composé **b1.***

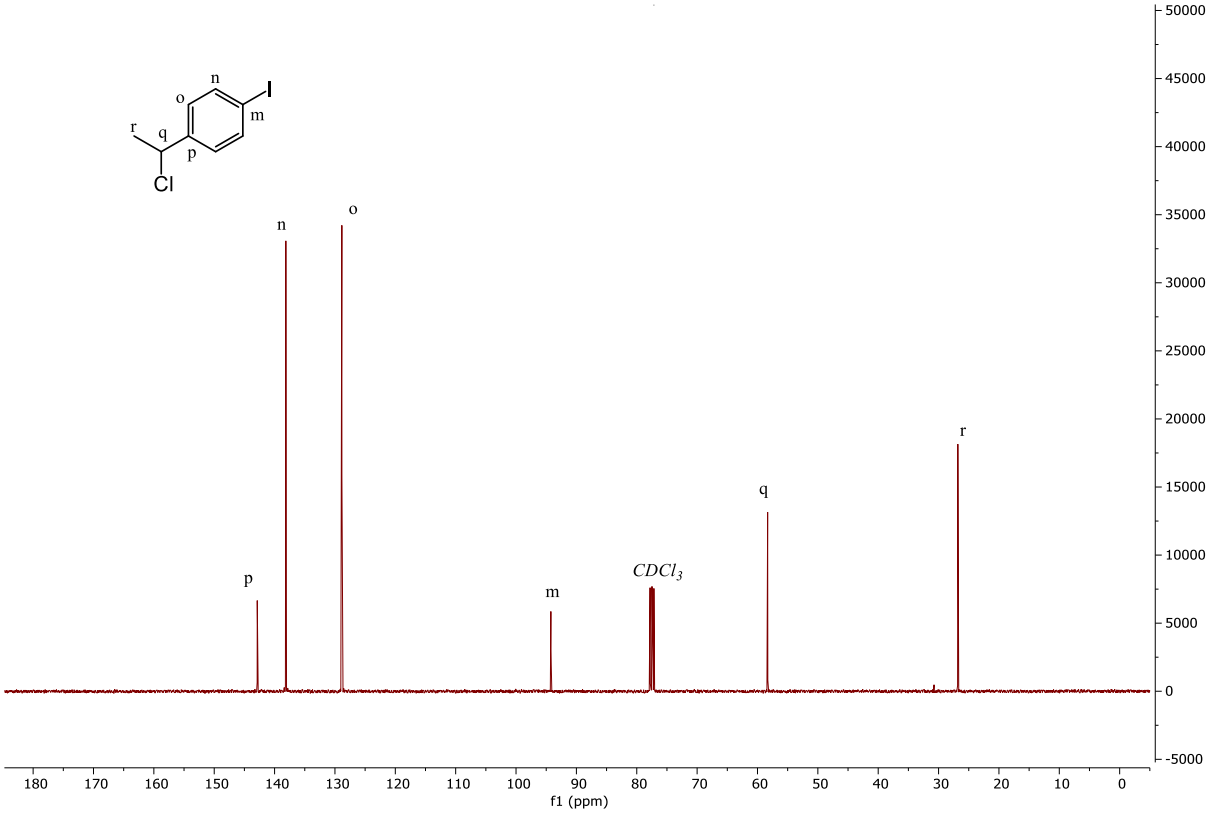


*Figure PE.3. 4 : Spectre RMN ^{13}C du composé **b1.***

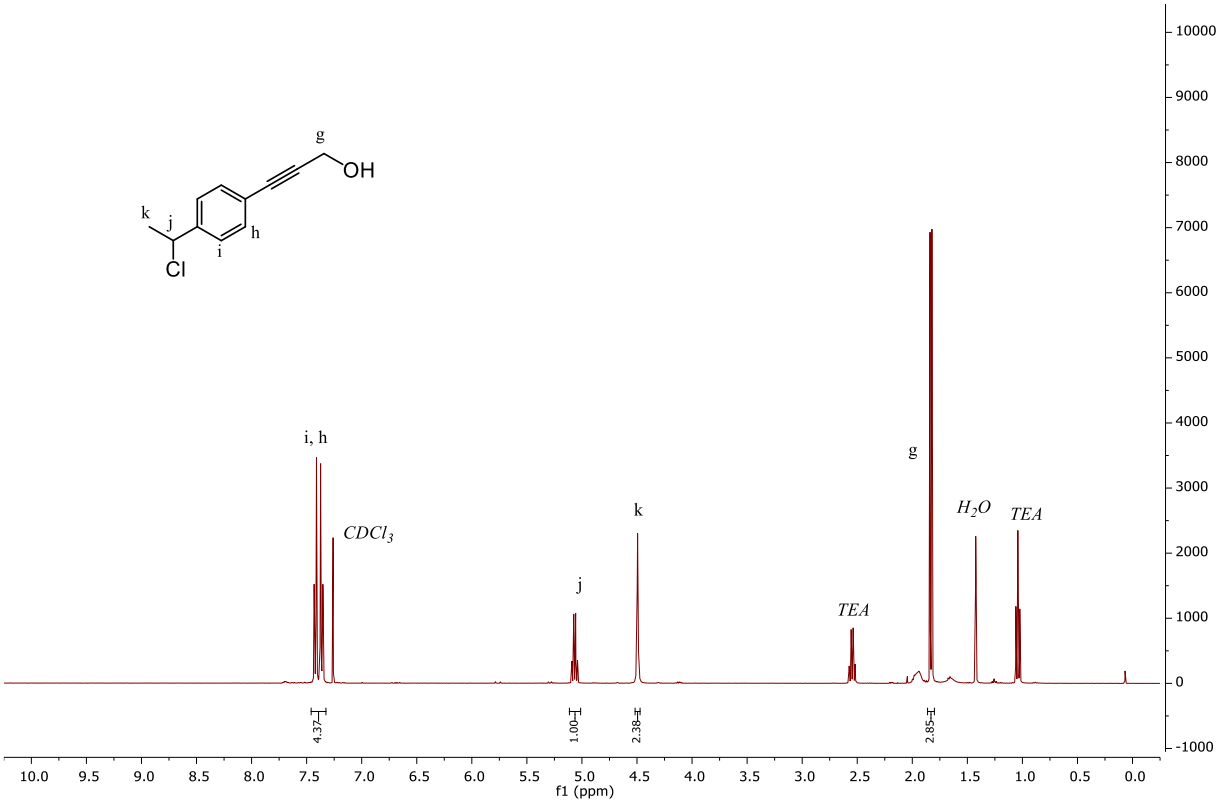


*Figure PE.3. 5 : Spectre RMN ^{1}H du composé **c1**.*

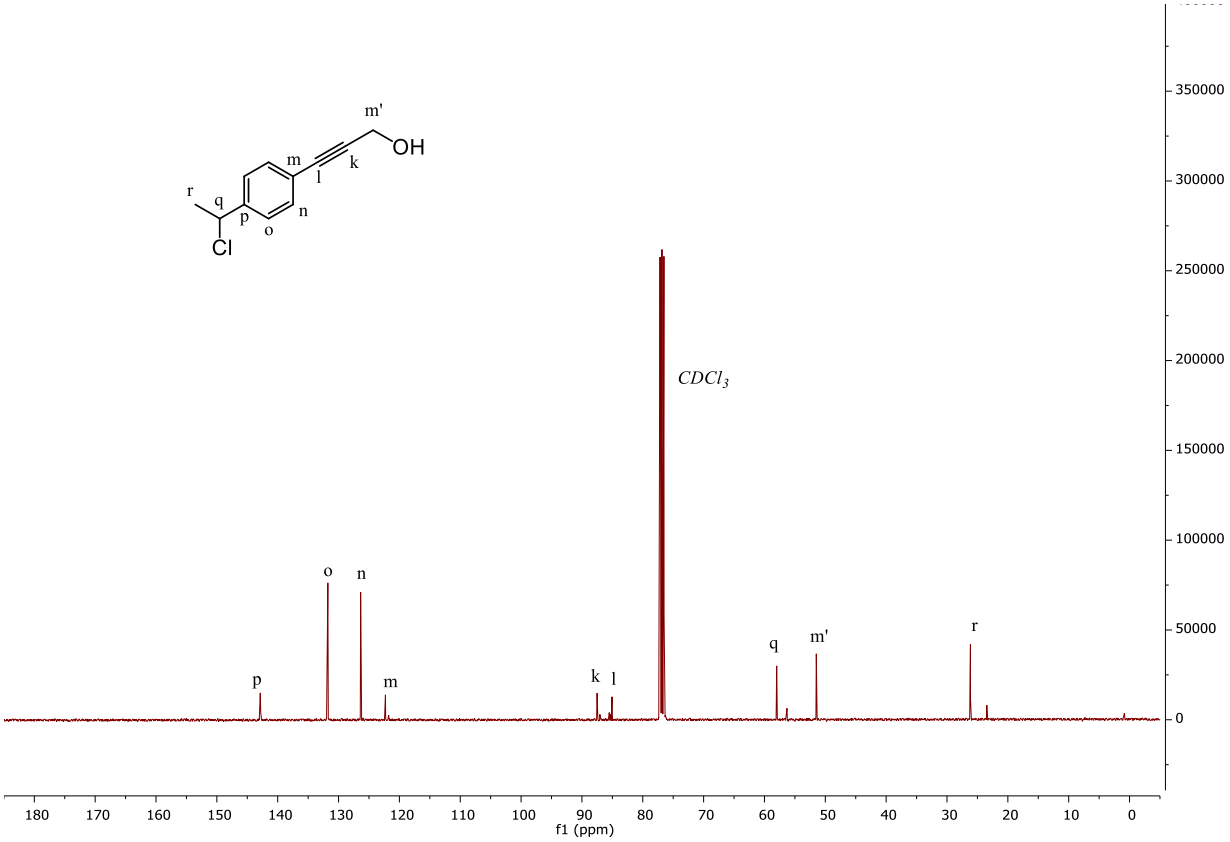


*Figure PE.3. 6 : Spectre RMN ^{13}C du composé **c1**.*

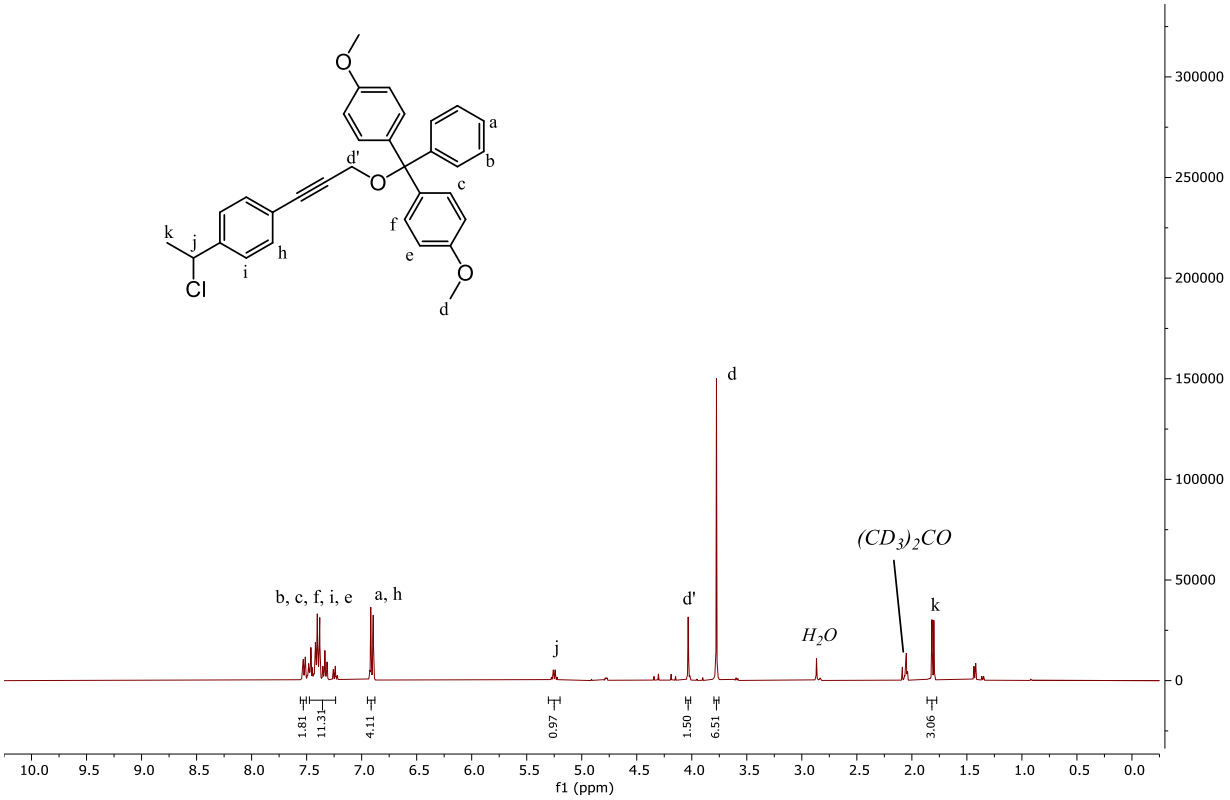


*Figure PE.3. 7 : Spectre RMN ^{1}H du composé **d1**.*

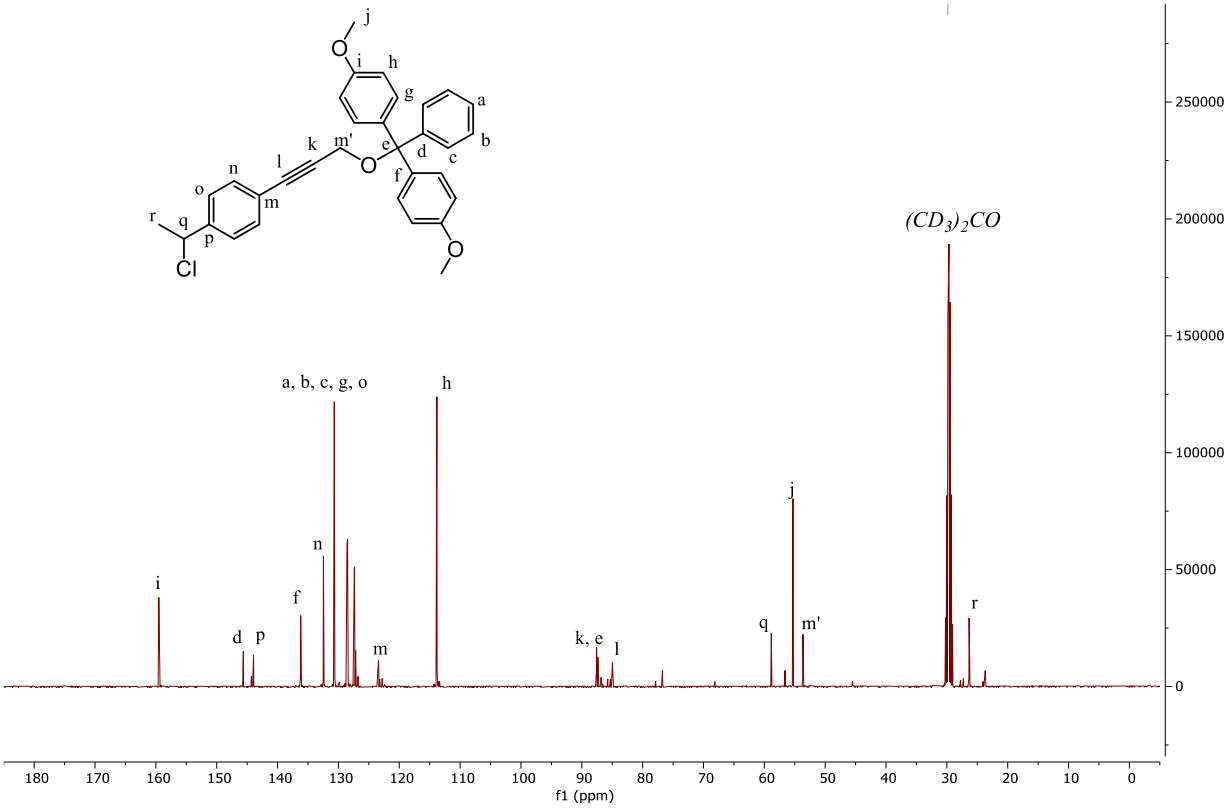


*Figure PE.3. 8 : Spectre RMN ^{13}C du composé **d1**.*

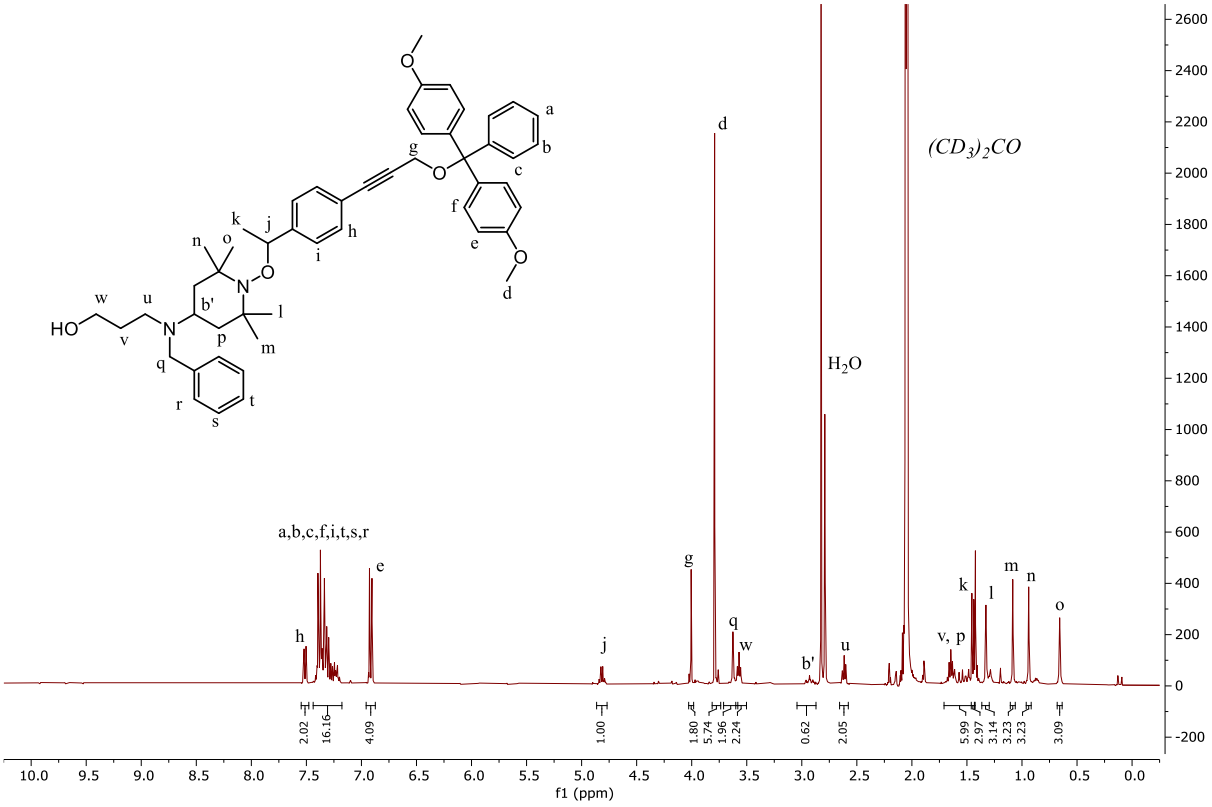


*Figure PE.3. 9 : Spectre RMN ^{1}H du composé **e1**.*

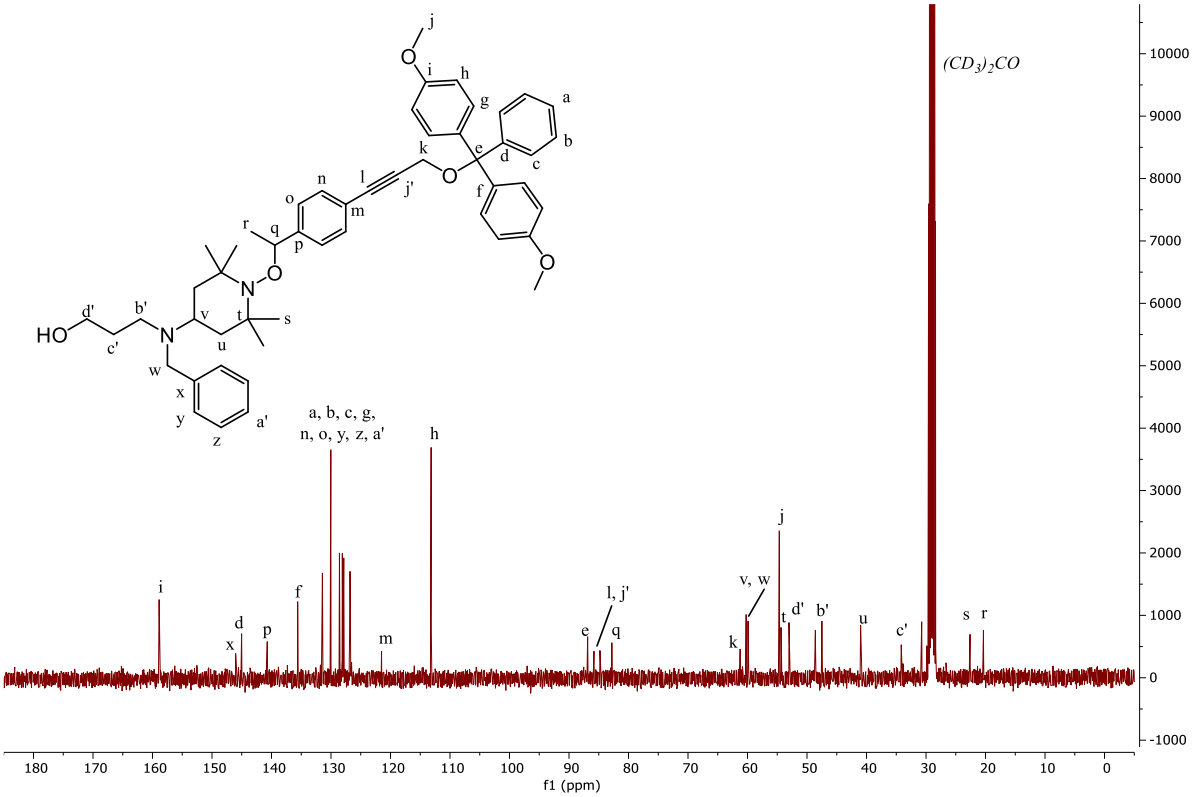


*Figure PE.3. 10 : Spectre RMN ^{13}C du composé **e1**.*

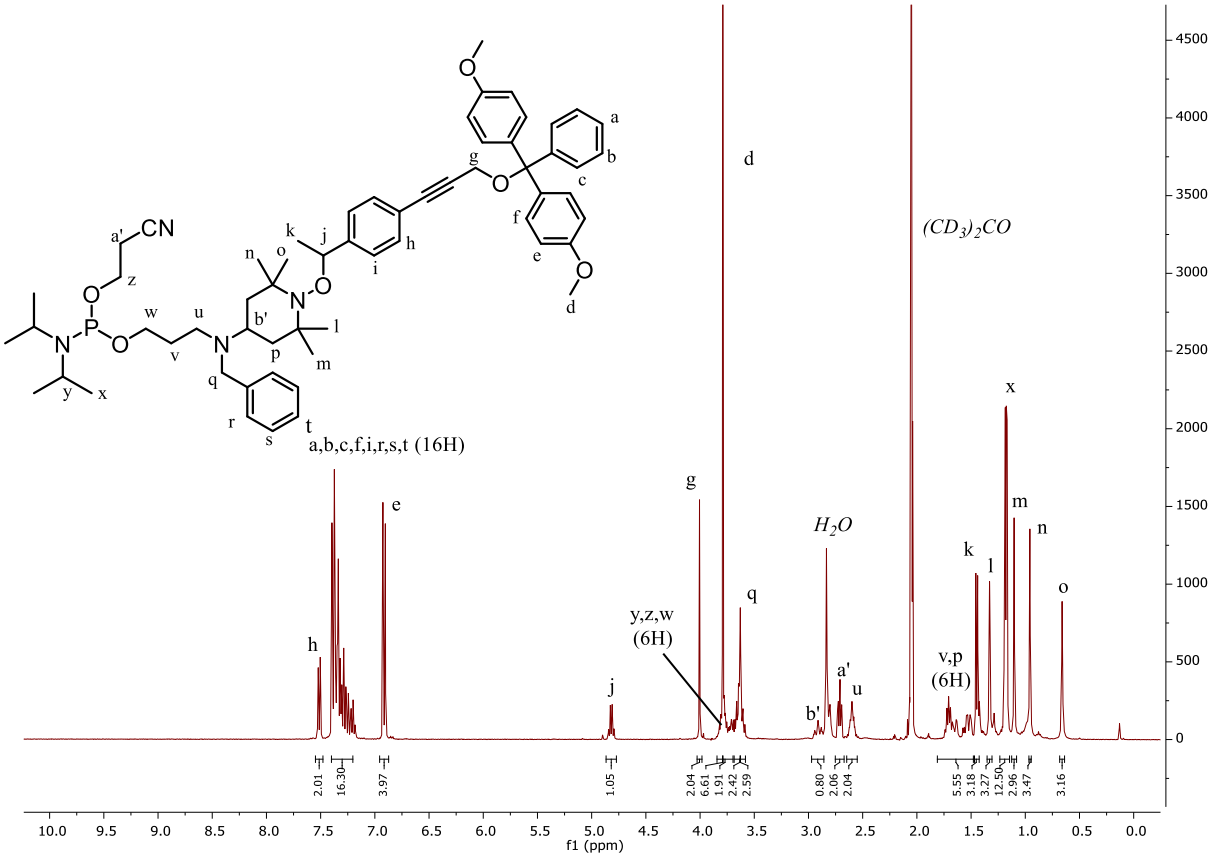


*Figure PE.3. 11 : Spectre RMN ^{1}H du composé **TISP**.*

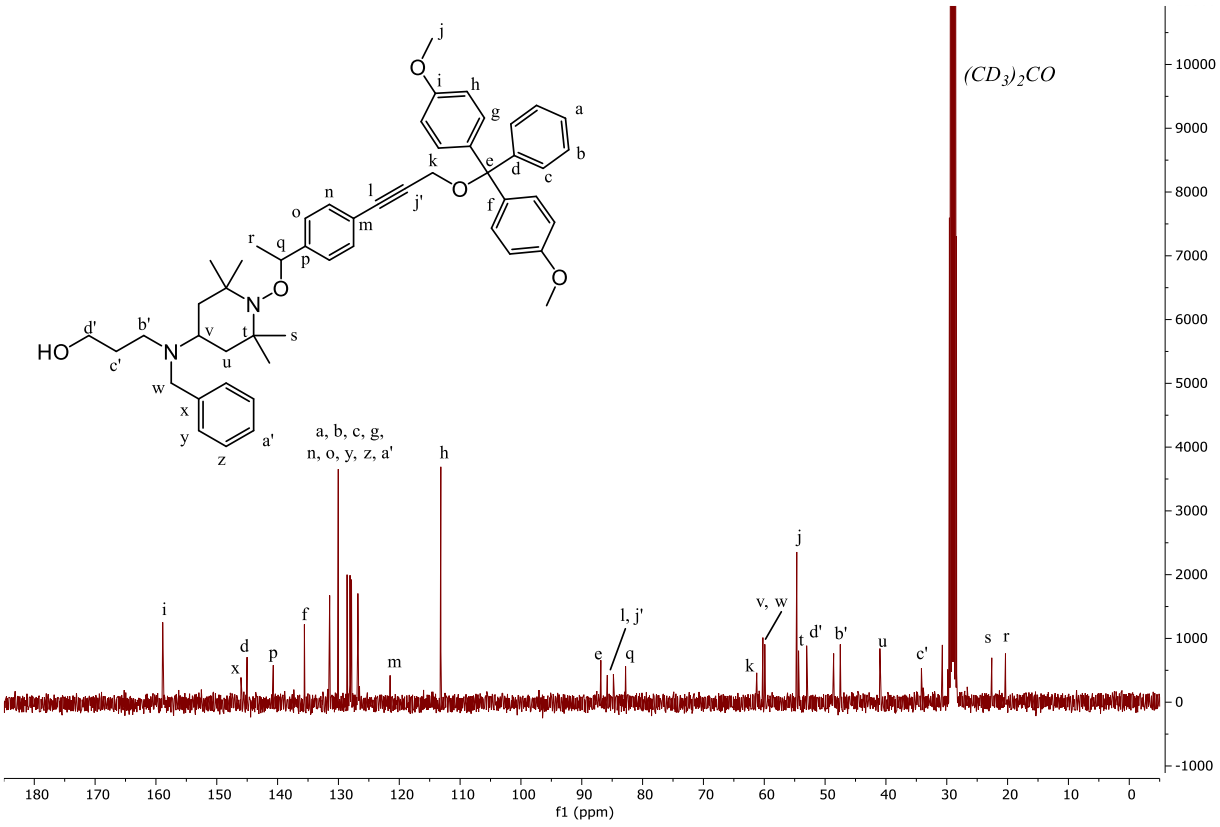


*Figure PE.3. 12 : Spectre RMN ^{13}C du composé **TISP**.*

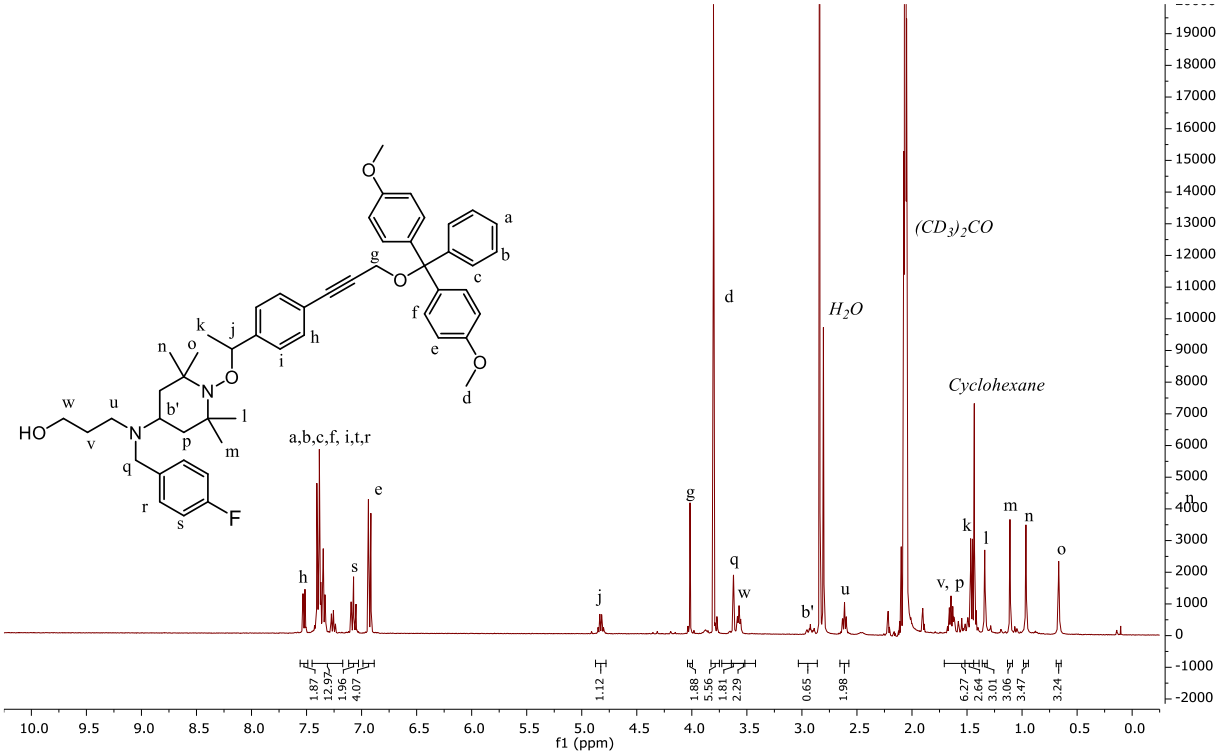


*Figure PE.3. 13 : Spectre RMN ^{1}H du composé **e1-F**.*

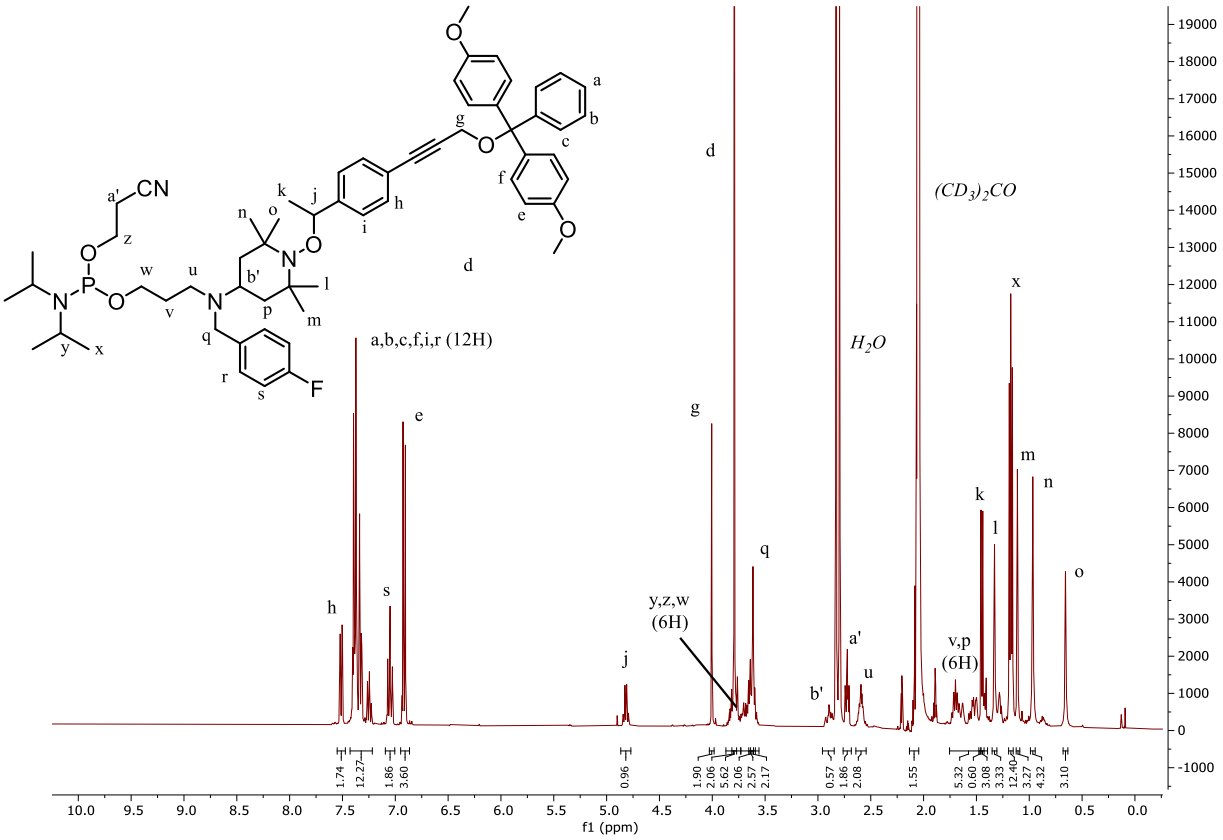


*Figure PE.3. 14 : Spectre RMN ^{1}H du composé **TISP-F**.*

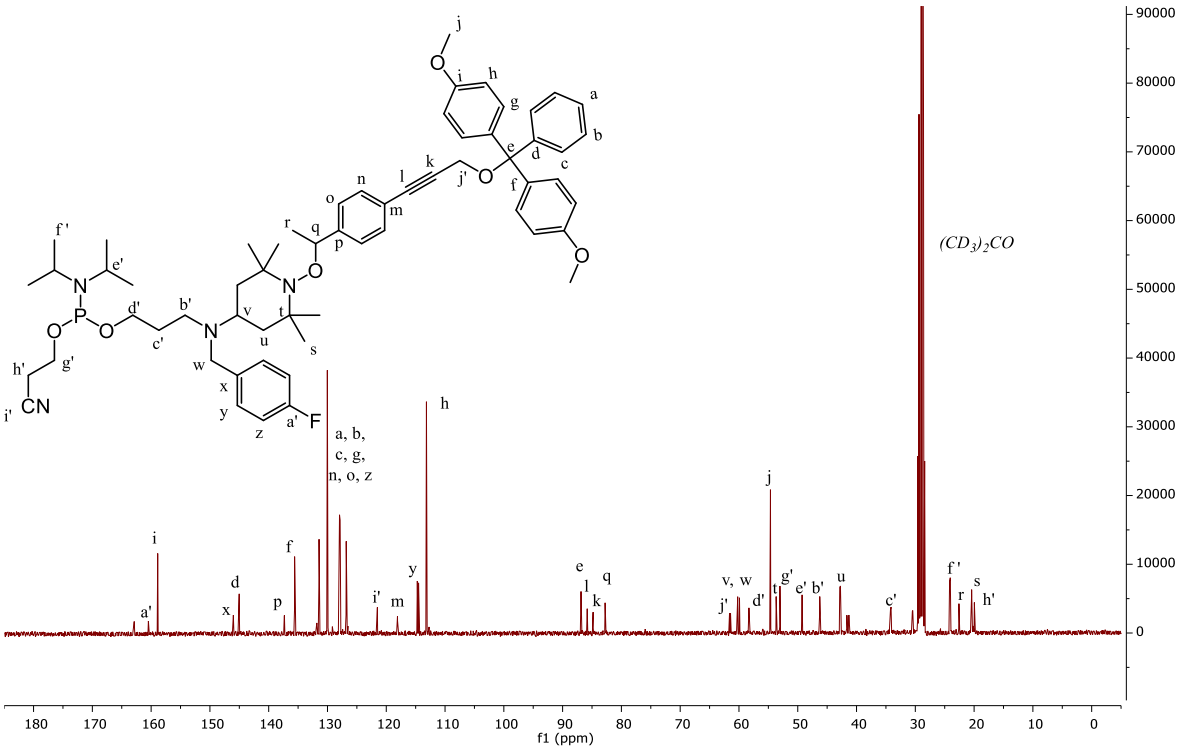


*Figure PE.3. 15 : Spectre RMN ^{13}C du composé **TISP-F**.*

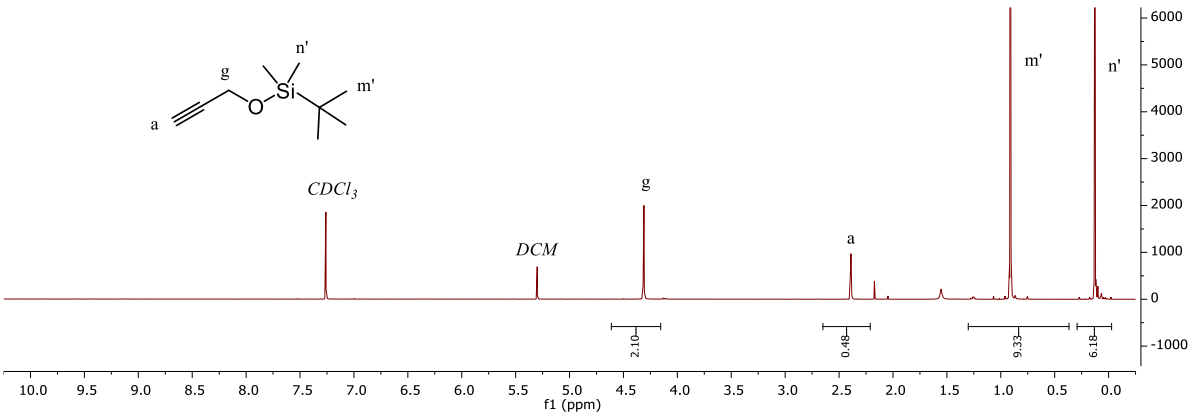


*Figure PE.3. 16 : Spectre RMN ^{1}H du composé **a2**.*

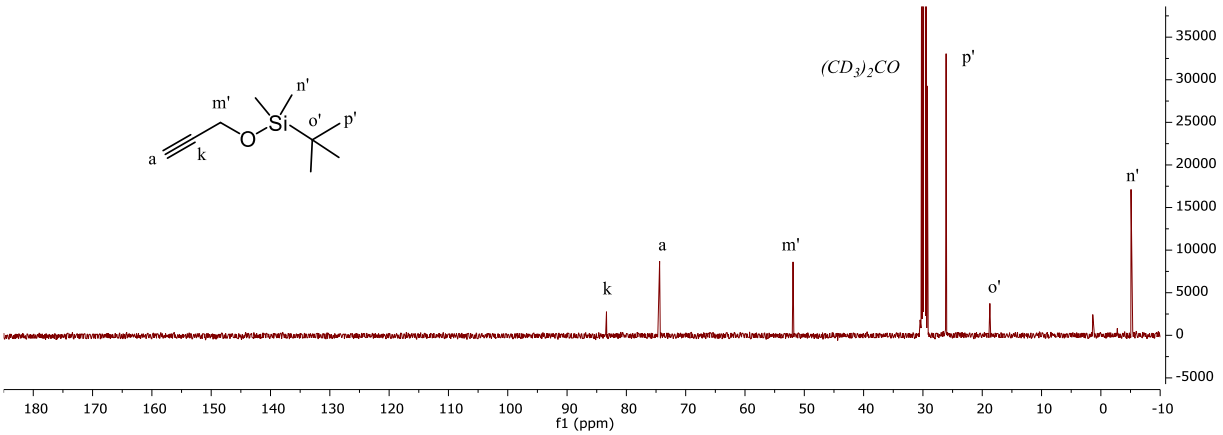


*Figure PE.3. 17 : Spectre RMN ^{13}C du composé **a2**.*

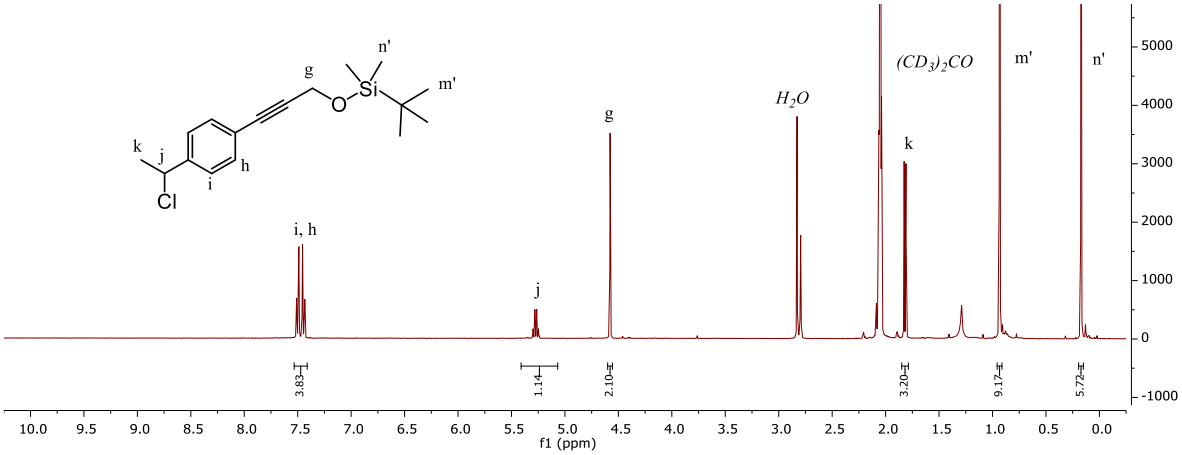


*Figure PE.3. 18 : Spectre RMN ^{1}H du composé **b2**.*

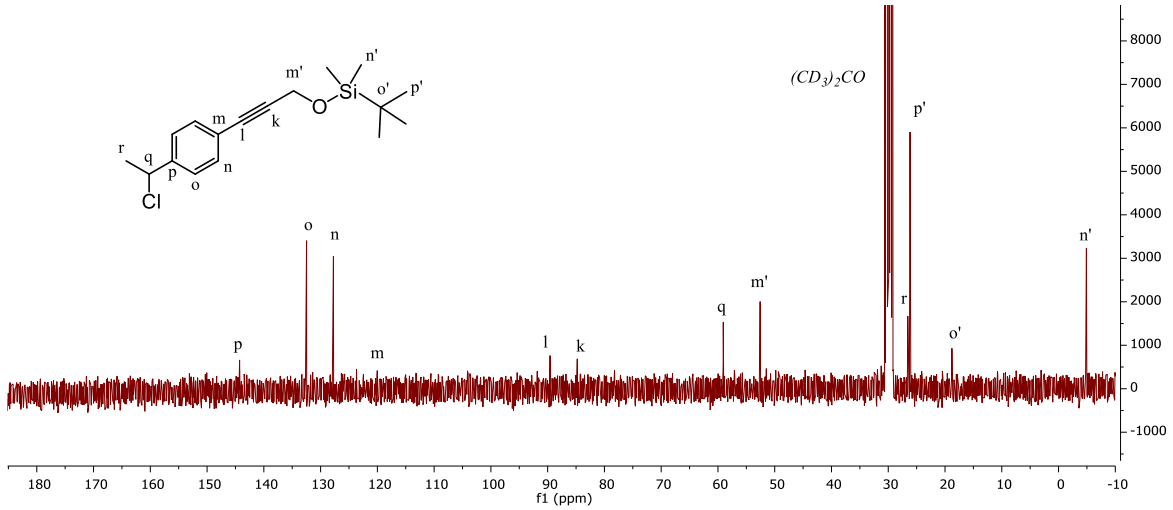


*Figure PE.3. 19 : Spectre RMN ^{13}C du composé **b2**.*

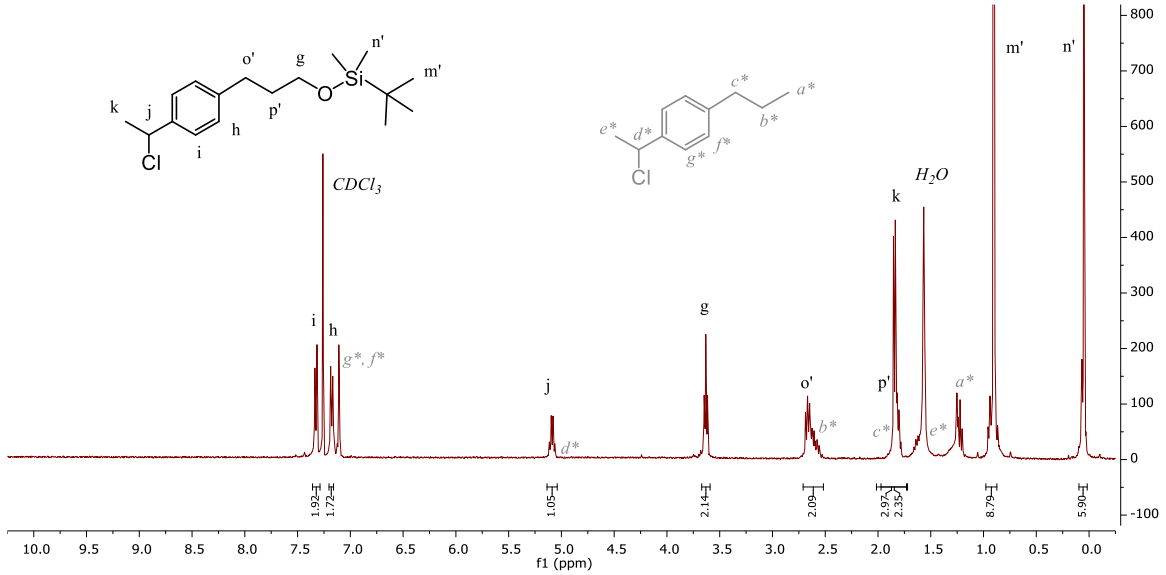


*Figure PE.3. 20 : Spectre RMN ^{1}H du composé **c2**. *En gris, sous-produit de la réaction.*

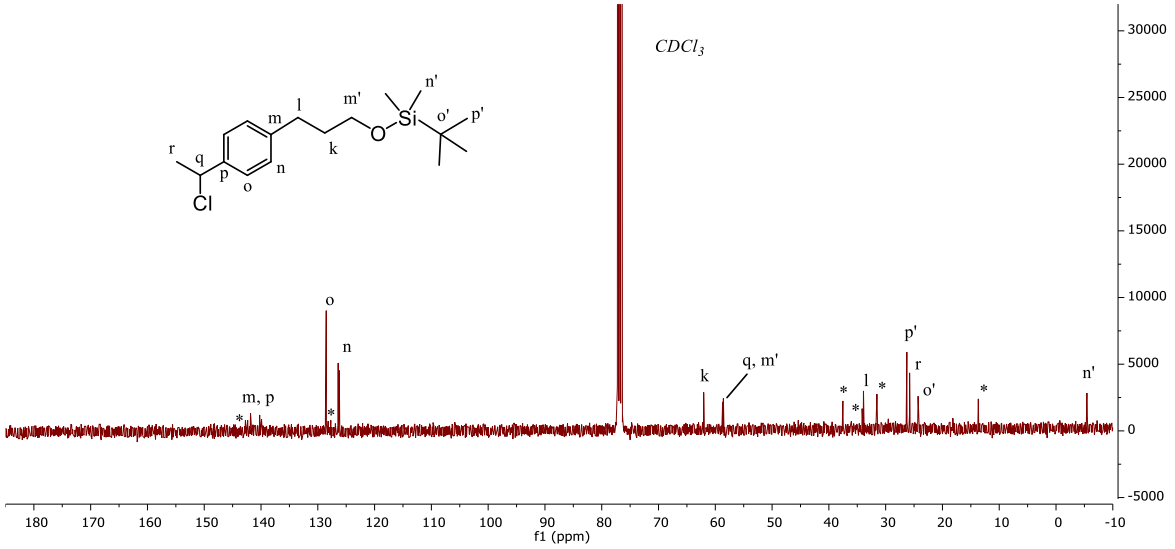


*Figure PE.3. 21 : Spectre RMN ^{13}C du composé **c2**.*

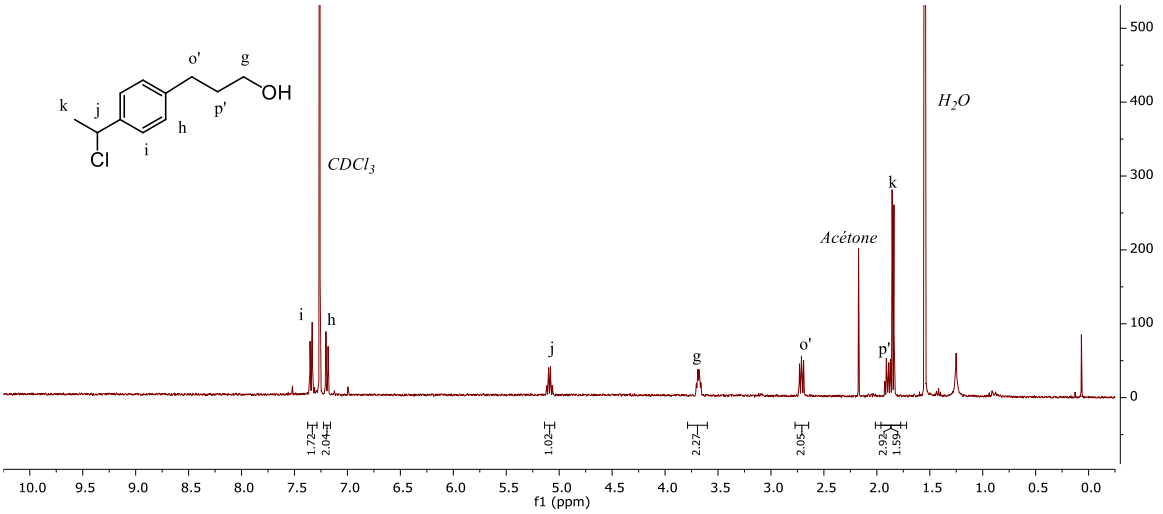


*Figure PE.3. 22 : Spectre RMN ^{1}H du composé **d2**.*

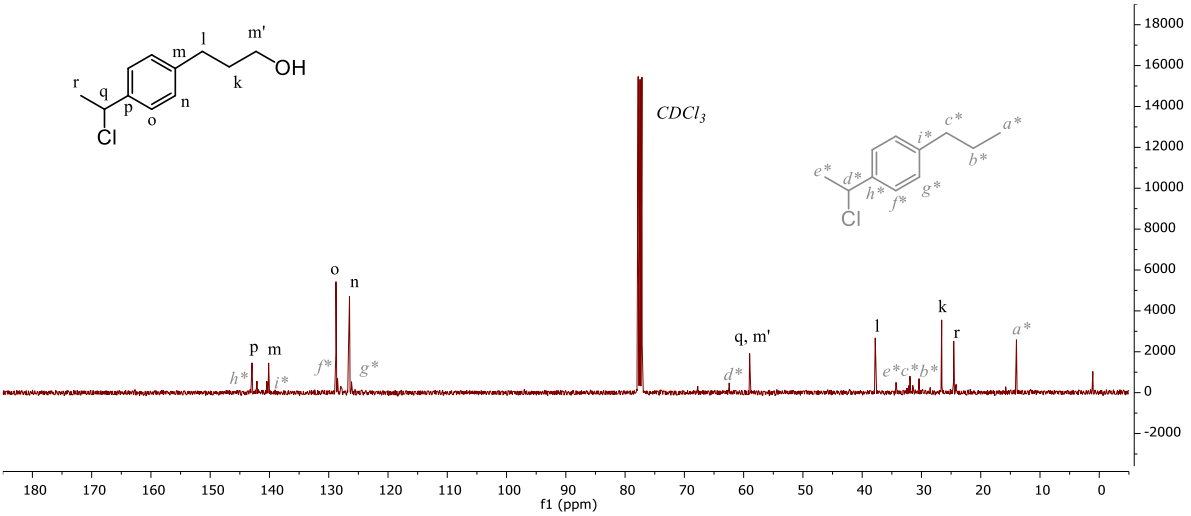


*Figure PE.3. 23 : Spectre RMN ^{13}C du composé **d2**. Brut de la réduction alcyne à partir de **c1**, montrant l'impureté majeure identifiée en gris*.*

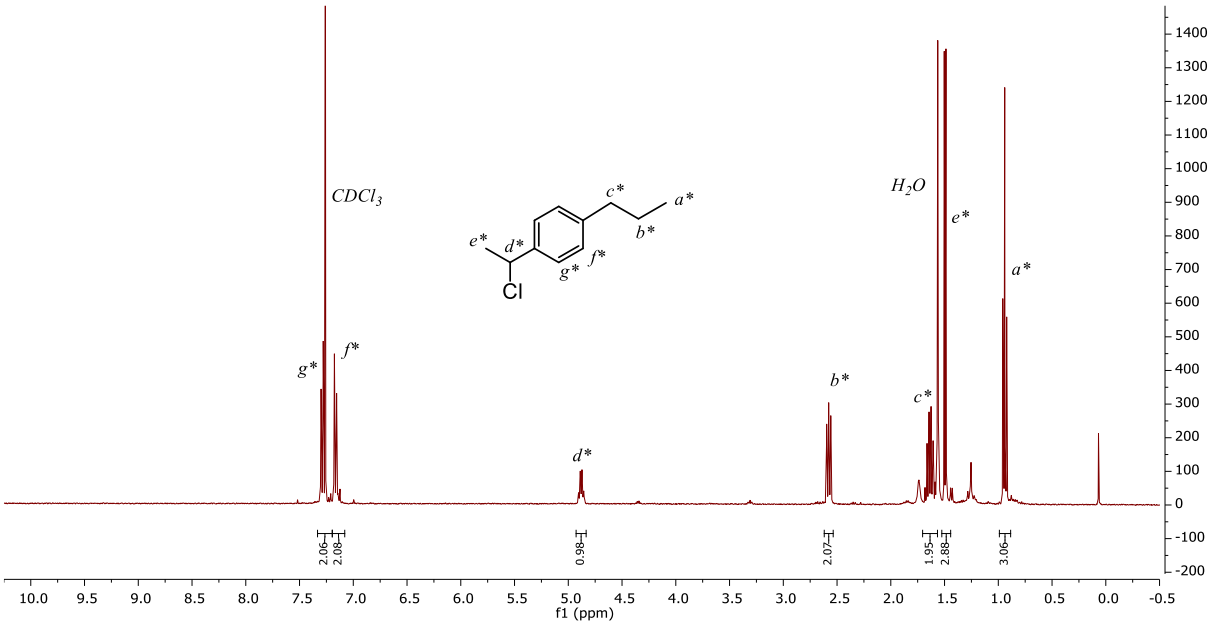


*Figure PE.3. 24 : Spectre RMN ^1H du sous-produit **SP** de la réaction menant à l'intermédiaire **d2**.*

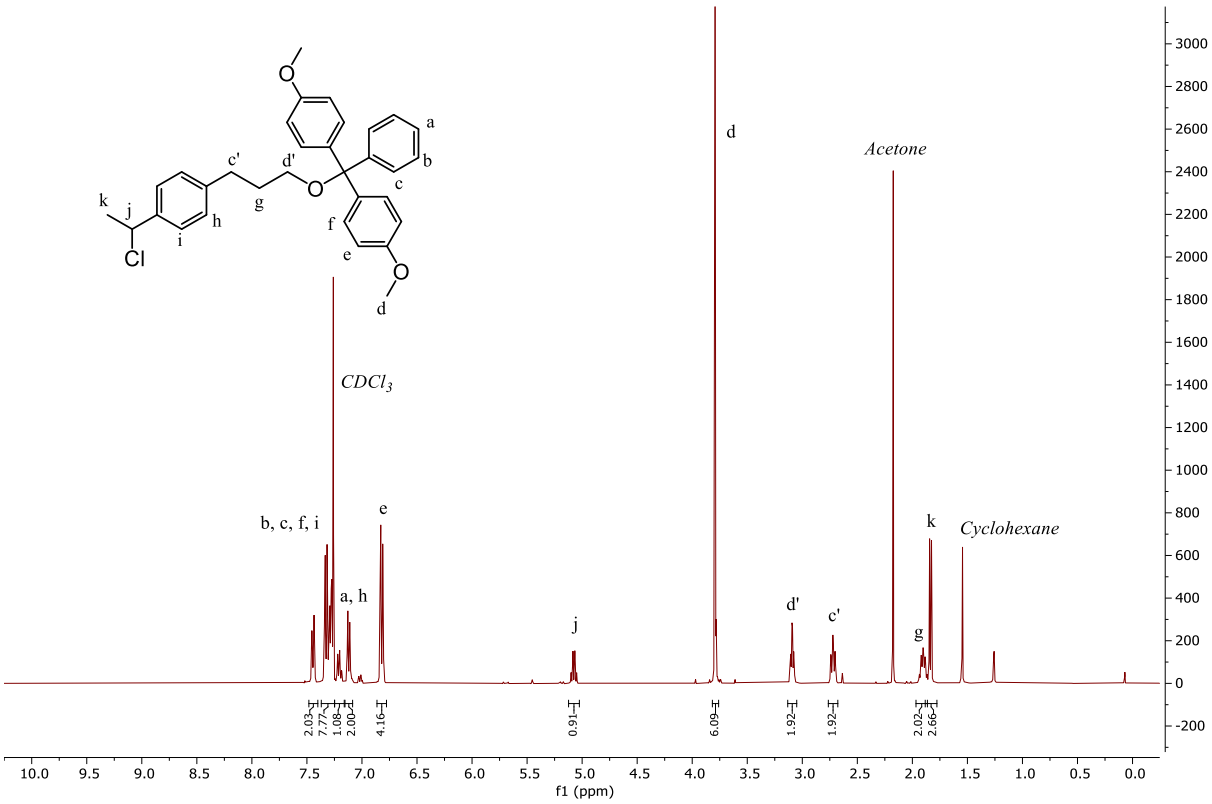


*Figure PE.3. 25 : Spectre RMN ^1H du composé **e2**.*

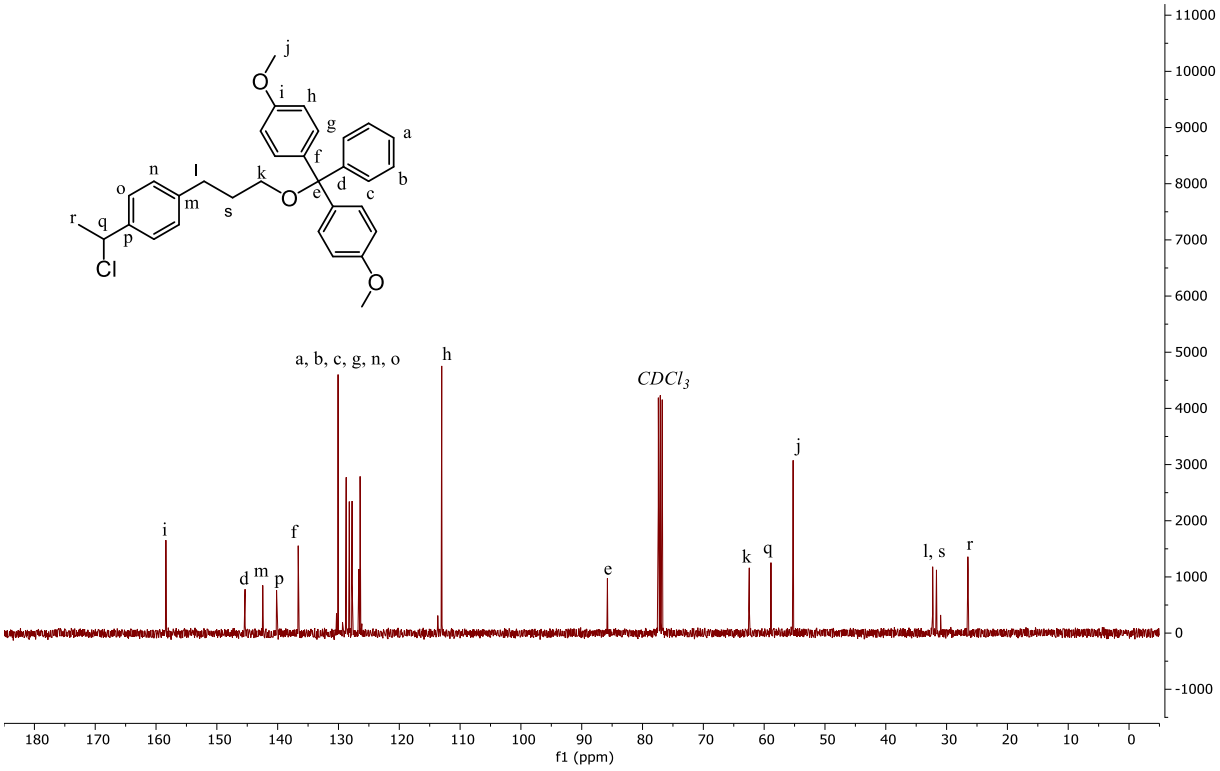


*Figure PE.3. 26 : Spectre RMN ^{13}C du composé **e2**.*

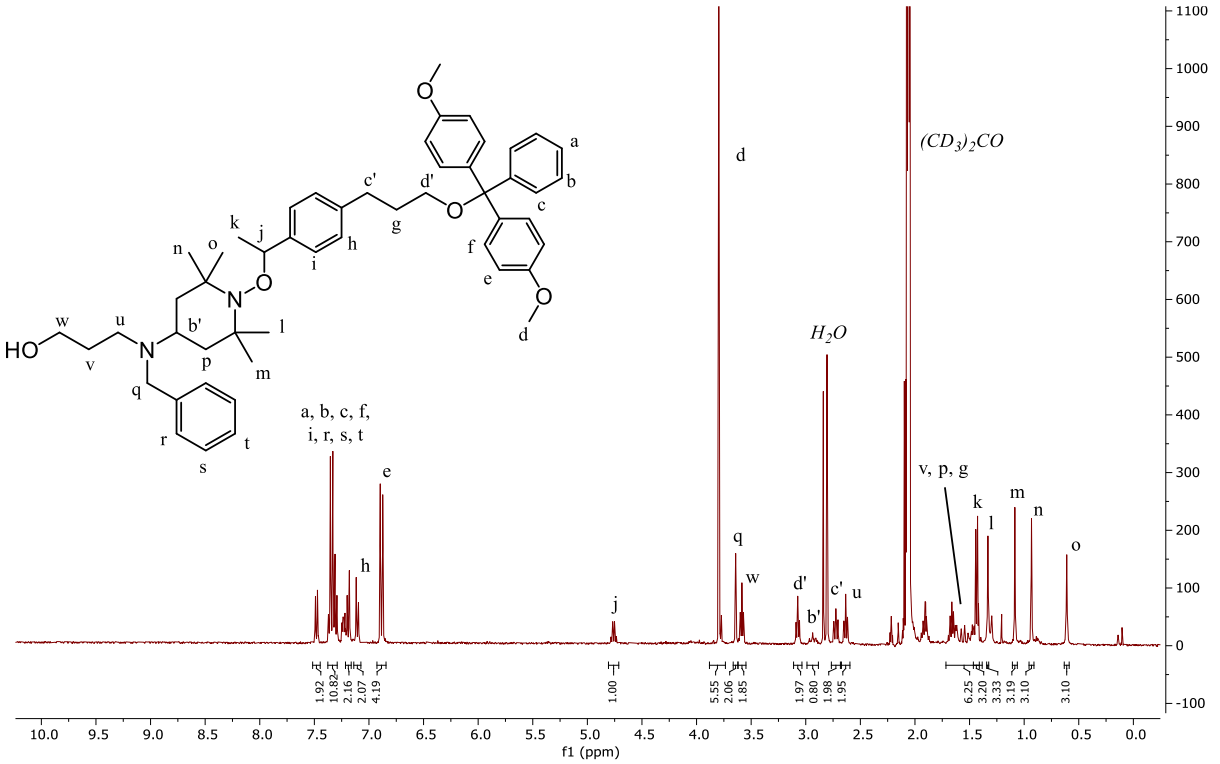


*Figure PE.3. 27 : Spectre RMN ^{1}H du composé **g2**.*

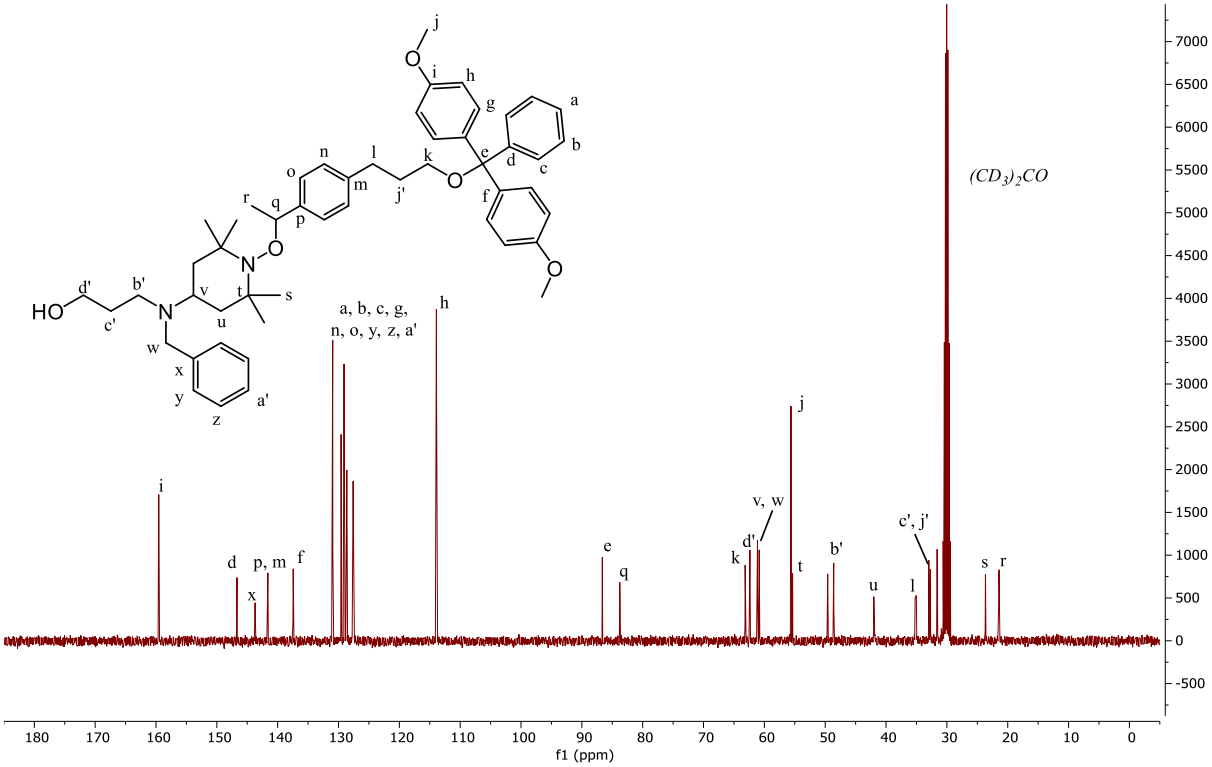


*Figure PE.3. 28 : Spectre RMN ^{13}C du composé **g2**.*

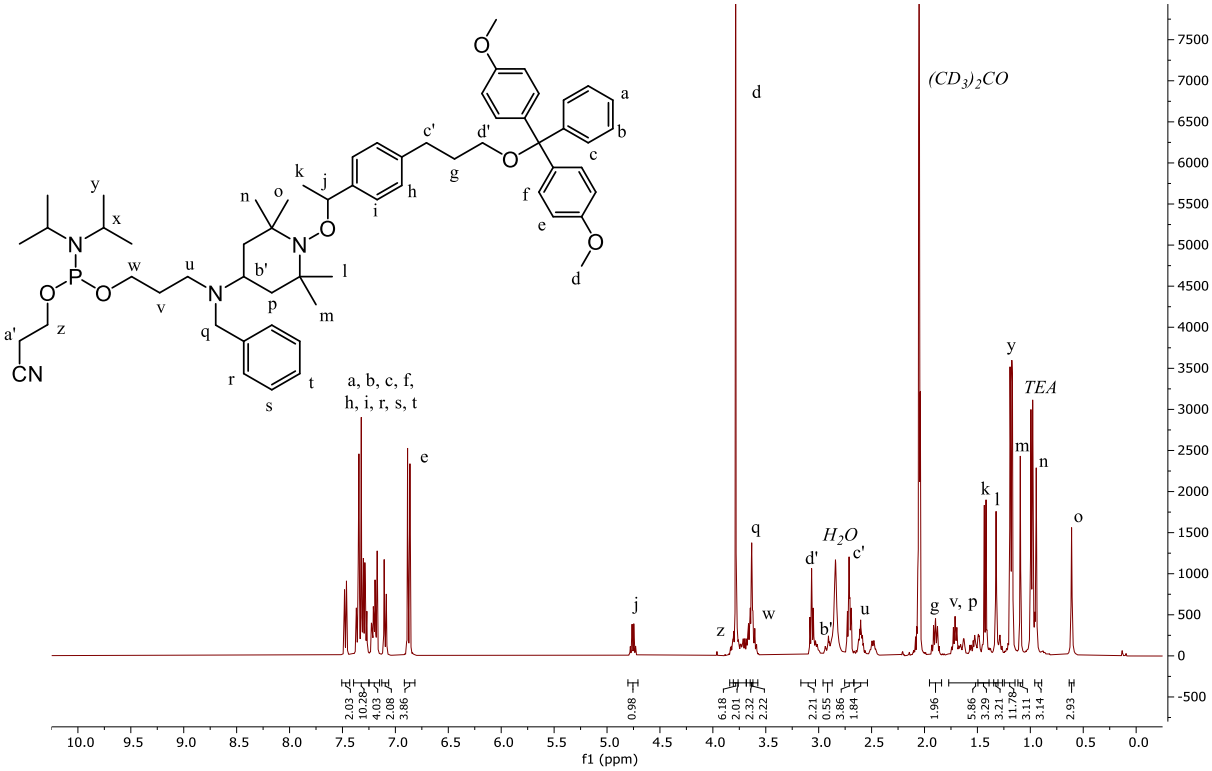


*Figure PE.3. 29 : Spectre RMN ^{1}H du composé **TISP2**.*

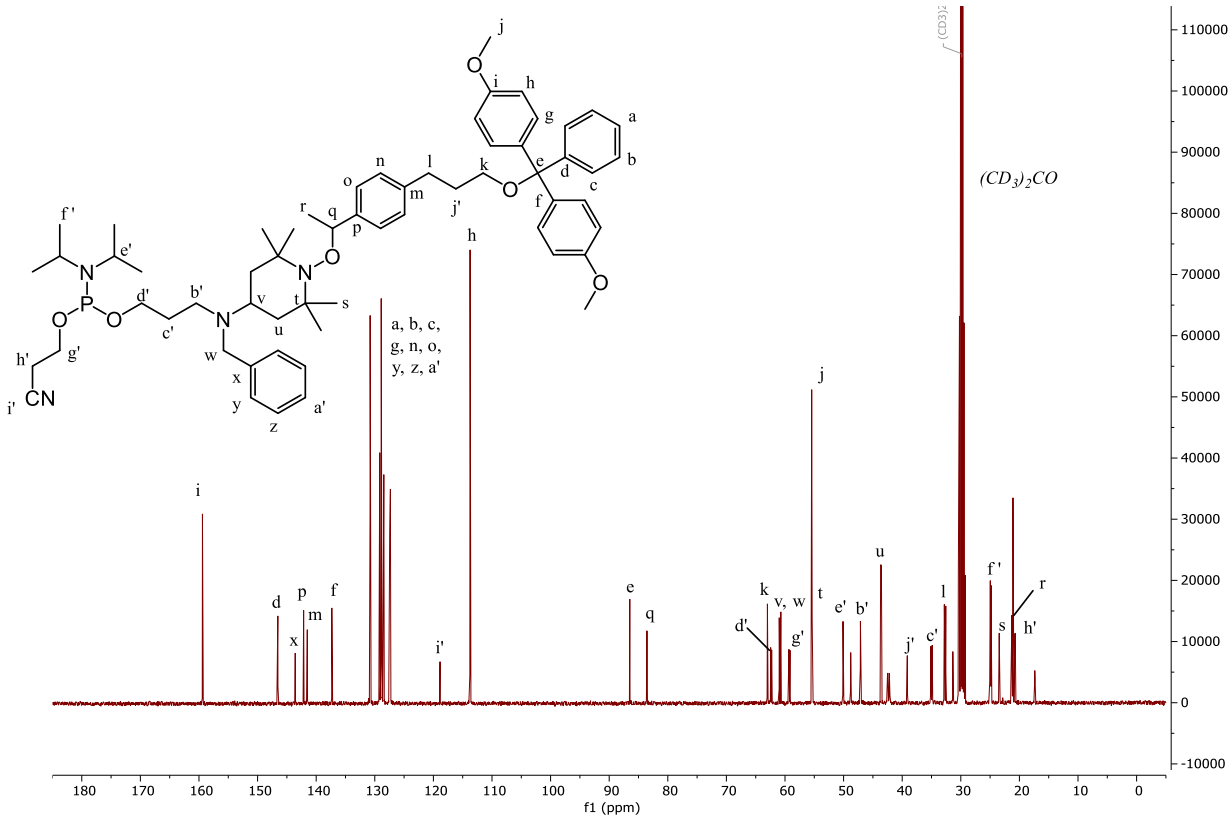


*Figure PE.3. 30 : Spectre RMN ^{13}C du composé **TISP2**.*

3.1.1. RMN ^{31}P TISP et TISP2

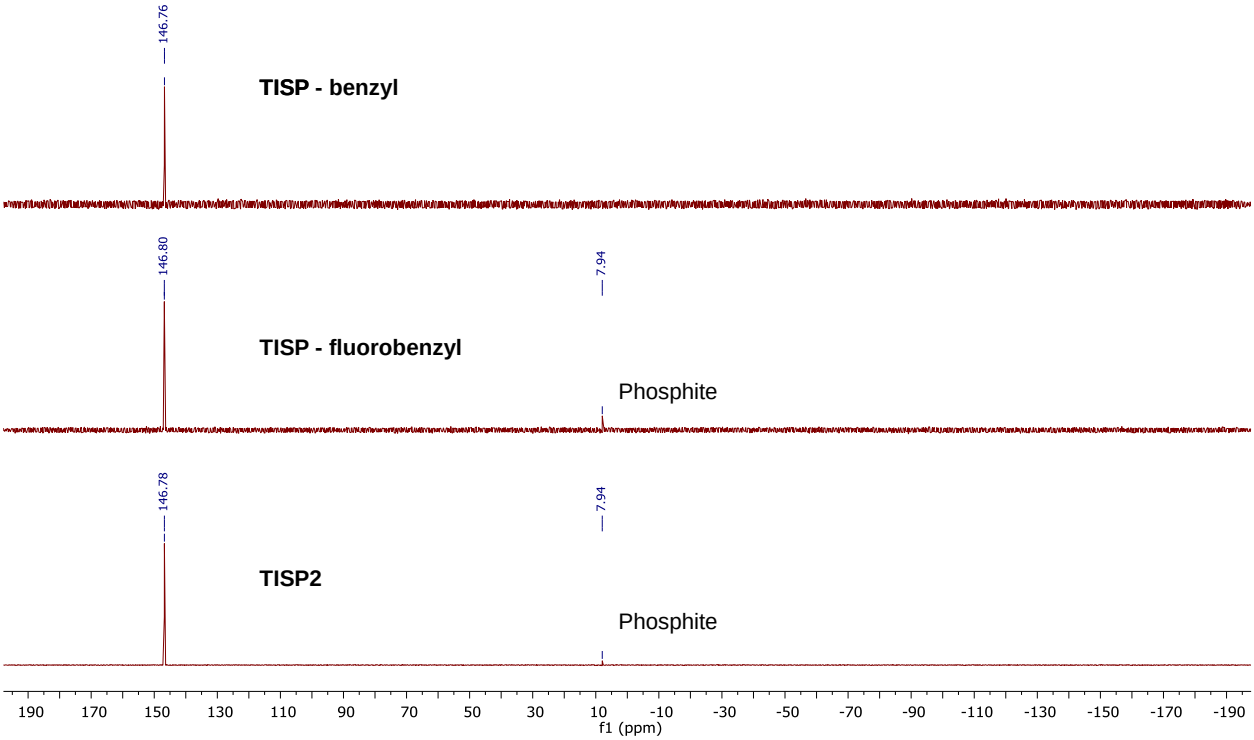


*Figure PE.3. 31 : Spectres RMN ^{31}P des composés **TISP**, **TISP-F** et **TISP2**.*

3.2. TISP3

3.2.1. RMN ^{1}H et ^{13}C

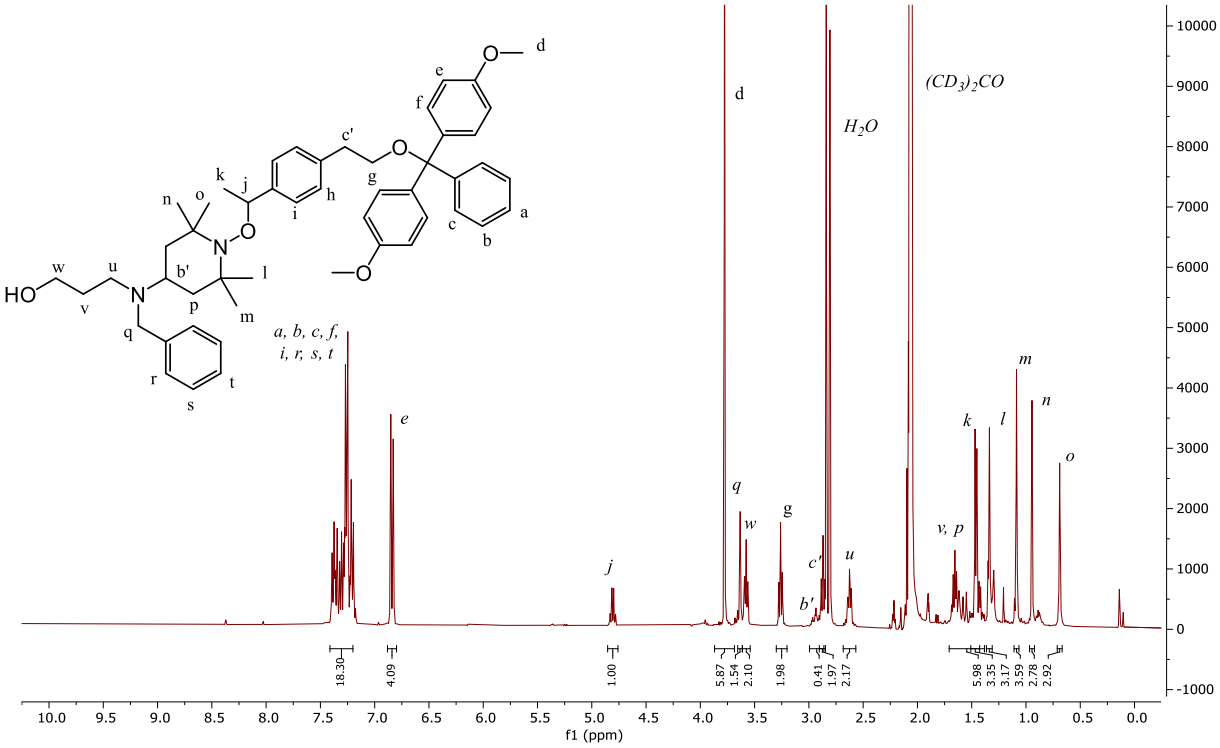


*Figure PE.3. 32 : Spectre RMN ^{1}H du composé **f2**.*

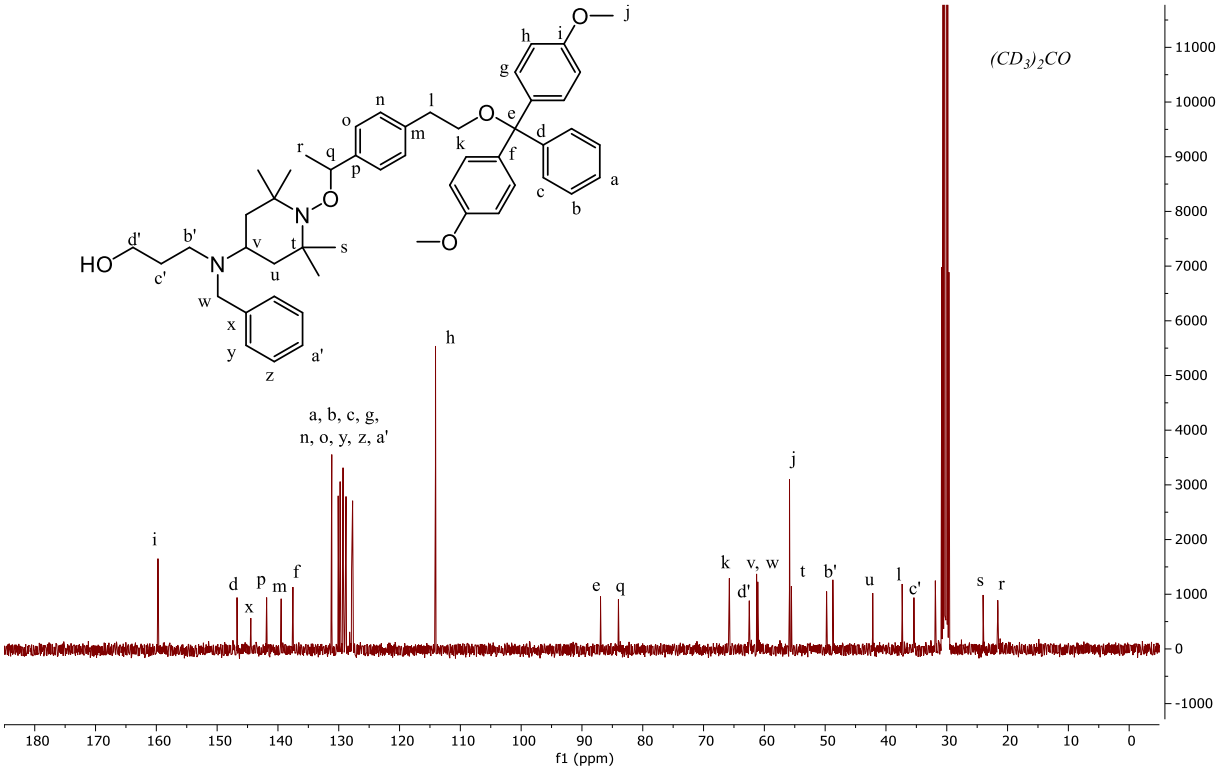


*Figure PE.3. 33 : Spectre RMN ^{13}C du composé **f2**.*

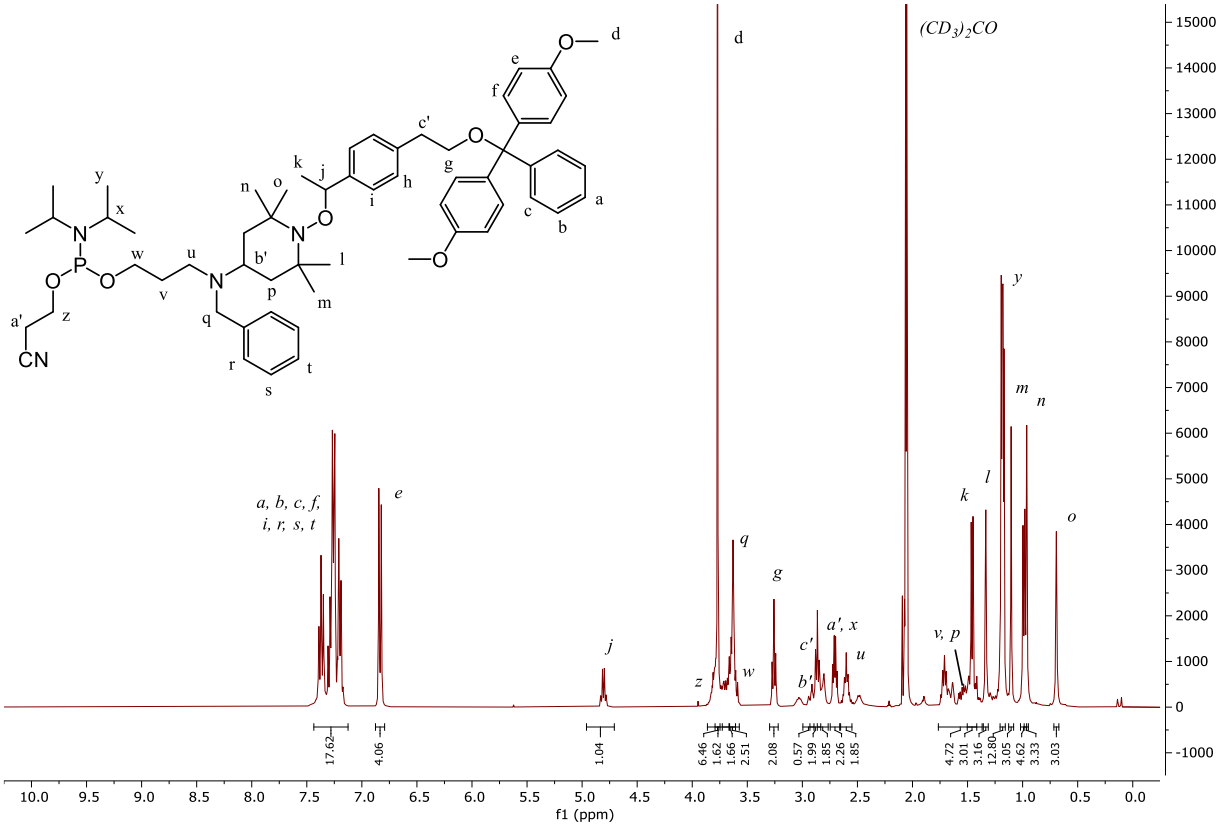


*Figure PE.3. 34 : Spectre RMN ^{1}H de l'Espaceur **TISP3-2**.*

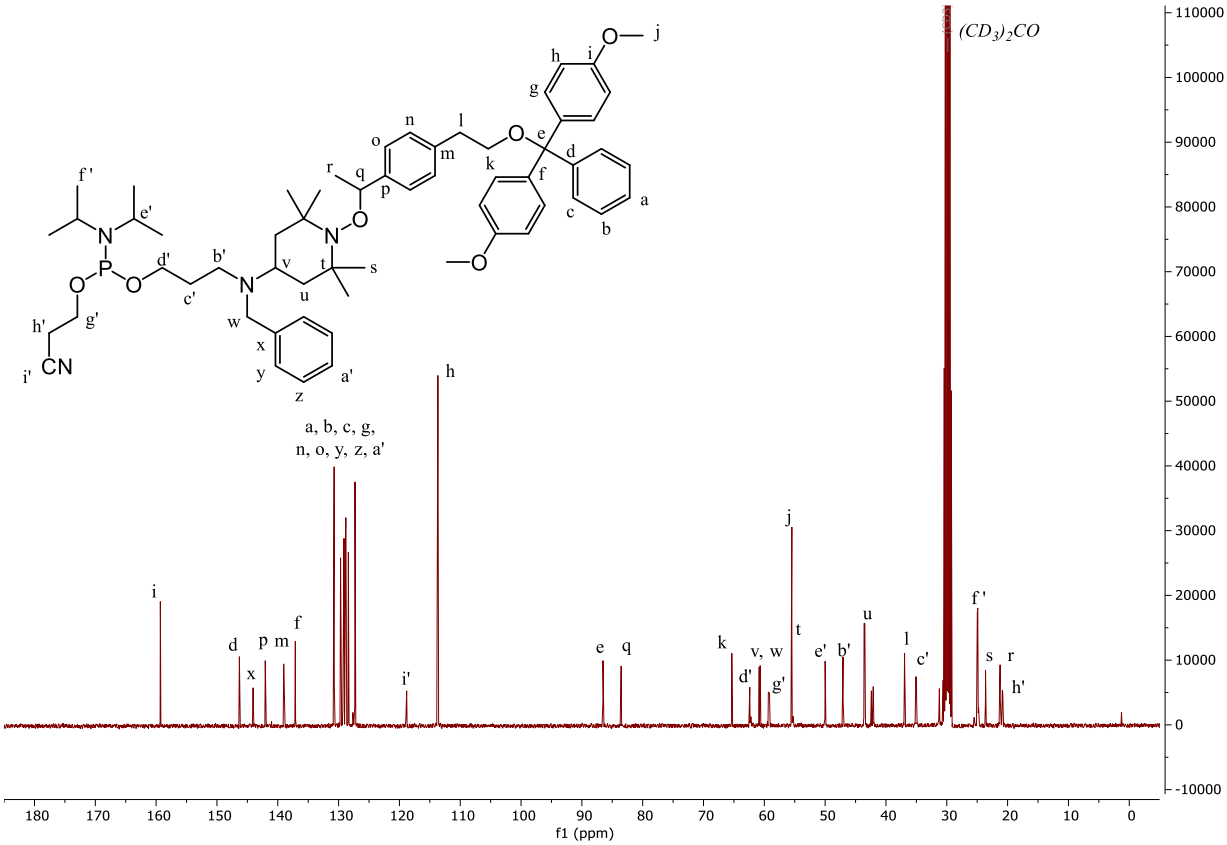


*Figure PE.3. 35 : Spectre RMN ^{13}C de l'Espaceur **TISP3-2**.*

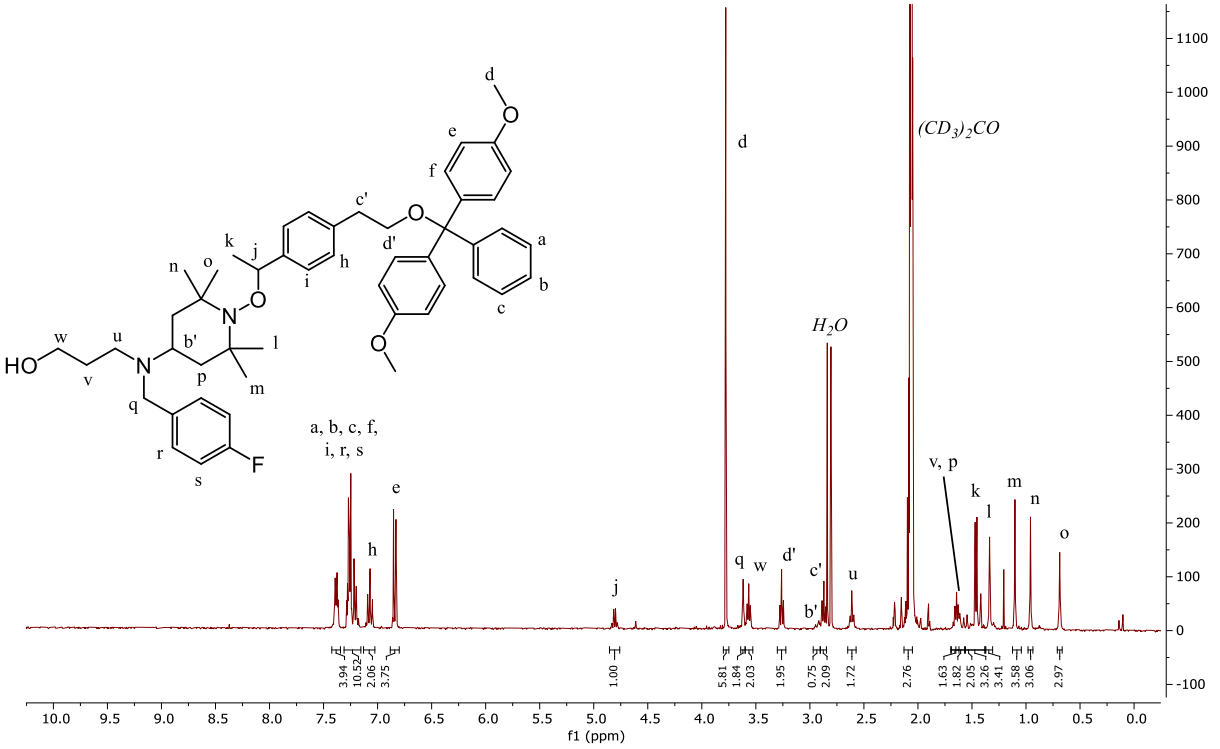


*Figure PE.3. 36 : Spectre RMN ^{1}H du composé **f3**.*

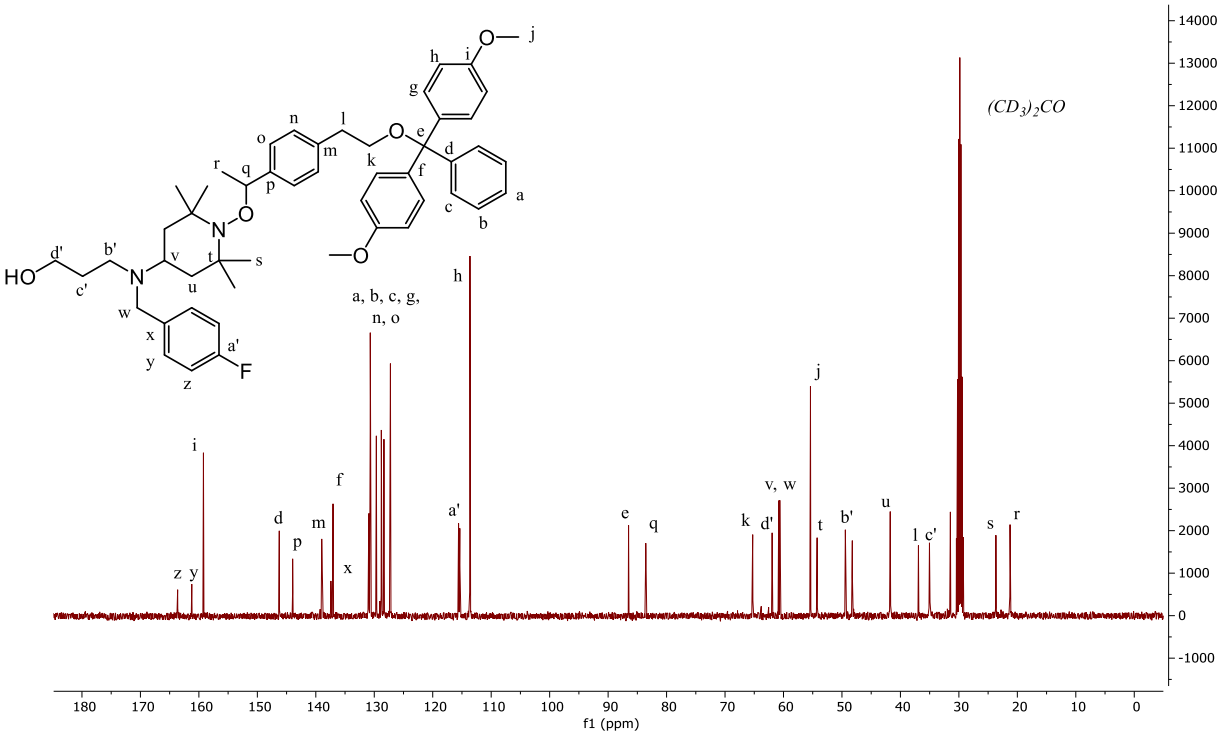


*Figure PE.3. 37 : Spectre RMN ^{13}C du composé **f3**.*

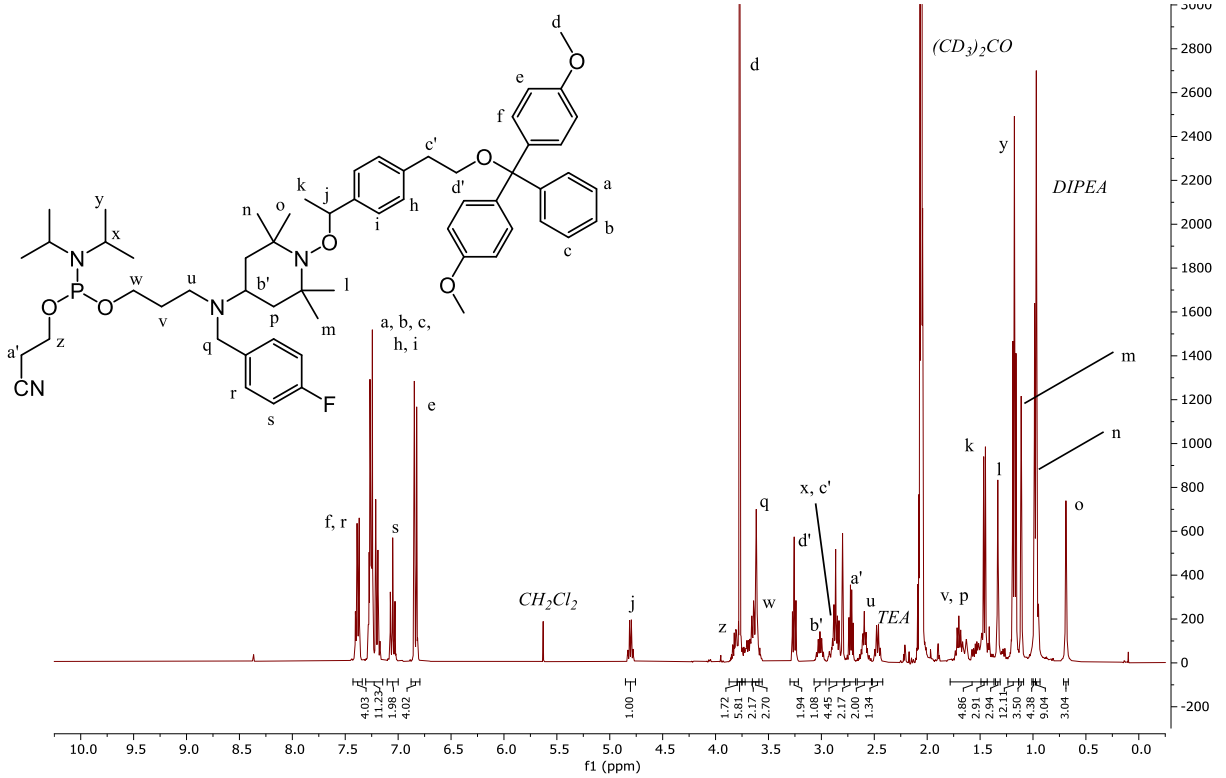


*Figure PE.3. 38 : Spectre RMN ^{1}H de l'Espaceur **TISP3-3**.*

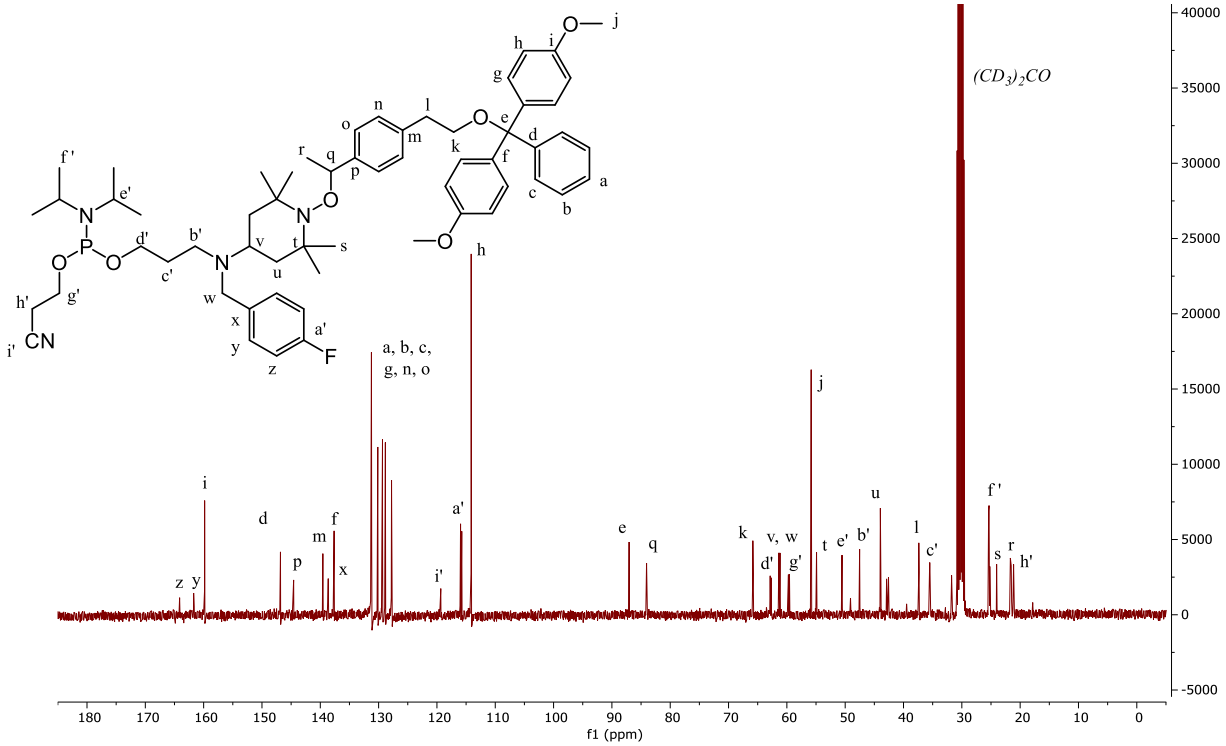


*Figure PE.3. 39 : Spectre RMN ^{13}C de l'Espaceur **TISP3-3**.*

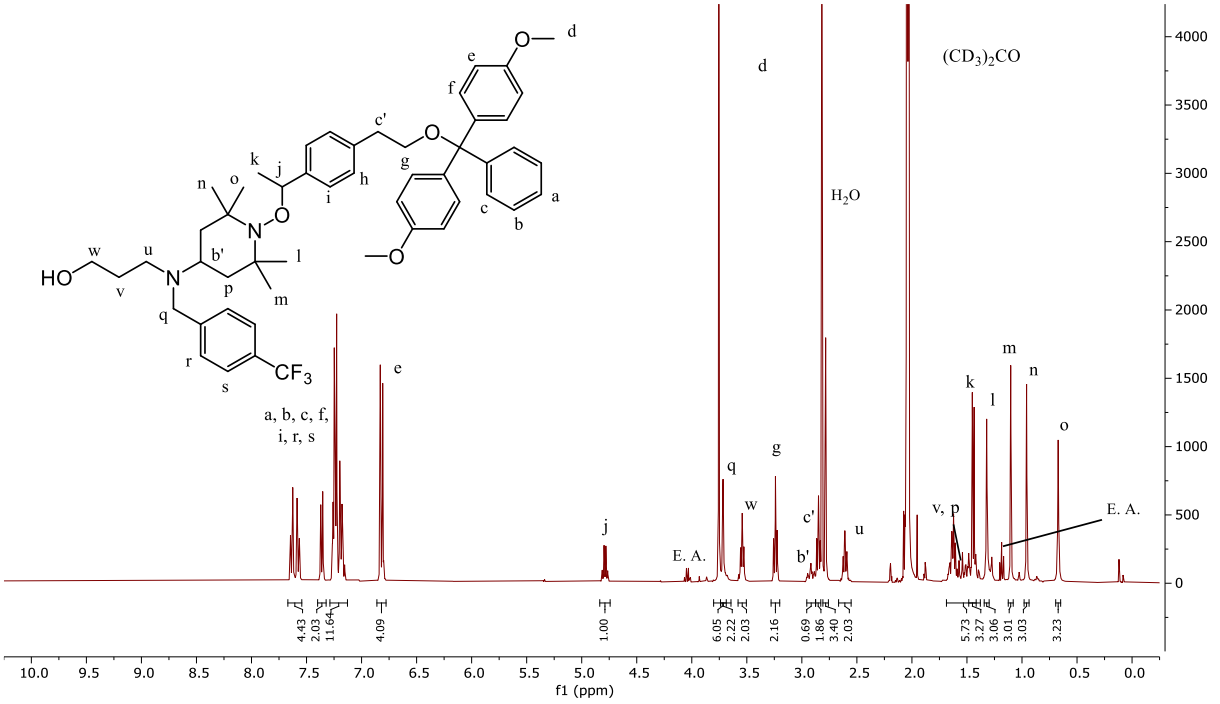


*Figure PE.3. 40 : Spectre RMN ^{1}H du composé **f4**.*

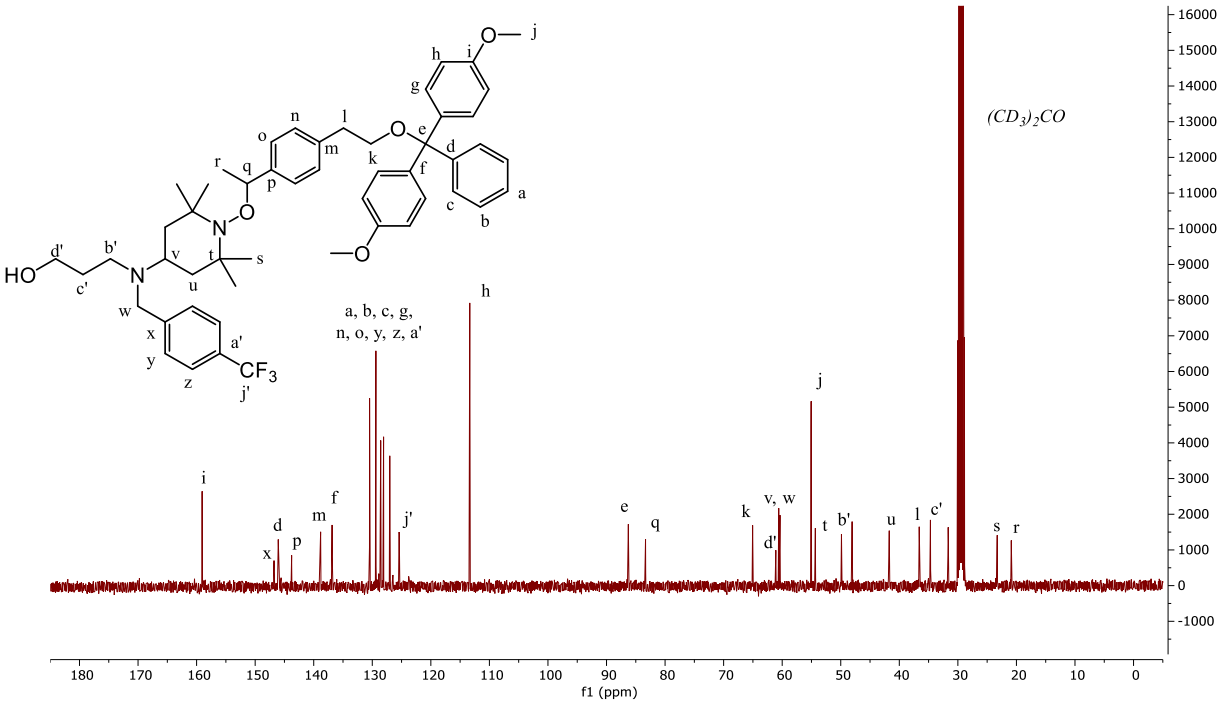


*Figure PE.3. 41 : Spectre RMN ^{13}C du composé **f4**.*

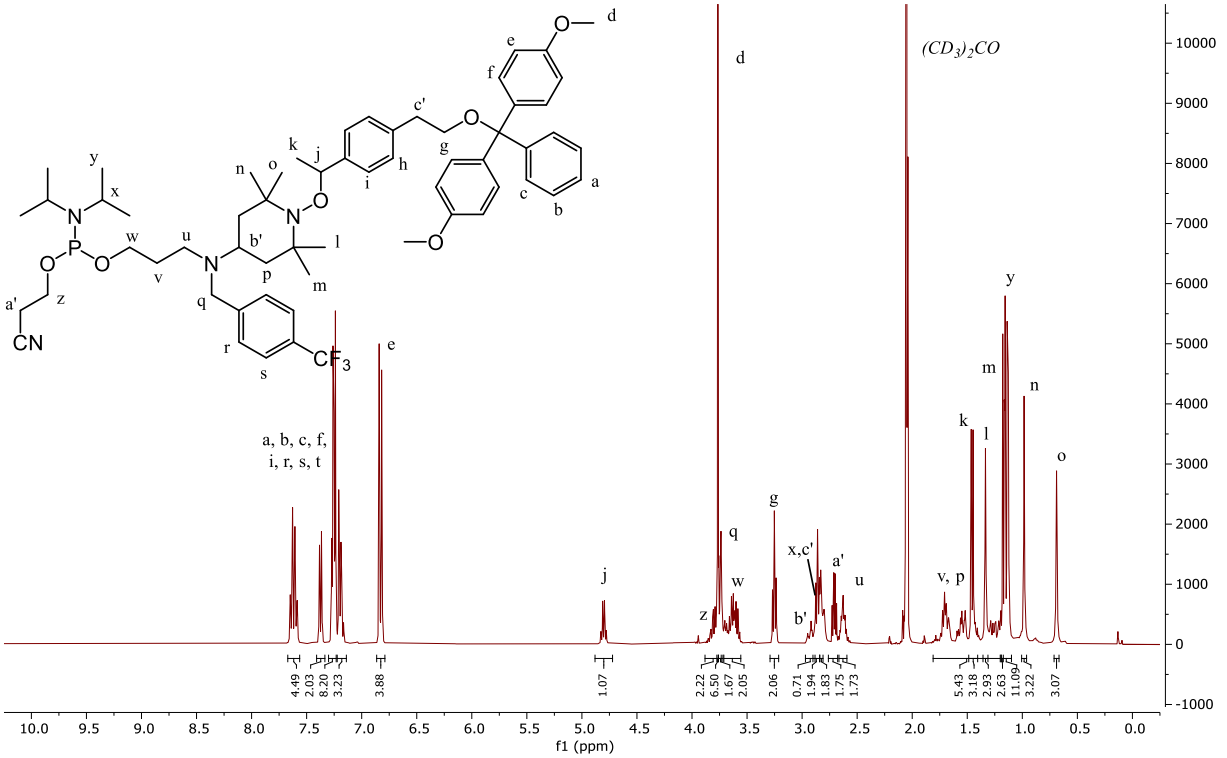


*Figure PE.3. 42 : Spectre RMN ^{1}H de l'Espaceur **TISP3-4**.*

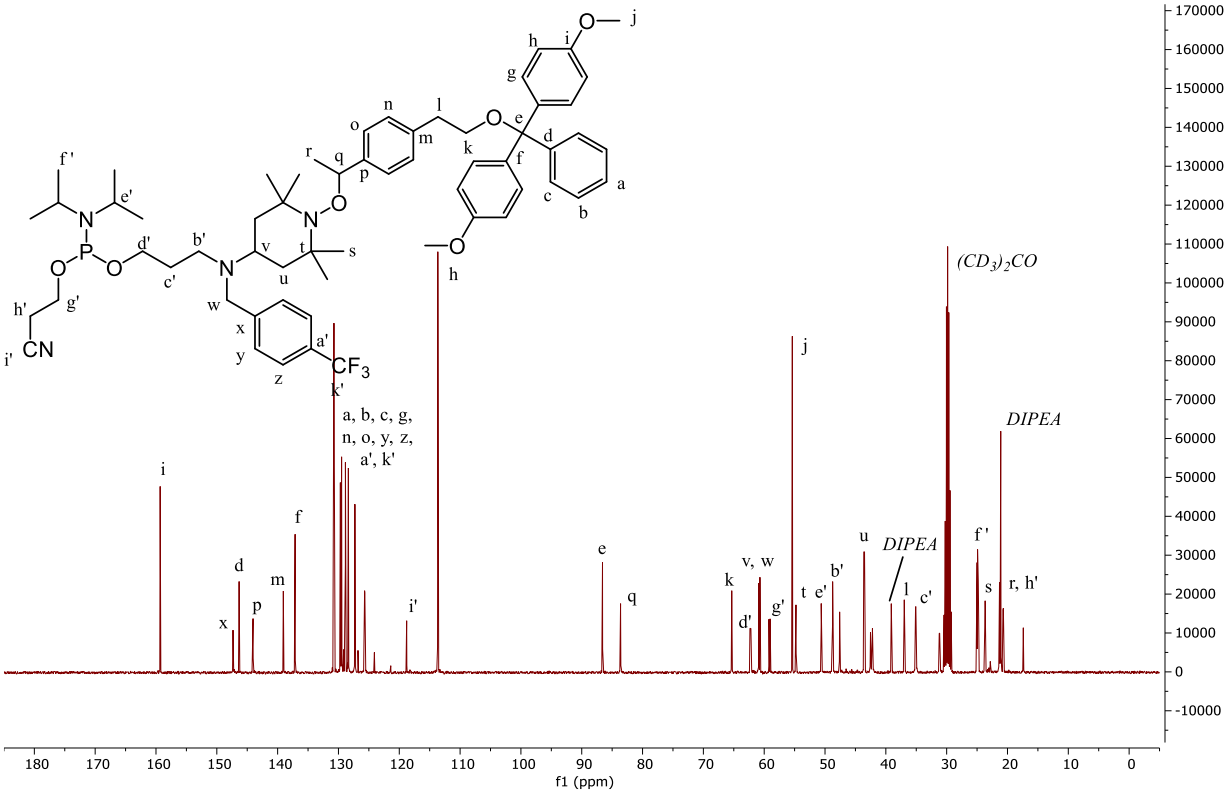


*Figure PE.3. 43 : Spectre RMN ^{13}C de l'Espaceur **TISP3-4**.*

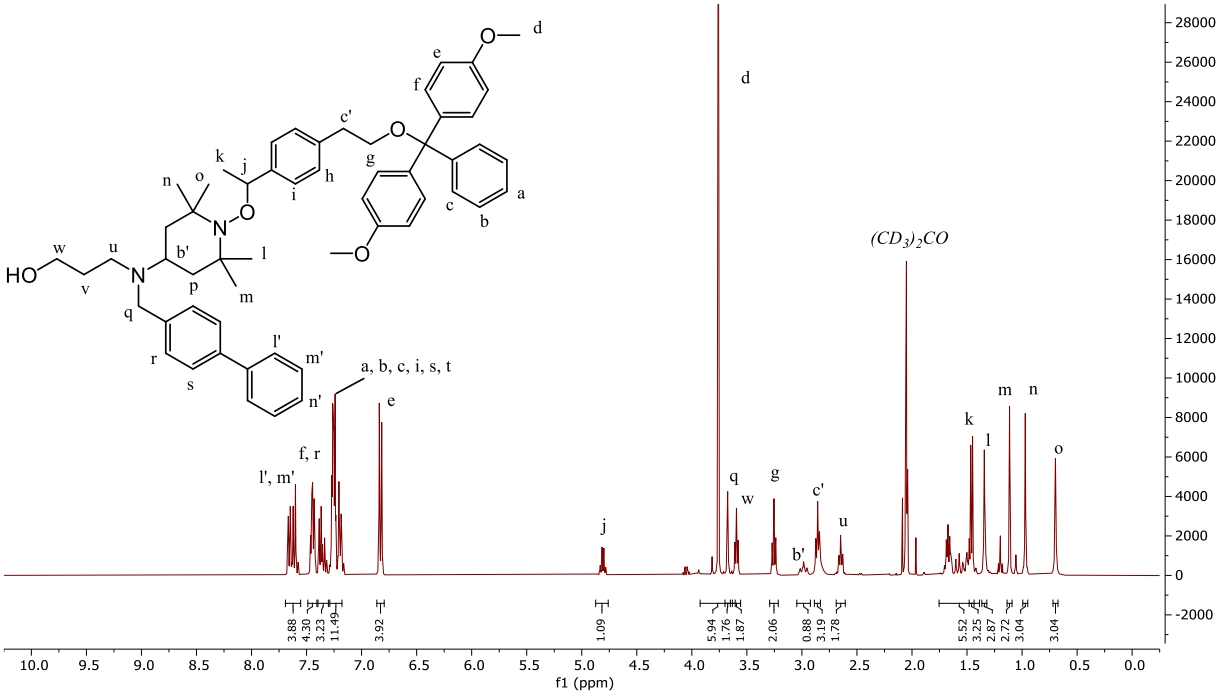


*Figure PE.3. 44 : Spectre RMN ^{1}H du composé **f5**.*

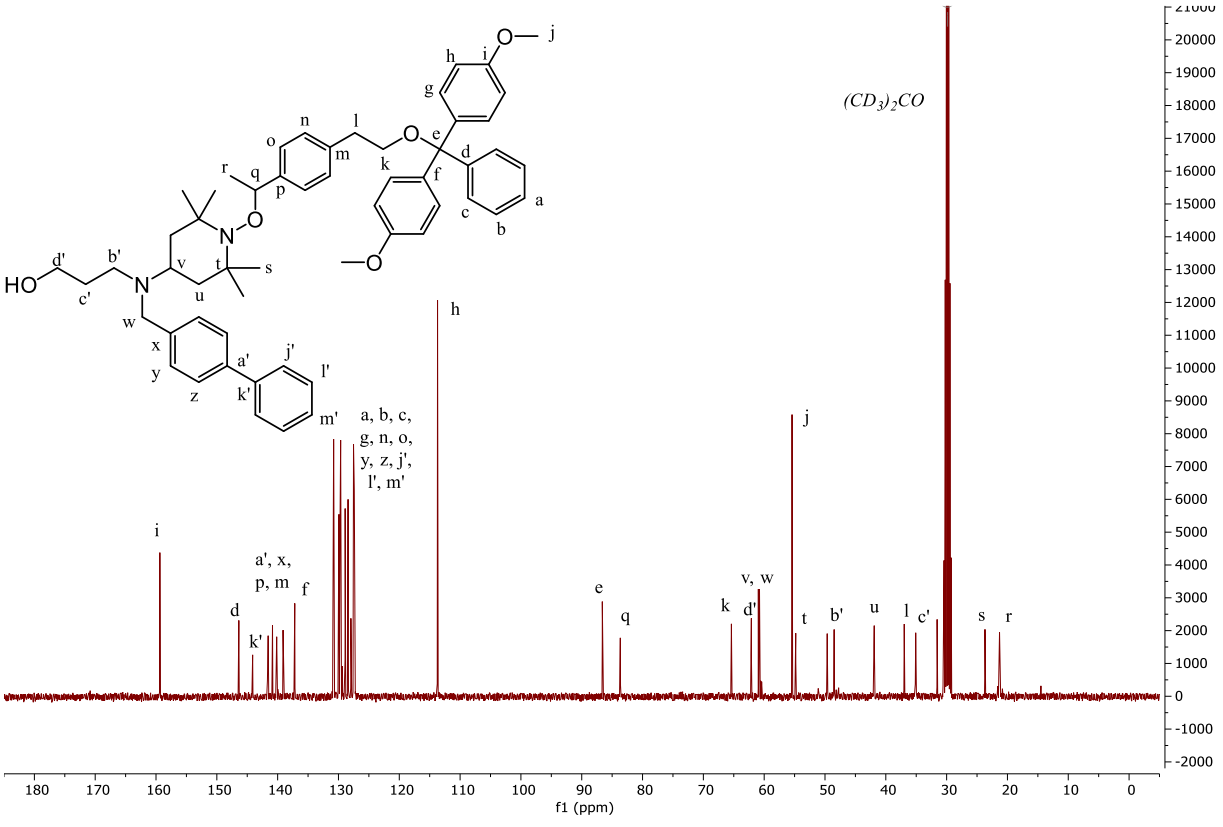


*Figure PE.3. 45 : Spectre RMN ^{13}C du composé **f5**.*

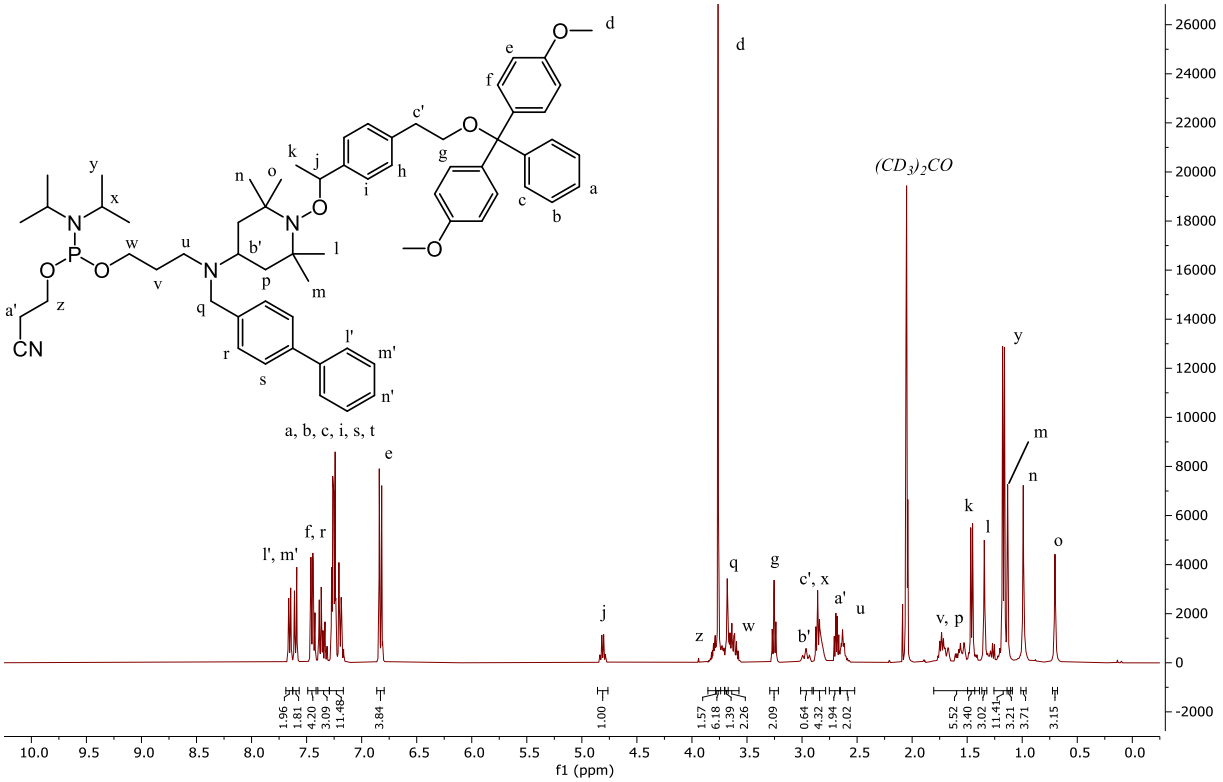


*Figure PE.3. 46 : Spectre RMN ^{1}H de l'Espaceur **TISP3-5**.*

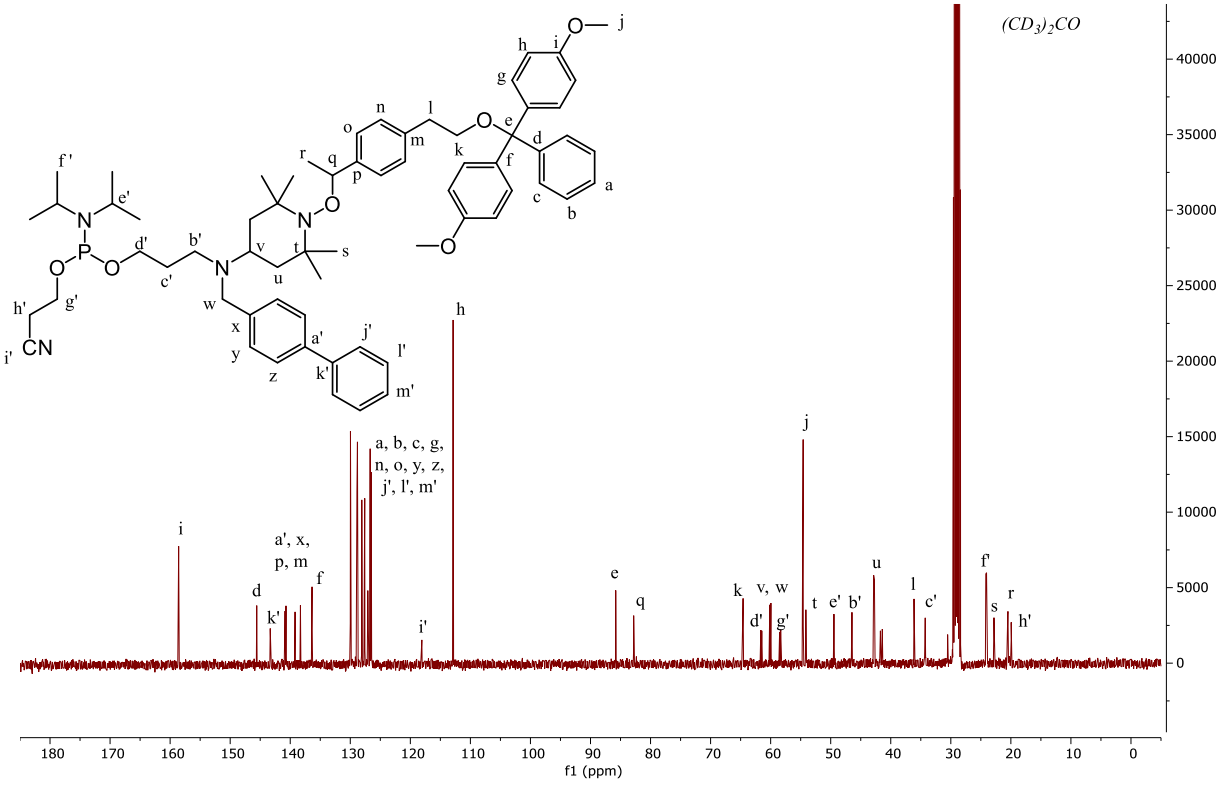


*Figure PE.3. 47 : Spectre RMN ^{13}C de l'Espaceur **TISP3-5**.*

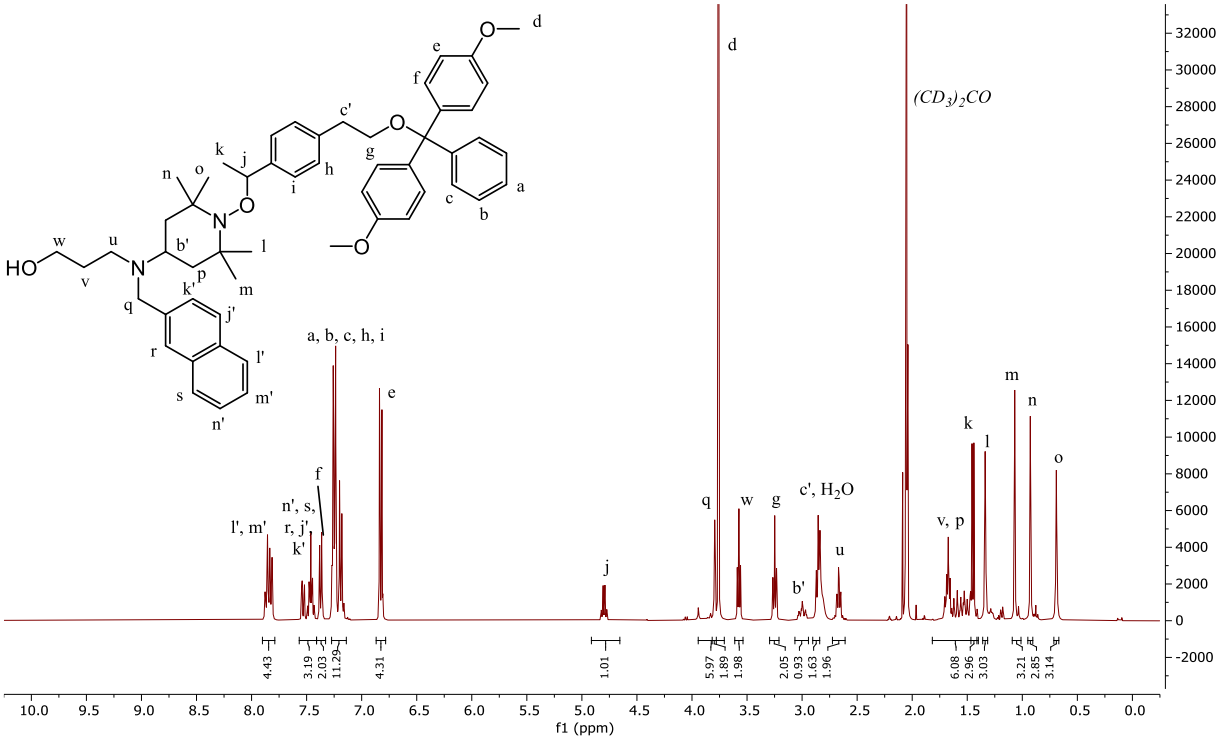


*Figure PE.3. 48 : Spectre RMN ^{1}H du composé **f6**.*

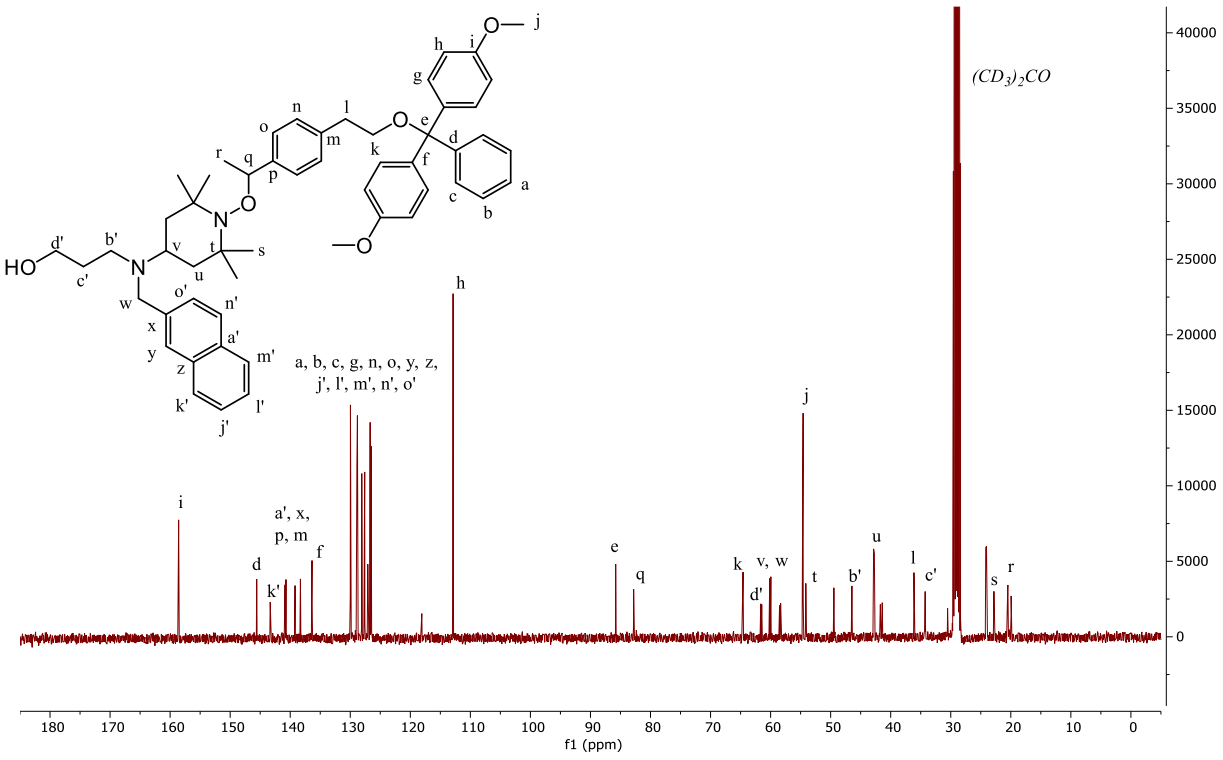


*Figure PE.3. 49 : Spectre RMN ^{13}C du composé **f6**.*

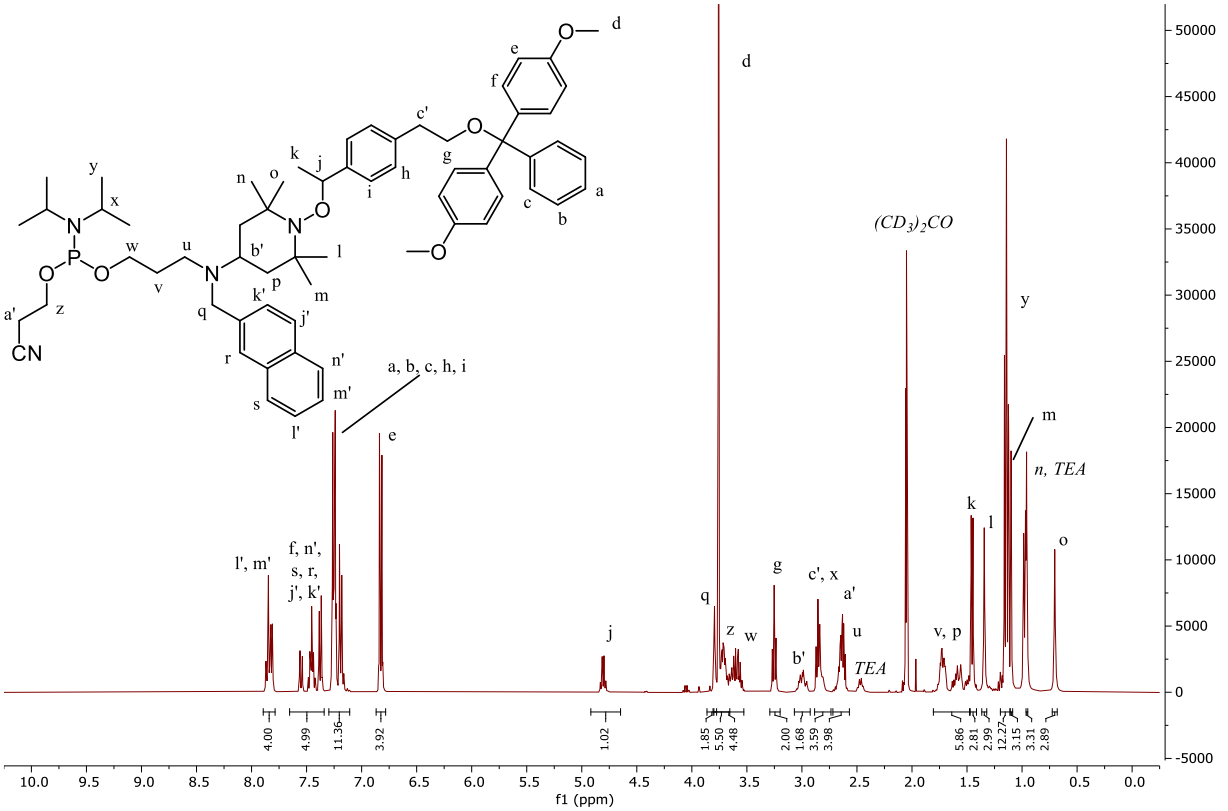


*Figure PE.3. 50 : Spectre RMN ^{1}H de l'Espaceur **TISP3-6**.*

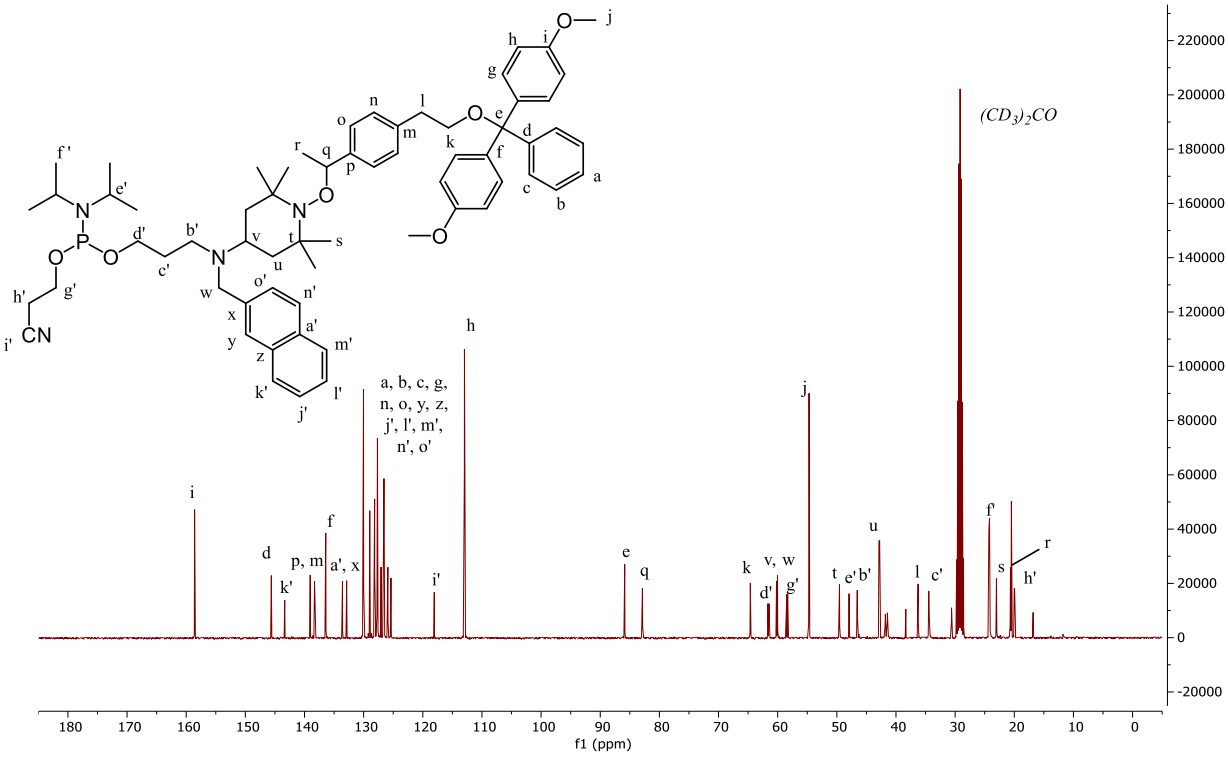


*Figure PE.3. 51 : Spectre RMN ^{13}C de l'Espaceur **TISP3-6**.*

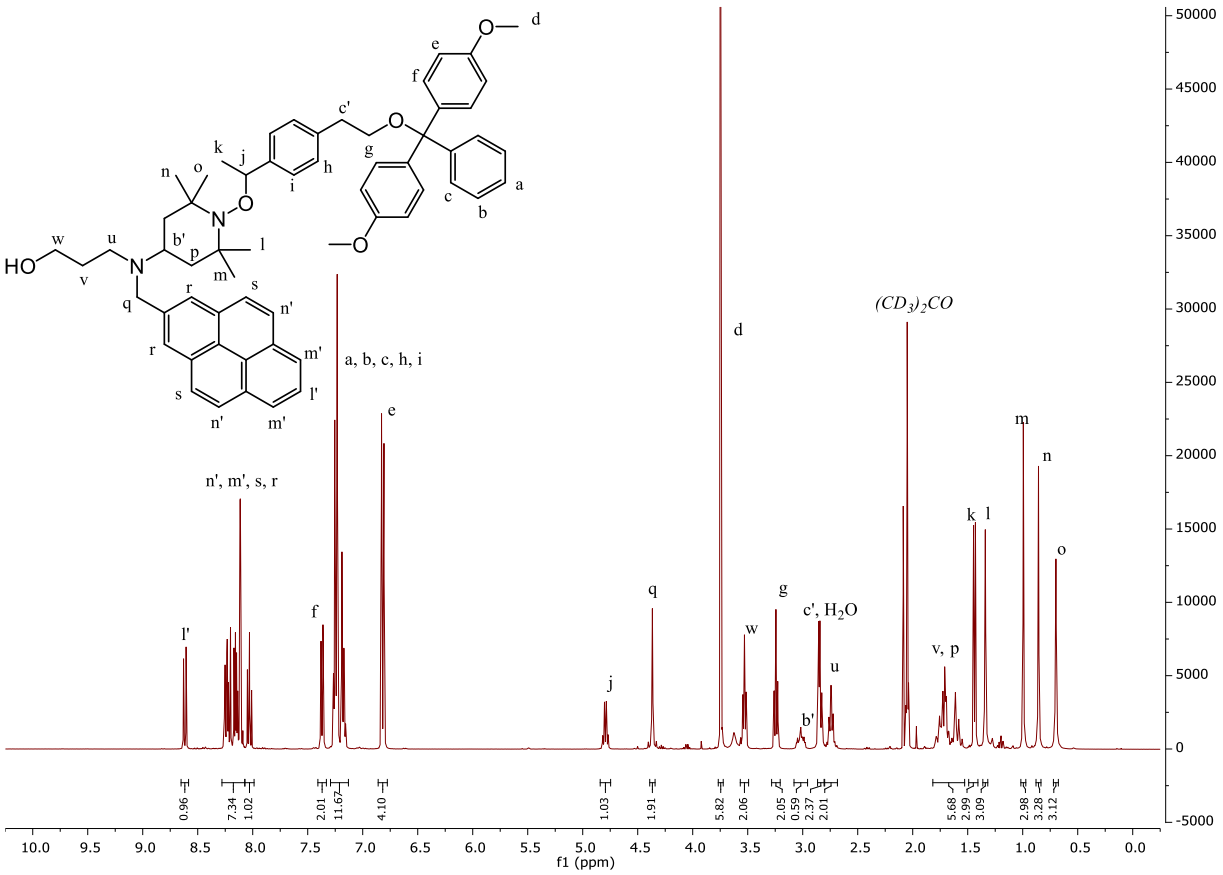


*Figure PE.3. 52 : Spectre RMN ^{1}H du composé **f7**.*

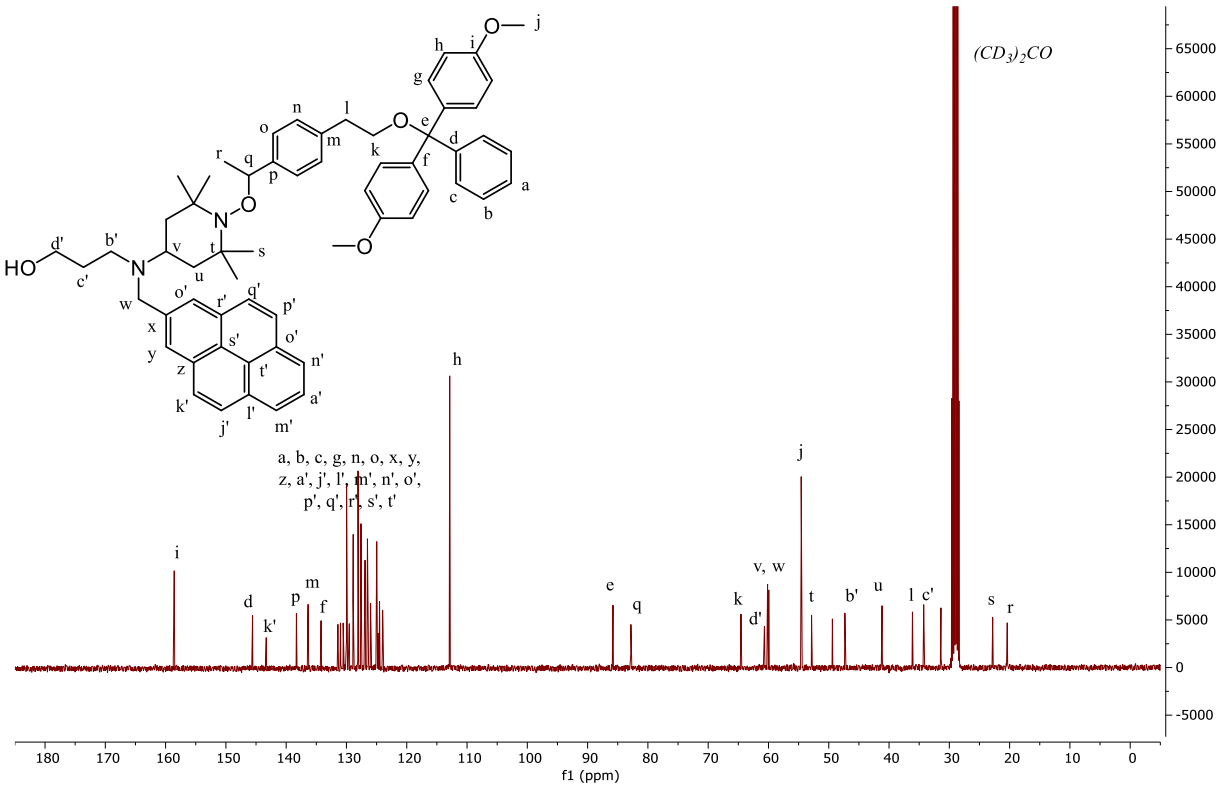


*Figure PE.3. 53 : Spectre RMN ^{13}C du composé **f7**.*

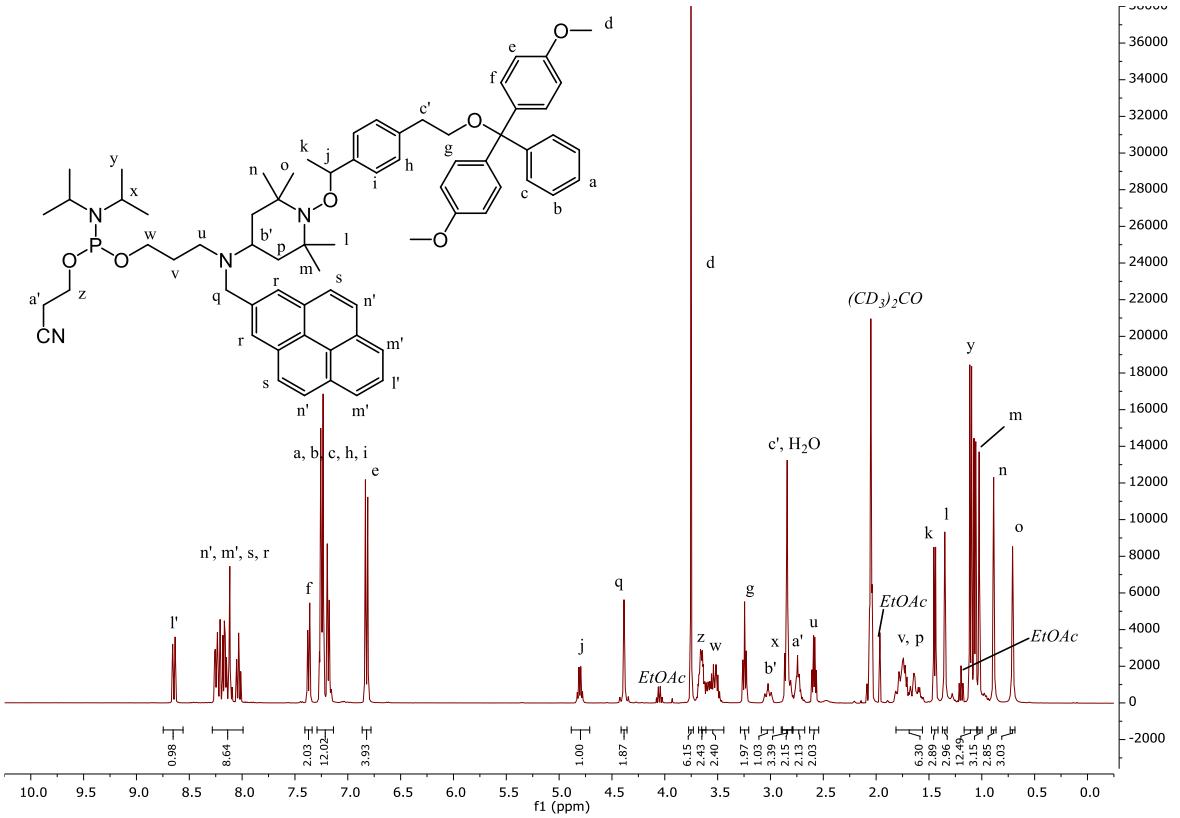


*Figure PE.3. 54 : Spectre RMN ^{1}H de l'Espaceur **TISP3-7**.*

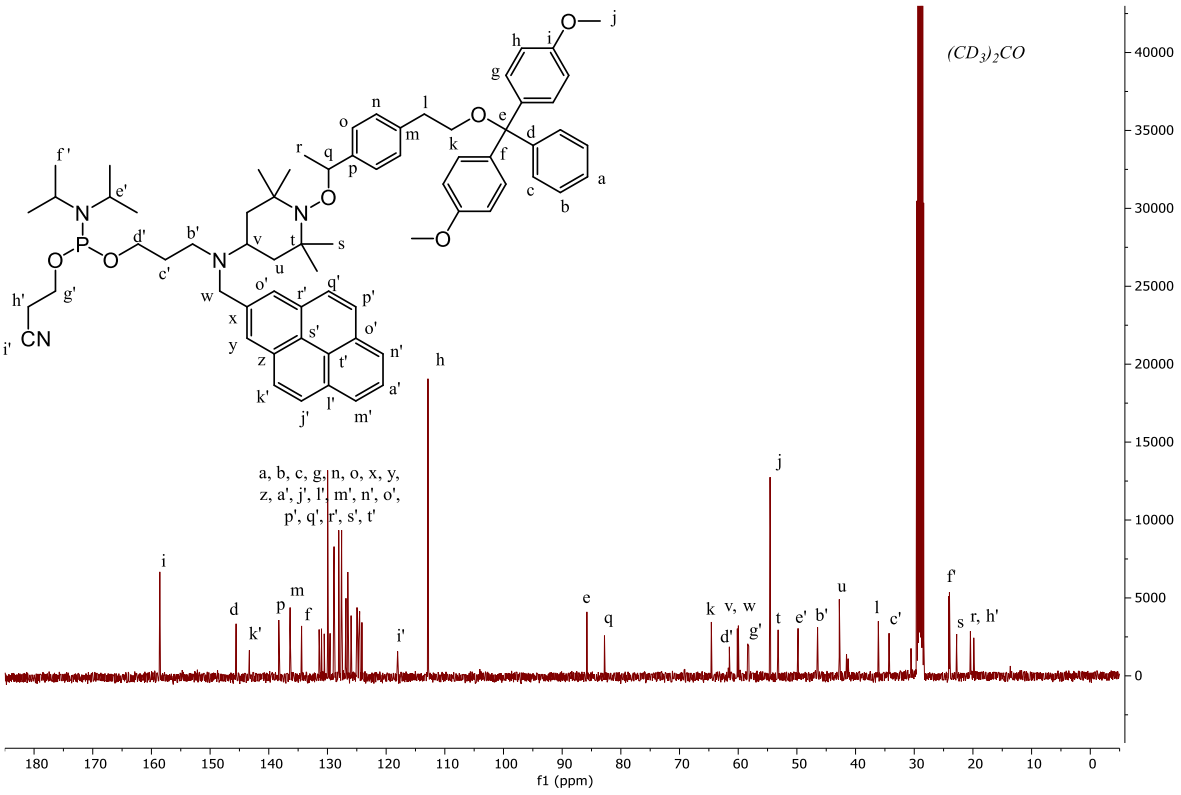


*Figure PE.3. 55 : Spectre RMN ^{13}C de l'Espaceur **TISP3-7**.*

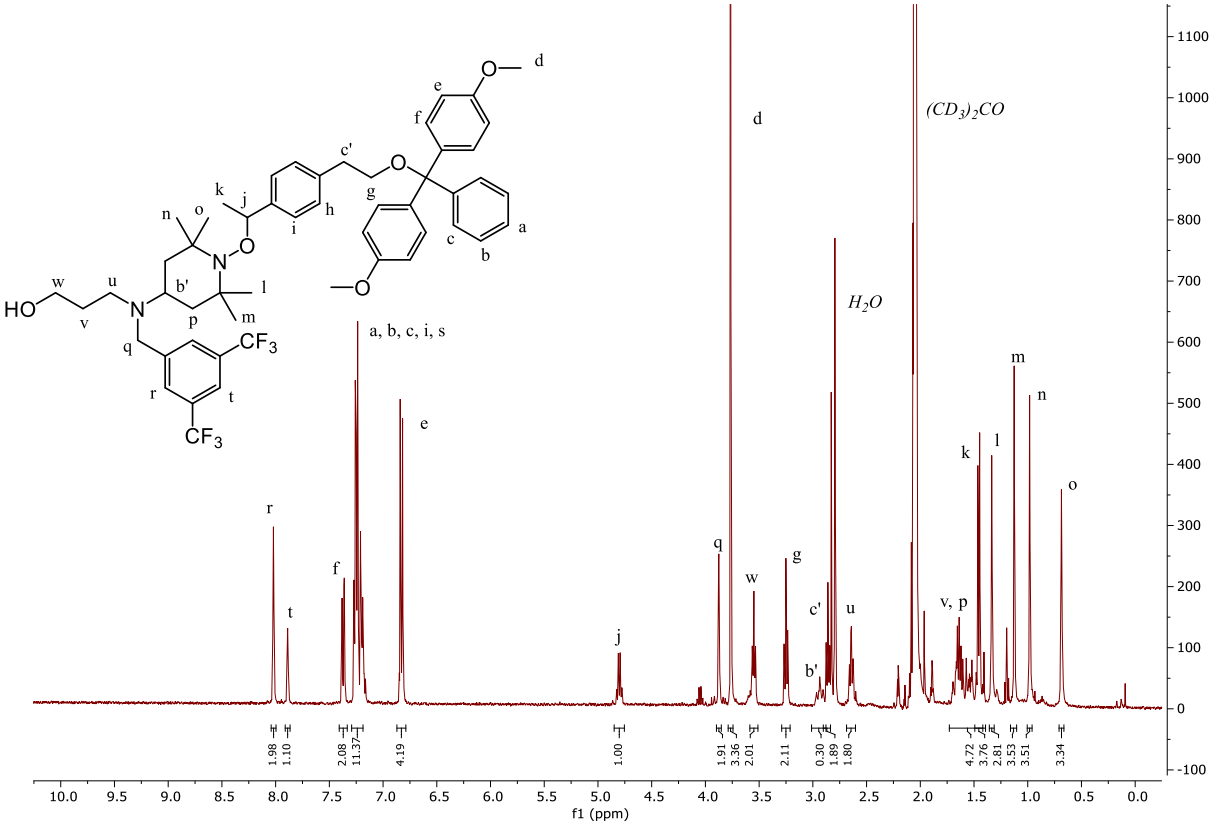


*Figure PE.3. 56 : Spectre RMN ^{1}H du composé **f8**.*

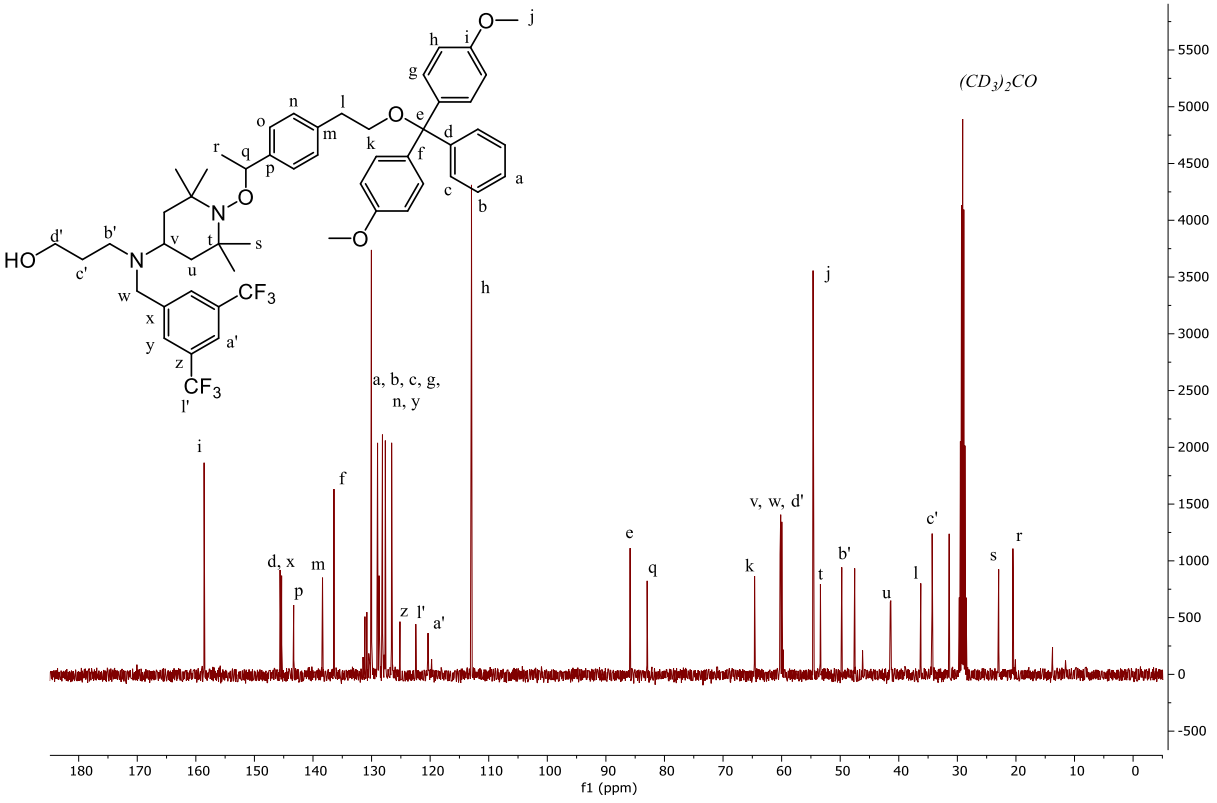


*Figure PE.3. 57 : Spectre RMN ^{13}C du composé **f8**.*

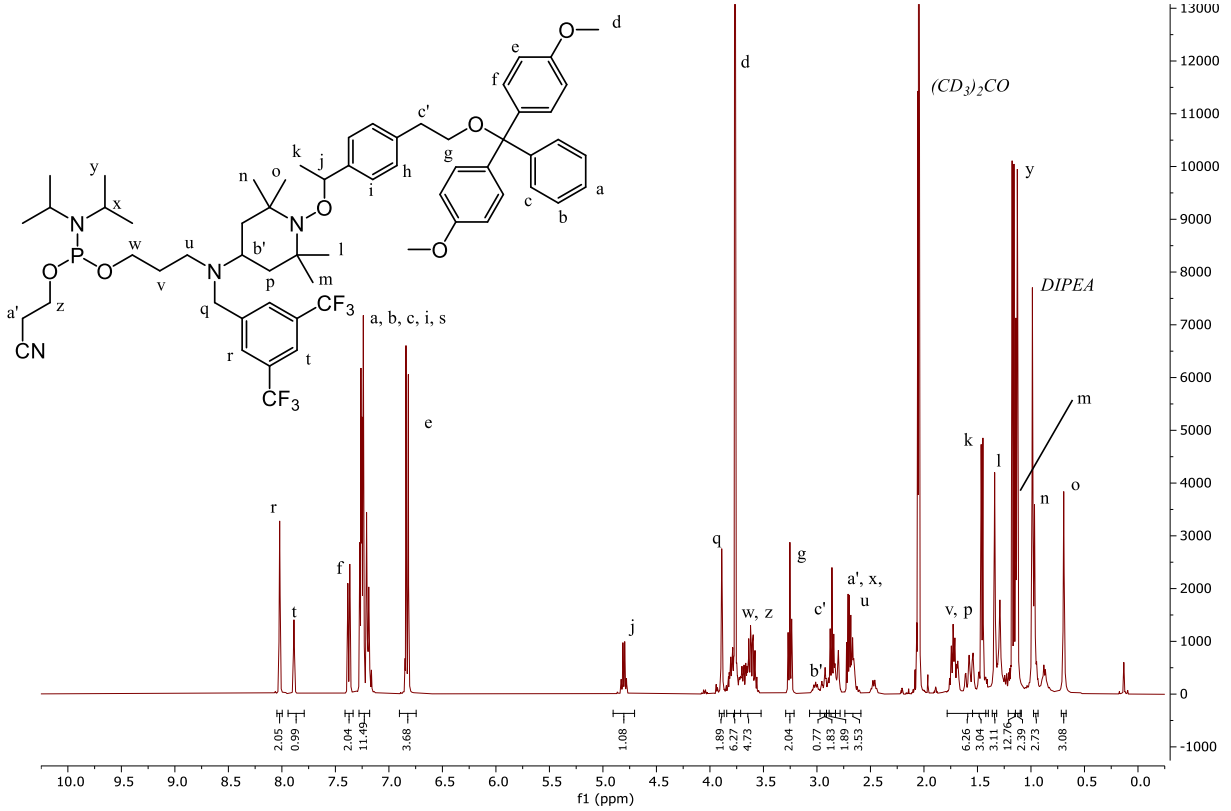


*Figure PE.3. 58 : Spectre RMN ^{1}H de l'Espaceur **TISP3-8**.*

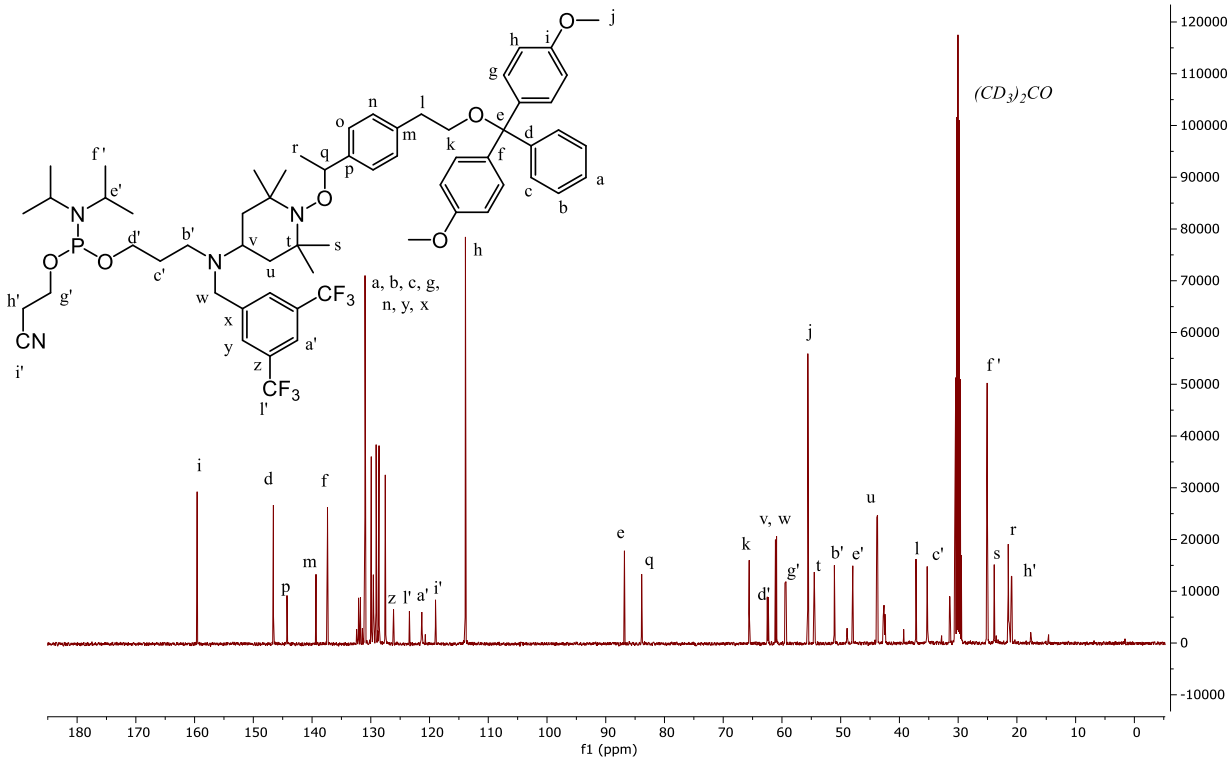


*Figure PE.3. 59 : Spectre RMN ^{13}C de l'Espaceur **TISP3-8**.*

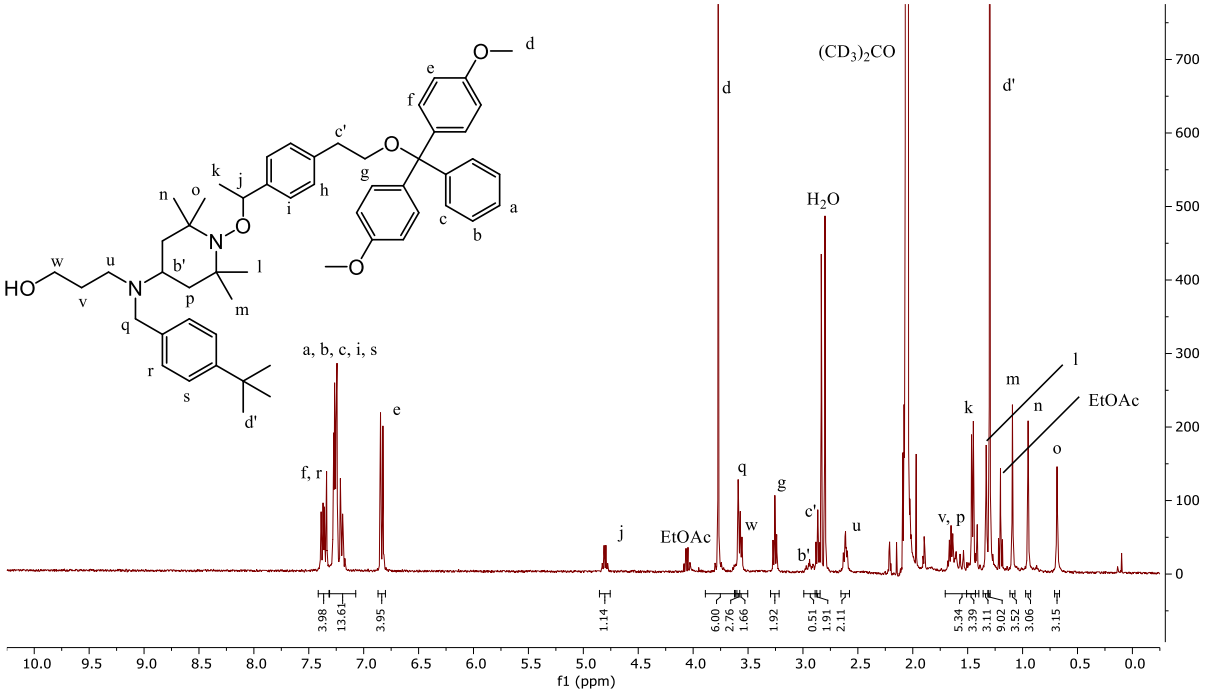


*Figure PE.3. 60 : Spectre RMN ^{1}H du composé **f9**.*

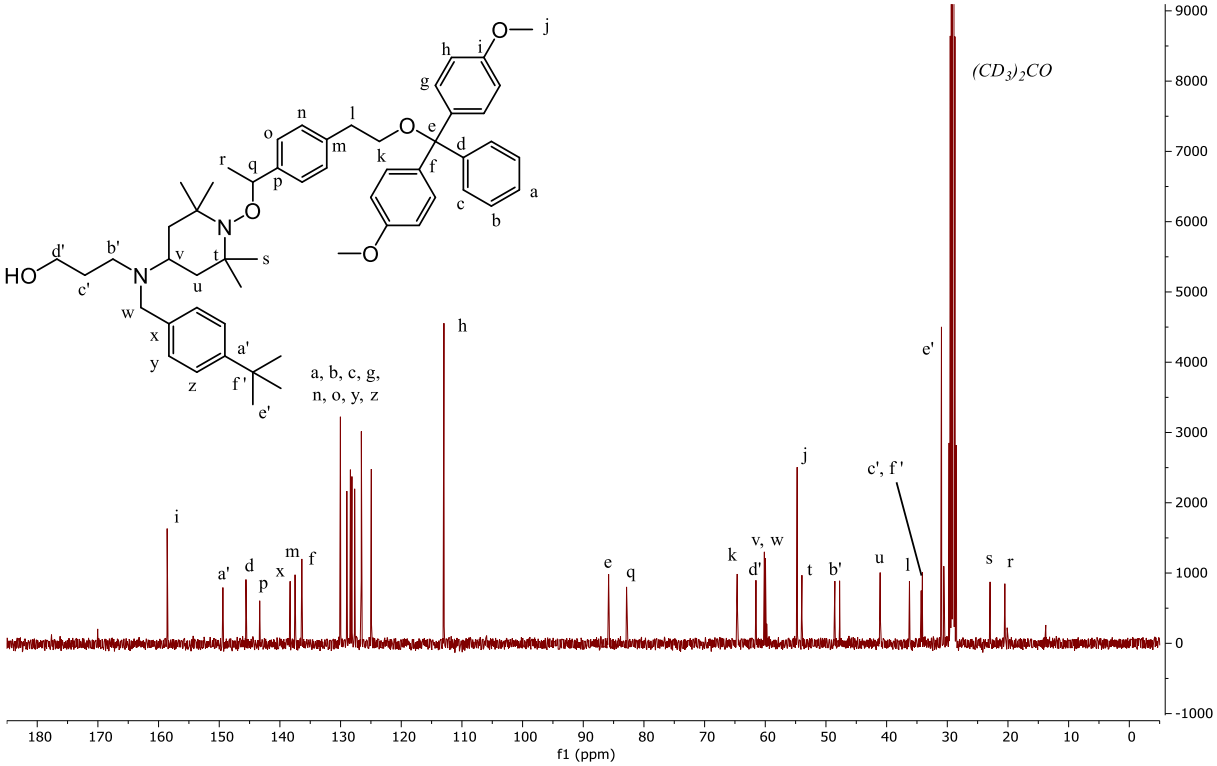


*Figure PE.3. 61 : Spectre RMN ^{13}C du composé **f9**.*

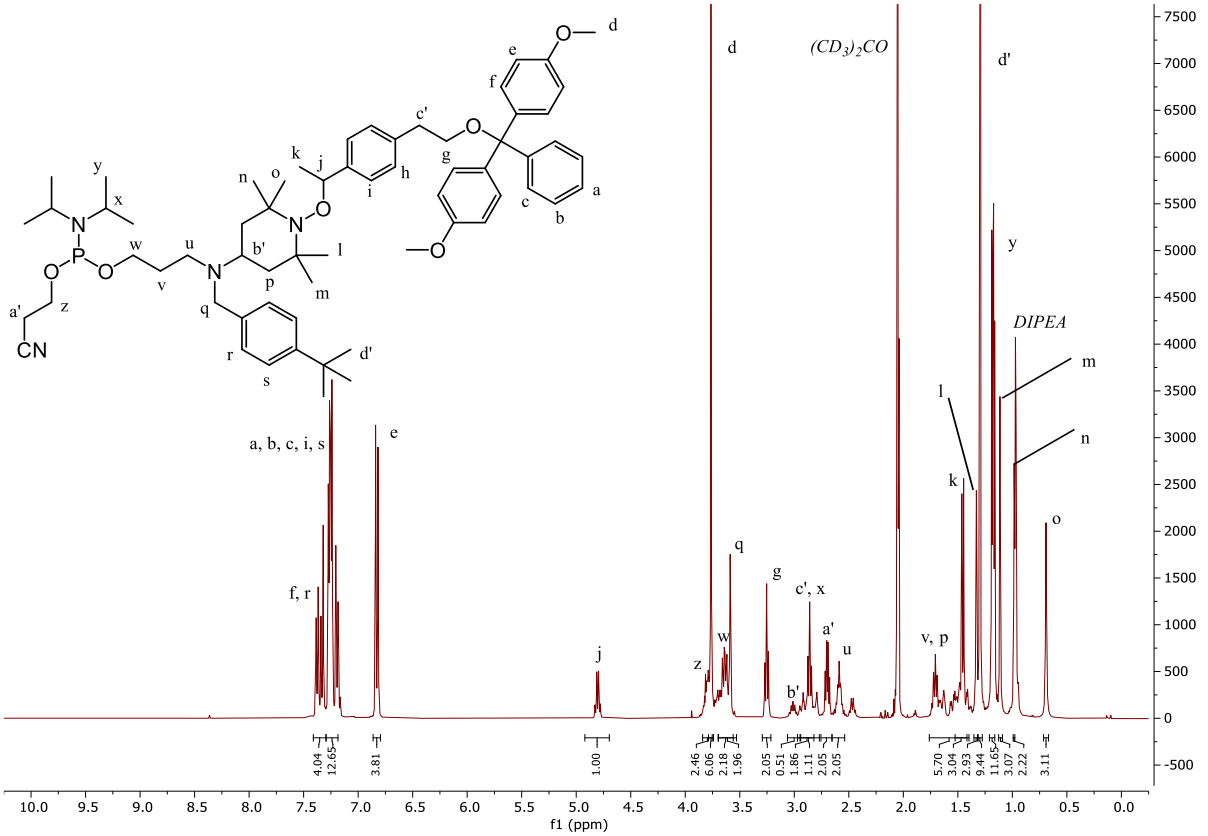


*Figure PE.3. 62 : Spectre RMN ^{1}H de l'Espaceur **TISP3-9**.*

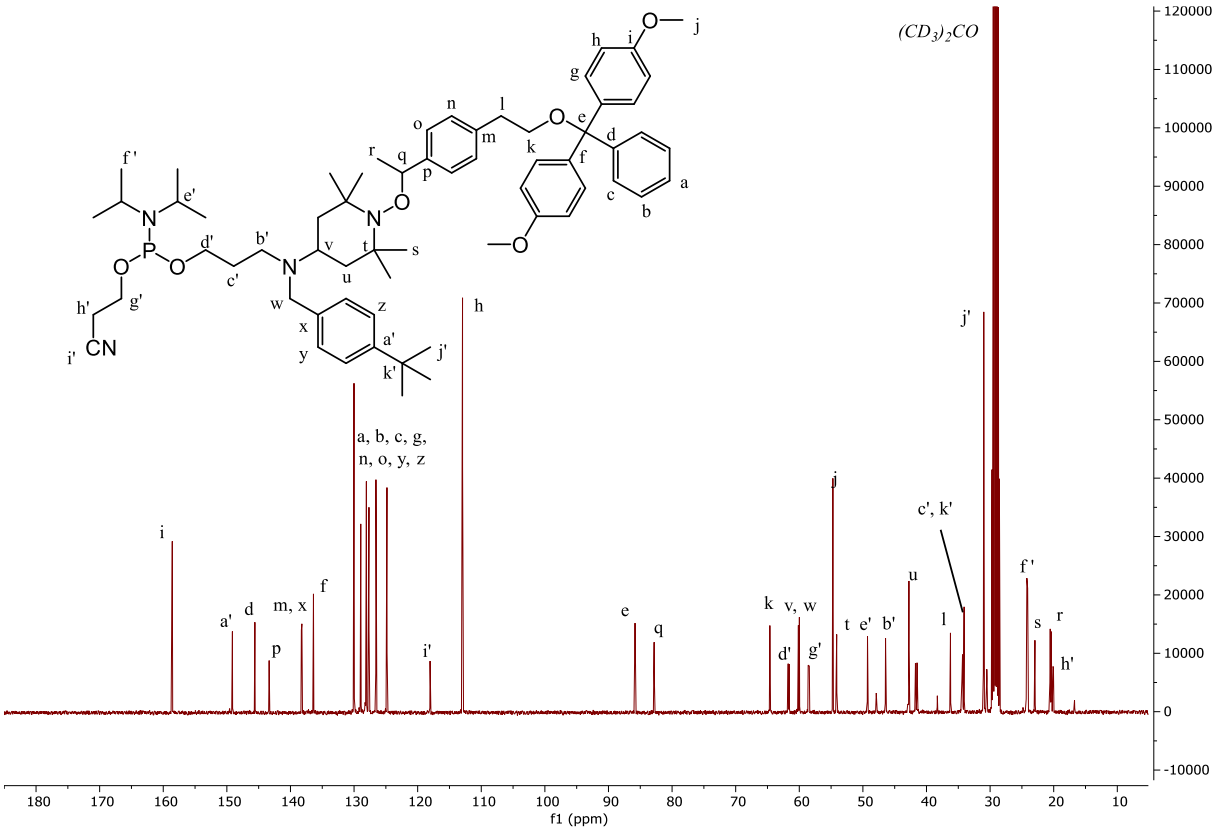


*Figure PE.3. 63 : Spectre RMN ^{13}C de l'Espaceur **TISP3-9**.*

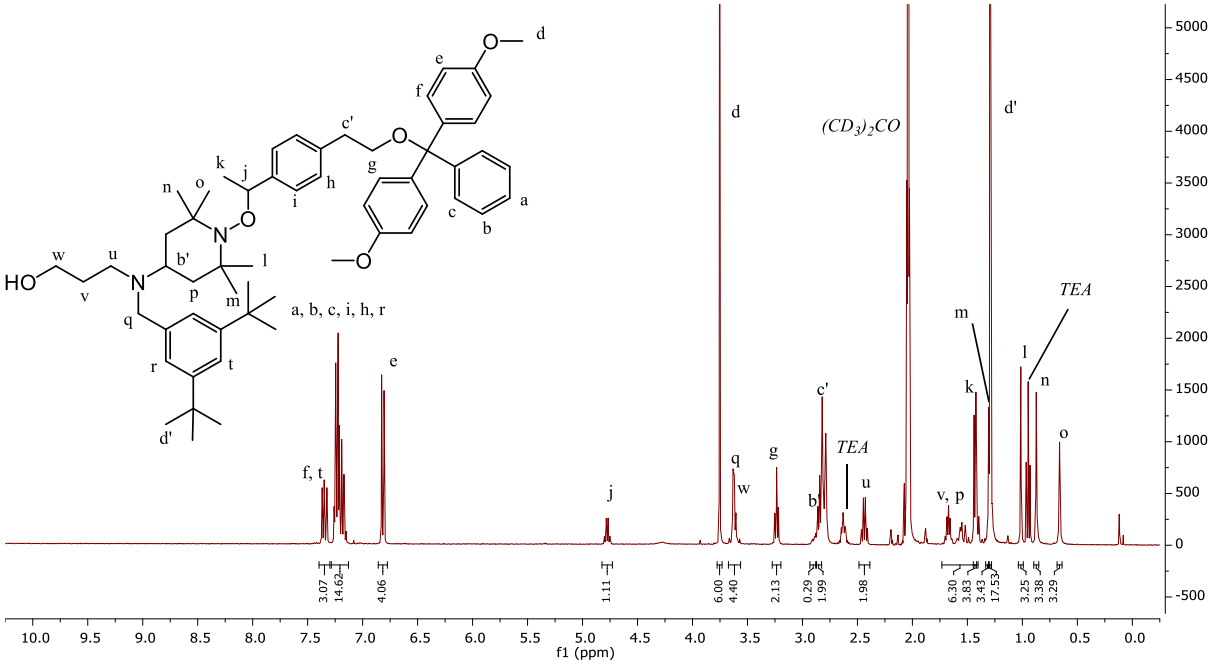


*Figure PE.3. 64 : Spectre RMN ^{1}H du composé **f10**.*

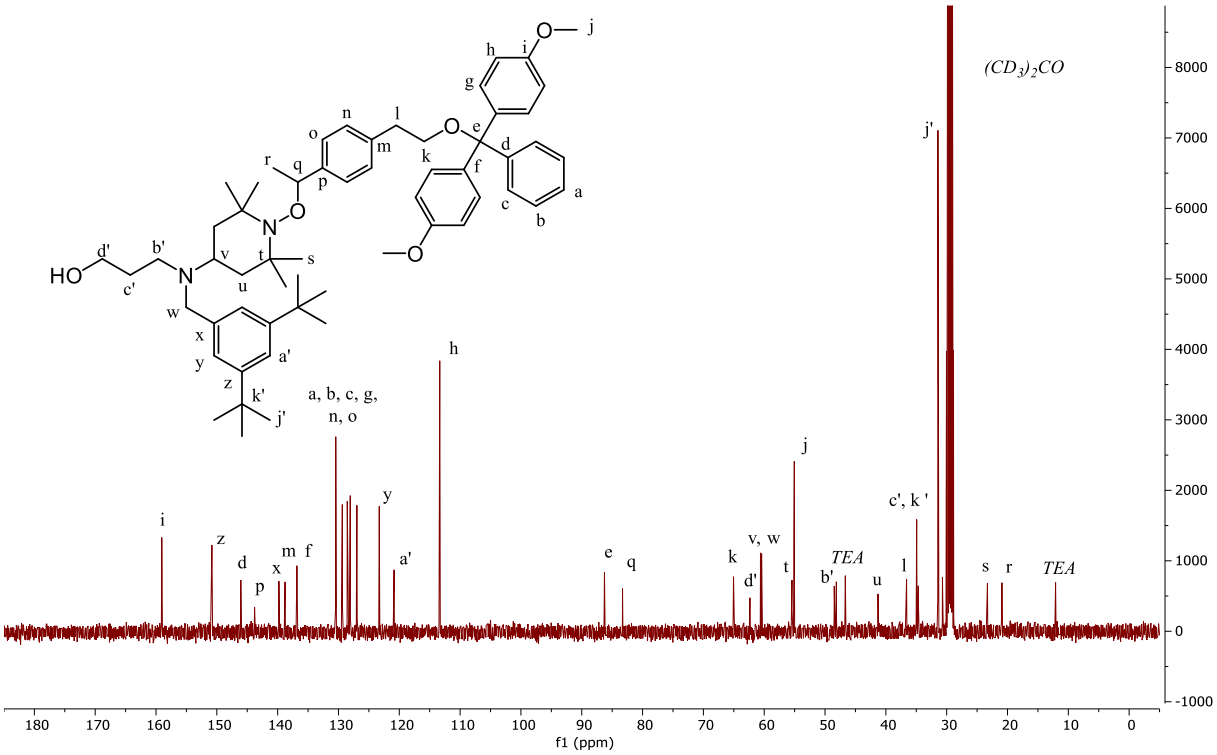


*Figure PE.3. 65 : Spectre RMN ^{13}C du composé **f10**.*

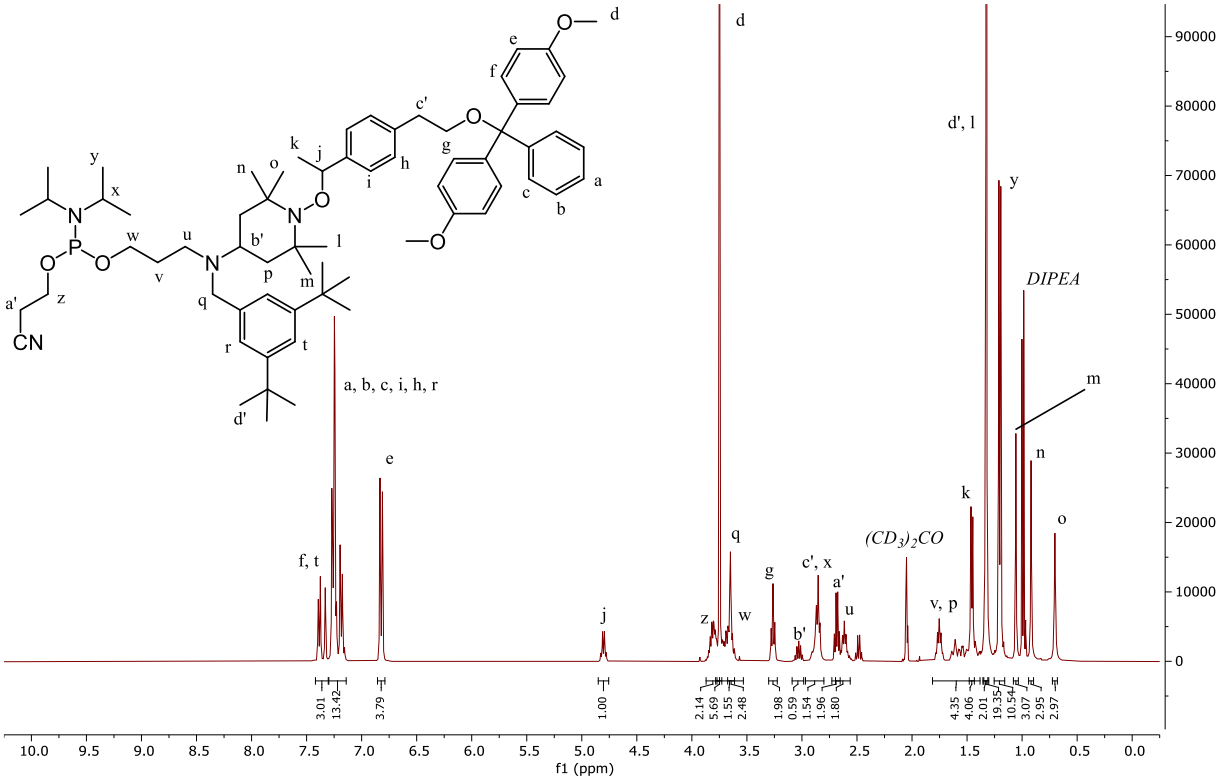


*Figure PE.3. 66 : Spectre RMN ^{1}H de l'Espaceur **TISP3-10**.*

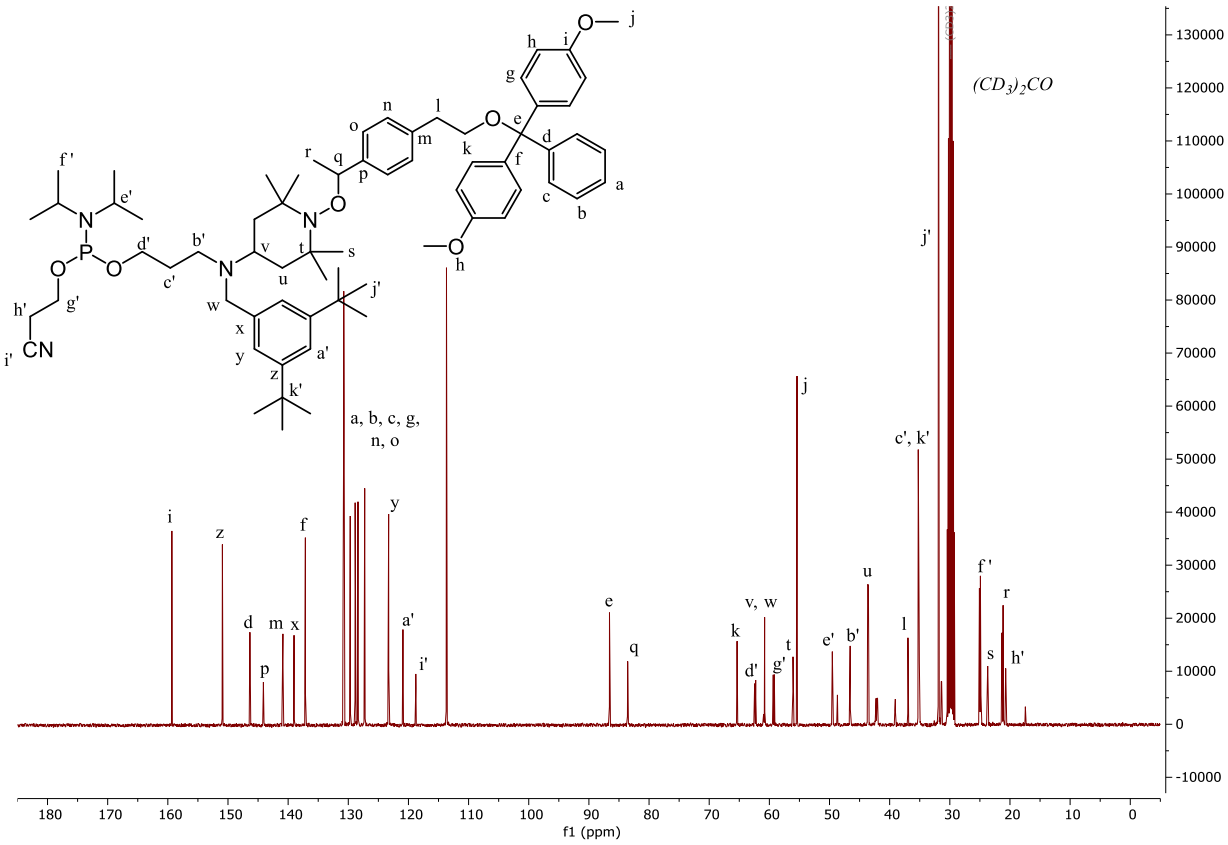


*Figure PE.3. 67 : Spectre RMN ^{13}C de l'Espaceur **TISP3-10**.*

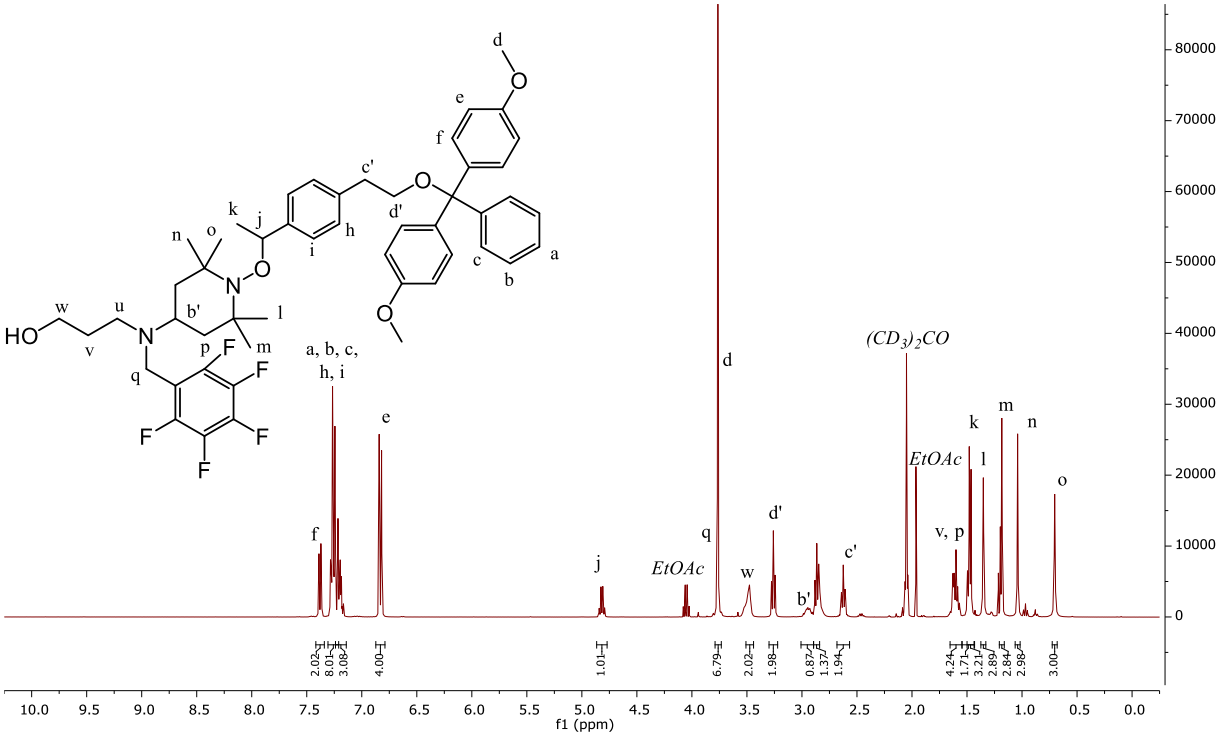


*Figure PE.3. 68 : Spectre RMN ^{1}H du composé **f11**.*

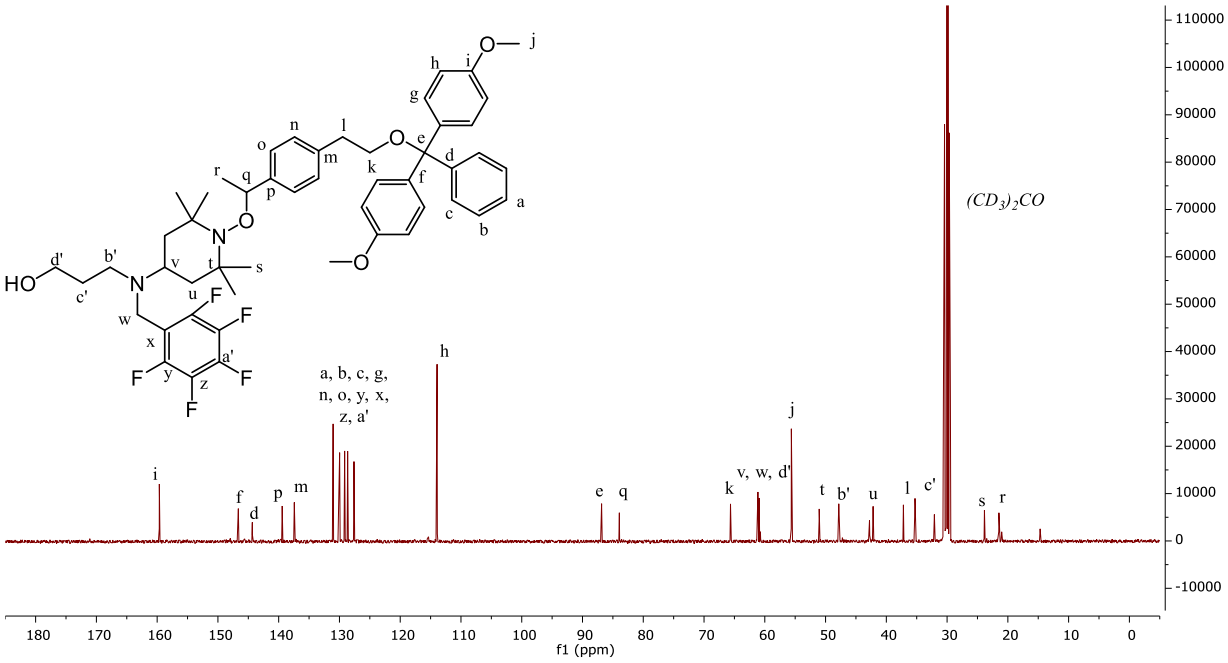


*Figure PE.3. 69 : Spectre RMN ^{13}C du composé **f11**.*

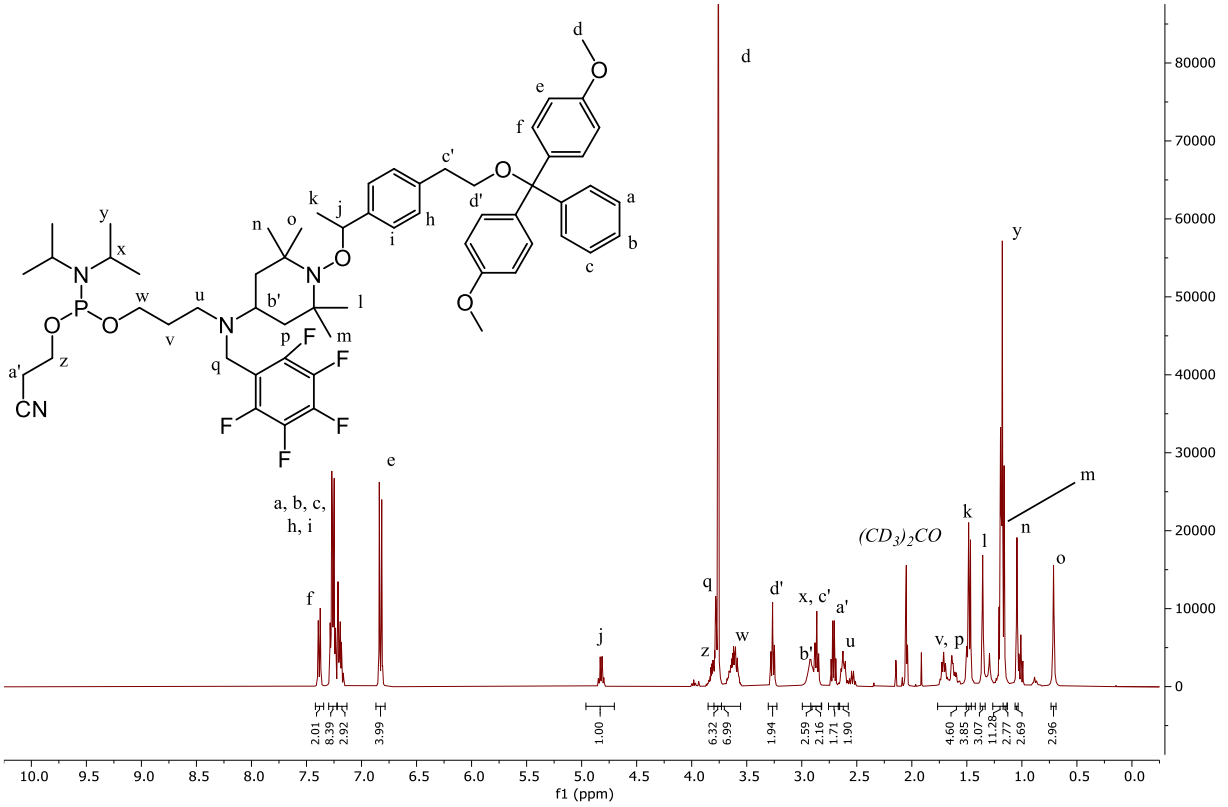


*Figure PE.3. 70 : Spectre RMN ^{1}H de l'Espaceur **TISP3-11**.*

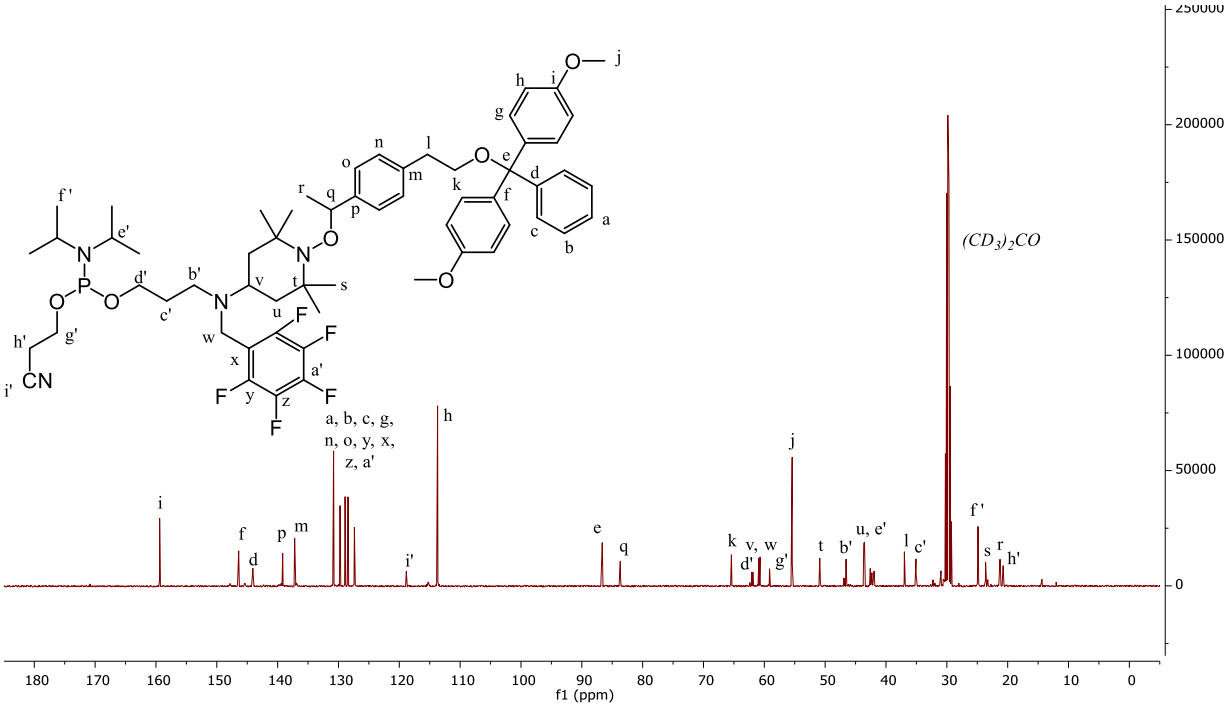


*Figure PE.3. 71 : Spectre RMN ^{13}C de l'Espaceur **TISP3-11**.*

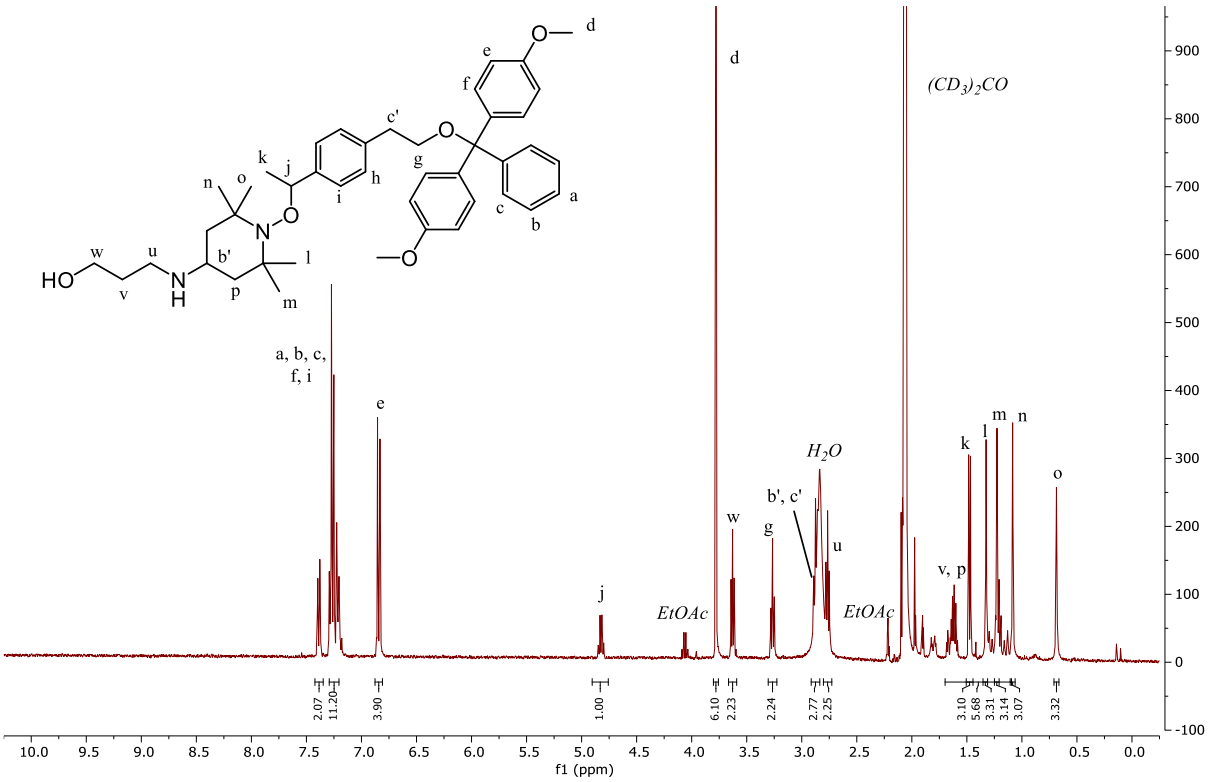


*Figure PE.3. 72 : Spectre RMN ^{1}H du composé **f12**.*

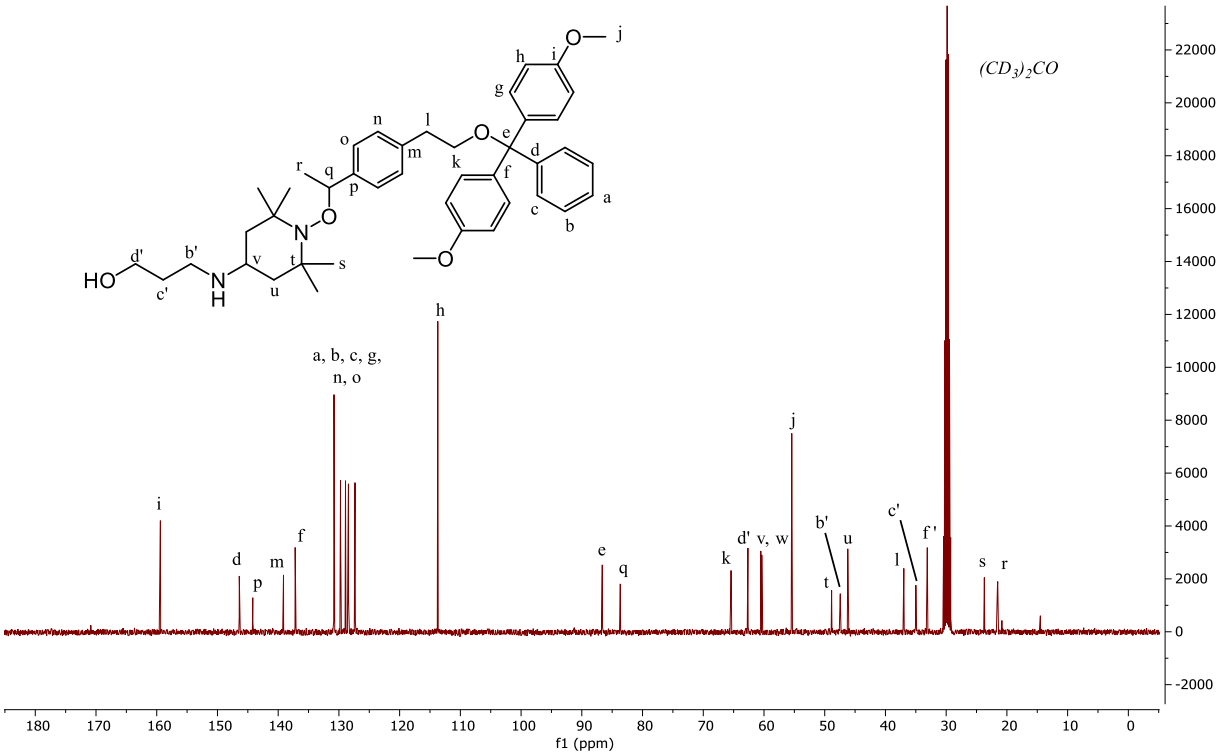


*Figure PE.3. 73 : Spectre RMN ^{13}C du composé **f12**.*

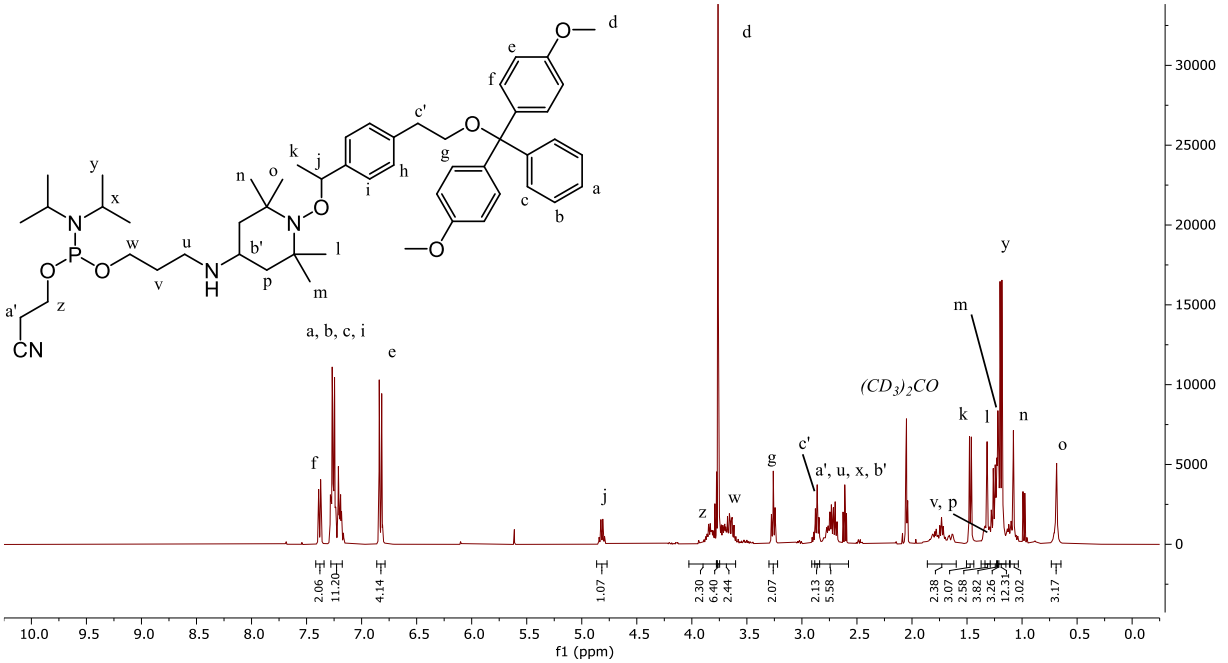


*Figure PE.3. 74 : Spectre RMN ^{1}H de l'Espaceur **TISP3-12**.*

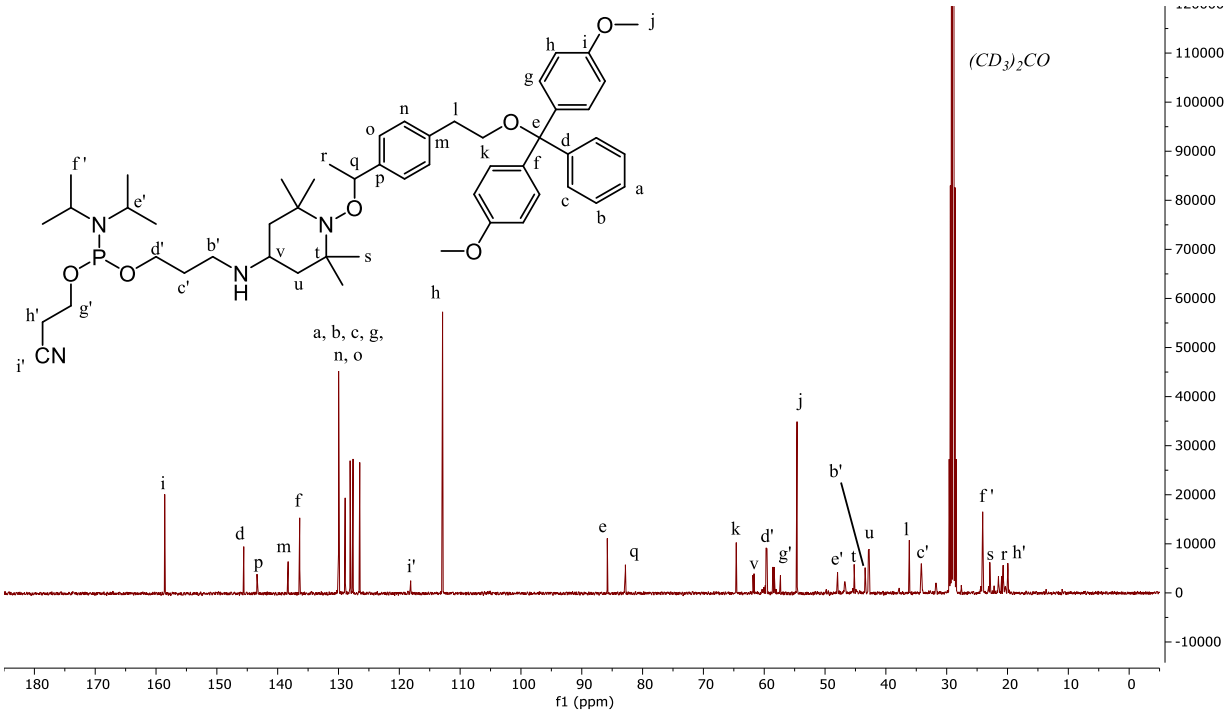


*Figure PE.3. 75 : Spectre RMN ^{13}C de l'Espaceur **TISP3-12**.*

3.2.2. RMN ^{31}P

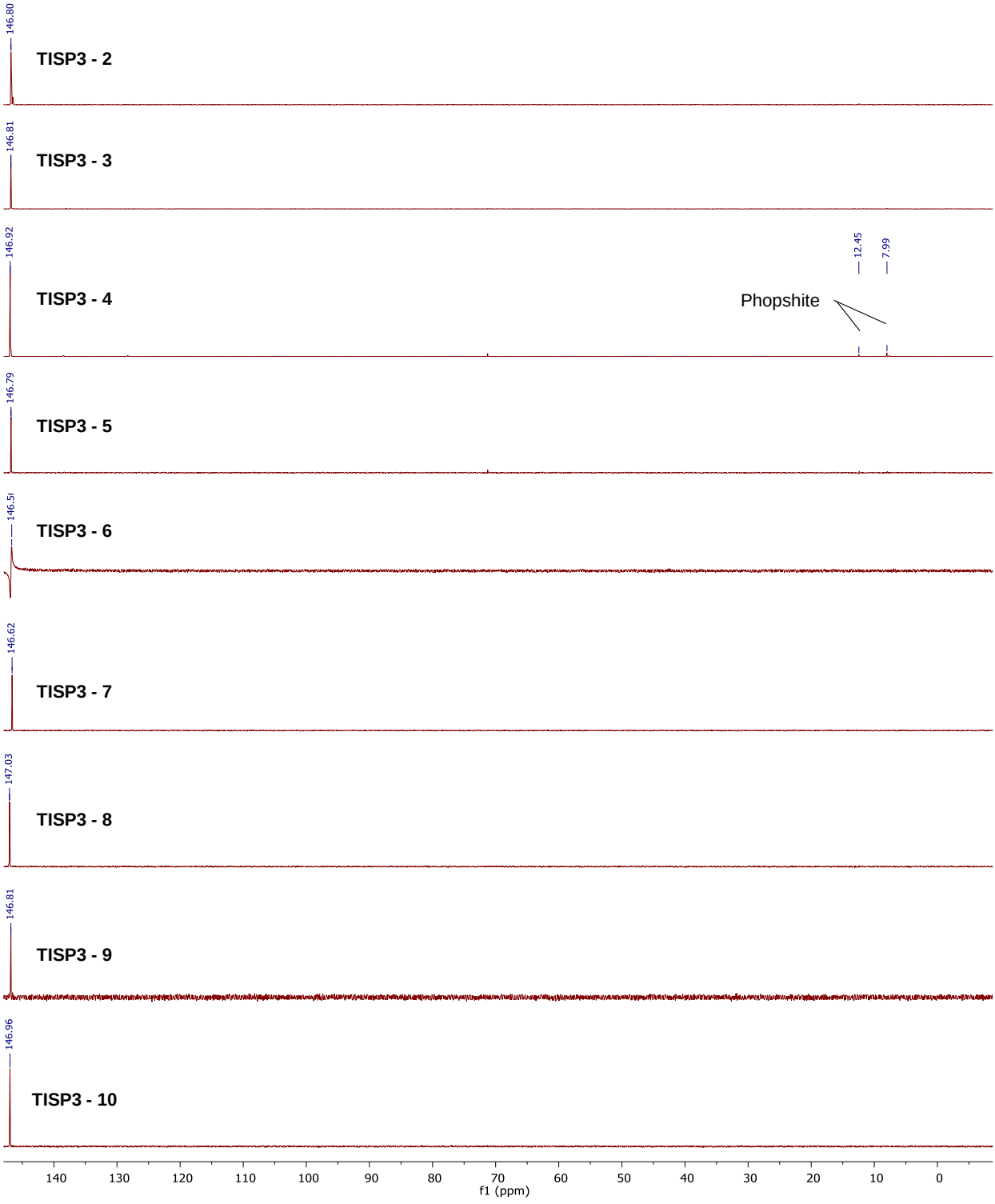


*Figure PE.3. 76 : Spectres RMN ^{31}P des Espaceurs **TISP3-2-10**.*

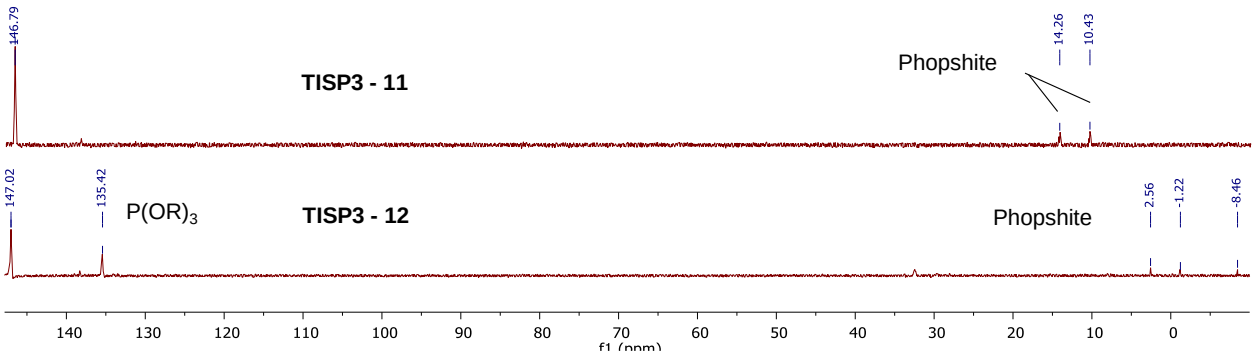


*Figure PE.3. 77 : Spectres RMN ^{31}P des Espaceurs **TISP3-11-12**.*

3.2.3. RMN ^{19}F

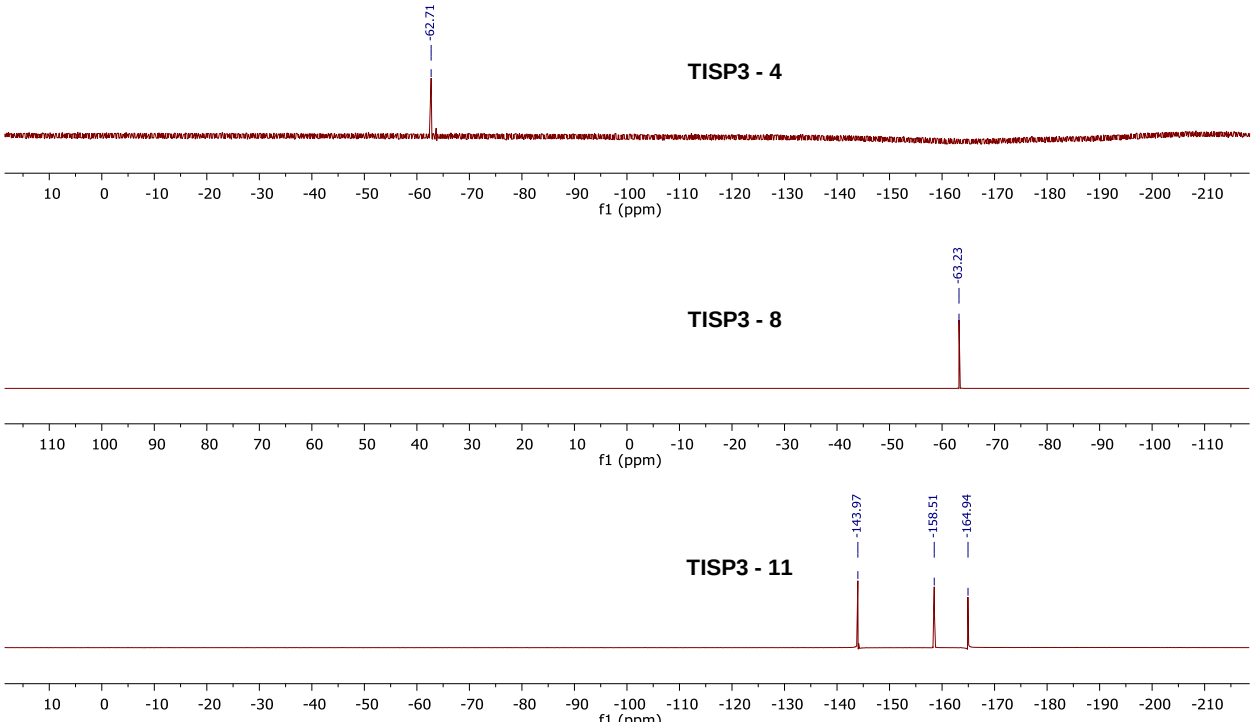


*Figure PE.3. 78 : Spectres RMN ^{19}F des Espaceurs **TISP3-4**, **TISP3-8** et **TISP3-11**.*

3.3. TISP4

3.3.1. RMN ^{1}H et ^{13}C

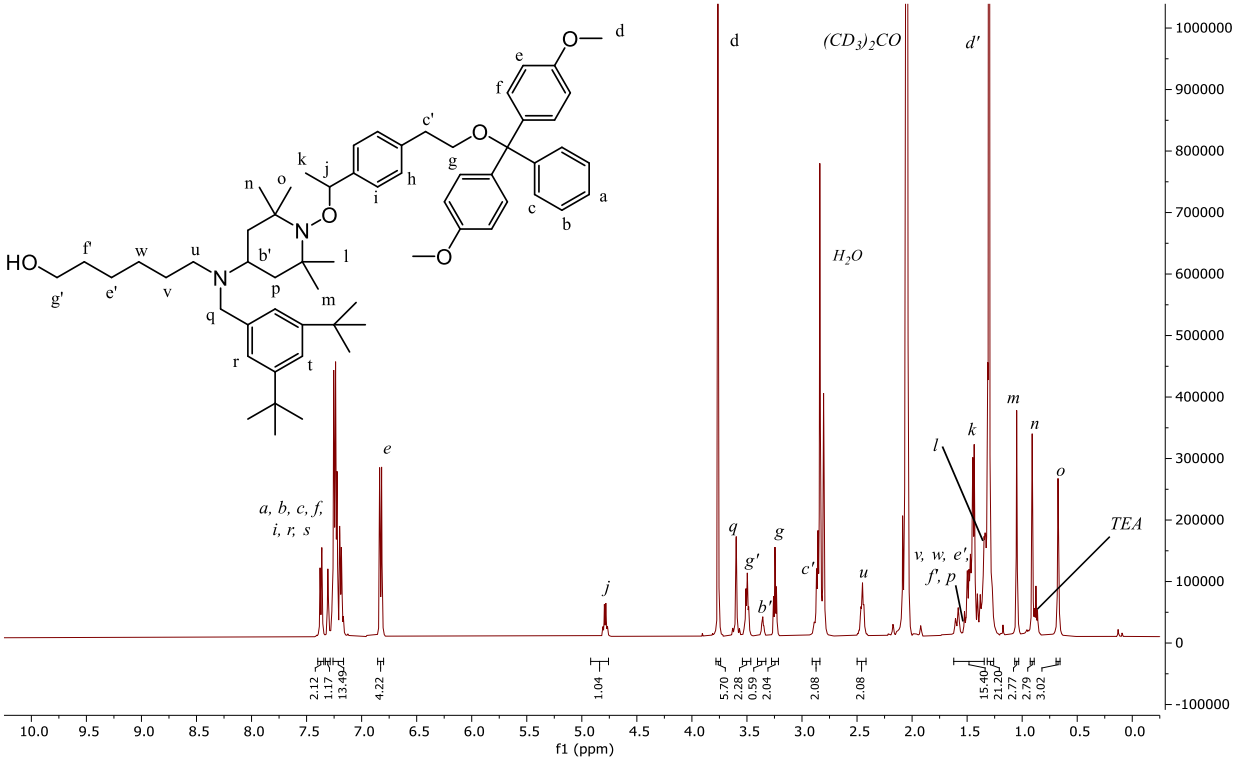


*Figure PE.3. 79 : Spectre RMN ^{1}H du composé **f13**.*

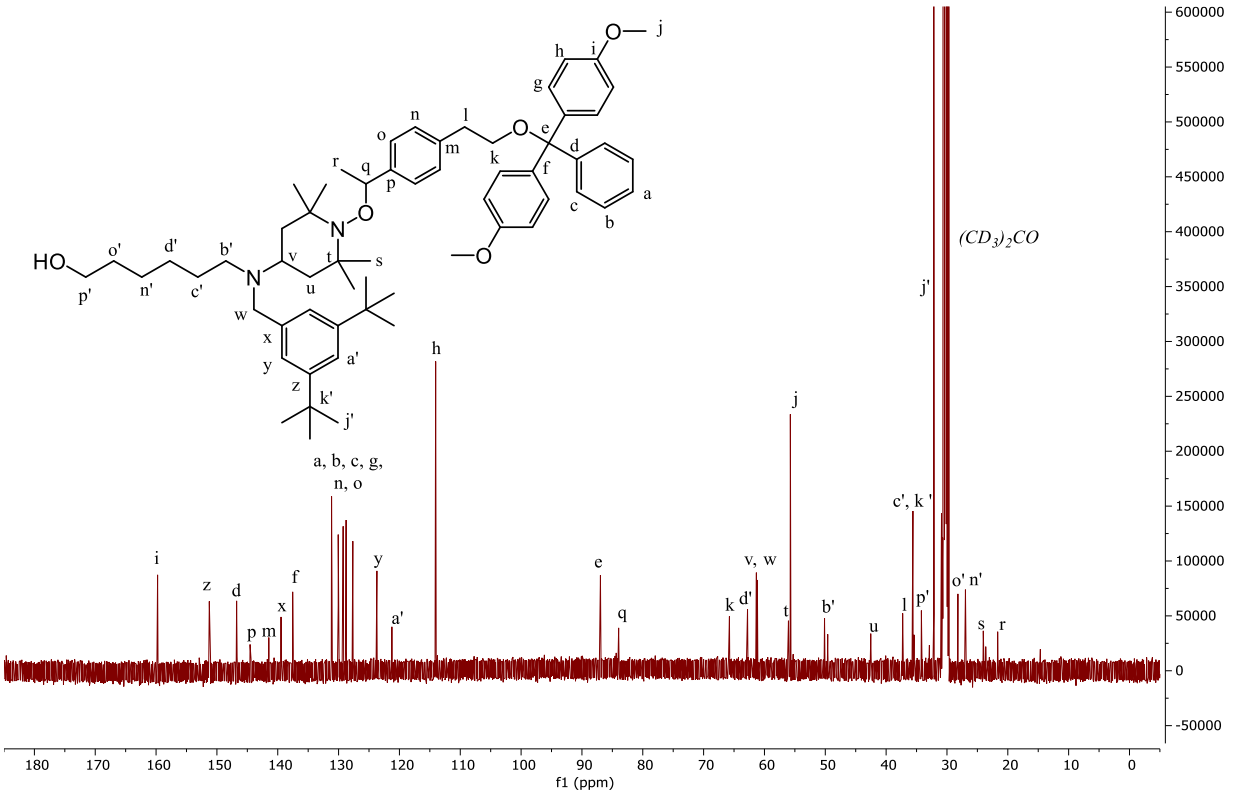


*Figure PE.3. 80 : Spectre RMN ^{13}C du composé **f13**.*

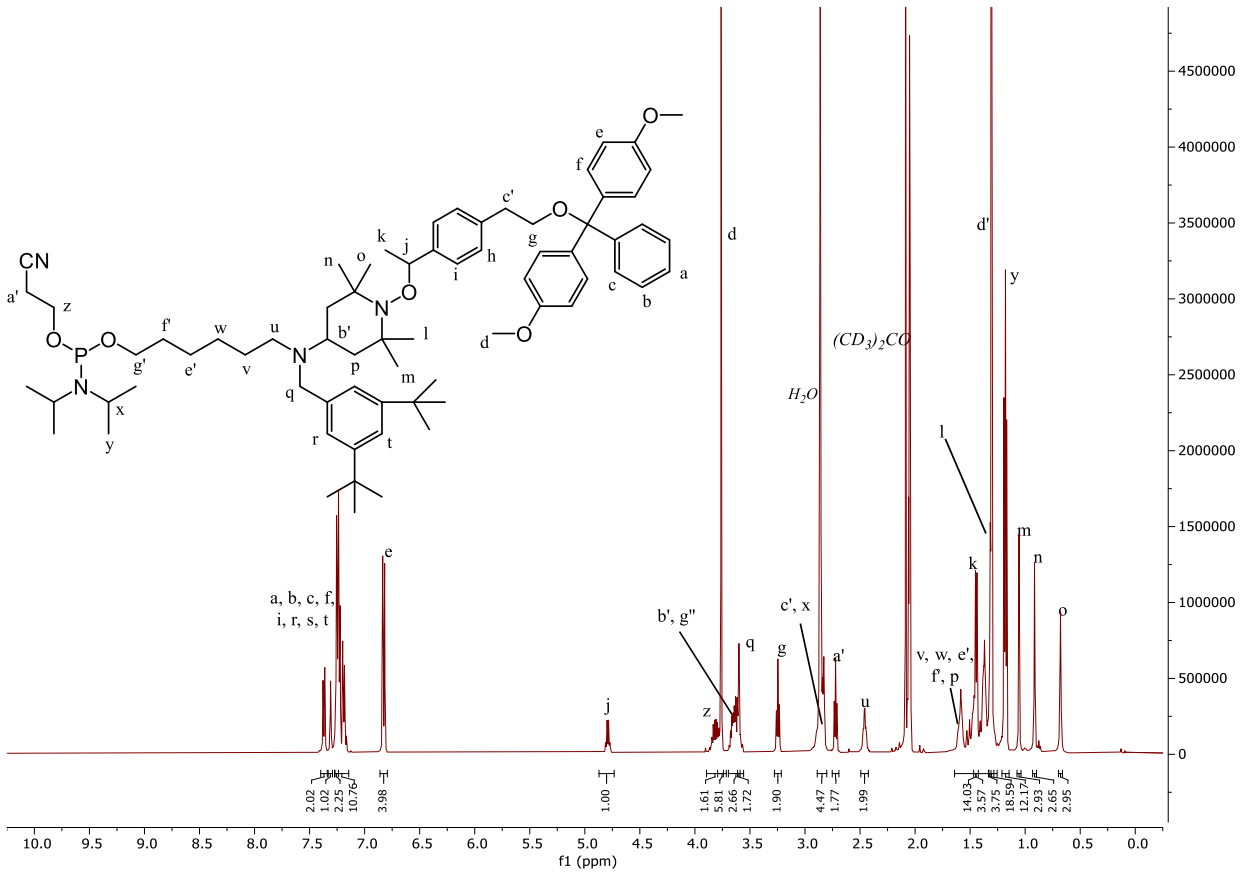


*Figure PE.3. 81 : Spectre RMN ^{1}H de l'Espaceur **TISP4**.*

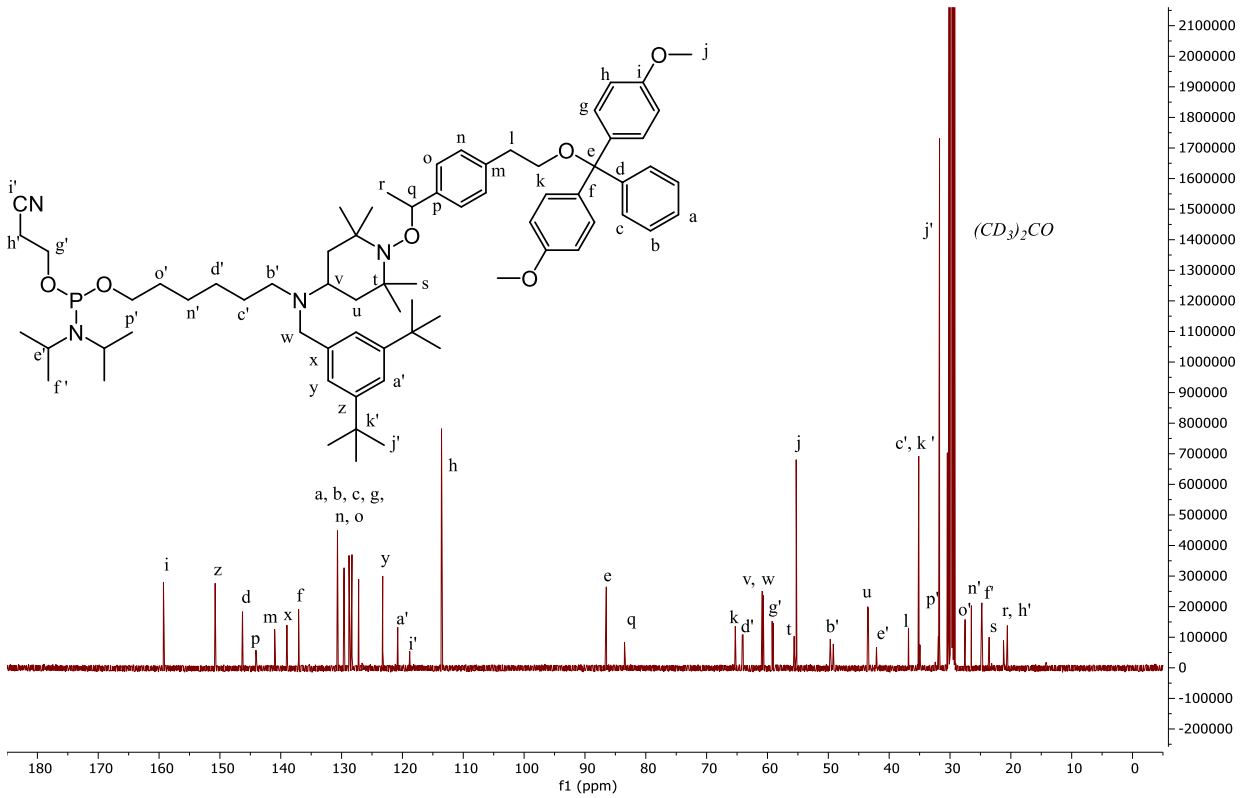


*Figure PE.3. 82 : Spectre RMN ^{13}C de l'Espaceur **TISP4**.*

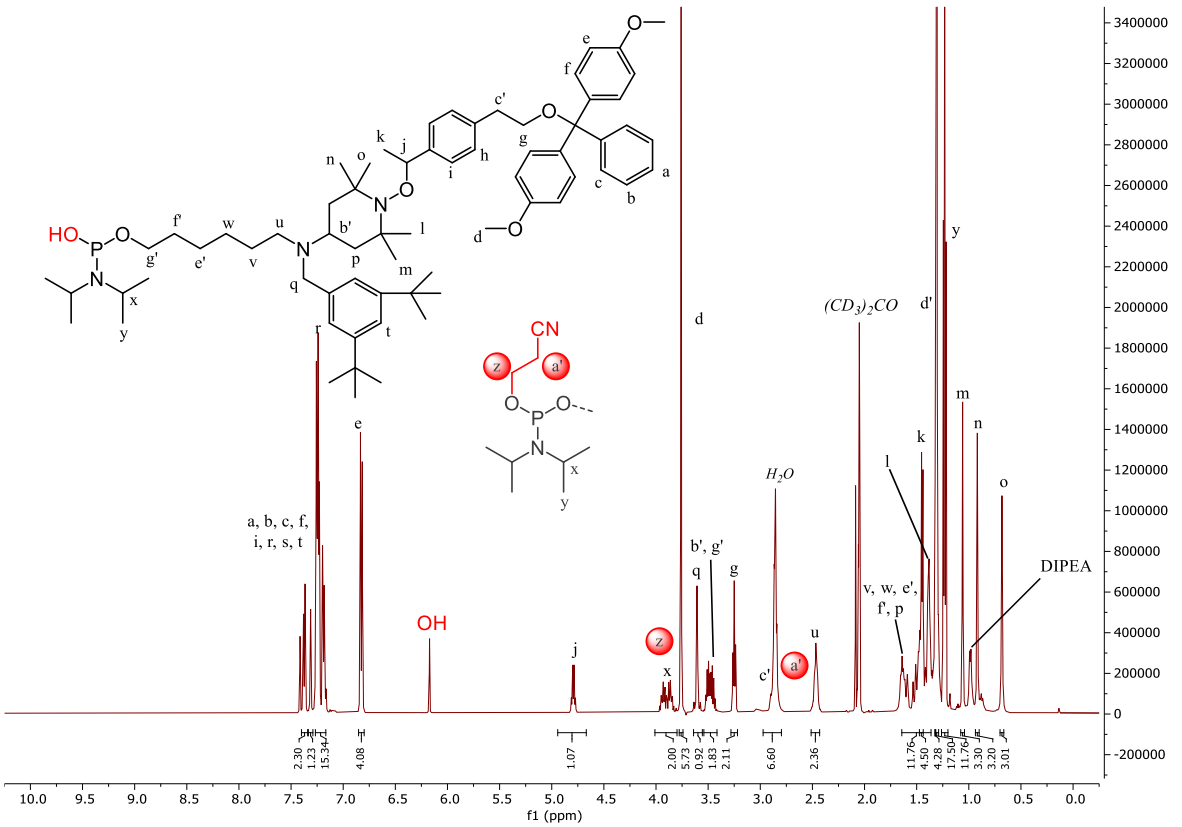


Figure PE.3. 83 : Spectre RMN ^{1}H du ***sous-produit phosphoramidite*** *identifié lors de la synthèse de* ***TISP4****.*

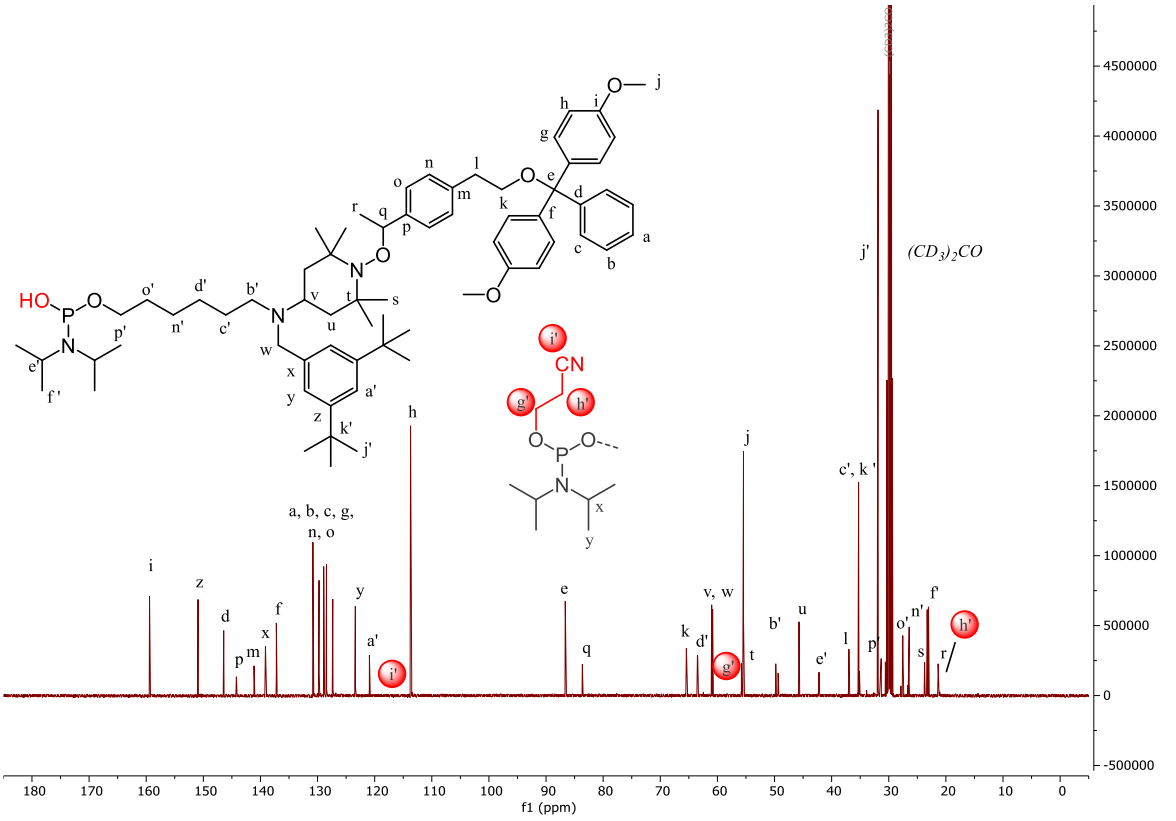


Figure PE.3. 84 : Spectre RMN ^{13}C du ***sous-produit phosphoramidite*** *identifié lors de la synthèse de* ***TISP4****.*

3.3.2. RMN ^{31}P

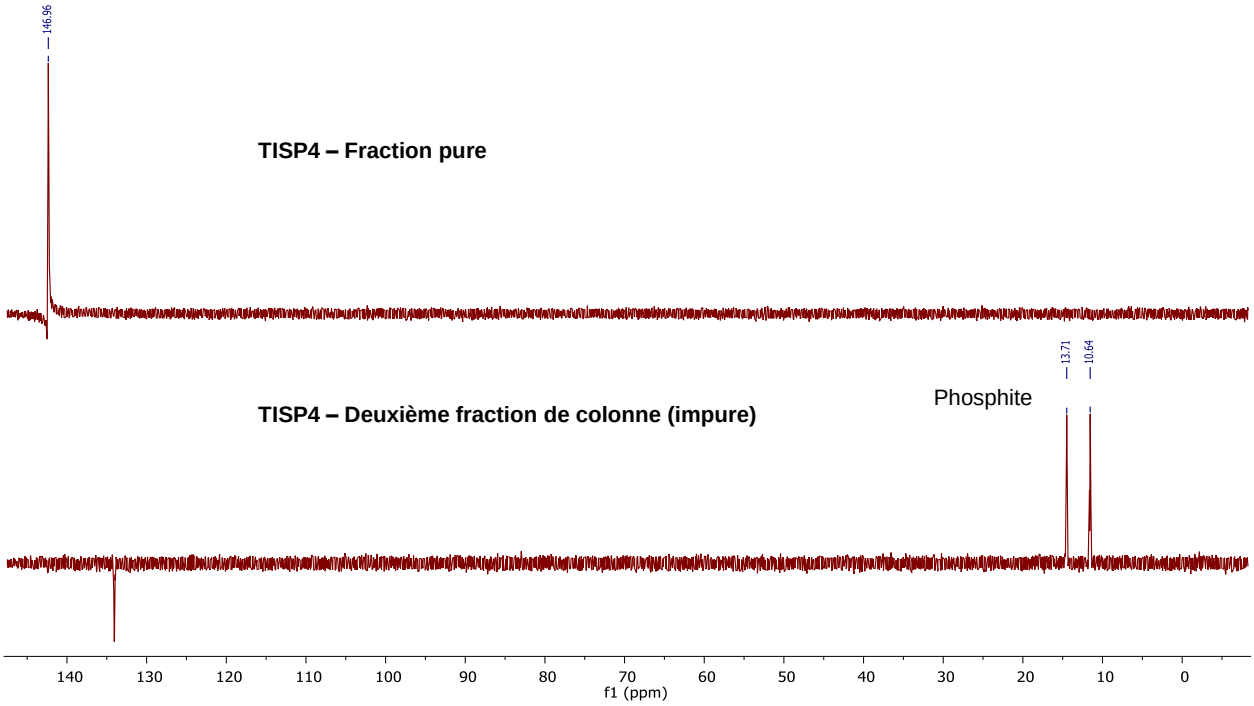
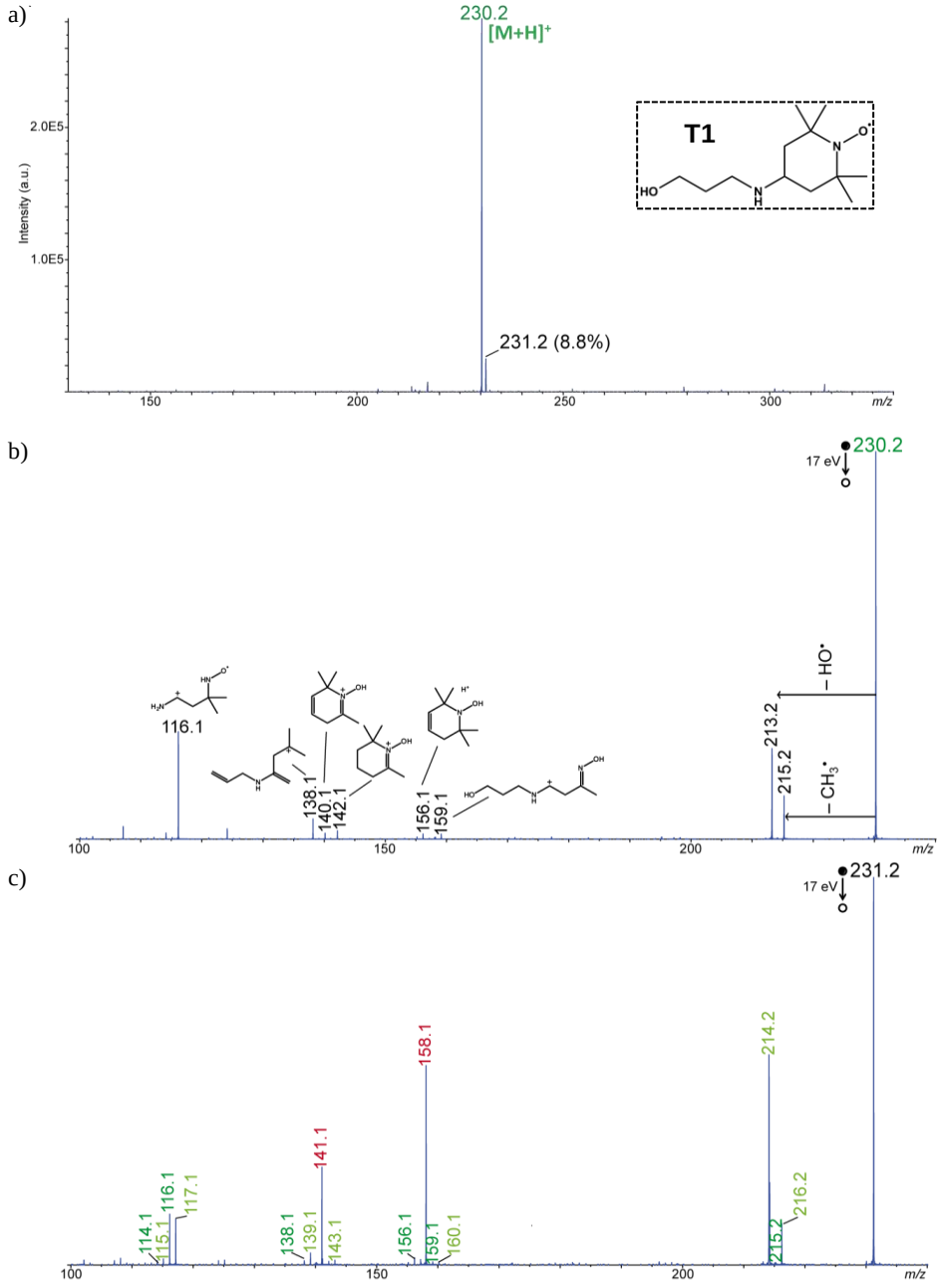


*Figure PE.3. 85 : Spectres RMN ^{31}P du composé **TISP4** et de son sous-produit.*

4. Spectres MS

4.1. Amino-TEMPO T1-T13



L'intensité relative du pic à *m/z* 231.2 par rapport au pic à *m/z* 230.2 (indiquée en noir, est comparable au rapport isotopique théorique (14.4%), ce qui suggère que l'ion *m/z* 231.2 correspondrait au seul pic ^{13}C du nitroxide protoné. Si l'ion *m/z* 231.2 correspond exclusivement au nitroxide protoné contenant un ^{13}C, ses fragments sont attendus i) à la même valeur *m/z* (en vert foncé) et/ou ii) à *m/z* +1 (en vert clair) comparée aux fragments de l'ion *m/z* 230.2 (selon que ces fragments ne contiennent plus ou contiennent encore le ^{13}C). En revanche, si l'ion *m/z* 231.2 correspond aussi à l'hydroxylamine protonée, de nouveaux fragments (en rouge) apparaissent.

*Figure PE.4. 1 : Spectres MS du composé **T1**.*
*a) Spectres ESI-MS enregistrés en mode positif, spectre ESI-MS/MS de l'ion m/z b) 230.2 et c) 231.2 pour l'échantillon **T1**.*

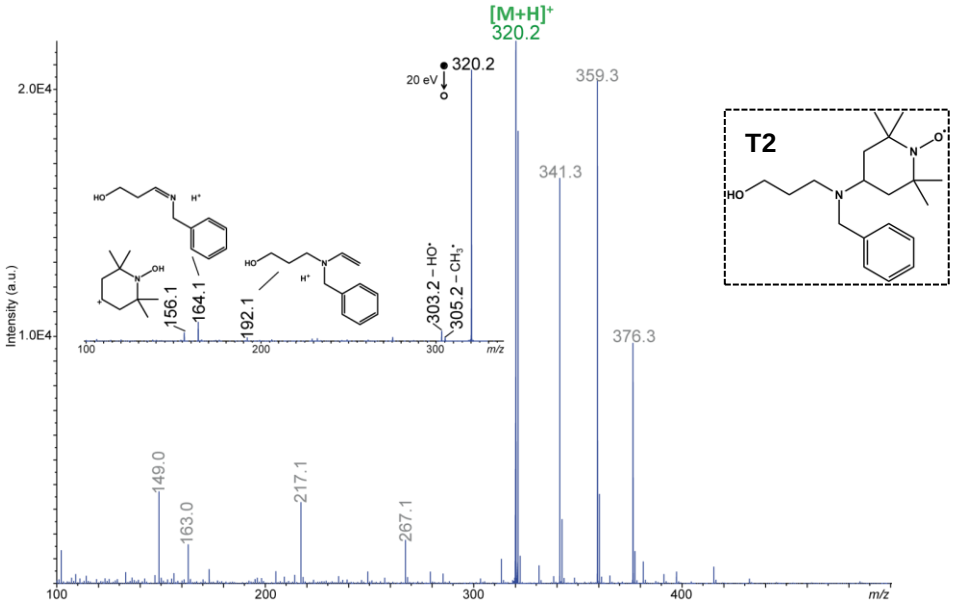


*Figure PE.4. 2 : Spectres MS du composé **T2**.*
*Spectre ESI-MS enregistré en mode positif pour l'échantillon **T2** et spectre MS/MS de l'ion m/z 320.2 en inset. (les ions annotés en gris font partie du bruit de fond chimique).*

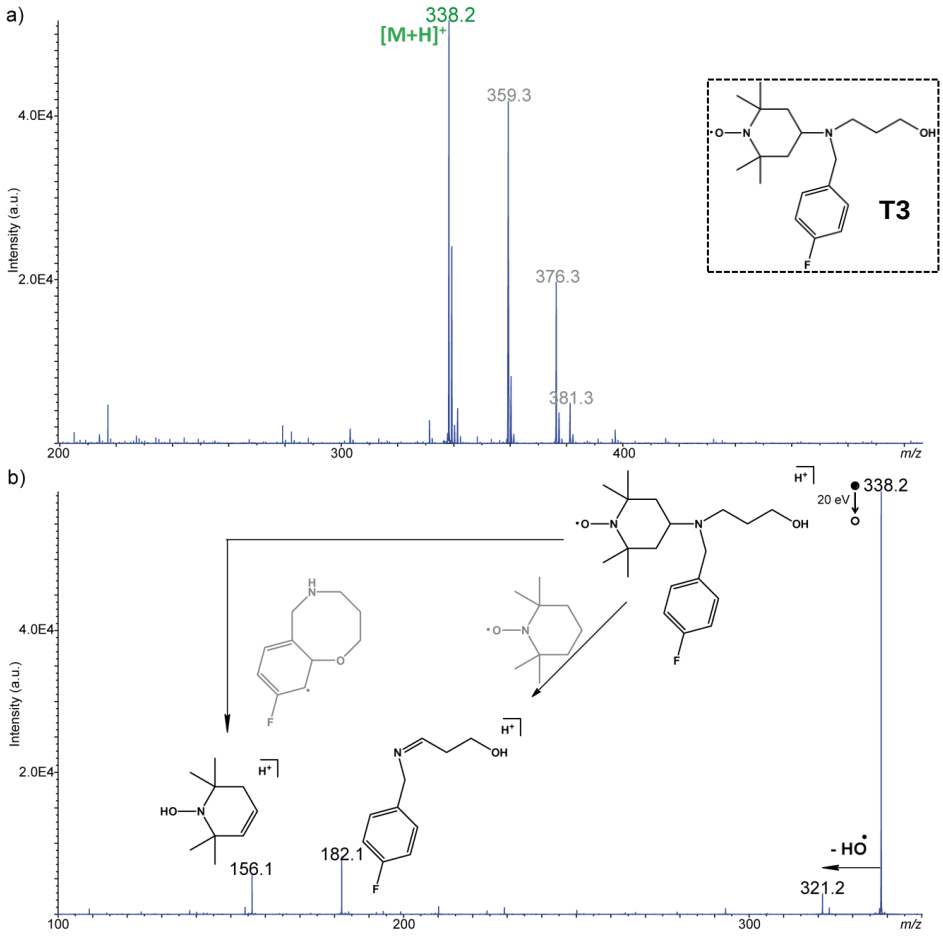


*Figure PE.4. 3 : Spectres MS du composé **T3.***
*a) Spectre ESI-MS en mode positif de l'échantillon **T3**. Les ions annotés en gris sont typiques du bruit de fond chimique dans cette gamme m/z. b) Spectre ESI-MS/MS de l'ion [M+H]+ à m/z 338.2 dans l'échantillon **T3**.*

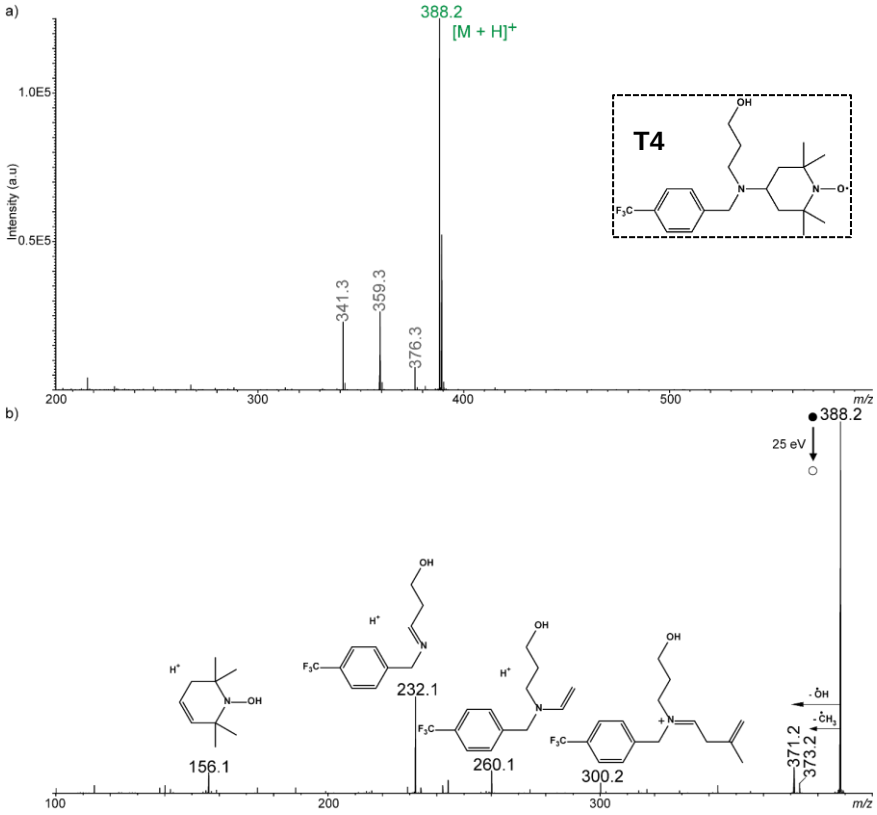


*Figure PE.4. 4 : Spectres MS du composé **T4**.*
*a) Spectre ESI(+)-MS de l'échantillon **T4** et b) spectre MS/MS de l'ion [M+H]+ à m/z 388.2.*

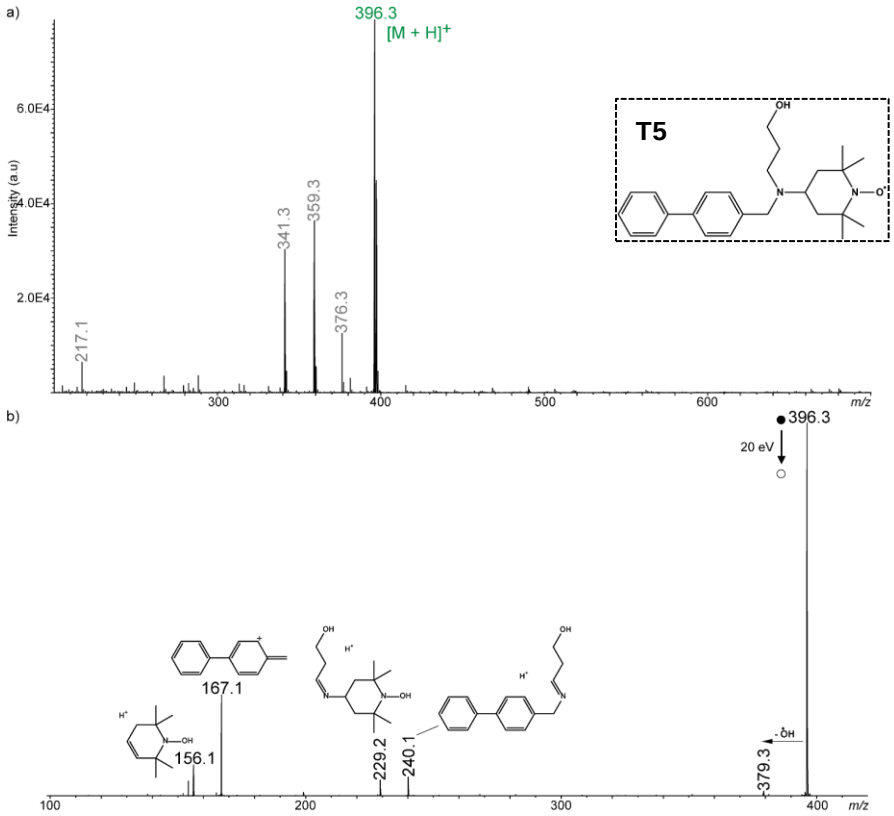


*Figure PE.4. 5 : Spectres MS du composé **T5**.*
*a) Spectre ESI(+)-MS de l'échantillon **T5** et b) spectre MS/MS de l'ion [M+H]+ à m/z 396.3.*

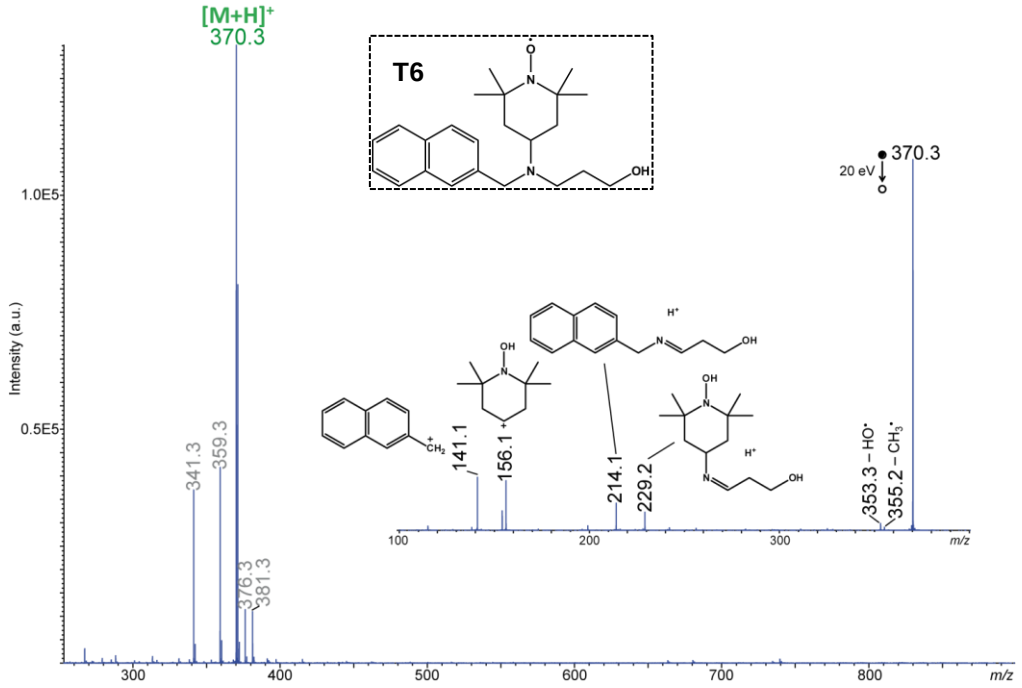


*Figure PE.4. 6 : Spectres MS du composé **T6**.*
*Spectre ESI-MS enregistré en mode positif pour l'échantillon **T6** et spectre MS/MS de l'ion m/z 370.3 en inset. (les ions annotés en gris font partie du bruit de fond chimique).*

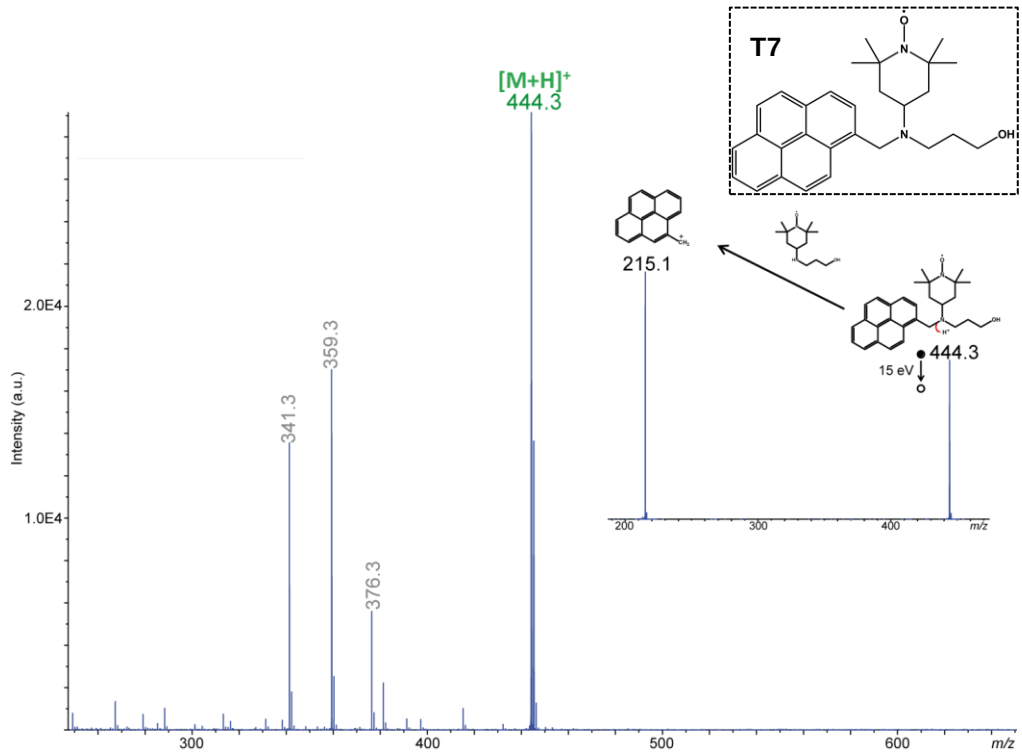


*Figure PE.4. 7 : Spectres MS du composé **T7**.*
*Spectre ESI-MS enregistré en mode positif pour l'échantillon **T7** et spectre MS/MS de l'ion m/z 444.3 en inset. (les ions annotés en gris font partie du bruit de fond chimique).*

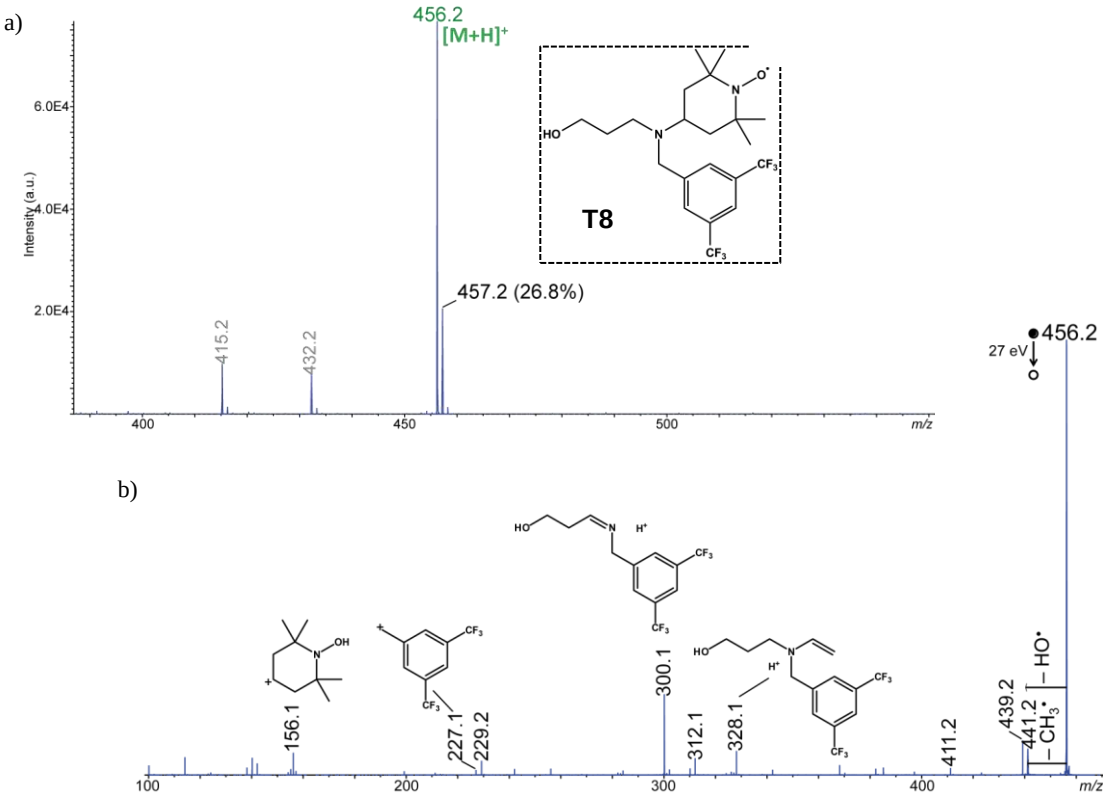


*Figure PE.4. 8 : Spectres MS du composé **T8**.*
*a) Spectre ESI-MS enregistré en mode positif et b) Spectre ESI-MS/MS de l'ion à m/z 456.2 pour l'échantillon **T8** (les ions annotés en gris font partie du bruit de fond chimique).*

L'intensité relative du pic à m/z 457.2 par rapport au pic à m/z 456.2 (indiquée en noir, Figure 1) est comparable au rapport isotopique théorique (24.5%), ce qui suggère que l'ion m/z 457.2 correspondrait au seul pic 13C du nitroxide protoné. Néanmoins, dans le spectre MS/MS de cet ion (non montré), l'abondance d'un fragment (m/z 328.1) est trop élevée pour être compatible avec le seul changement isotopique 12C/13C et indiquerait donc la présence de l'hydroxylamine.

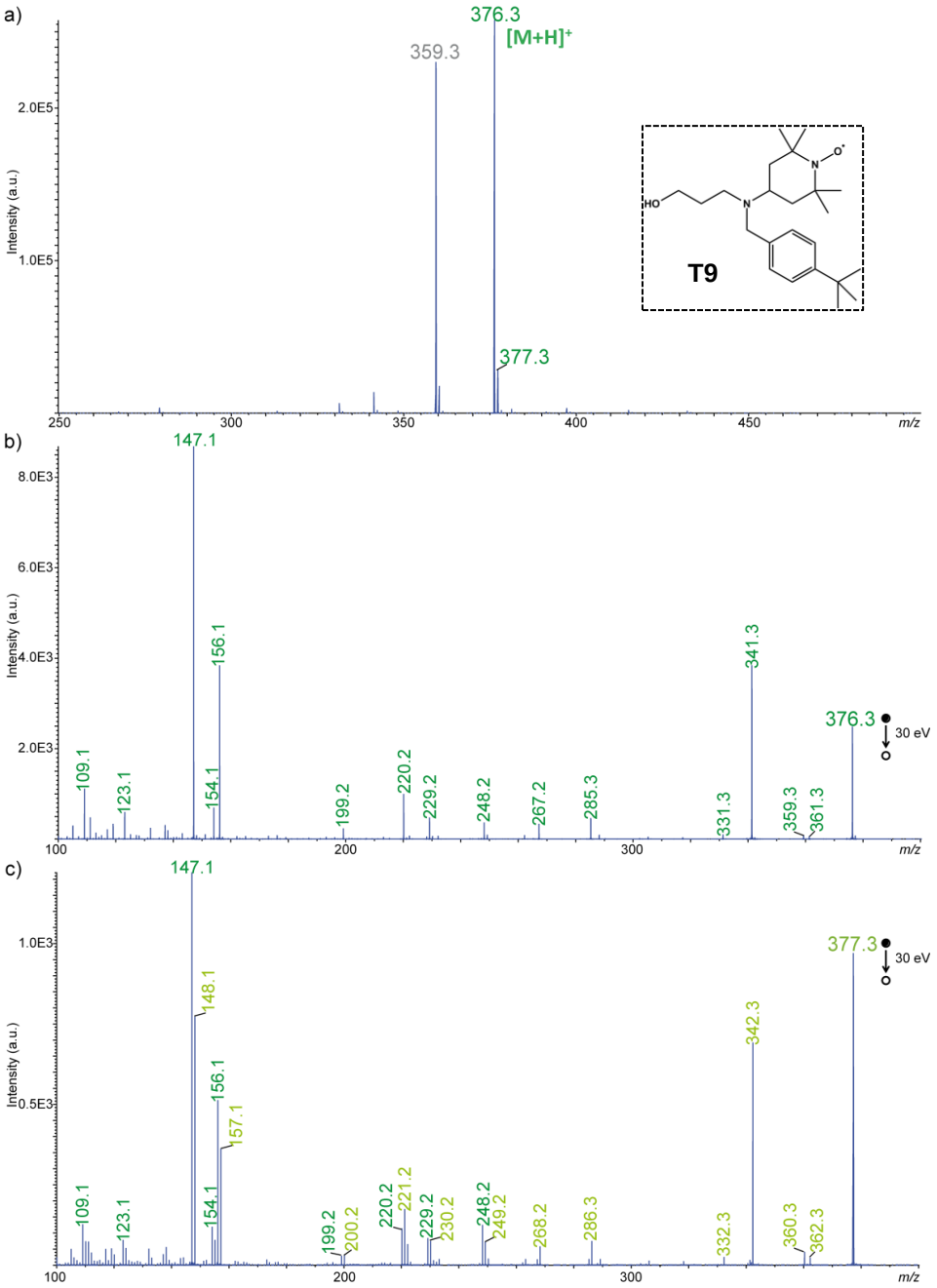


*Figure PE.4. 9 : Spectres MS du composé **T9**.*
*a) Spectre ESI-MS enregistré en mode positif pour l'échantillon **T9** (les ions annotés en gris font partie du bruit de fond chimique). Spectres MS/MS des ions b) m/z 376.3 et c) m/z 377.3.*

Le spectre de masse enregistré en mode positif montre bien le composé recherché sous sa forme protonée [M+H]+ à m/z 376.3 (Figure 1a). Cette attribution n'a pas pu être validée par mesures de masse exacte du fait de l'interférence de l'adduit ammonium correspondant à la molécule protonée à m/z 359.3(en gris, Figure 1a). Néanmoins, le spectre MS/MS de l'ion m/z 376.3 (Figure 1b) montre bien l'élimination des radicaux CH3• et HO• (mexp = 17.0051 Da vs mthéo = 17.0027 Da ≠ m(NH3) = 17.0265 Da) symptomatique du radical TEMPO. Par ailleurs, l'intensité de l'ion m/z 377.3 est compatible avec le seul pic 13C de l'ion ciblé, comme confirmé par la comparaison des spectres MS/MS enregistrés pour ces deux ions (Figure 1b-c). En d'autres termes, ces données n'apportent aucune indication de la présence de la forme hydroxylamine du nitroxyde ciblé.

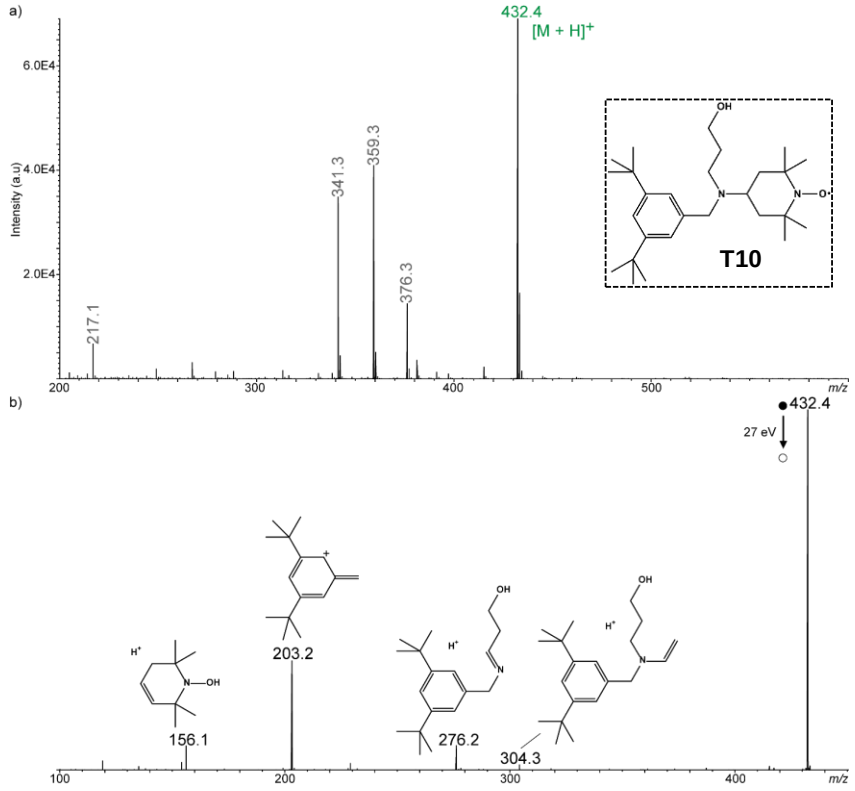


*Figure PE.4. 10 : Spectres MS du composé **T10**.*
*a) Spectre ESI(+)-MS de l'échantillon **T10** et b) spectre MS/MS de l'ion [M+H]+ à m/z 432.4.*

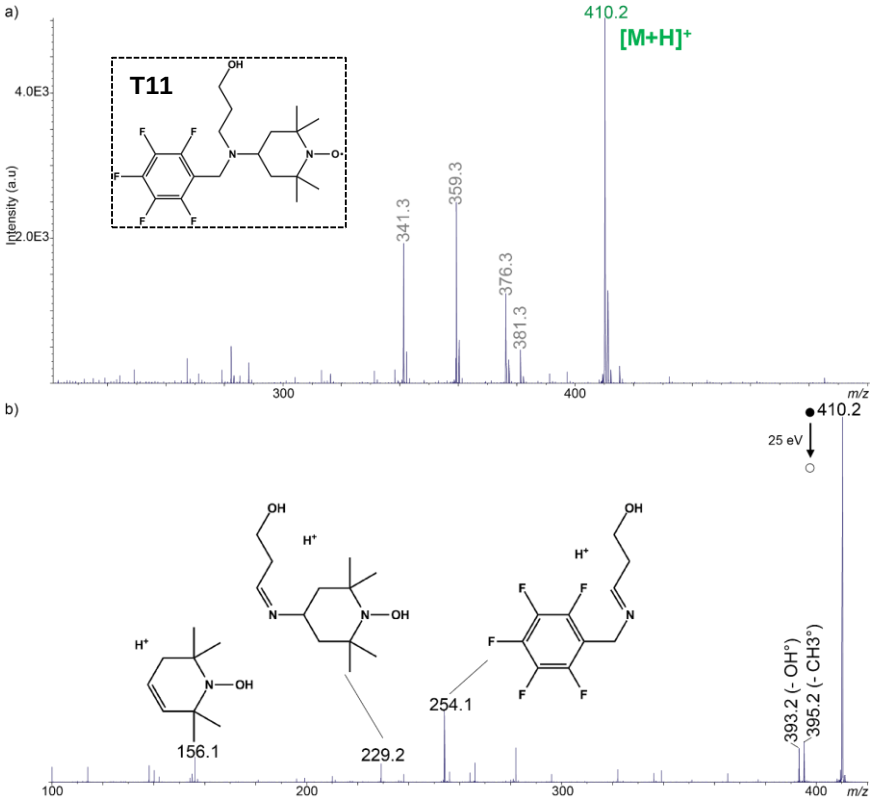


*Figure PE.4. 11 : Spectres MS du composé **T11**.*
*a) Spectre ESI(+)-MS de l'échantillon **T11** (les pics annotés en gris appartiennent au bruit de fond chimique) et b) spectre MS/MS de l'ion [M+H]+ à m/z 410.2.*

4.2. $PdCl_2(PPh_3)_2$

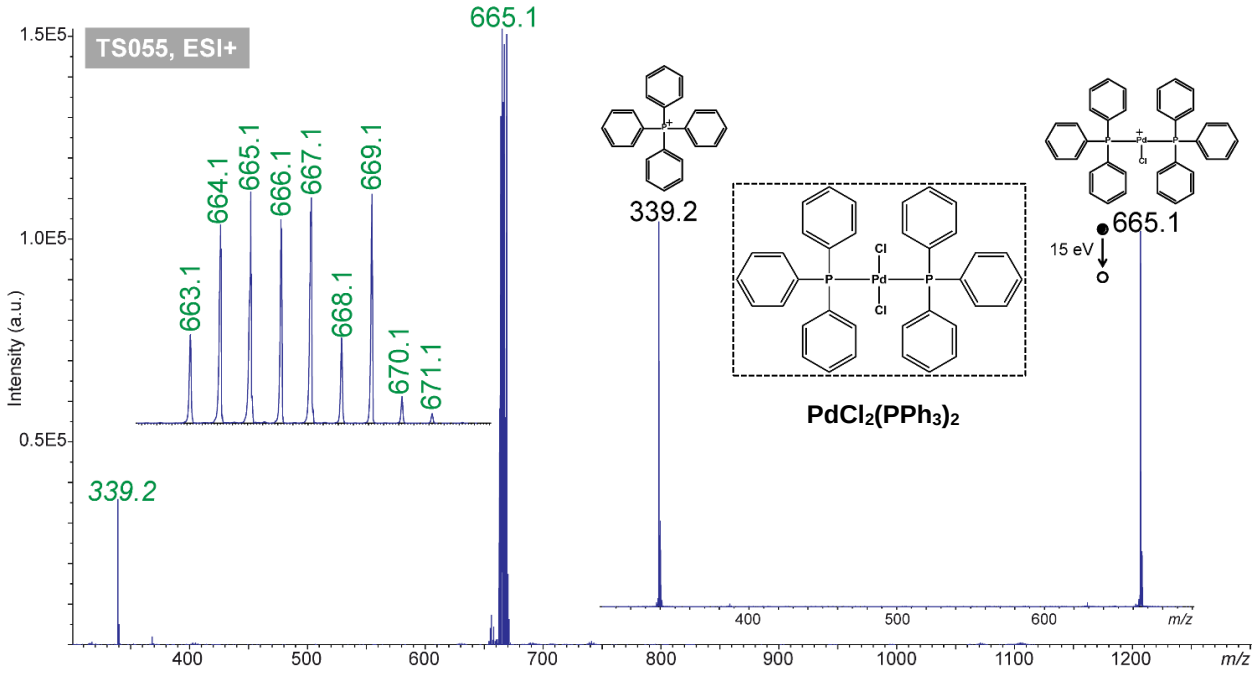


*Figure PE.4. 12 : Spectres MS du catalyseur $PdCl_2(PPh_3)_2$ utilisé pour la synthèse des intermédiaires **c1** et **b2**.*
Spectre ESI(+)-MS de l'échantillon $PdCl_2(PPh_3)_2$ et spectre MS/MS de l'ion [M – Cl]+ à m/z 665.1.

4.3. TISP et TISP2

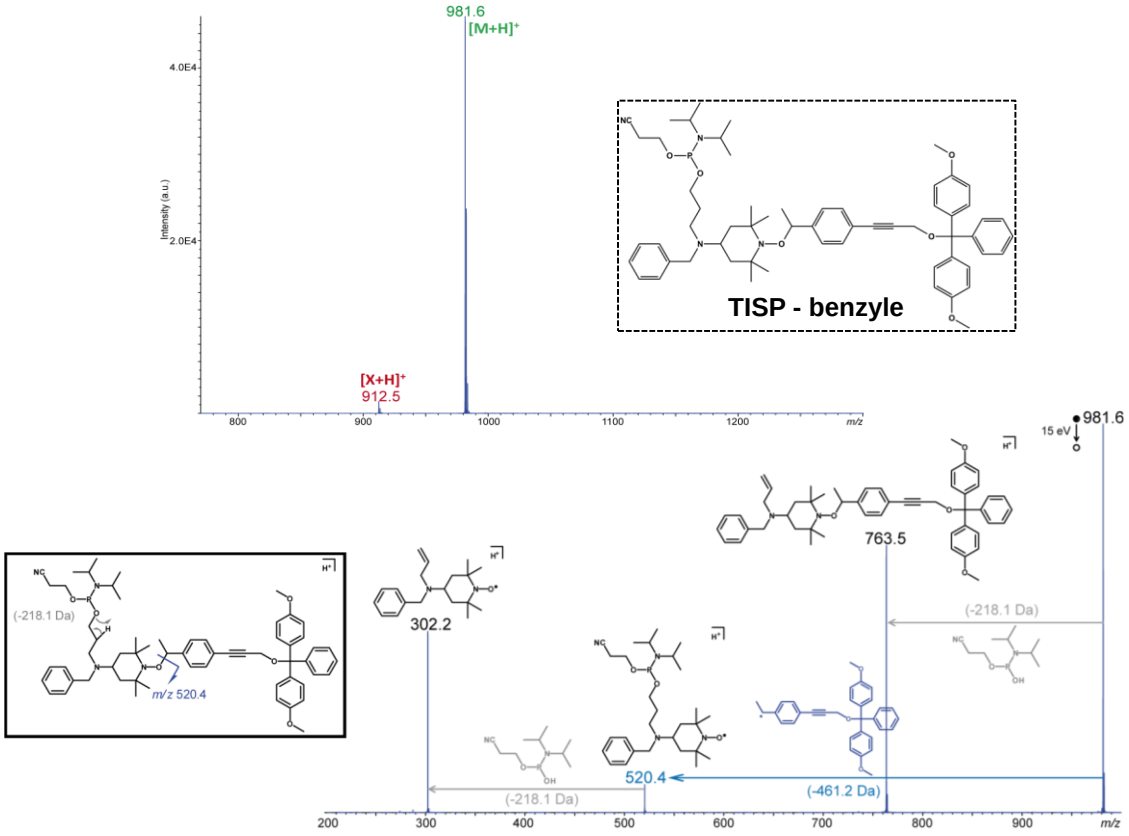


*Figure PE.4. 13 : Spectres MS de l'Espaceur **TISP**.*
*Spectres ESI-MS en mode positif et ESI-MS/MS de l'ion [M+H]+ à m/z 981.6 dans l'échantillon **TISP**.*

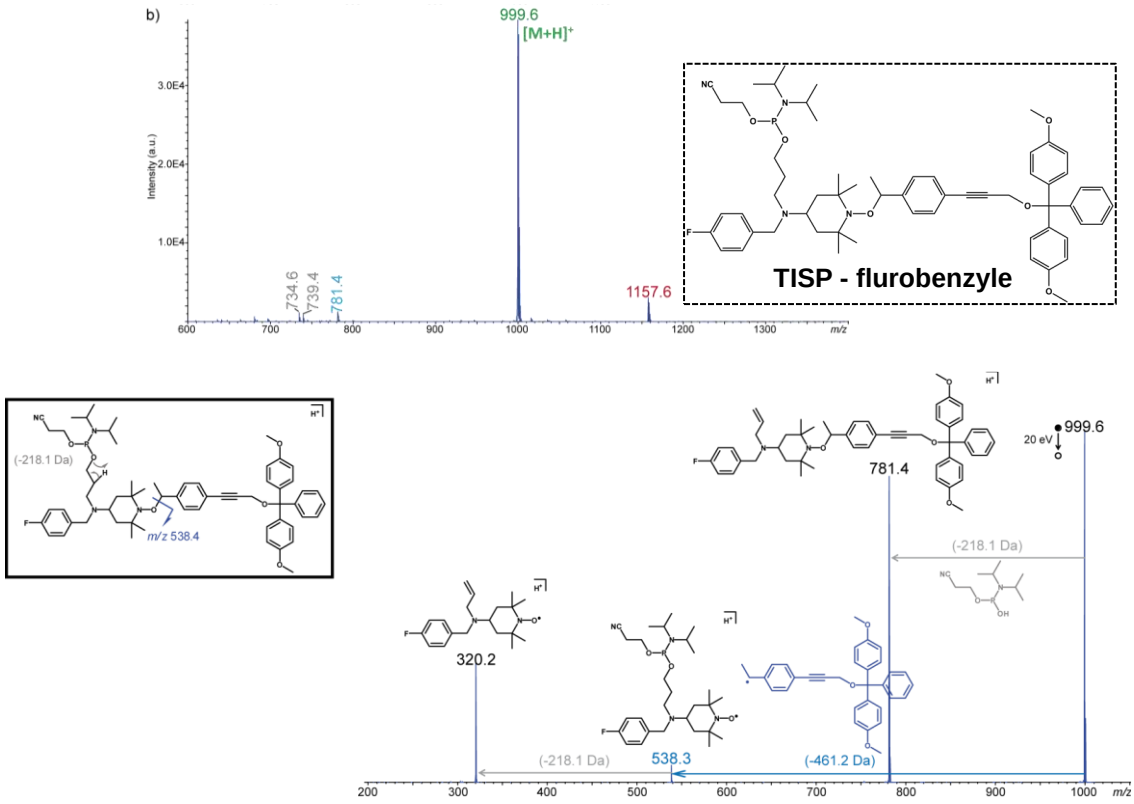


*Figure PE.4. 14 : Spectres MS de l'Espaceur **TISP-F**.*
*Spectres ESI-MS en mode positif et ESI-MS/MS de l'ion [M+H]+ à m/z 999.6 dans l'échantillon **TISP-F**.*

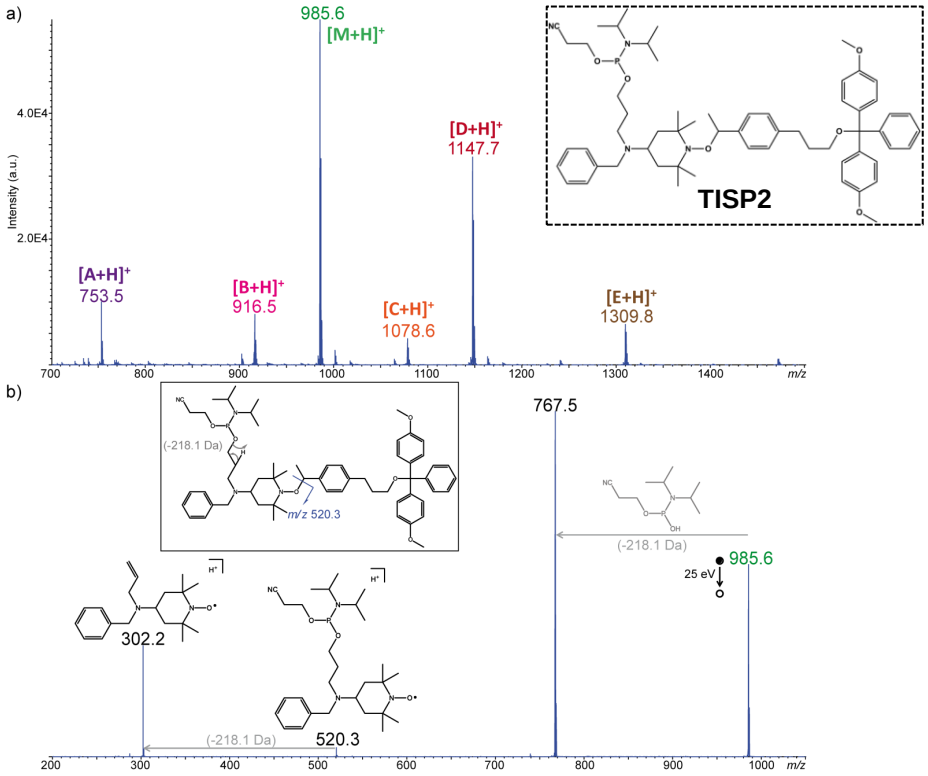


*Figure PE.4. 15 : Spectres MS de l'Espaceur **TISP2**.*
*a) Spectre ESI(+)-MS de l'échantillon **TISP2** b) spectre MS/MS de l'ion [M+H]+ à m/z 985.6. Les impuretés sont analysées en MS/MS dans la Figure PE.4. 16.*

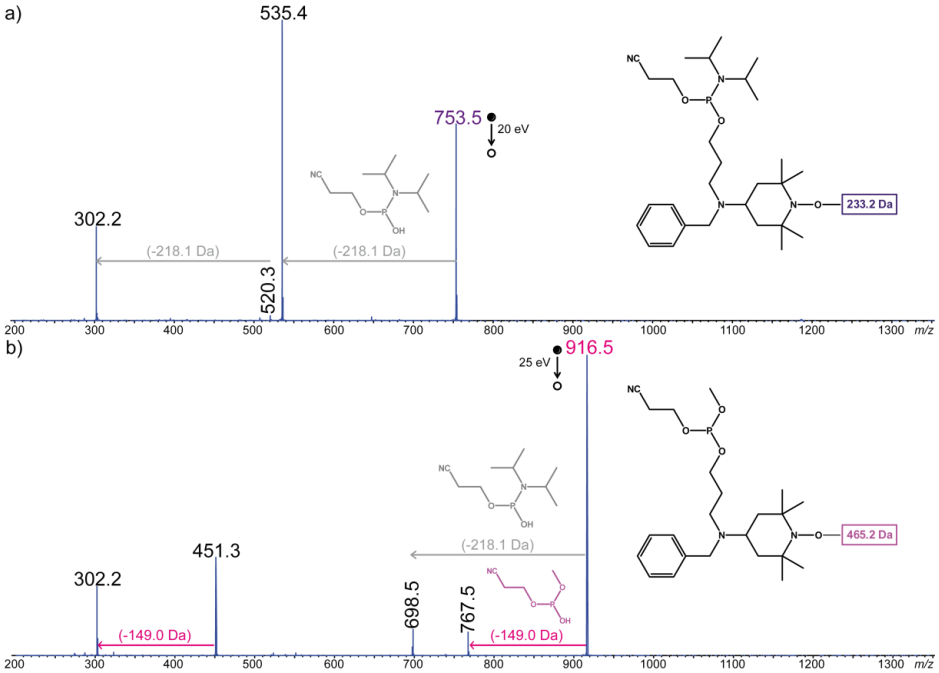


*Figure PE.4. 16 : Spectres ESI-MS/MS des impuretés observées dans la synthèse de l'Espaceur **TISP2**.*
Spectres ESI-MS/MS des ions a) m/z 753.5, b) m/z 916.5,

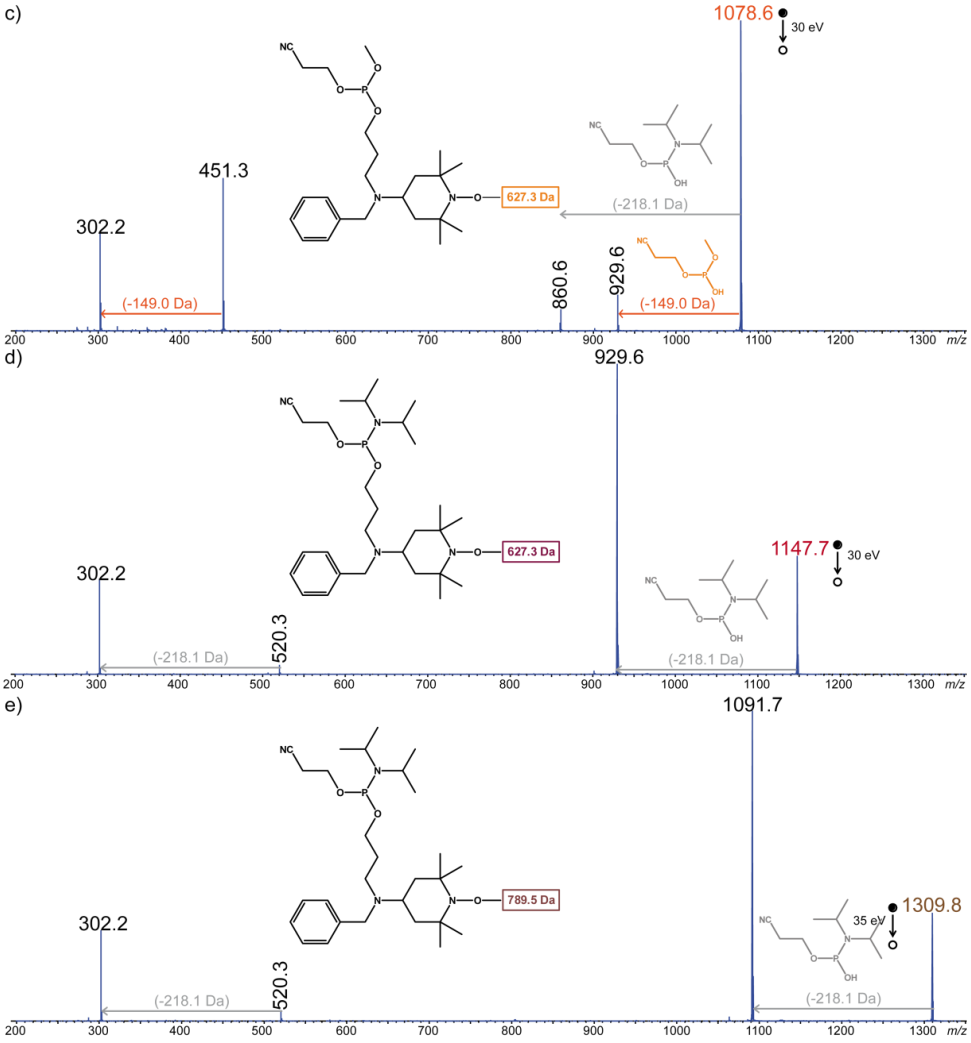


*Figure PE.4. 16-Suite : Spectres ESI-MS/MS des impuretés observées dans la synthèse de l'Espaceur **TISP2**.*
*Spectres ESI-MS/MS des ions c) m/z 1078.6, d) m/z 1147.7 et e) m/z 1309.8 détectés dans l'échantillon **TISP2**.*

4.4. TISP3

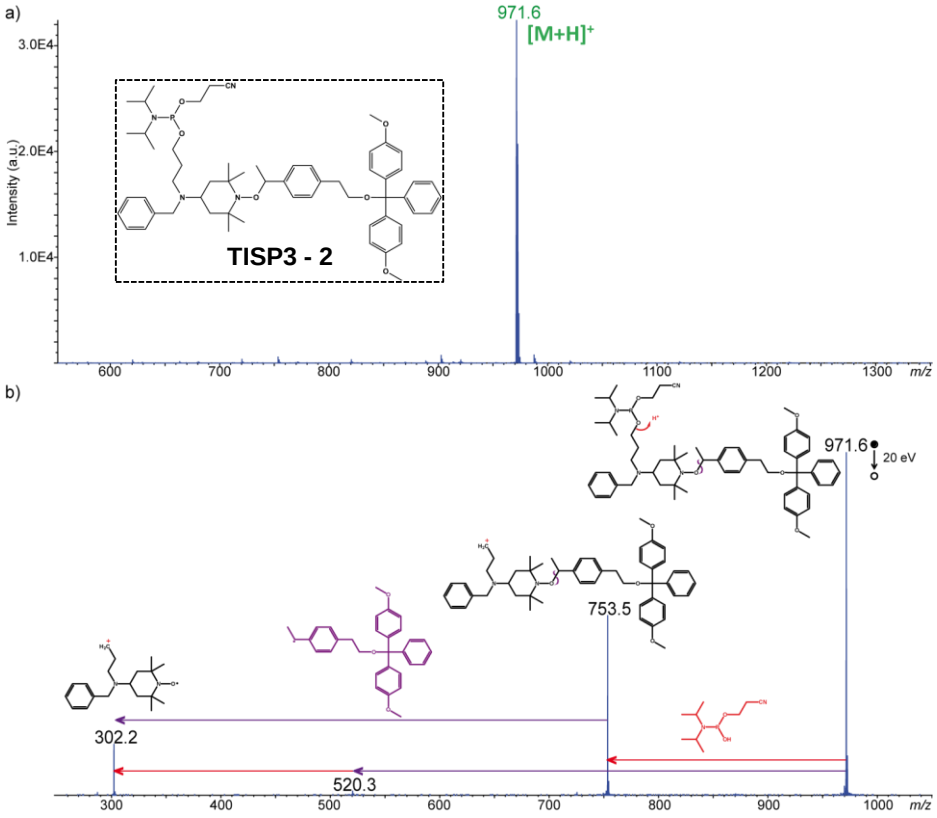
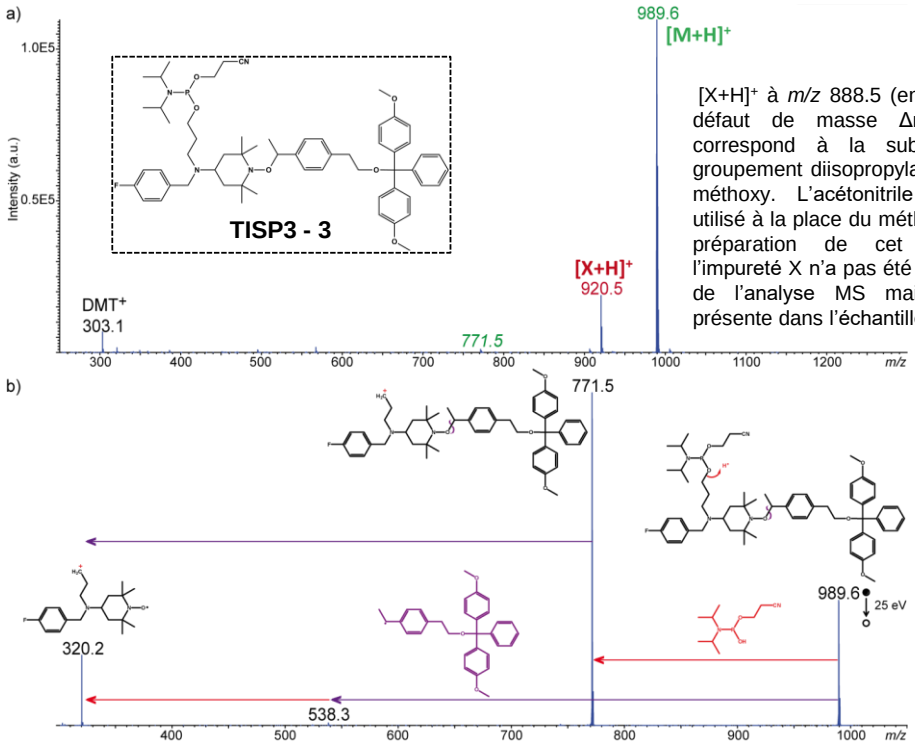
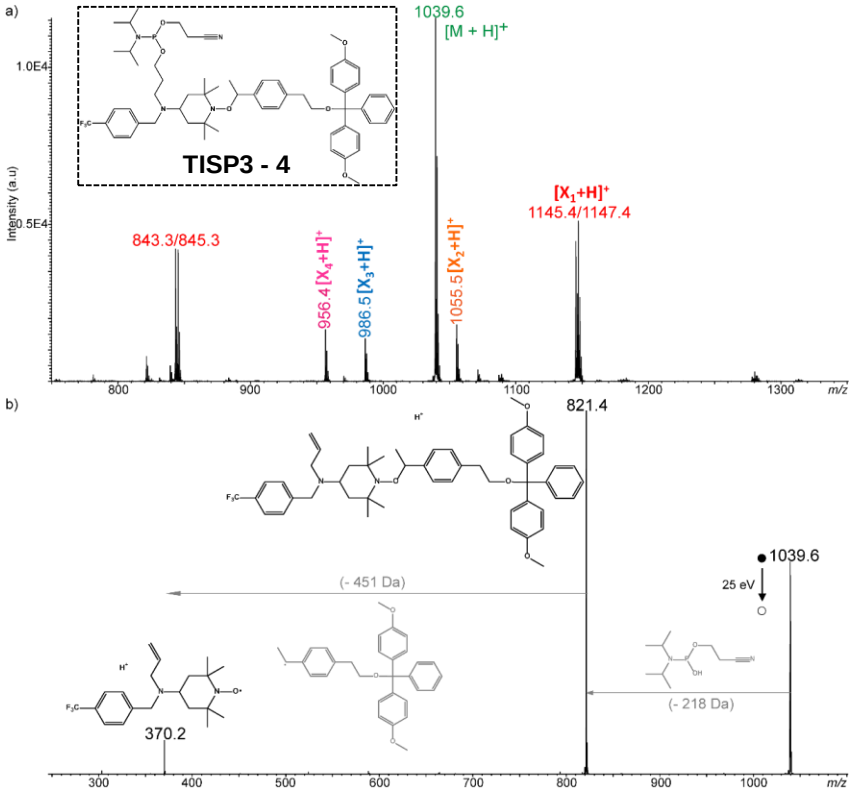


*Figure PE.4. 17 : Spectres MS de l'Espaceur **TISP3-2**.*
*a) Spectre ESI(+)-MS de l'échantillon **TISP3-2** b) spectre MS/MS de l'ion [M+H]+ à m/z 971.6.*



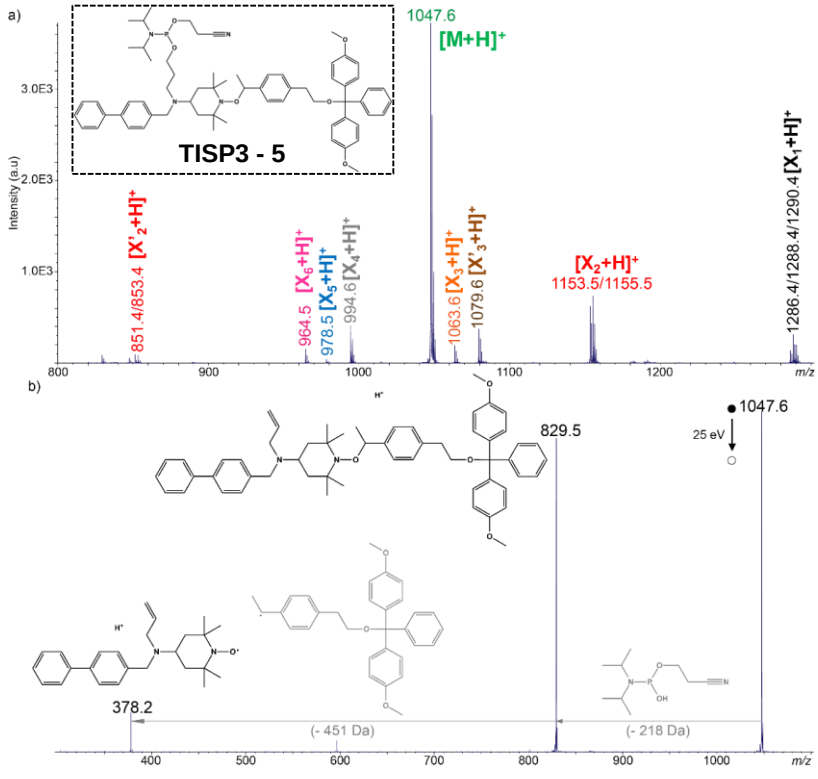
[X+H]+ à *m/z* 888.5 (en rouge) : ce défaut de masse Δm –69 Da correspond à la substitution du groupement diisopropylamine par un méthoxy. L'acétonitrile ayant été utilisé à la place du méthanol pour la préparation de cet échantillon, l'impureté X n'a pas été générée lors de l'analyse MS mais est bien présente dans l'échantillon.

*Figure PE.4. 18 : Spectres MS de l'Espaceur **TISP3-3**.*
*a) Spectre ESI(+)-MS de l'échantillon **TISP3-3** b) spectre MS/MS de l'ion [M+H]+ à m/z 989.6.*



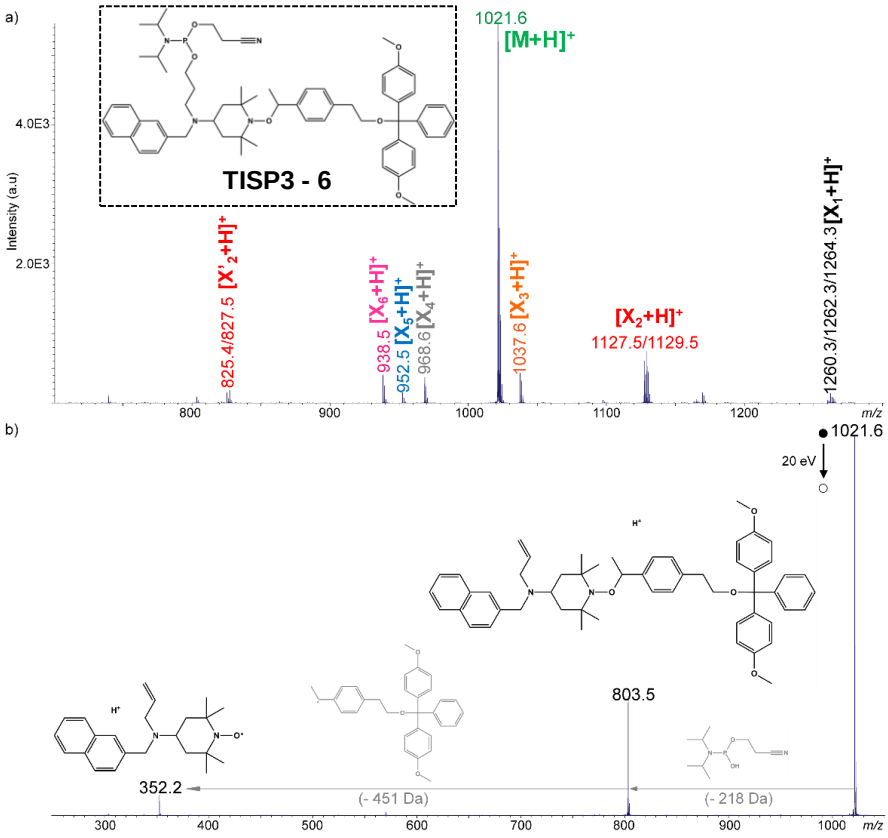
Cinq impuretés majeures sont observées. X_1 à *m/z* 1145.4 (en rouge) présente un excès de masse Δm +105.9 Da et son massif isotopique indique un atome Br : sa dissociation conduit à la perte du radical 451 Da ainsi qu'à la formation de l'ion DMT^+, indiquant que l'excès de masse est localisé dans la partie phosphoramidite. X'_1 à *m/z* 843.3 provient de la fragmentation en source de X_1 (-DMT+H). L'excès de masse Δm +15.9 Da de X_2 à *m/z* 1055.5 (en orange) correspondrait à un atome d'oxygène. L'impureté X_3 à *m/z* 986.5 (en bleu) a un défaut de masse Δm –53 Da correspondant à l'absence du groupement CH_2-CH_2-CN tandis que le défaut de masse Δm –83 Da de l'impureté X_4 à *m/z* 956.4 (en rose) n'a pas pu être caractérisé.

*Figure PE.4. 19 : Spectres MS de l'Espaceur **TISP3-4**.*
*a) Spectre ESI(+)-MS de l'échantillon **TISP3-4** b) spectre MS/MS de l'ion [M+H]+ à m/z 1039.6.*



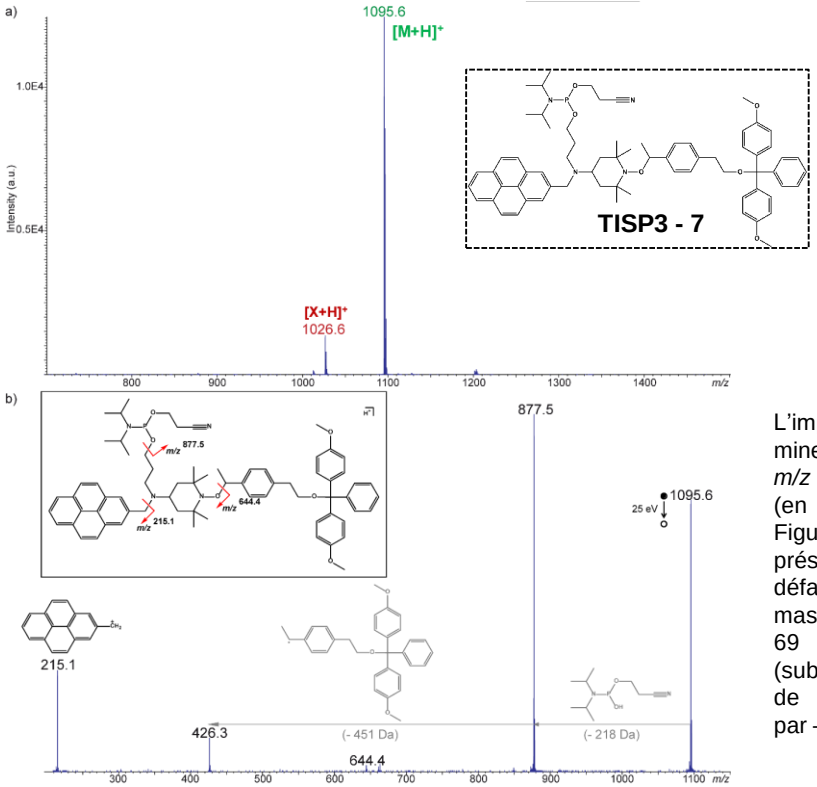
Sept impuretés mineures sont observées et leur dissociation indique un défaut structural localisé dans la partie phosphoramidite. X_1 à *m/z* 1286.4 (en noir) présente un excès de masse Δm +238.8 Da et contient deux atomes Br. X_2 à *m/z* 1153.5 (en rouge) présente un excès de masse Δm +105.9 Da et contient un atome Br : sa dissociation en source conduit à X'_2 à *m/z* 851.4. Les excès de masse Δm +31.9 Da de X'_3 à *m/z* 1079.6 (en marron) et Δm +15.9 Da de X_3 à *m/z* 1063.6 (en orange) correspondraient respectivement à deux et un atome d'oxygène. L'impureté X_3 à *m/z* 994.6 (en gris) a un défaut de masse Δm –53 Da (absence du groupement CH_2-CH_2-CN). L'impureté X_5 à *m/z* 978.5 (en bleu) a un défaut de masse Δm –69 Da (substitution de $-N(iPr)_2$ par –OMe). Le défaut de masse Δm –83 Da de X_6 à *m/z* 964.5 (en rose) n'a pas pu être caractérisé.

*Figure PE.4. 20 : Spectres MS de l'Espaceur **TISP3-5**.*
*a) Spectre ESI(+)-MS de l'échantillon **TISP3-5** b) spectre MS/MS de l'ion [M+H]+ à m/z 1047.6.*



Sept impuretés mineures sont également observées et leur dissociation indique un défaut structural localisé dans la partie phosphoramidite. X_1 à *m/z* 1260.3 (en noir) présente un excès de masse Δm +238.8 Da et contient deux atomes Br. X_2 à *m/z* 1127.5 (en rouge) présente un excès de masse Δm +105.9 Da et contient un atome Br : sa dissociation en source conduit à X'_2 à *m/z* 825.4. L'excès de masse Δm +15.9 Da de X_3 à *m/z* 1037.6 (en orange) correspondrait à un atome d'oxygène. L'impureté X_4 à *m/z* 968.6 (en gris) a un défaut de masse Δm –53 Da (absence du groupement CH_2-CH_2-CN). L'impureté X_5 à *m/z* 952.5 (en bleu) a un défaut de masse Δm –69 Da (substitution de –$N(iPr)_2$ par –OMe). Le défaut de masse Δm –83 Da de X_6 à *m/z* 938.5 (en rose) n'a pas pu être caractérisé.

*Figure PE.4. 21 : Spectres MS de l'Espaceur **TISP3-6**.*
*a) Spectre ESI(+)-MS de l'échantillon **TISP3-6** b) spectre MS/MS de l'ion [M+H]+ à m/z 1021.6.*



L'impureté mineure X à *m/z* 1026.6 (en rouge, Figure 1a) présente un défaut de masse Δm –69 Da (substitution de –$N(iPr)_2$ par –OMe).

*Figure PE.4. 22 : Spectres MS de l'Espaceur **TISP3-7**.*
*a) Spectre ESI(+)-MS de l'échantillon **TISP3-7** b) spectre MS/MS de l'ion [M+H]+ à m/z 1095.6.*

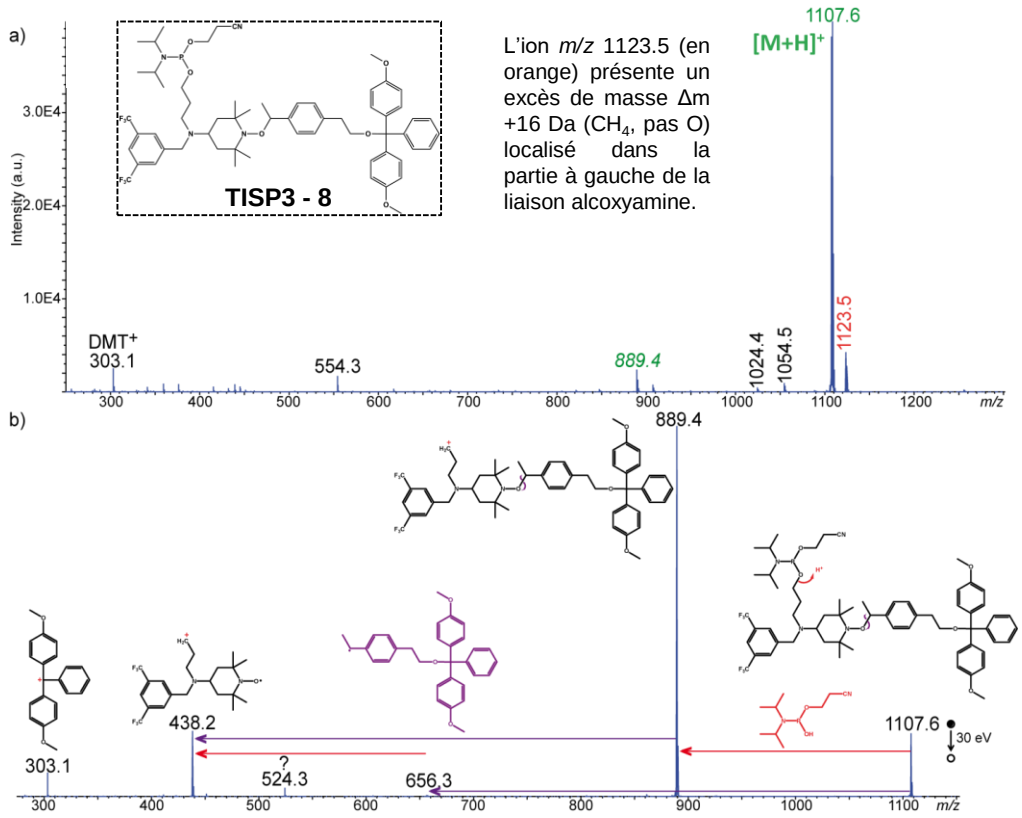


*Figure PE.4. 23 : Spectres MS de l'Espaceur **TISP3-8**.*
*a) Spectre ESI(+)-MS de l'échantillon **TISP3-8** b) spectre MS/MS de l'ion [M+H]+ à m/z 1107.6.*

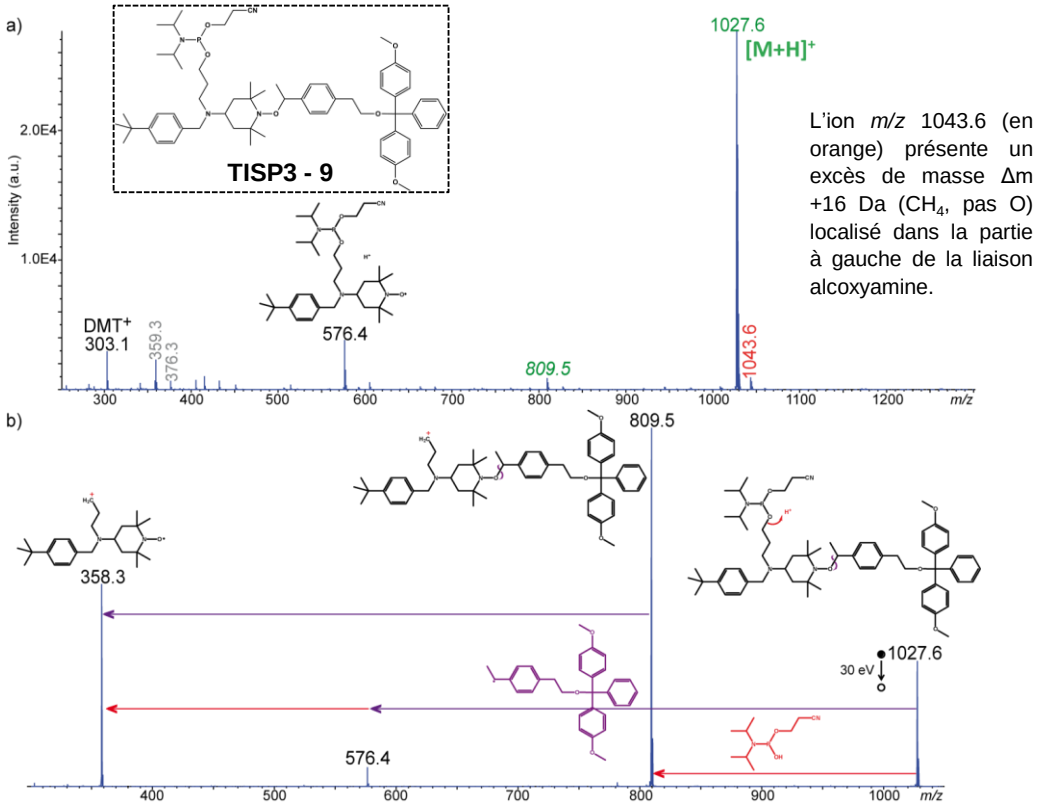
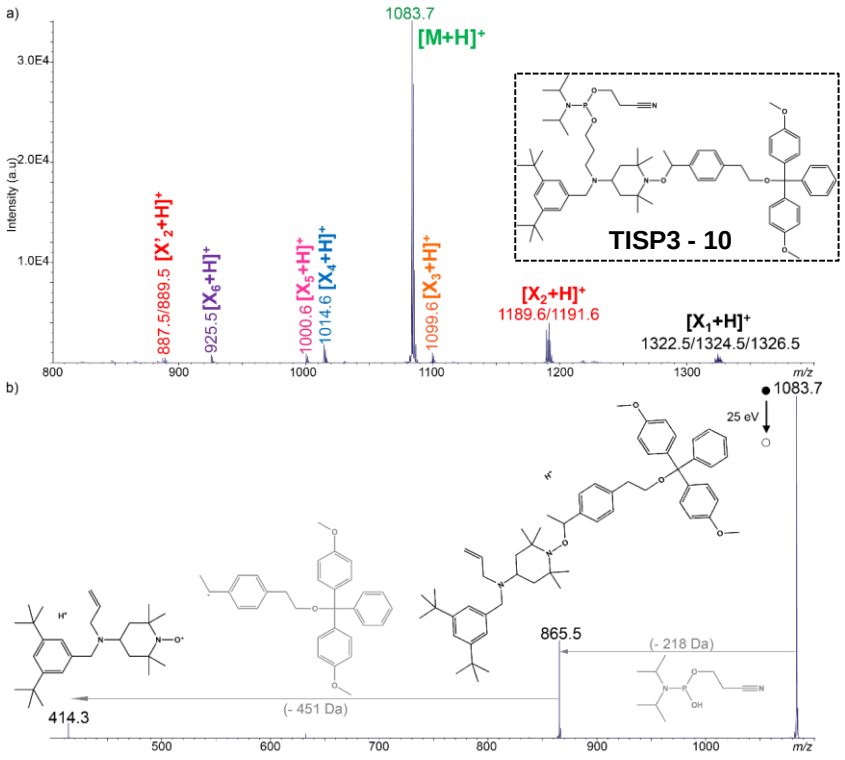
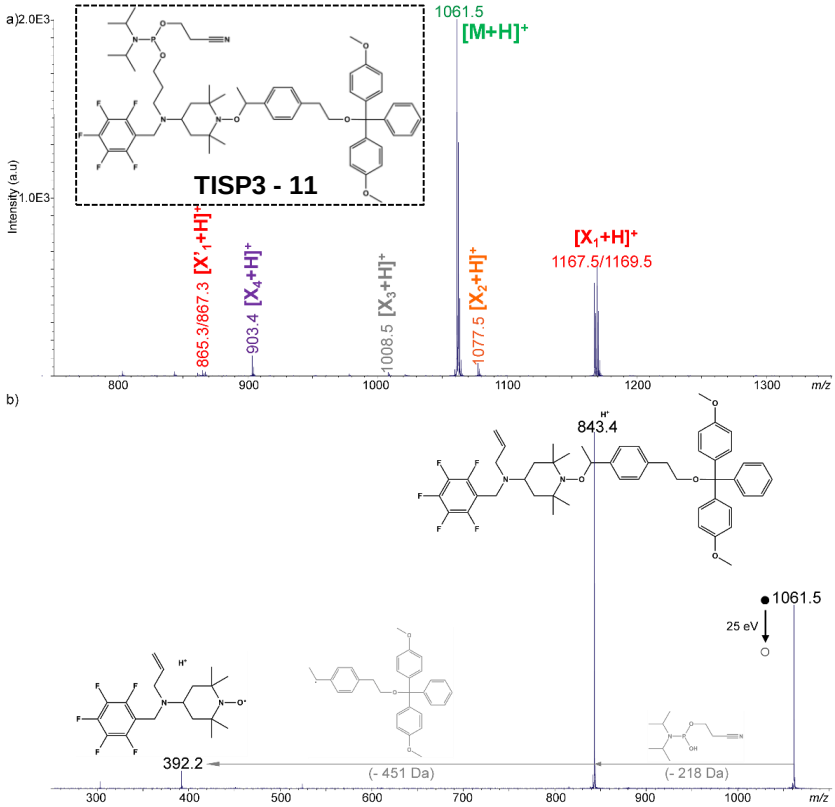


*Figure PE.4. 24 : Spectres MS de l'Espaceur **TISP3-9**.*
*a) Spectre ESI(+)-MS de l'échantillon **TISP3-9** b) spectre MS/MS de l'ion [M+H]+ à m/z 1027.6.*



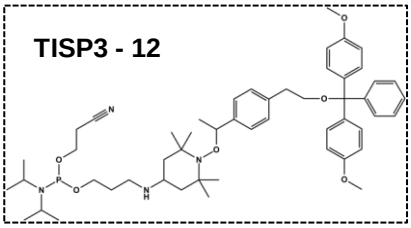
Sept impuretés mineures sont observées et leur dissociation indique un défaut structural localisé dans la partie phosphoramidite. X_1 à *m/z* 1322.5 (en noir) présente un excès de masse Δm +238.8 Da et contient deux atomes Br. X_2 à *m/z* 1189.6 (en rouge) présente un excès de masse Δm +105.9 Da et contient un atome Br : sa dissociation en source conduit à X'_2 à *m/z* 887.5. L'excès de masse Δm +15.9 Da de X_3 à *m/z* 1099.6 (en orange) correspondrait à un atome d'oxygène. L'impureté X_4 à *m/z* 1014.6 (en bleu) a un défaut de masse Δm –69 Da correspondant à la substitution de $–N(iPr)_2$ par –OMe. X_5 à *m/z* 1000.6 (Δm –83 Da, en rose) et X_6 à *m/z* 925.5 (Δm –158 Da, en violet) n'ont pas pu être caractérisées.

*Figure PE.4. 25 : Spectres MS de l'Espaceur **TISP3-10**.*
*a) Spectre ESI(+)-MS de l'échantillon **TISP3-10** b) spectre MS/MS de l'ion [M+H]+ à m/z 1083.7.*



Cinq impuretés mineures sont observées et leur dissociation indique un défaut structural localisé dans la partie phosphoramidite. X_1 à *m/z* 1167.5 (en rouge) présente un excès de masse Δm +105.9 Da et contient un atome Br : sa dissociation en source conduit à X'_1 à *m/z* 865.3. L'excès de masse Δm +15.9 Da de X_2 à *m/z* 1077.5 (en orange) correspondrait à un atome d'oxygène. L'impureté X_3 à *m/z* 1008.5 (en gris) a un défaut de masse Δm –53 Da (absence du groupement CH_2-CH_2-CN). Le défaut de masse Δm –158 Da de X_4 à *m/z* 903.4 (en violet) n'a pas pu être caractérisé

*Figure PE.4. 26 : Spectres MS de l'Espaceur **TISP3-11**.*
*a) Spectre ESI(+)-MS de l'échantillon **TISP3-11** b) spectre MS/MS de l'ion [M+H]+ à m/z 1061.5.*



Ce spectre est dominé par l'ion *m/z* 780.4 (en rouge), à qui il manque le groupement diisopropylamine ; comme les données MS/MS montrent que cet ion est également un fragment du composé attendu, il reste difficile de trancher entre la présence de ce composé dans l'échantillon ou sa formation en source.

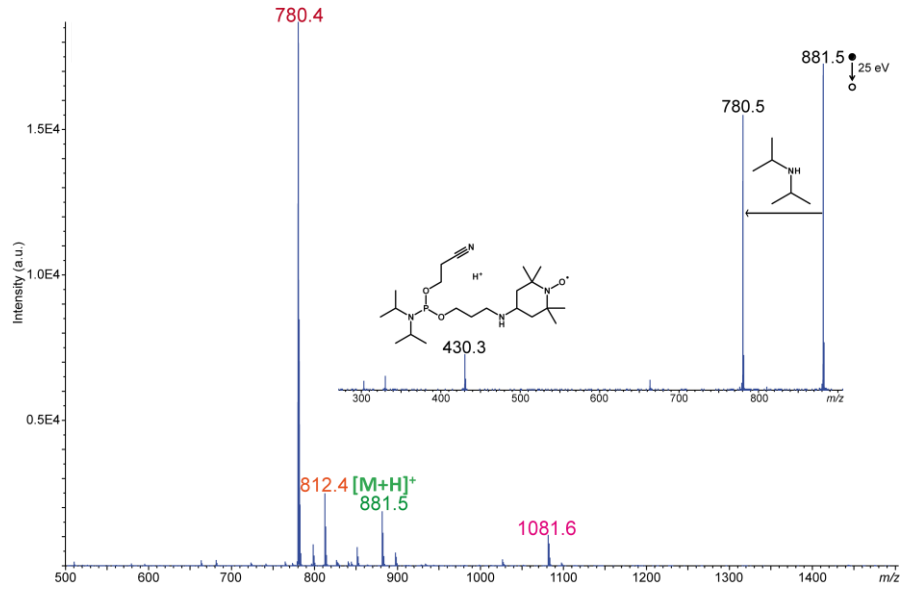


*Figure PE.4. 27 : Spectres MS de l'Espaceur **TISP3-12**.*
*a) Spectre ESI(+)-MS de l'échantillon **TISP3-12** b) spectre MS/MS de l'ion [M+H]+ à m/z 881.5.*

4.5. TISP4

4.5.1. T12-T13

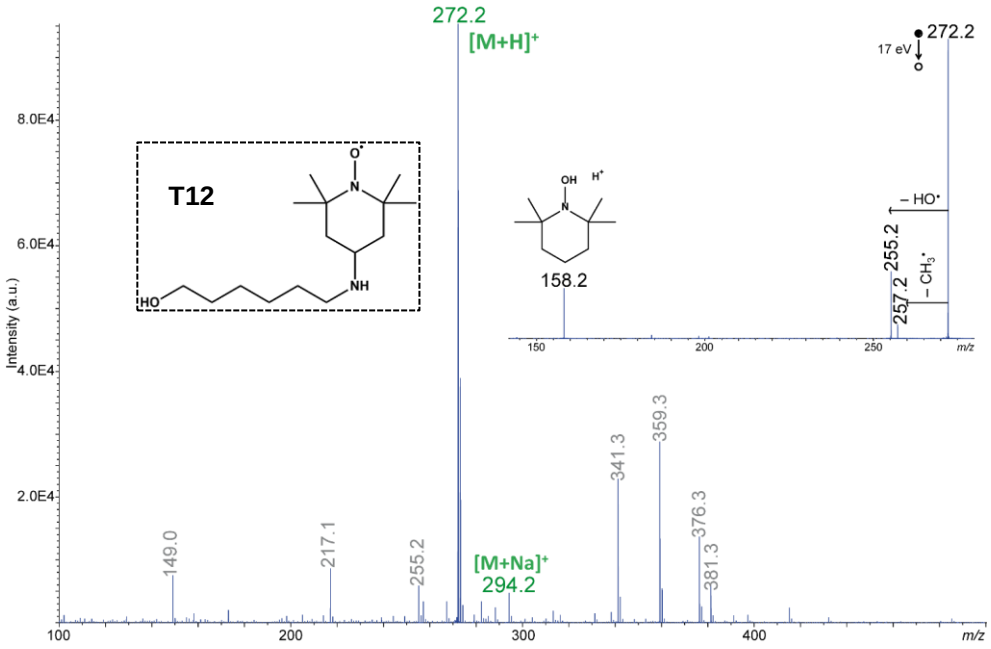


*Figure PE.4. 28 : Spectres MS du composé **T12**.*
*Spectre ESI-MS enregistré en mode positif pour l'échantillon **T12** et spectre MS/MS de l'ion m/z 272.2 en inset. Les ions annotés en gris font partie du bruit de fond chimique de la source ESI.*

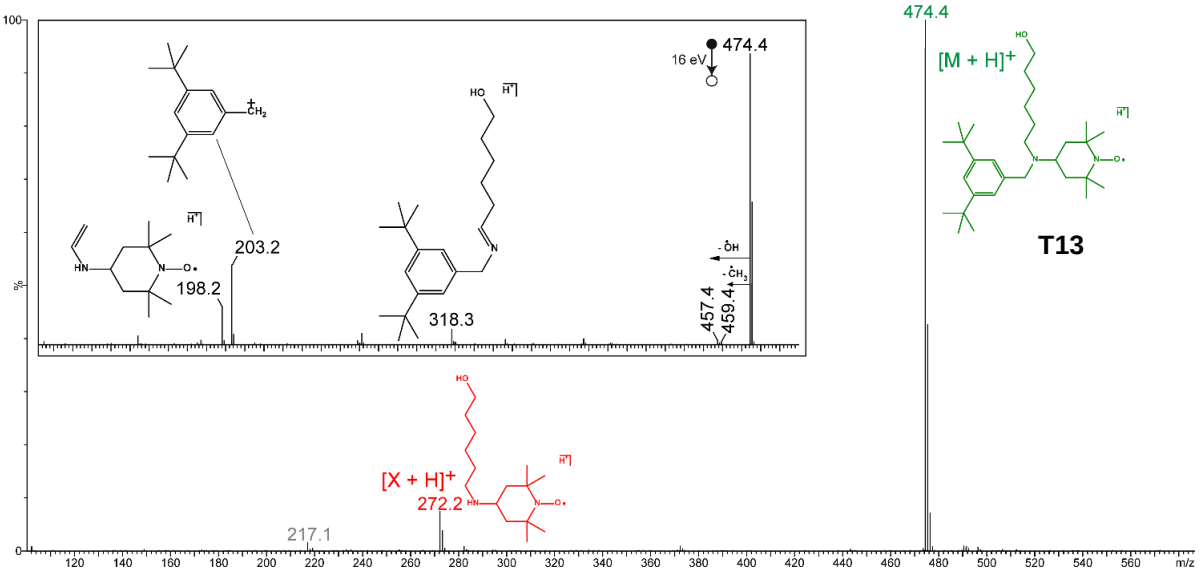


*Figure PE.4. 29 : Spectres MS du composé **T13**.*
*Spectre ESI(+)-MS de l'échantillon **T13** et spectre MS/MS de l'ion [M+H]+ à m/z 474.4 en inset.*

4.5.2. f12 et Espaceur TISP4

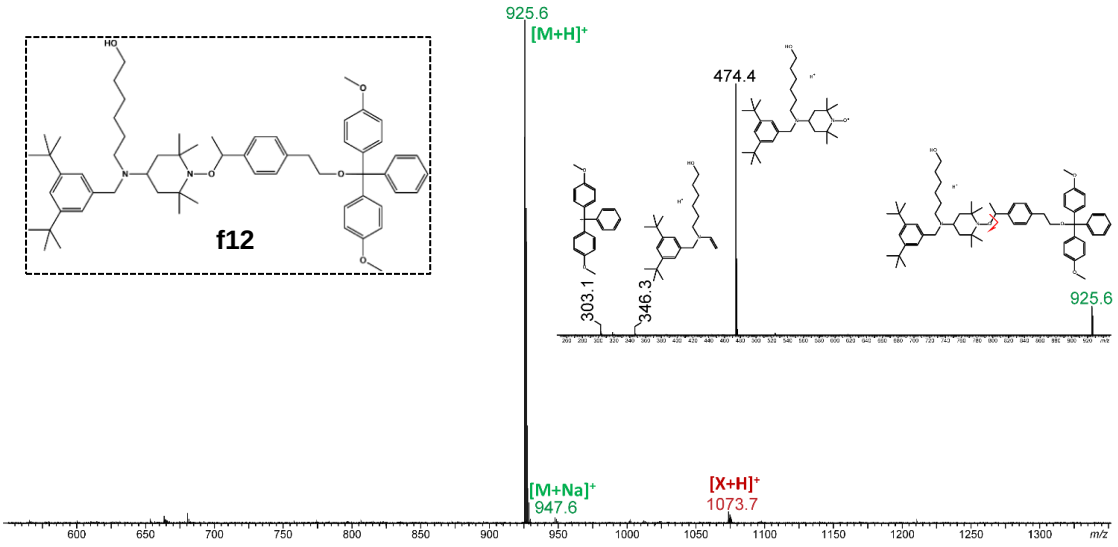


*Figure PE.4. 30 : Spectres MS du composé **f12**.*
*Spectre ESI(+)-MS de l'échantillon **f12** et spectre MS/MS de l'ion [M+H]+ à m/z 925.6 en inset. Un signal de très faible abondance indique la présence d'une molécule X de masse 1072.7 Da (en rouge, Figure 1) : les données MS/MS indiquent que l'excès de masse Δm = +148 Da est localisé à droite de la liaison alcoxyamine.*

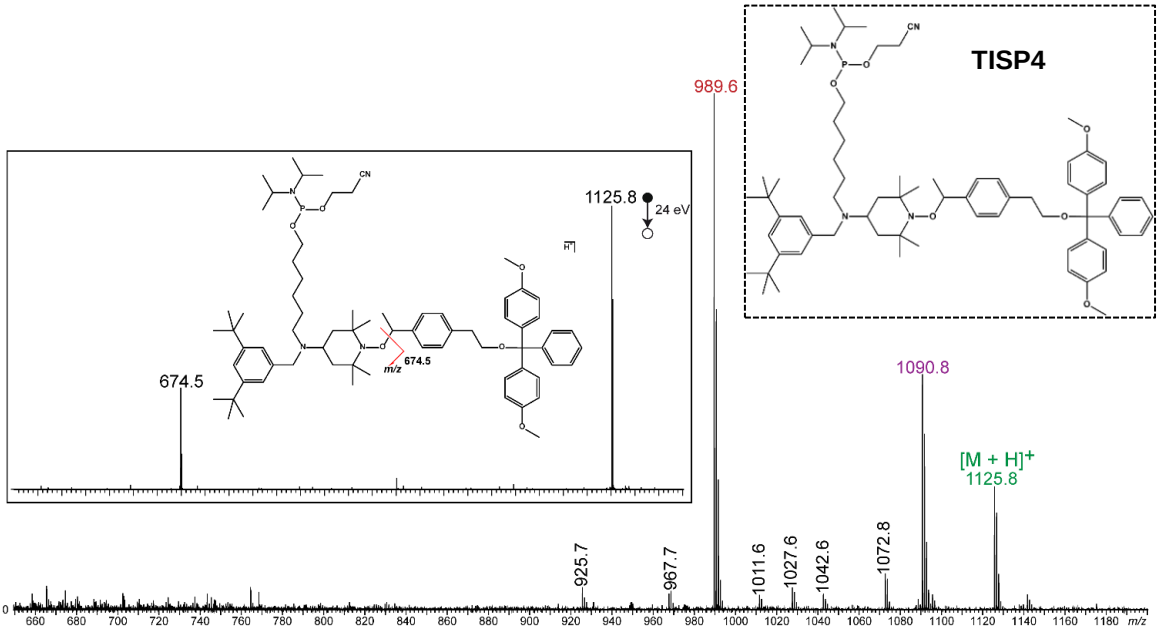


*Figure PE.4. 31 : Spectres MS de l'Espaceur **TISP4**.*
*Spectre ESI(+)-MS de l'échantillon **TISP4** et spectre MS/MS de l'ion [M+H]+ à m/z 1125.8 en inset.*

Le spectre de masse enregistré en mode positif pour cet échantillon montre bien le composé recherché sous sa forme protonée, $[M+H]^+$ à m/z 1125.8 (en vert, Figure PE.4. 31). Cette attribution est validée par des mesures de masse exacte (non montrée ici) et par MS/MS.

Cependant, les signaux majoritaires sont observés à m/z 989.6 (Δm –136.2 Da, en rouge) et m/z 1090.8 (Δm –35 Da, en violet). Les données MS/MS (non montrées) indique que ces deux ions possèdent le groupement DMT (observé comme fragment à m/z 303.1) et que l'ion m/z 989.6 est un fragment de l'ion m/z 1090.8 (soit une perte de 101.2 Da). Une attribution de ces impuretés a pu être proposée dans le Chapitre 3 (partie 5.1).

4.6. Données complémentaires sur les polymères

4.6.1. TISP (Chapitre 2)

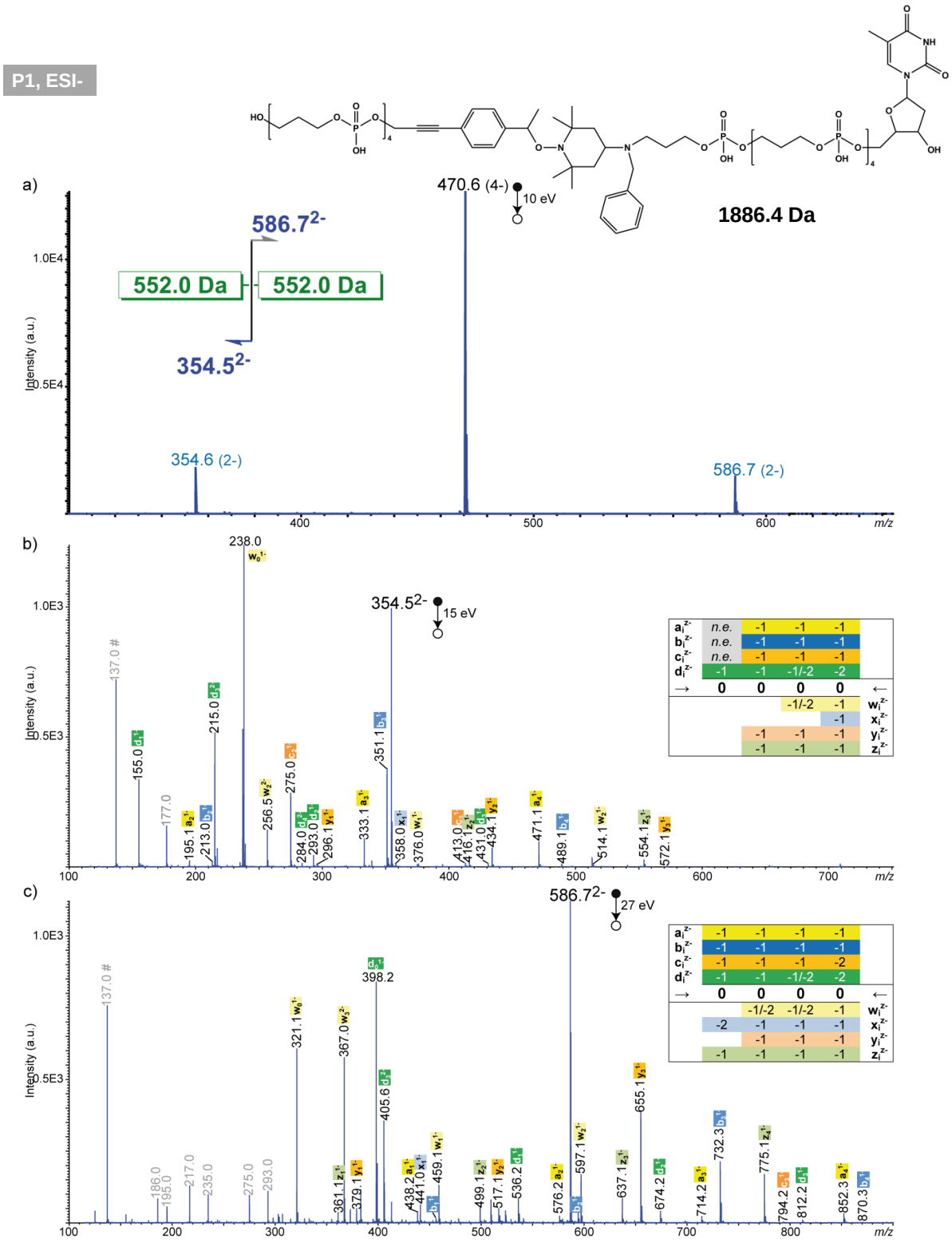


*Figure PE.4. 32 : Spectres MS de **P1**.*
*a) Spectre MS/MS de l'ion $[M – 4H]^{4–}$ à m/z 470.6 correspondant au polymère recherché dans l'échantillon **P1** et son schéma de fragmentation en inset (où l'encadré vert symbolise le bloc codant de masse 552.0 Da). Spectres pseudo-MS^3 des ions b) m/z 541.1 (3–) pour le séquençage du bloc 1 et c) m/z 553.8 (3–) pour le séquençage du bloc 2. Les tableaux en inset illustrent la couverture de*

séquence de chaque bloc. Les fragments secondaires sont annotés en gris, y compris les monomères déprotonés annotés par #.

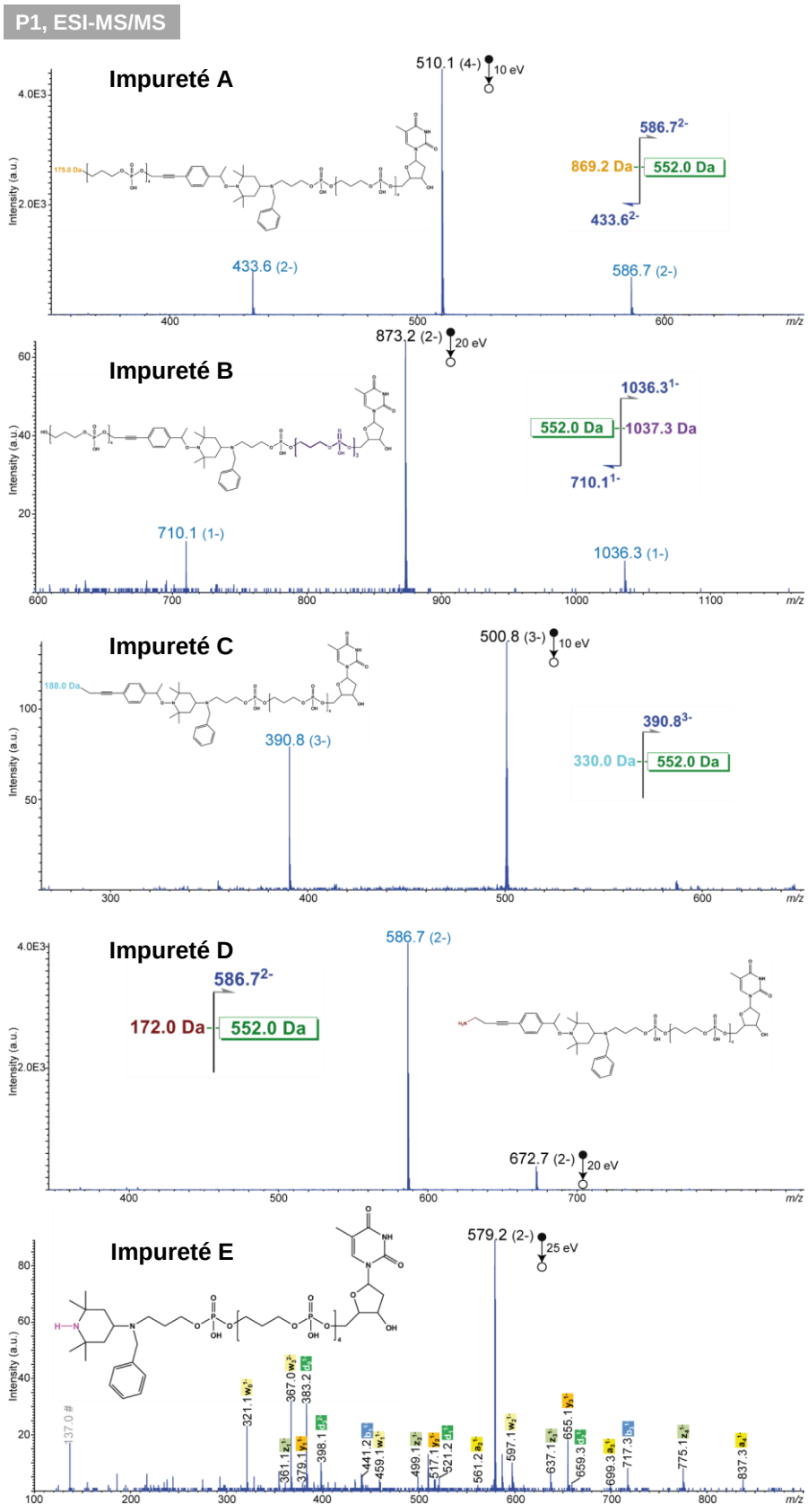


*Figure PE.4. 33 : Spectres MS/MS des impuretés observées dans **P1** (Figure PE.4. 32).*

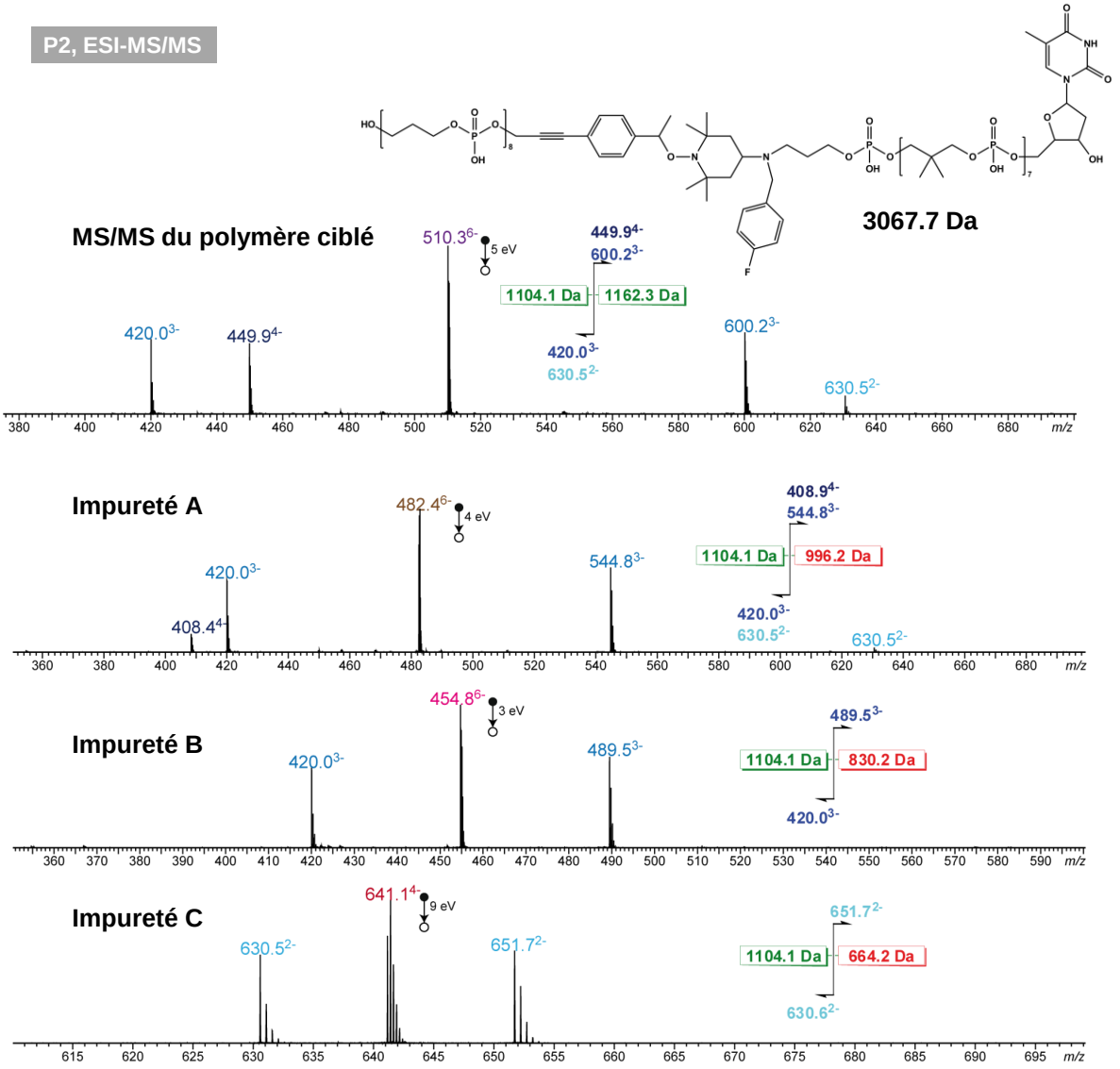


*Figure PE.4. 34 : Spectres MS/MS du polymère ciblé et des impuretés observées dans l'échantillons **P2**.*
Le spectre ESI-MS est disponible dans le Chapitre 2 (Figure II. 6). Spectres MS/MS des ions m/z 482.46– (bloc 1 : 8 M1, bloc 2 : 6 M3), m/z 454.96– (bloc 1 : 8 M1, bloc 2 : 5 M3) et m/z 641.14– (bloc 1 : 8 M1, bloc 2 : 4 M3). Dans les schémas de dissociation, les valeurs encadrées correspondent aux masses des blocs codants.

Les **impuretés A**, **B** et **C** correspondent à différents polymères où il manque 1 à 4 unités « 1 » dans le second bloc, comme validé par MS/MS pour les trois plus abondantes (Figure PE.4. 34).

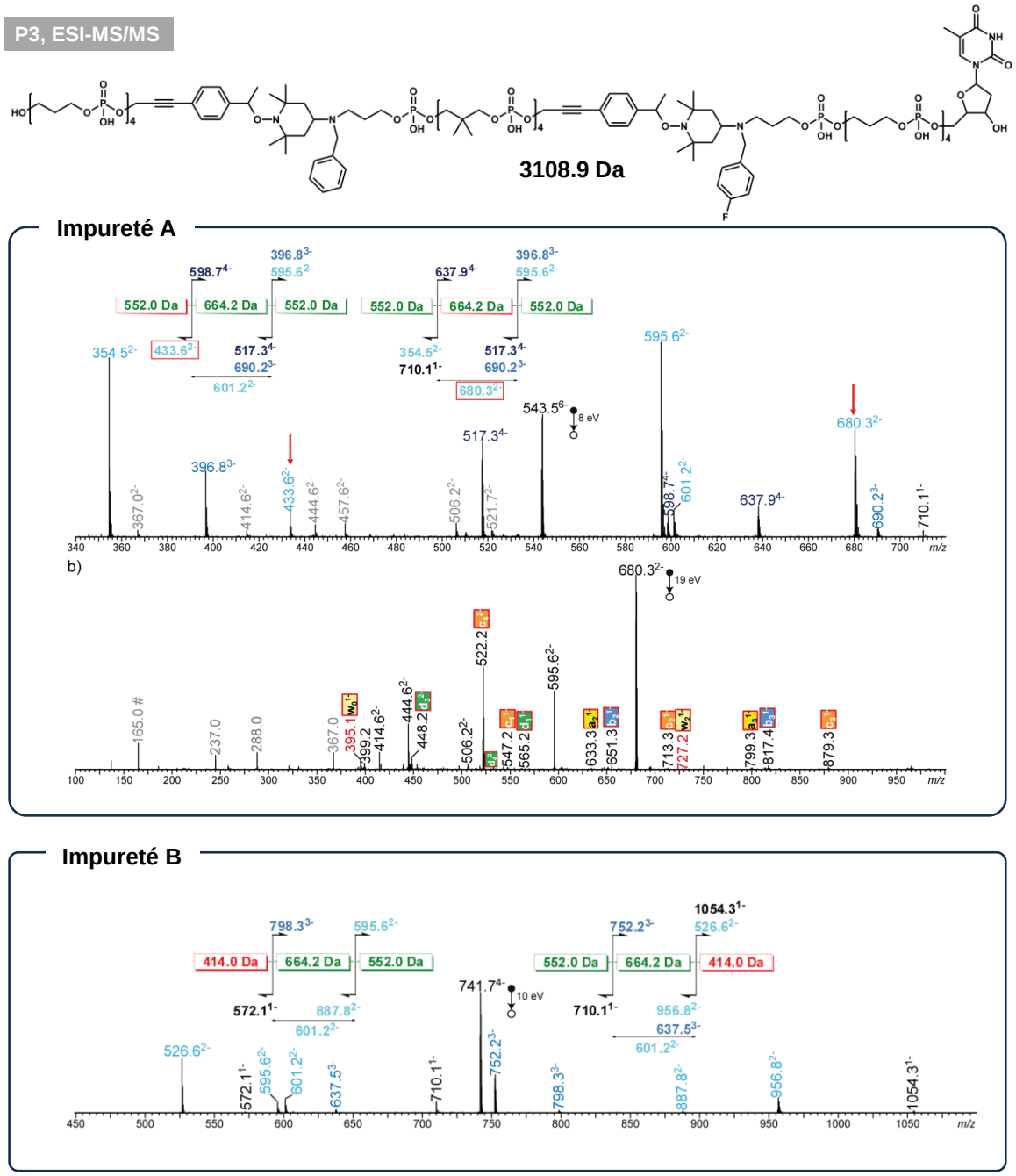
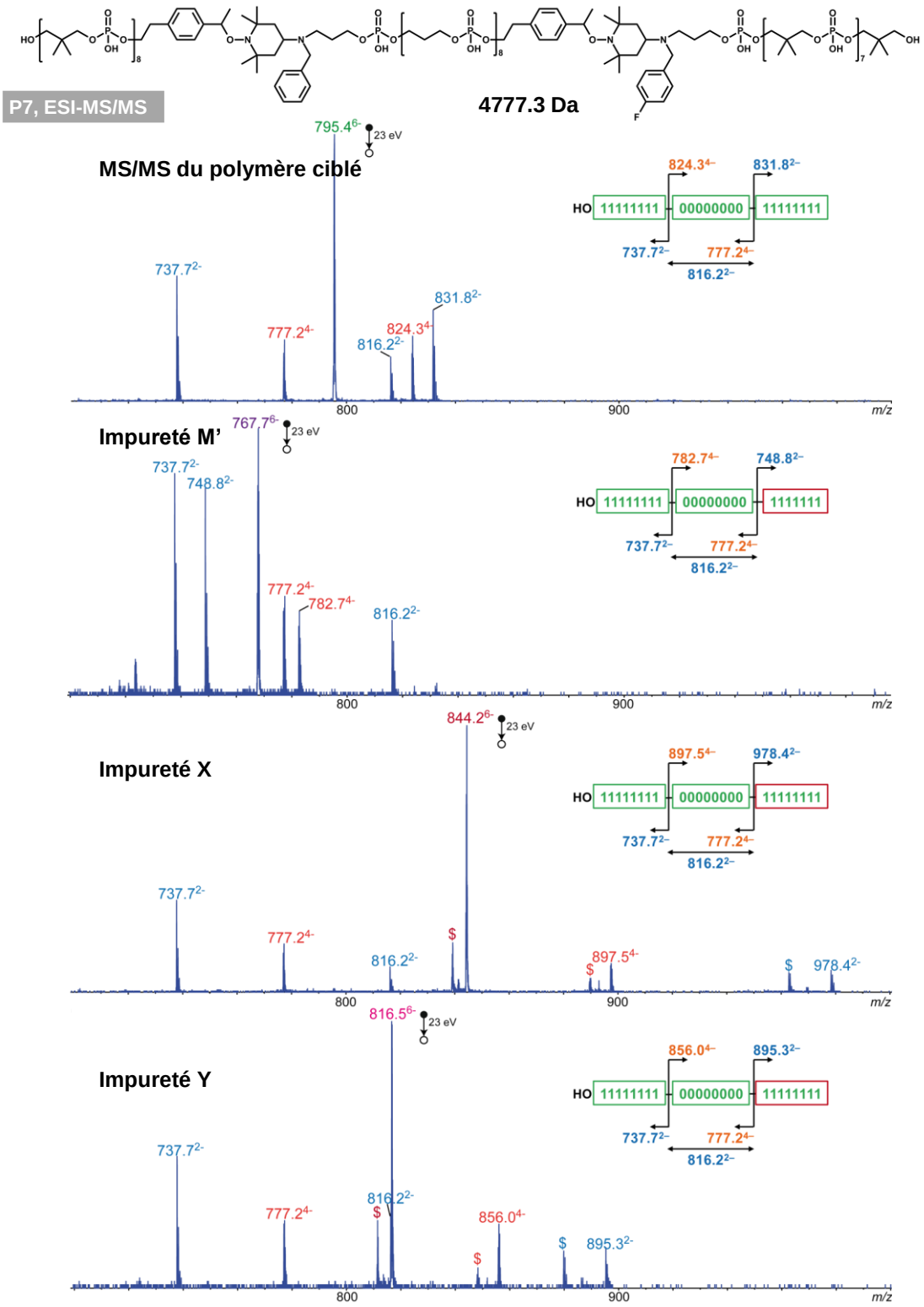


*Figure PE.4. 35 : Spectres MS/MS des impuretés observées dans **P3**.*
Le spectre ESI-MS est disponible dans le Chapitre 2 (Figure II. 8).

Impureté A : spectre MS/MS de l'ion $[A - 6H]^{6-}$ à *m/z* 543.5. Mêmes observations que pour l'analyse MS/MS du polymère ciblé (Chapitre 2 Figure II. 9), qui révèle que l'excès de masse Δm +158 Da est localisé dans le segment 1 ou 2 (les valeurs encadrées correspondent aux masses des blocs codants). Les fragments annotés en gris sont des fragments issus de ruptures de liaisons phosphate. b) Spectre pseudo-MS^3 de l'ion *m/z* 680.3^{2-} contenant le bloc 2 et le défaut structural. Les annotations encadrées de rouge pour les fragments de ce segment central indiquent que les ions a/b/c/d sont majoritairement détectés avec un excès de masse de 1 Da tandis que les ions w/x/y/z sont majoritairement détectés avec un défaut de masse de 1 Da. Néanmoins, seuls les ions w/x/y/z (en rouge) sont déplacés de +158 Da, ce qui indique que le défaut structural est localisé dans le "résidu alkyle" de l'alcoxyamine. Fragments internes en gris ; # monomères déprotonés (Figure PE.4. 35).

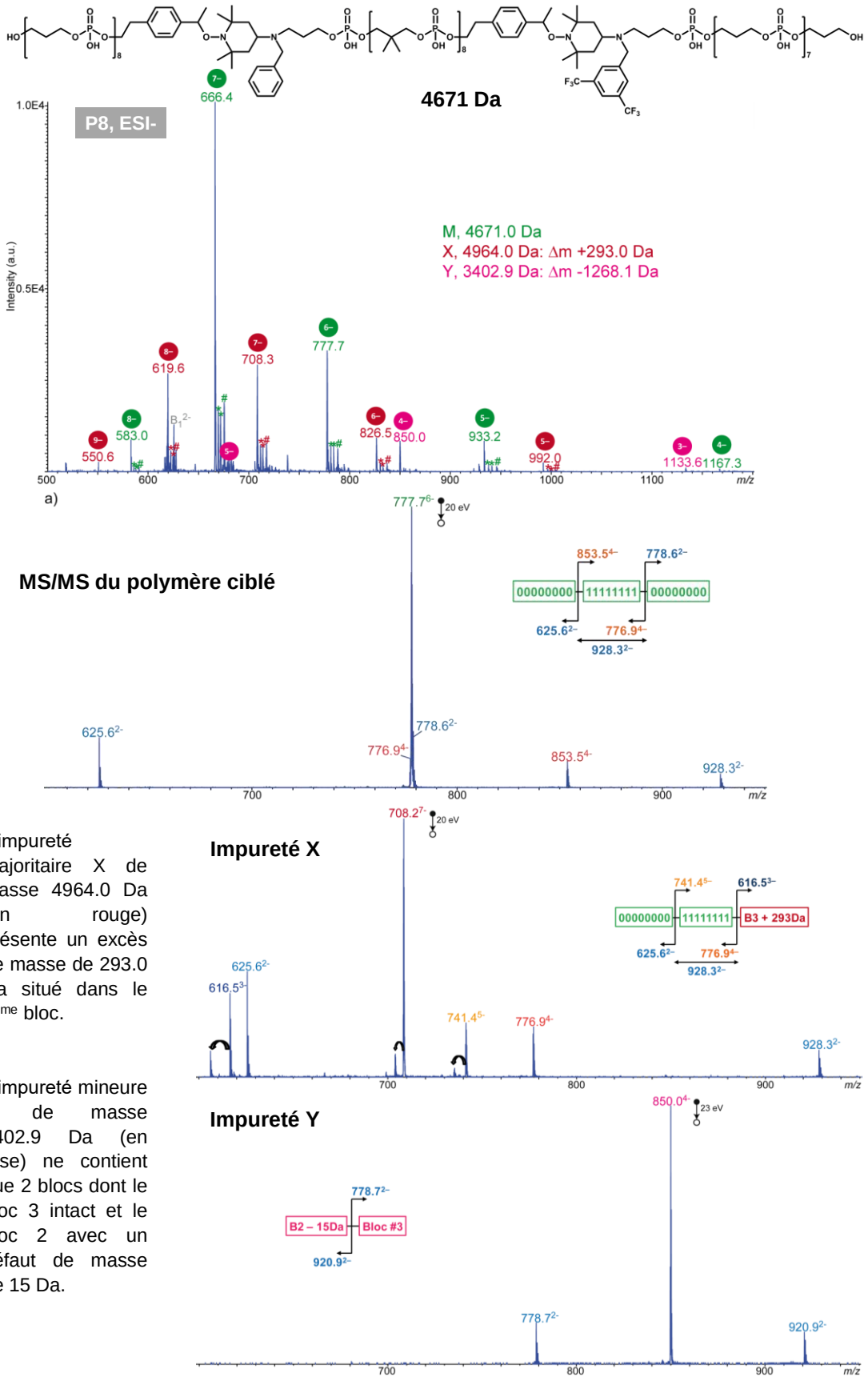
Impureté B : spectre MS/MS de l'ion $[B - 4H]^{4-}$ à *m/z* 741.7, qui révèle que le défaut de masse Δm -138 Da est localisé dans le bloc 1 ou le bloc 3 (les valeurs encadrées correspondent aux masses des blocs codants) (Figure PE.4. 35).

4.6.2. TISP3 (Chapitre 3)



L'impureté mineure M' de masse 4611.4 Da (en violet) présente un défaut de masse de 166 Da par rapport au polymère ciblé, suggérant l'absence d'une unité 1. Les données MS/MS indiquent que ce défaut de masse est dans 3ème bloc. **L'impureté majoritaire X** de masse 5070.4 Da (en rouge) présente un excès de masse de 293.1 Da situé dans le 3ème bloc. **L'impureté Y** de masse 4904.2 Da (en rose) présente un excès de masse de 126.9 Da situé dans le 3ème bloc (Figure PE.4. 36).

*Figure PE.4. 36 : Spectres MS/MS du polymère ciblé et des impuretés observées dans l'échantillon **P7**.*
*Le spectre ESI-MS est disponible dans le Chapitre 3 (Figure III. 7). Spectre MS/MS des ions [M – 6H]6– à m/z 795.4, [M' – 6H]$^{6-}$ à m/z 767.7, [X – 6H]$^{6-}$ à m/z 844.2 et [Y – 6H]$^{6-}$ à m/z 816.5 dans l'échantillon **P7**. $ désigne des fragments secondaires*



L'impureté majoritaire X de masse 4964.0 Da (en rouge) présente un excès de masse de 293.0 Da situé dans le 3ème bloc.

L'impureté mineure Y de masse 3402.9 Da (en rose) ne contient que 2 blocs dont le bloc 3 intact et le bloc 2 avec un défaut de masse de 15 Da.

*Figure PE.4. 37 : Spectres MS du polymère ciblé des impuretés observées dans l'échantillon **P8**. Spectre ESI-MS(-) et Spectre MS/MS des ions $[M - 6H]^{6-}$ à m/z 777.7, $[X - 7H]^{7-}$ à m/z 708.2 et $[Y - 4H]^{4-}$ à m/z 850.0 dans l'échantillon **P8**. Les flèches noires indiquent une perte de 31.0 Da à partir de tous les fragments contenant le bloc "Δm +293 Da".*

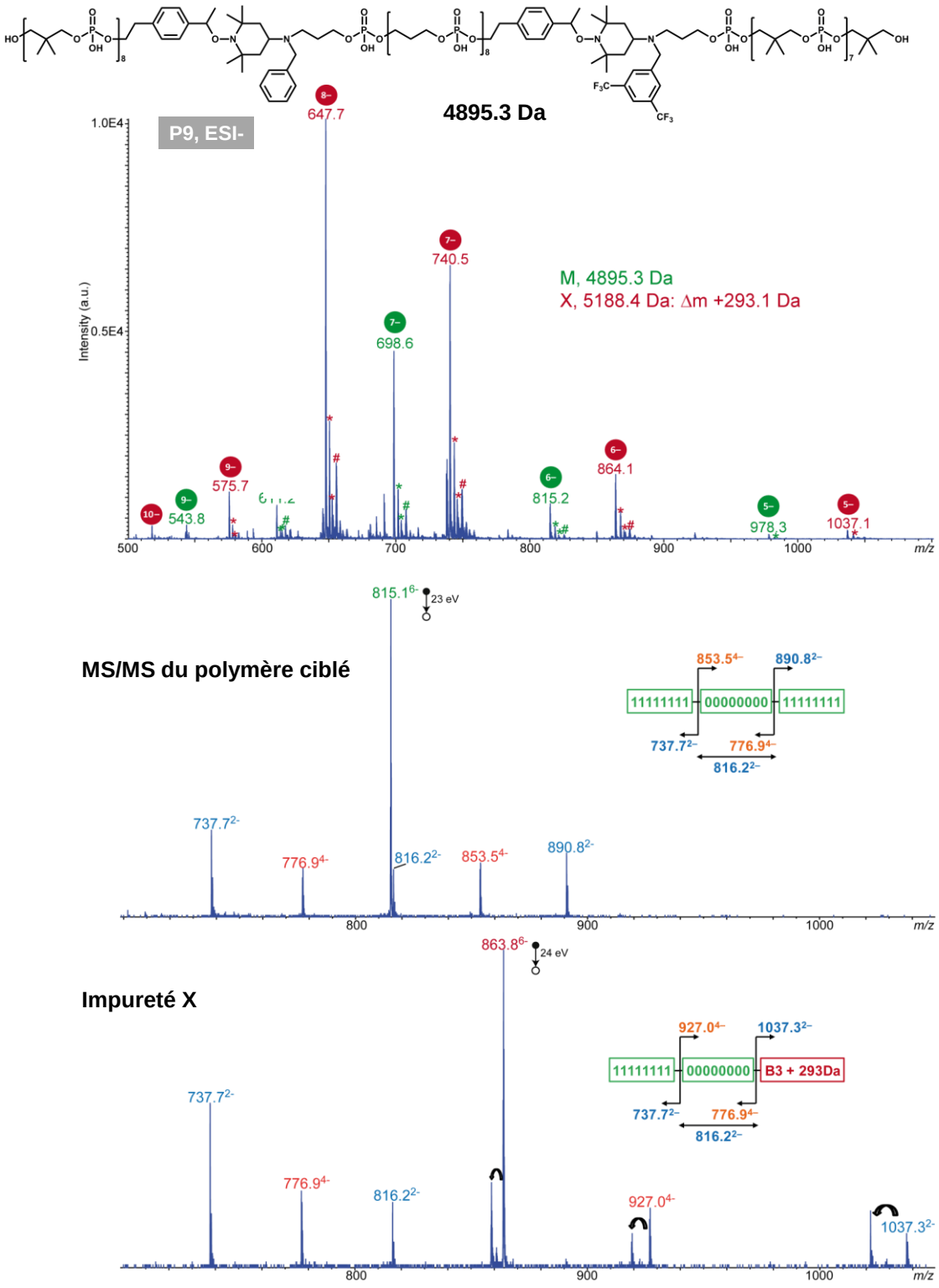


*Figure PE.4. 38 : Spectres MS du polymère ciblé et de l'impureté X observée dans l'échantillon **P9**. Spectre ESI-MS(-) et spectre MS/MS des ions $[M - 6H]^{6-}$ à m/z 815.2 et $[X - 6H]^{6-}$ à m/z 864.1 dans l'échantillon **P9**. Les pics repérés par un astérisque sont issus d'échanges H/Na et/ou H/K et ceux annotés par # correspondent à un échange H/Cu. Les pics non annotés correspondent à des fragments formés en source. Les flèches noires indiquent une perte de 31.0 Da à partir de tous les fragments contenant le bloc "Δm +293 Da".*

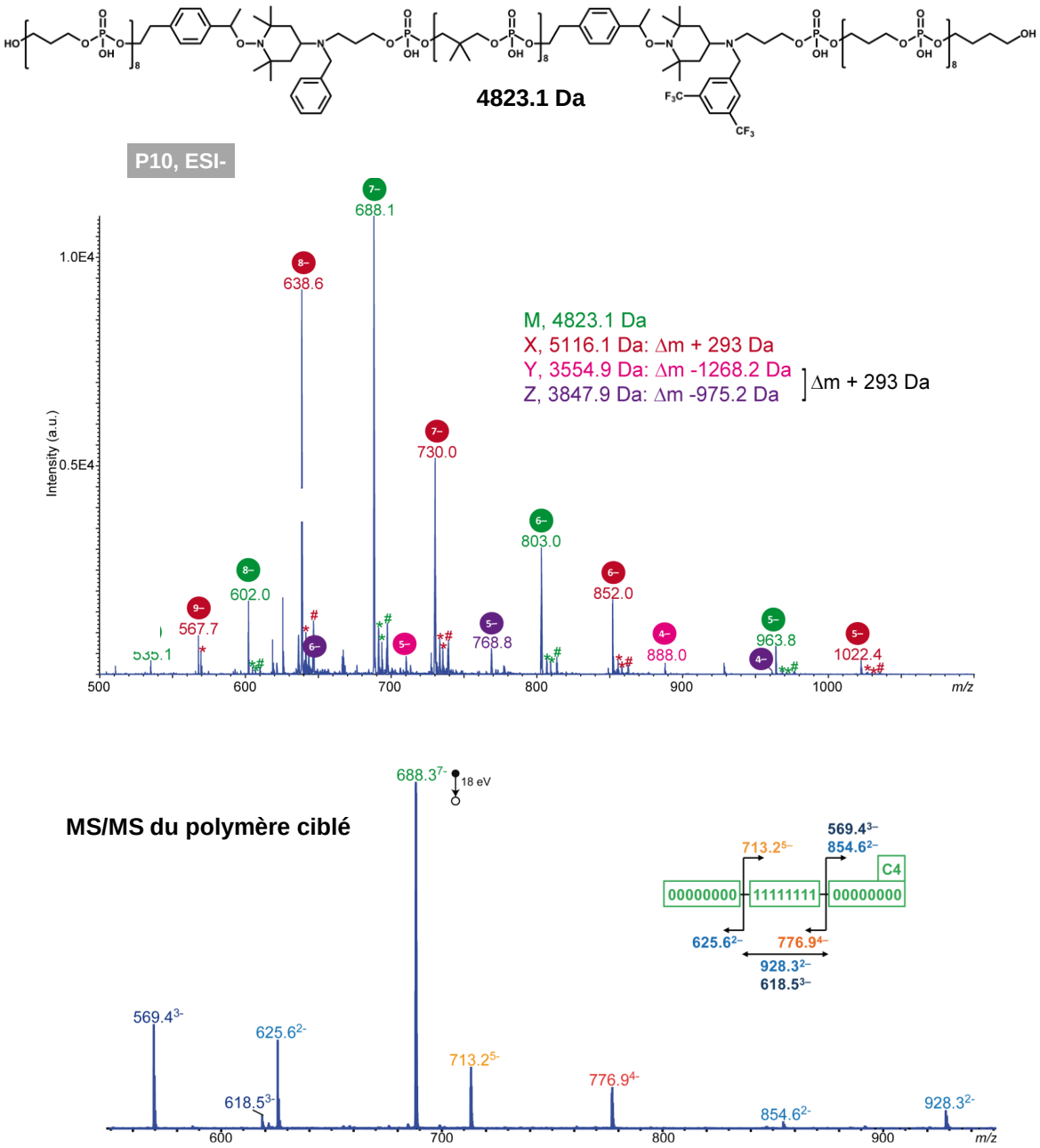
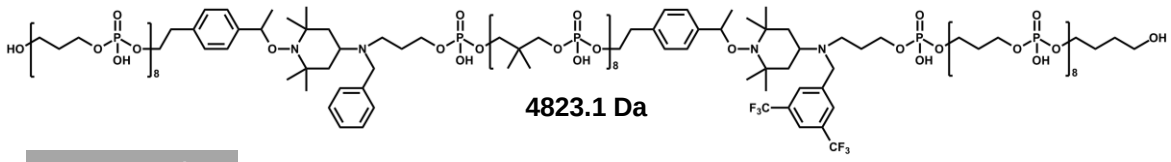


*Figure PE.4. 39 : Spectres MS de l'échantillon **P10**.*
*Spectre ESI-MS(-) et spectre MS/MS de l'ion $[M - 7H]^{7-}$ à m/z 688.3 dans l'échantillon **P10** correspondant au polymère ciblé.*



P10, ESI-MS/MS

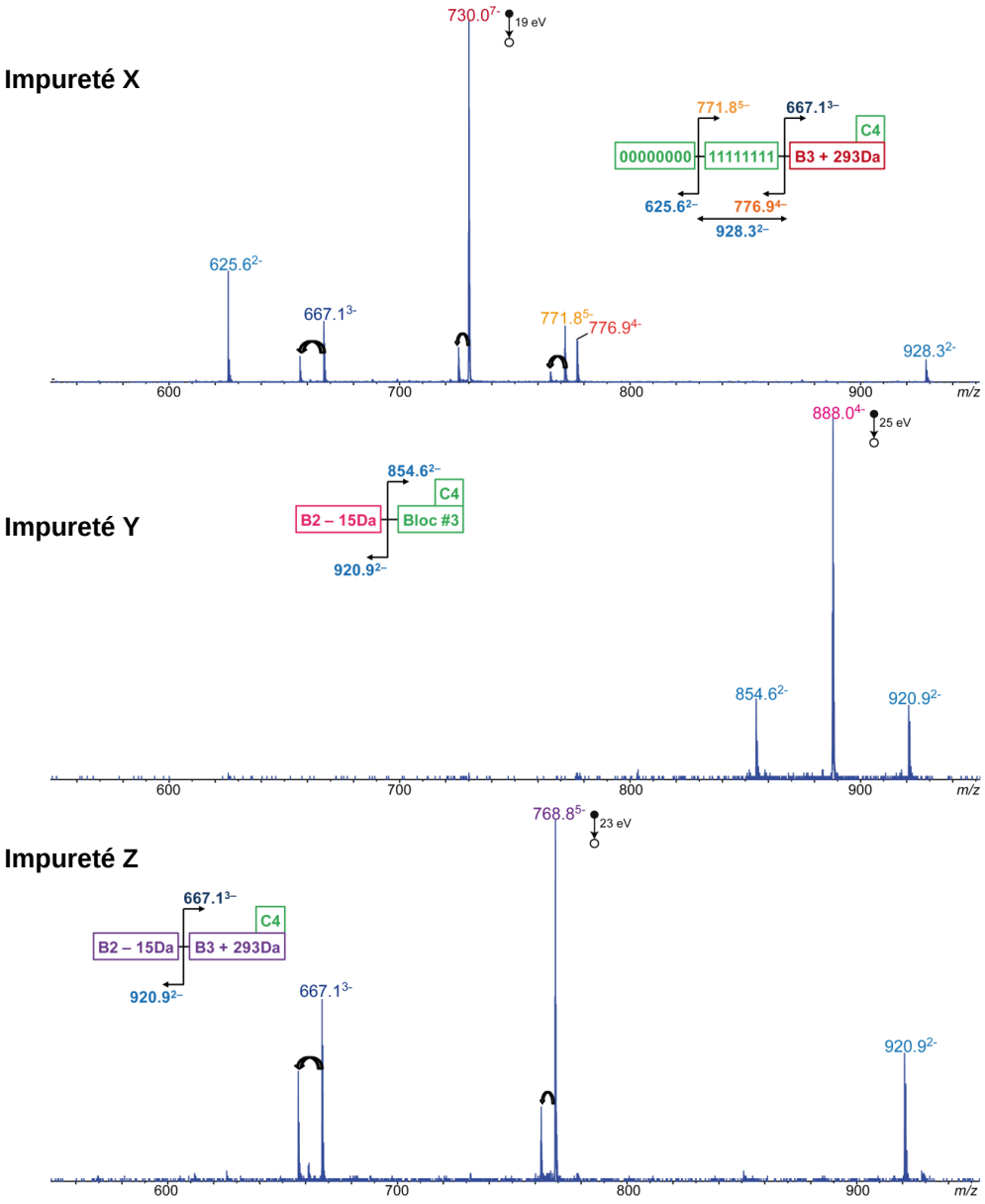


*Figure PE.4. 40 : Spectres MS/MS des impuretés observées dans l'échantillon **P10** (Figure PE.4. 39). Spectres MS/MS des ions $[X - 7H]^{7-}$ à m/z 730.0, $[Y - 4H]^{4-}$ à m/z 888.0, et $[Z - 5H]^{5-}$ à m/z 768.8 dans l'échantillon **P10**. Les flèches noires indiquent une perte de 31.0 Da à partir de tous les fragments contenant le bloc "Δm +293 Da".*

L'impureté majoritaire X de masse 5116.1 Da (en rouge) présente un excès de masse de 293.0 Da situé dans le 3ème bloc. **L'impureté mineure Y** de masse 3402.9 Da (en rose) ne contient que 2 blocs dont le bloc 3 intact et le bloc 2 avec un défaut de masse de 15 Da. **L'impureté mineure Z** de masse 3847.9 Da (en violet) présente un excès de masse de 293.0 Da, situé dans le dernier bloc, par rapport à Y (Figure PE.4. 40).

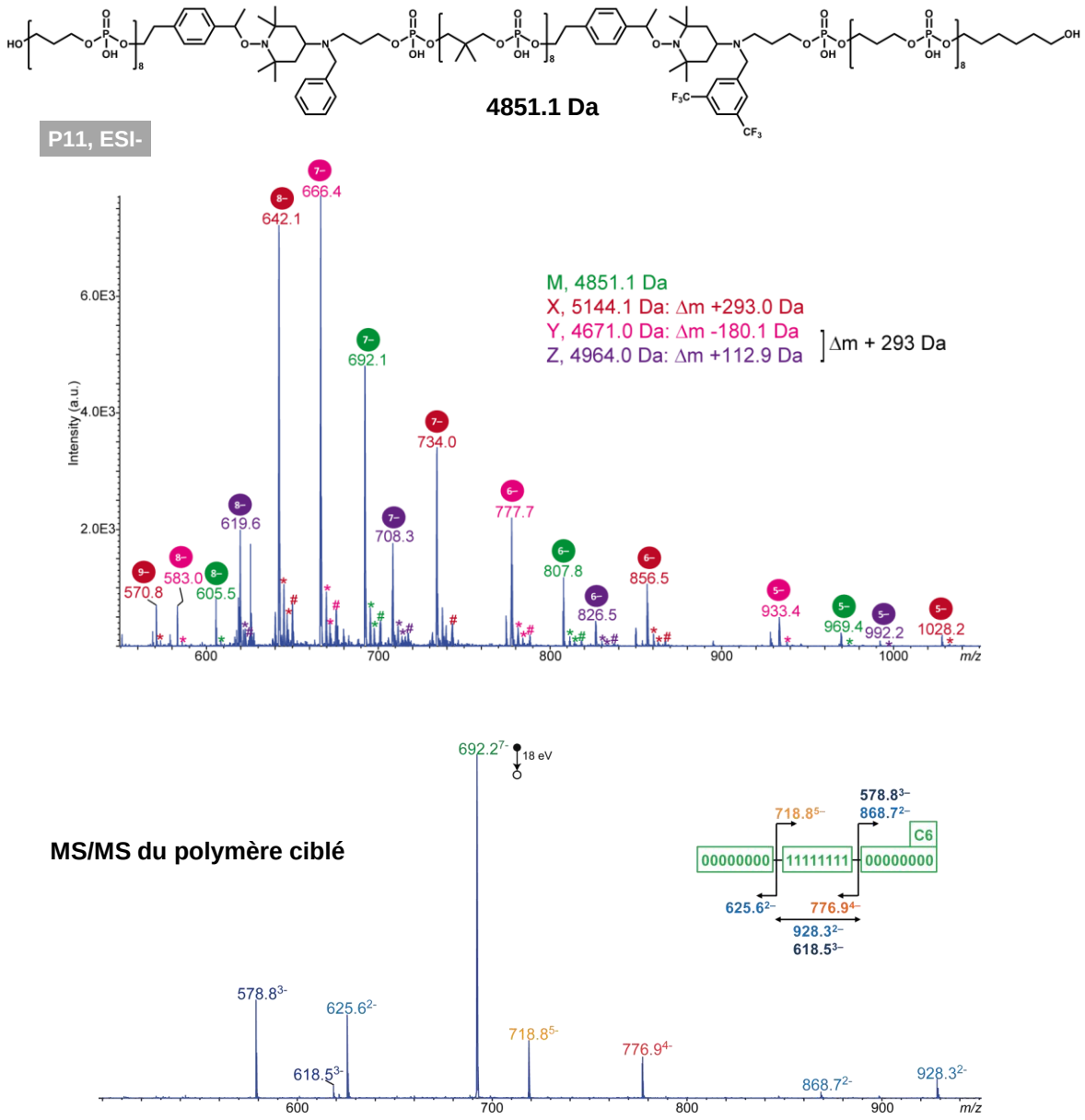


*Figure PE.4. 41 : Spectres MS de l'échantillon **P11**.*
*Spectre ESI-MS(-) et spectre MS/MS de l'ion $[M - 7H]^{7-}$ à m/z 692.2 dans l'échantillon **P11** correspondant au polymère ciblé.*

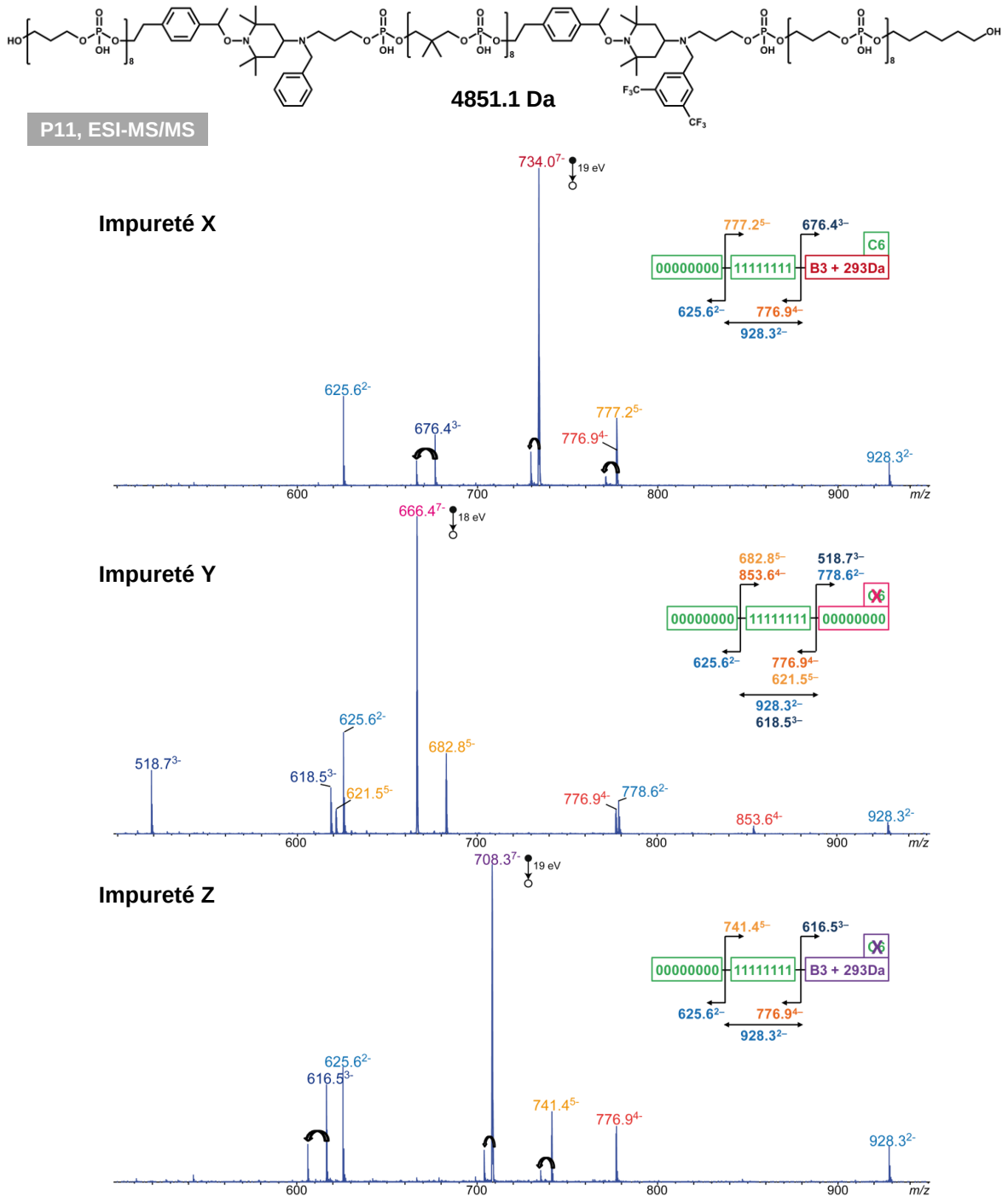
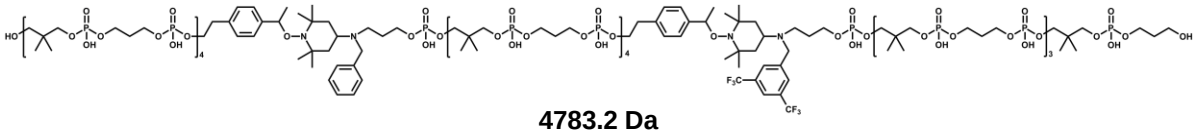


*Figure PE.4. 42 : Spectres MS/MS des impuretés observées dans l'échantillon **P11** (Figure PE.4. 41). Spectres MS/MS des ions $[X - 7H]^{7-}$ à m/z 734.0, $[Y - 7H]^{7-}$ à m/z 666.4, et $[Z - 7H]7^{-}$ à m/z 708.3 dans l'échantillon **P11**. Les flèches noires indiquent une perte de 31.0 Da à partir de tous les fragments contenant le bloc "Δm +293 Da".*

L'impureté X de masse 5144.1 Da (en rouge) présente un excès de masse de 293.0 Da situé dans le 3ème bloc. **L'impureté Y** de masse 4671.0 Da (en rose) présente un défaut de masse de 180.1 Da situé dans le 3ème bloc qui pourrait correspondre à un groupement phosphate lié au C6 (soit l'équivalent du polymère recherché dans l'échantillon **P8**). **L'impureté mineure Z** de masse 4964.0 Da (en violet) présente un excès de masse de 293.0 Da, situé dans le 3ème bloc, par rapport à Y (Figure PE.4. 42).



4783.2 Da

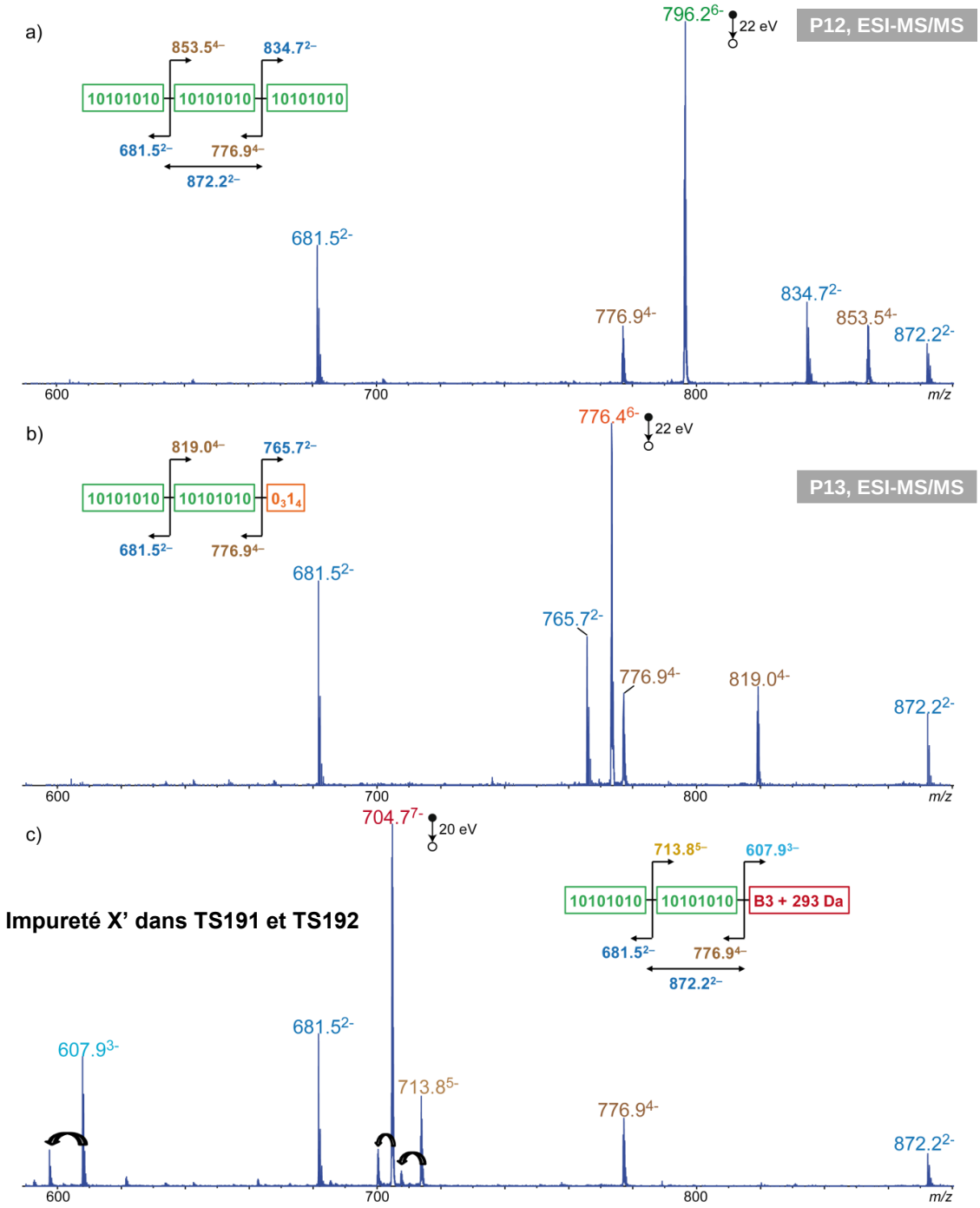


*Figure PE.4. 43 : Spectres MS/MS des différentes espèces observées dans les échantillons **P12** et **P13**.*
*Spectre MS/MS a) des ions $[M - 6H]^{6-}$ à m/z 796.2 dans l'échantillon **P12**, b) $[M' - 6H]^{6-}$ à m/z 776.4 et c) $[X' - 7H]^{7-}$ à m/z 704.7 dans l'échantillon **P13**. Les flèches noires indiquent une perte de 31.0 Da à partir de tous les fragments contenant le bloc 3 porteur du défaut "Δm +293 Da".*
Spectre ESI-MS des deux échantillons disponibles dans le Chapitre 3 (Figure III. 8).

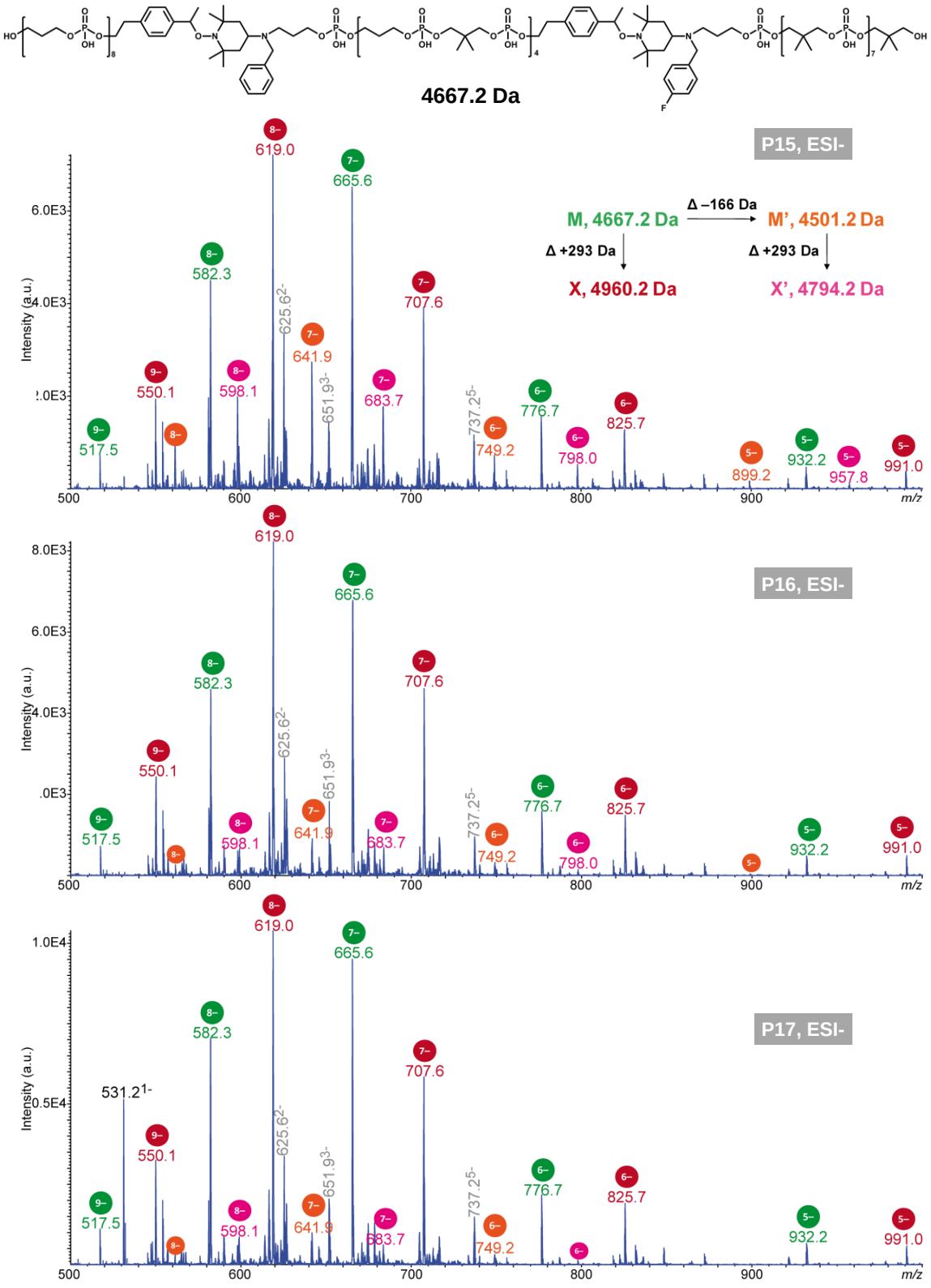


*Figure PE.4. 44 : Spectres ESI-MS en mode négatif des échantillons **P15**, **P16** et **P17**. Les pics annotés en gris (annotation non exhaustive) correspondent à des fragments formés en source.*

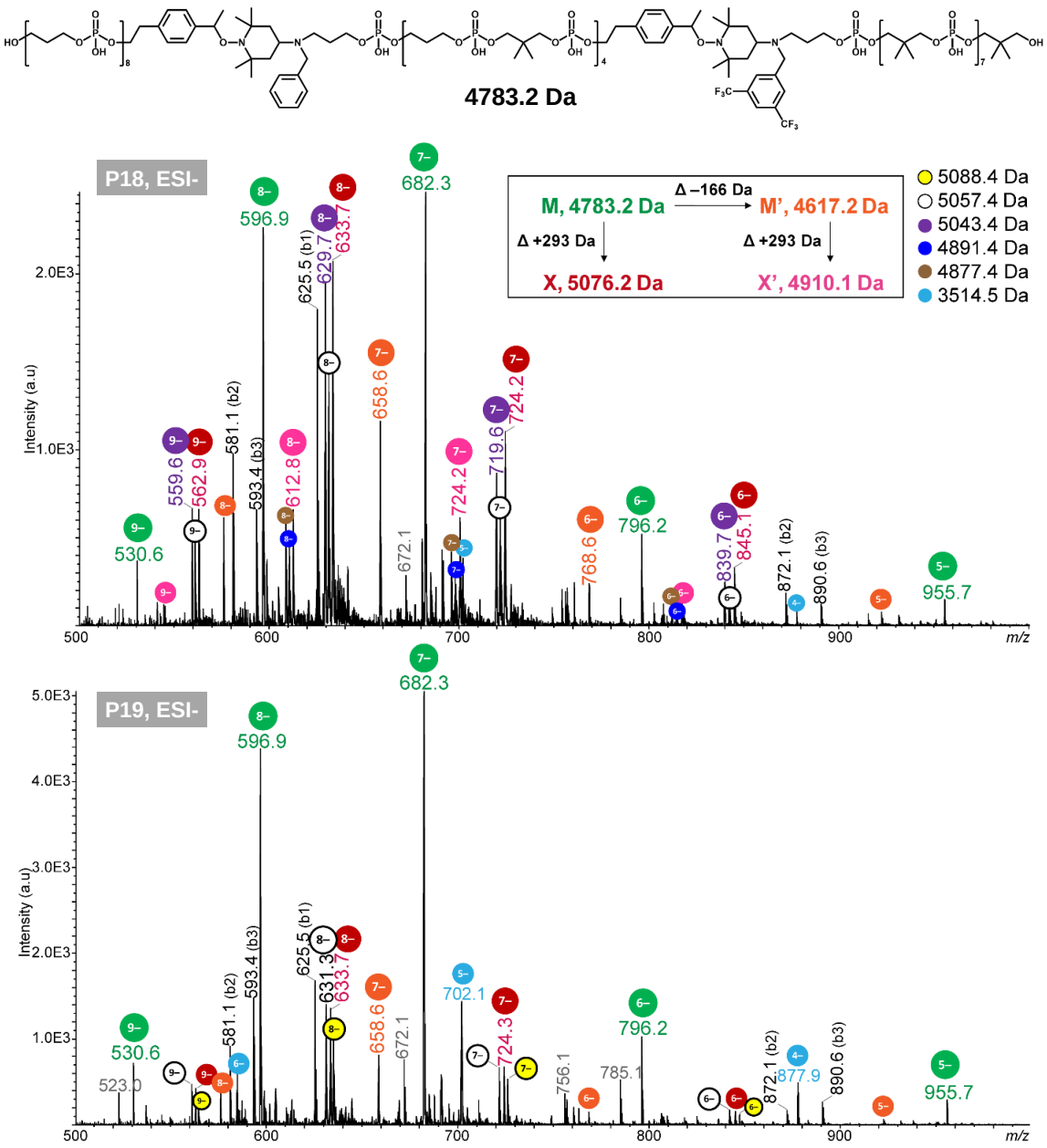


*Figure PE.4. 45 : Spectres ESI(–)-MS des échantillons **P18** et **P19**. Les blocs libérés en source sont annotés en noir et les pics annotés en gris sont des fragments de très faible état de charge.*

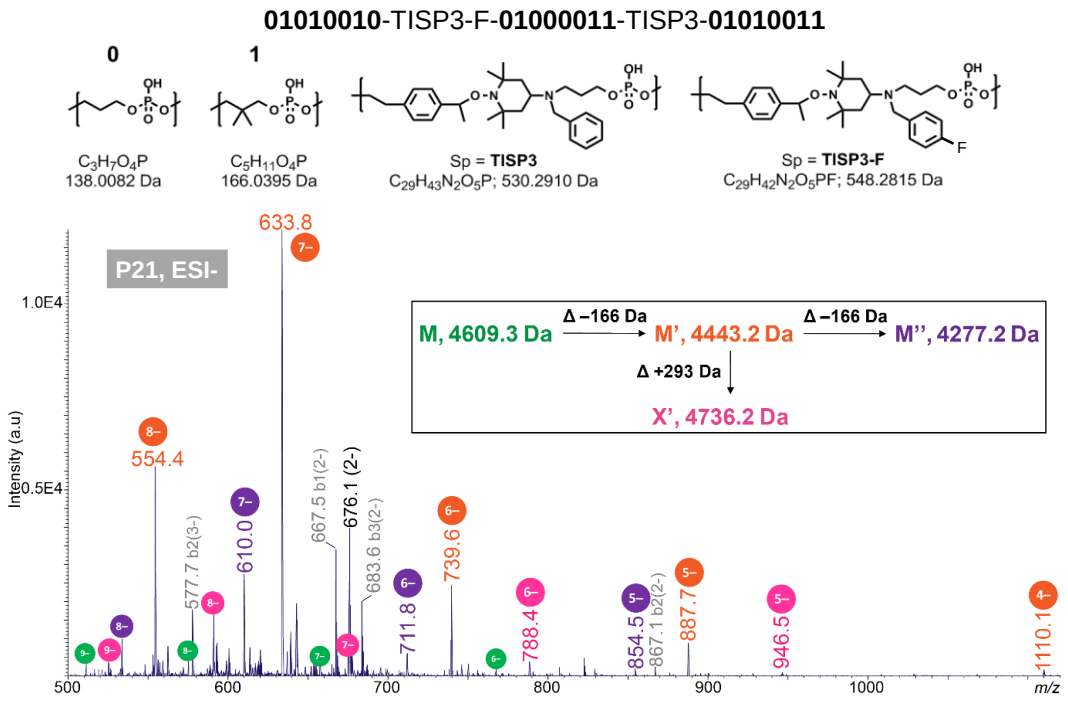


*Figure PE.4. 46 : Spectres ESI(–)-MS de l'échantillon **P21**. Les blocs libérés en source sont annotés en gris et les pics annotés en noir sont des fragments de très faible état de charge.*

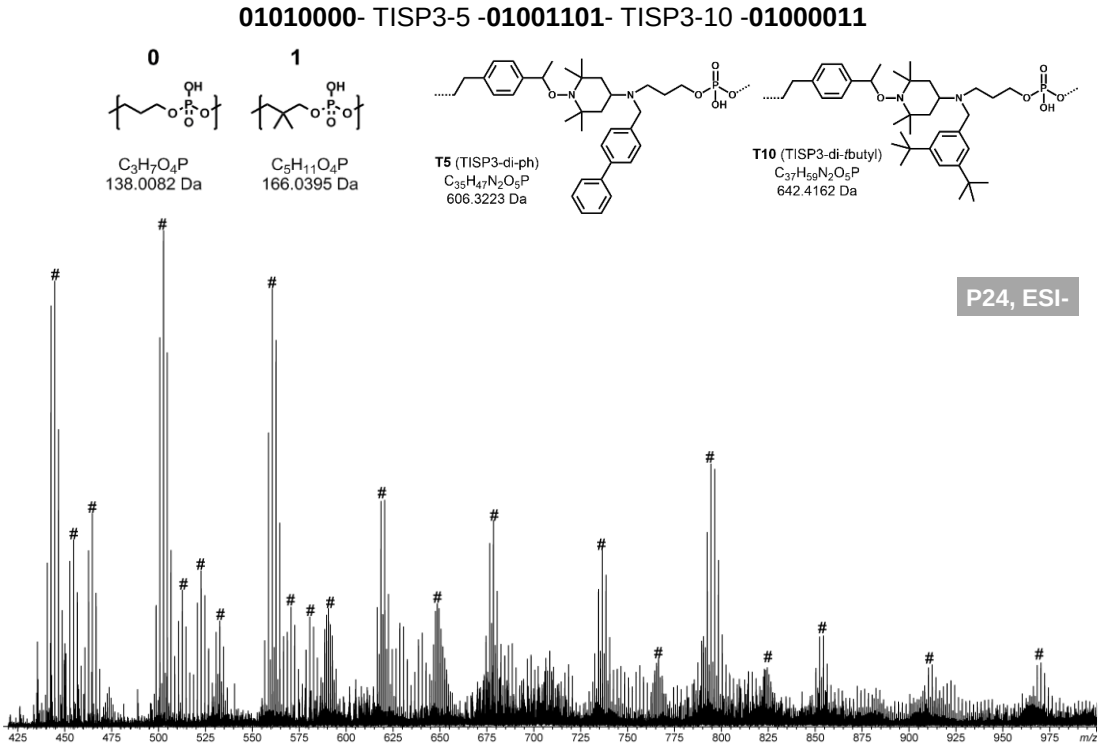


*Figure PE.4. 47 : Spectres ESI(–)-MS de l'échantillon **P24**. De nombreux échanges H/Na (annotés par un astérisque) sont observés. L'échantillon présente trop de sel de TCA pour être ionisé correctement.*

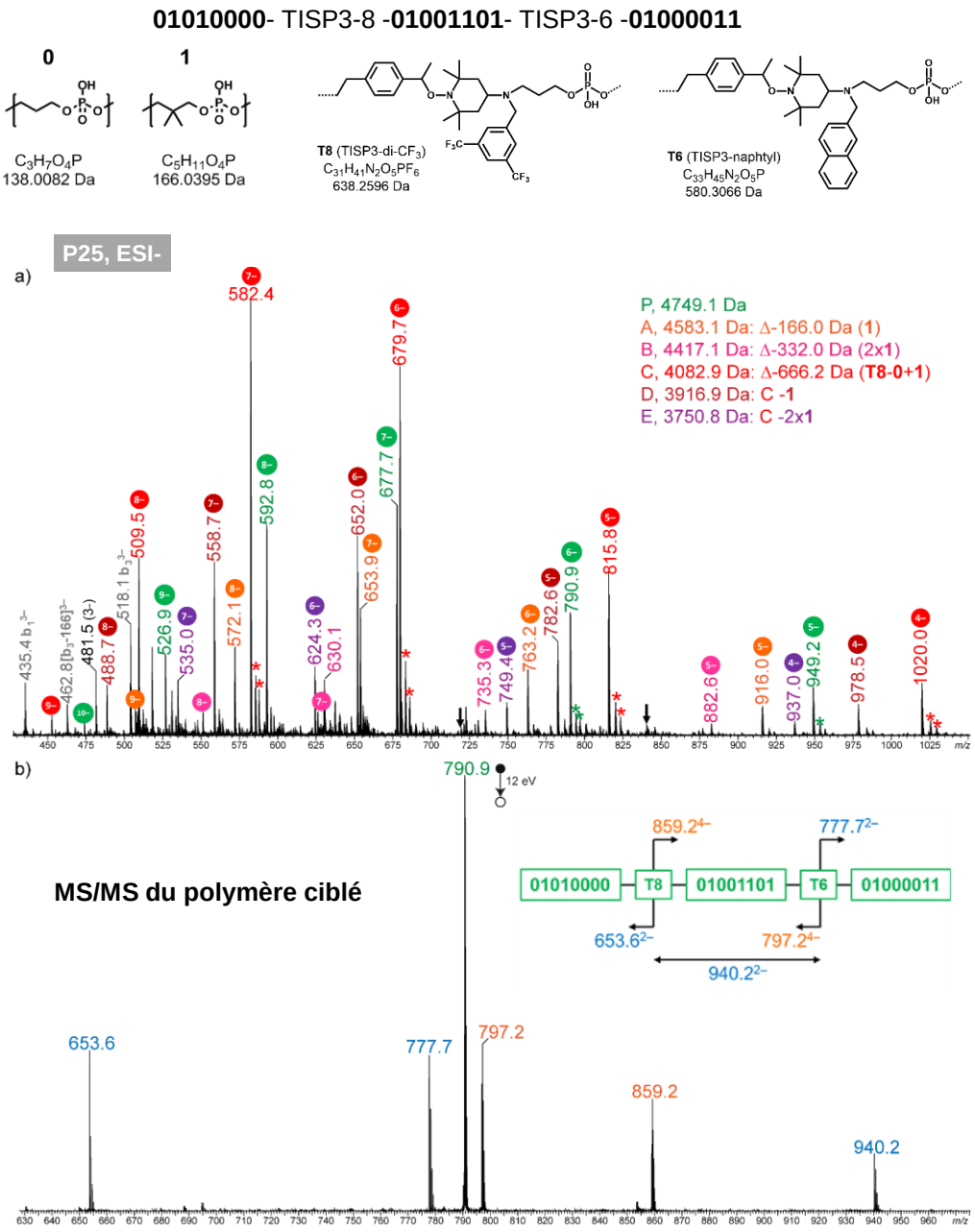


*Figure PE.4. 48 : Spectres MS de l'échantillon **P25**.*
*a) Spectre ESI(–)-MS et b) Spectre MS/MS de l'ion [P – 6H]6– à m/z 790.9 dans l'échantillon **P25**. Les astérisques indiquent des échanges H/Na et H/K. Les blocs libérés en source sont annotés en gris. Les impuretés observées sont décrites en inset.*

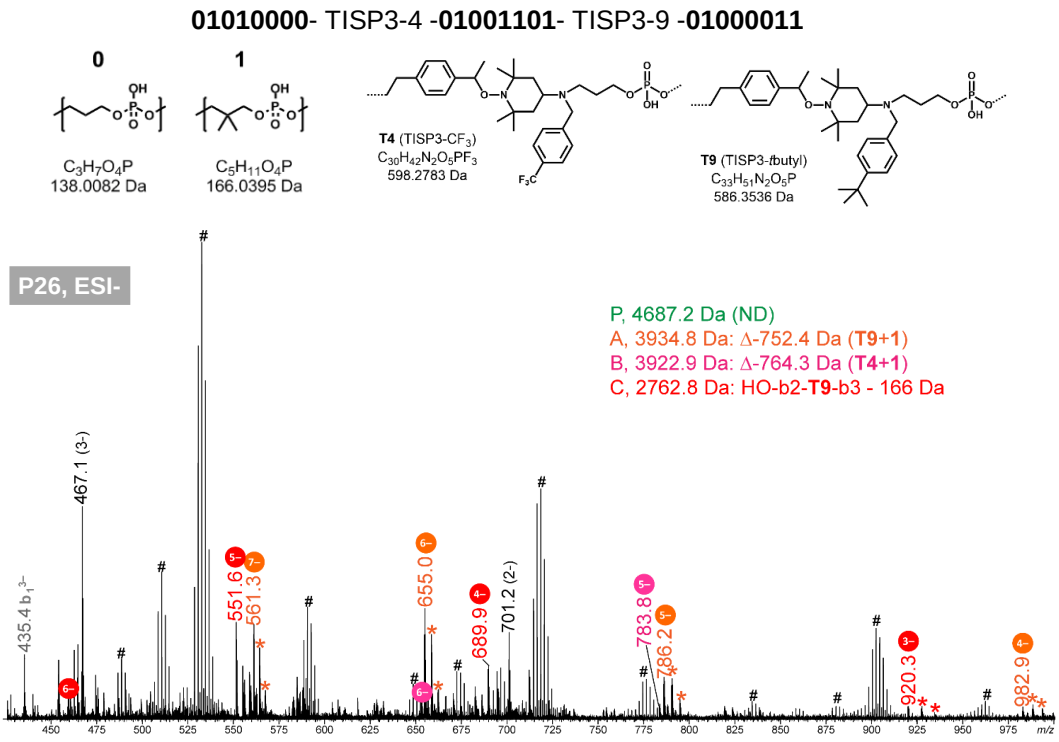


*Figure PE.4. 49 : Spectres ESI(–)-MS de l'échantillon **P26**.*
Les astérisques indiquent des échanges H/Na et H/K. Les blocs libérés en source sont annotés en gris. Les impuretés observées sont décrites en inset.

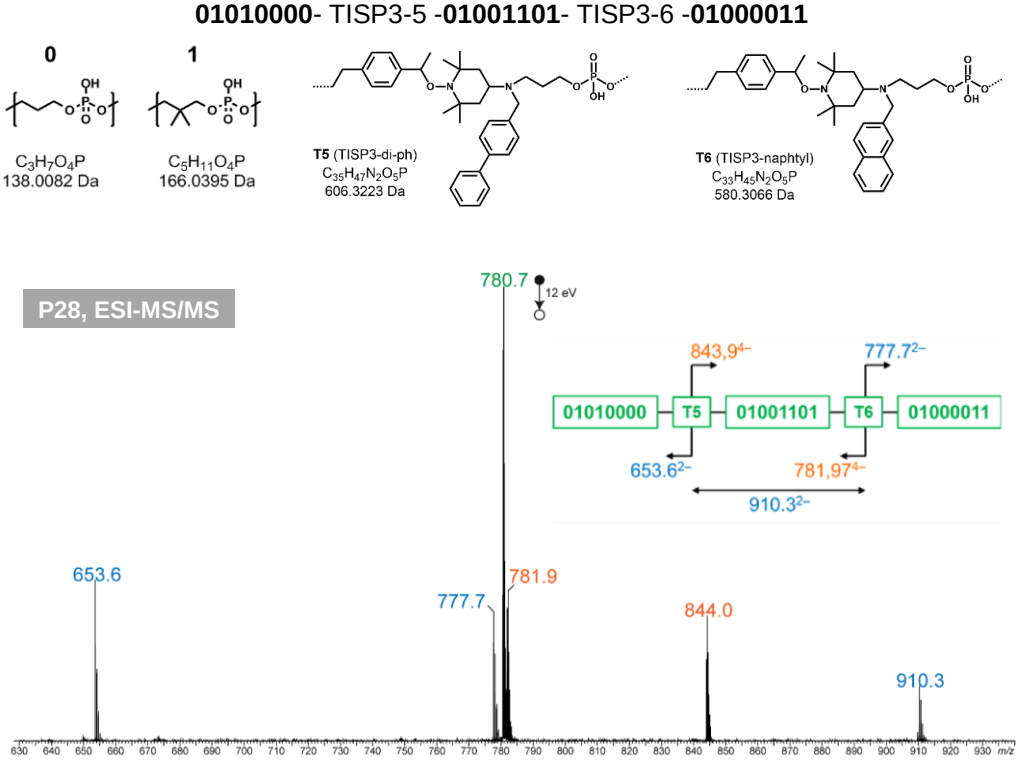


*Figure PE.4. 50 : Spectre MS/MS de l'ion $[P - 6H]^{6-}$ à m/z 780.7 dans l'échantillon **P28**.*
Le spectre ESI-MS(-) est disponible dans le Chapitre 3 (Figure III. 14).

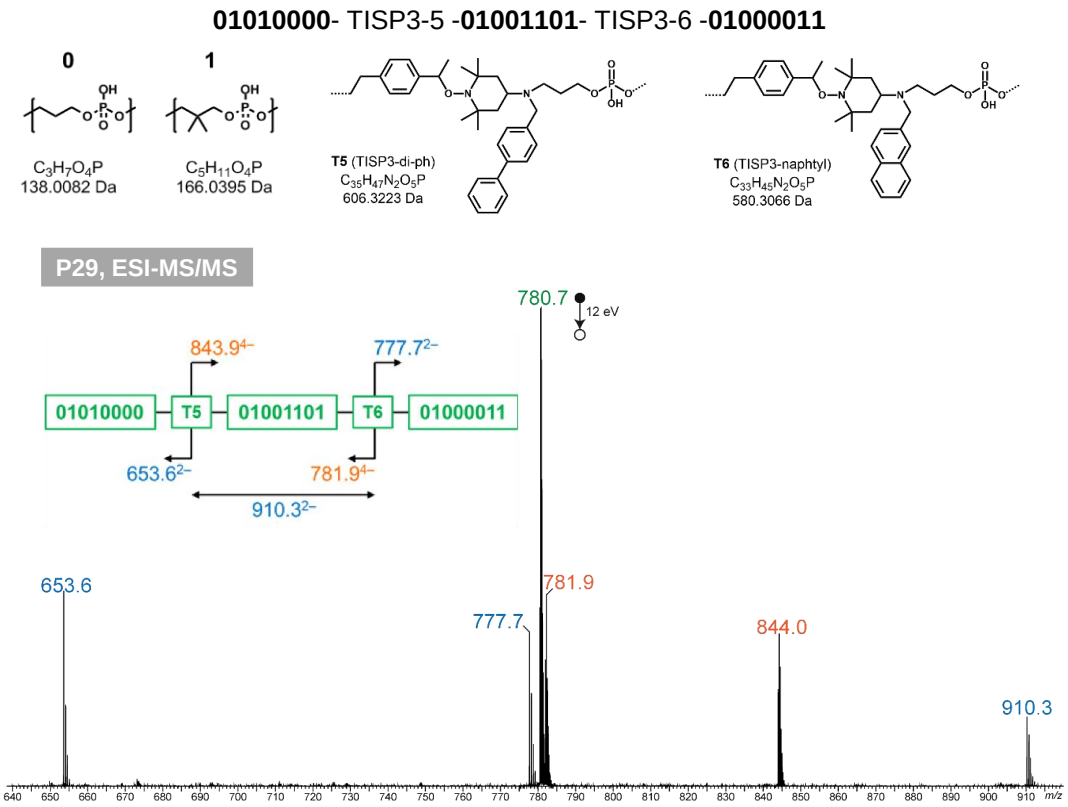


*Figure PE.4. 51 : Spectre MS/MS de l'ion [P – 6H]$^{6-}$ à m/z 780.7 dans l'échantillon **P29**. Le spectre ESI-MS(-) est disponible dans le Chapitre 3 (Figure III. 14).*

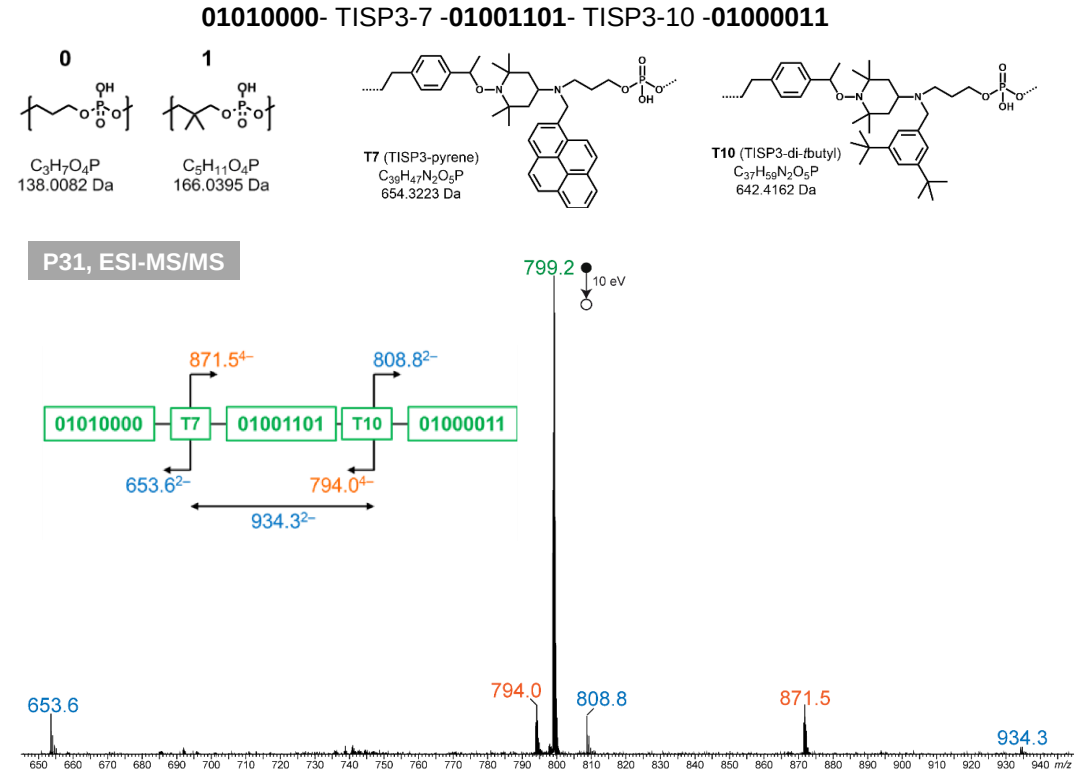


*Figure PE.4. 52 : Spectre MS/MS de l'ion [P – 6H]$^{6-}$ à m/z 799.2 dans l'échantillon **P31**. Le spectre ESI-MS(-) est disponible dans **le** Chapitre 3 (Figure III. 15).*

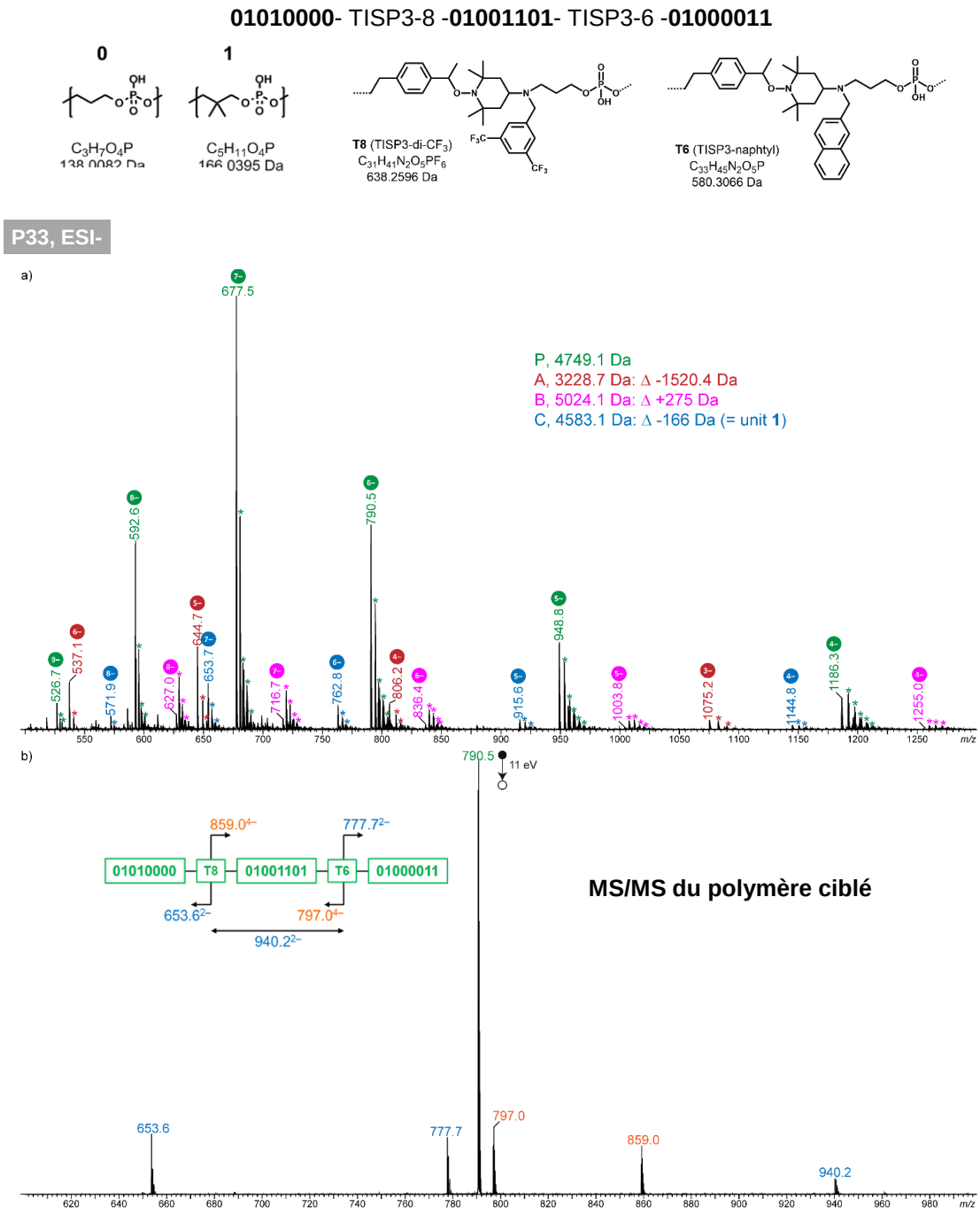


*Figure PE.4. 53 : Spectres MS de l'échantillon **P33**.*
*a) Spectre ESI(–)-MS et b) Spectre MS/MS de l'ion $[P - 6H]^{6-}$ à m/z 790.5 dans l'échantillon **P33**. Les astérisques désignent les échanges H/Na et H/K.*

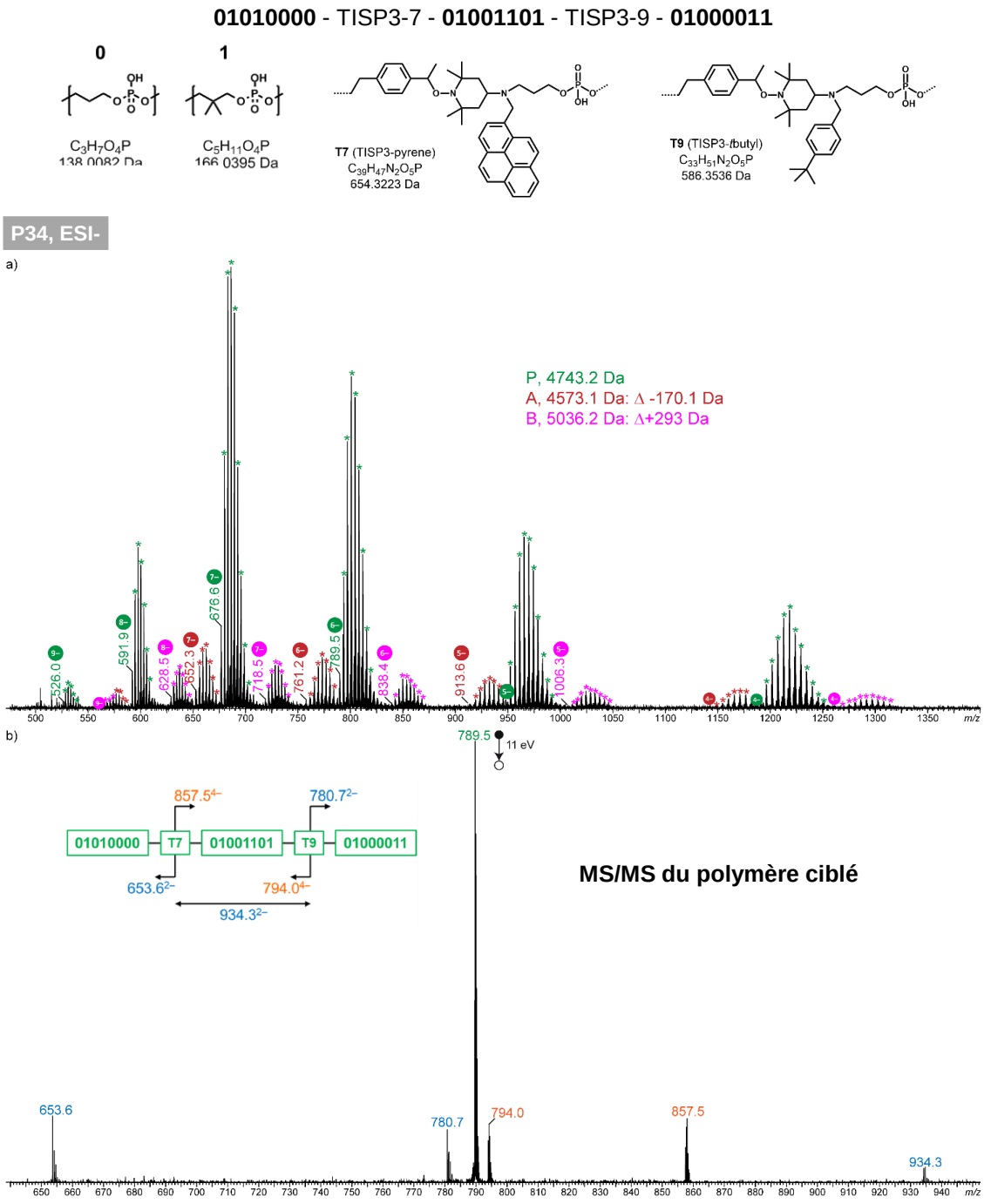
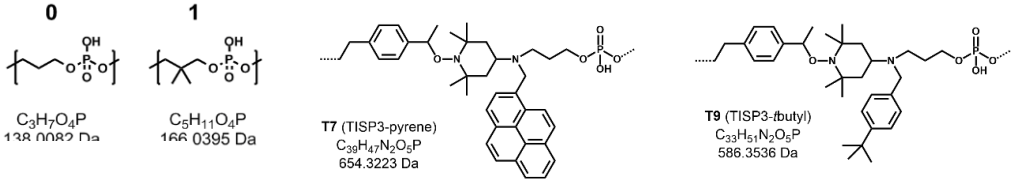


*Figure PE.4. 54 : Spectres MS de l'échantillon **P34**.*
*a) Spectre ESI(–)-MS et b) Spectre MS/MS de l'ion $[P - 6H]^{6-}$ à m/z 789.5.dans l'échantillon **P34**. Les astérisques désignent les échanges H/Na et H/K.*

01010000 - TISP3-7 - **01001101** - TISP3-9 - **01000011**



P34, ESI-MS/MS **(Pseudo-MS³)**

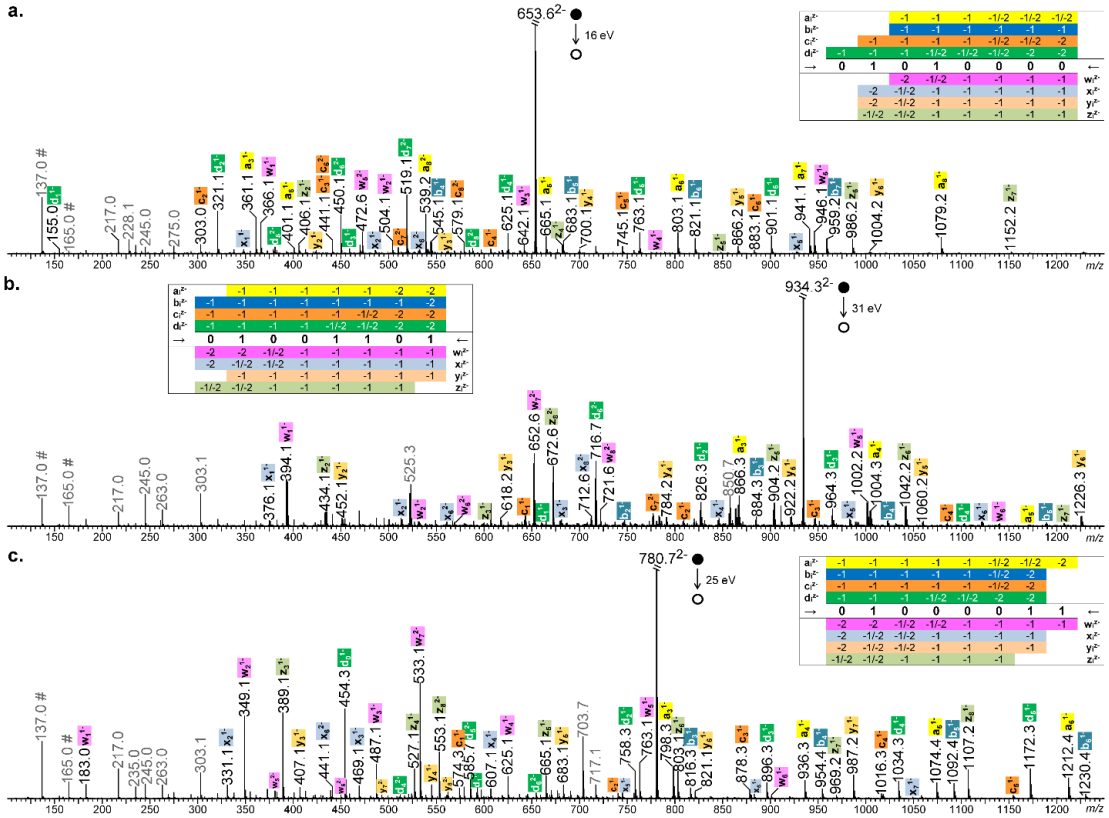


*Figure PE.4. 55 : Séquençage pseudo-MS³ du polymère ciblé dans l'échantillon **P34** (Figure PE.4. 54).*
a) Bloc 1 à m/z 653,6, (b) bloc 2 à m/z 934,3, et (c) bloc 3 à m/z 780,7, avec la couverture de séquence représentée dans des tableaux intégrés. Les pics annotés en gris correspondent à des fragments internes, y compris des unités déprotonées (marquées avec #) à m/z 137,0 pour 0 et à m/z 165,0 pour 1.

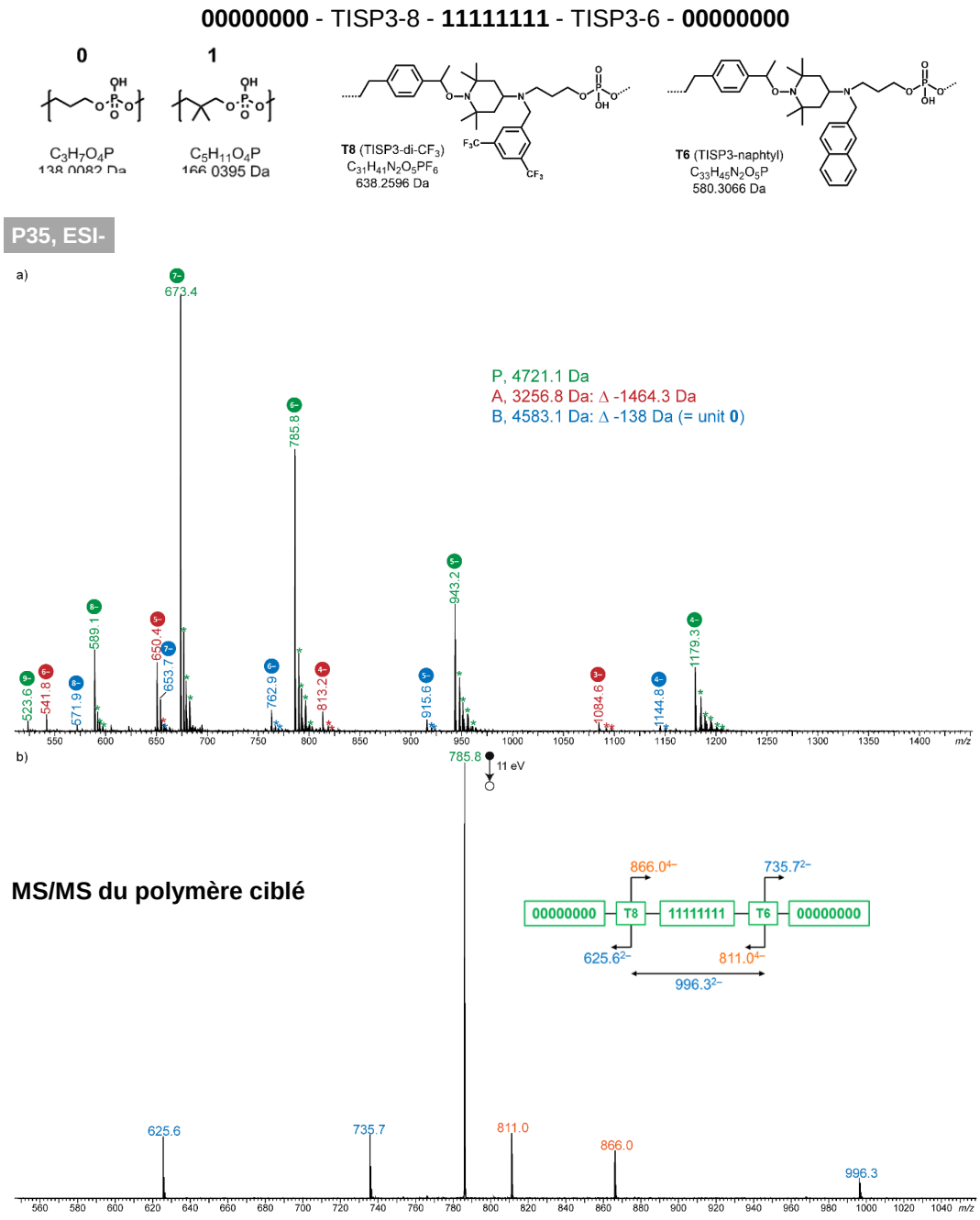


*Figure PE.4. 56 : Spectres MS de l'échantillon **P35**.*
*a) Spectre ESI(–)-MS et b) Spectre MS/MS de l'ion $[P - 6H]^{6-}$ à m/z 785.8 dans l'échantillon **P35**. Les astérisques désignent les échanges H/Na et H/K.*

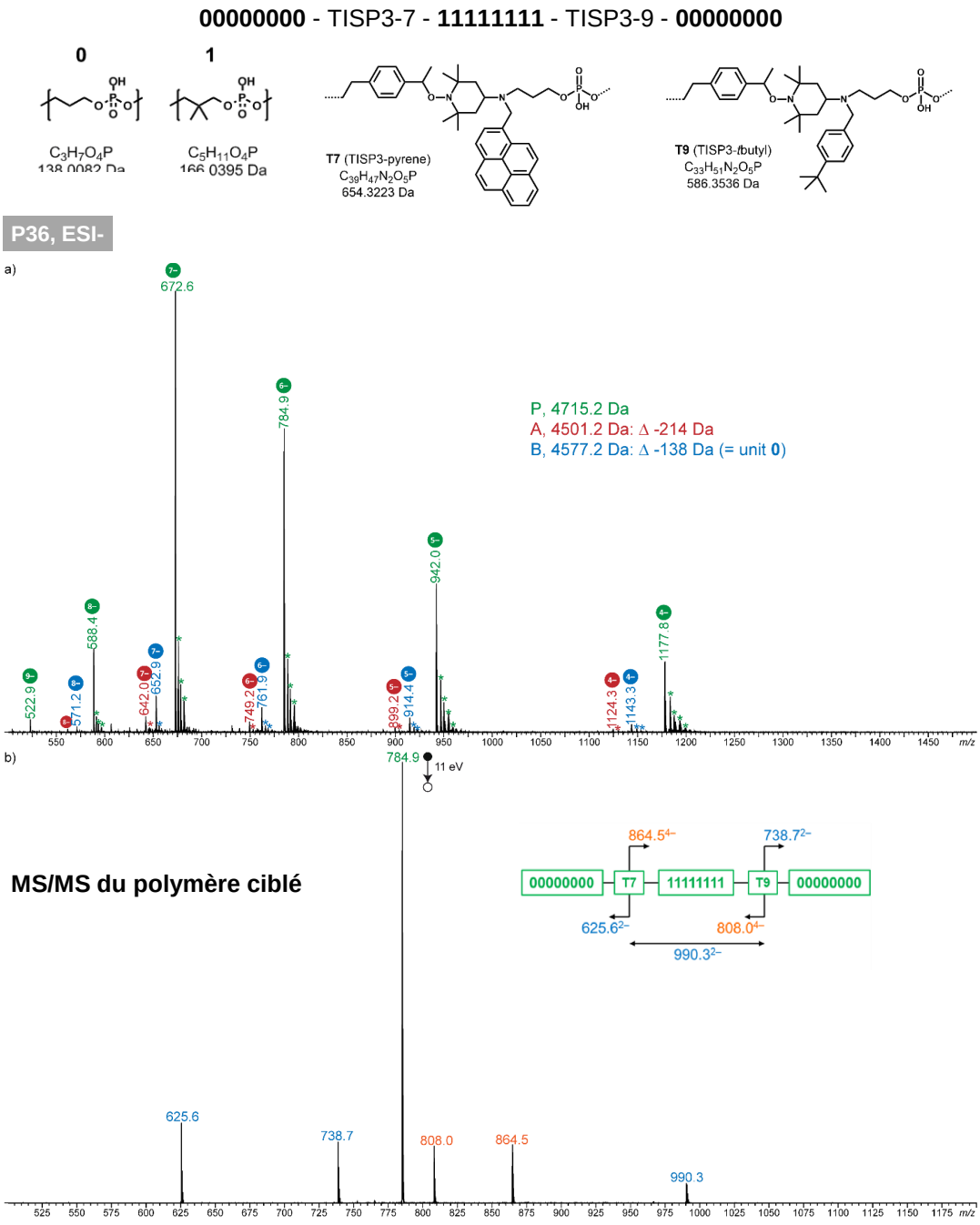
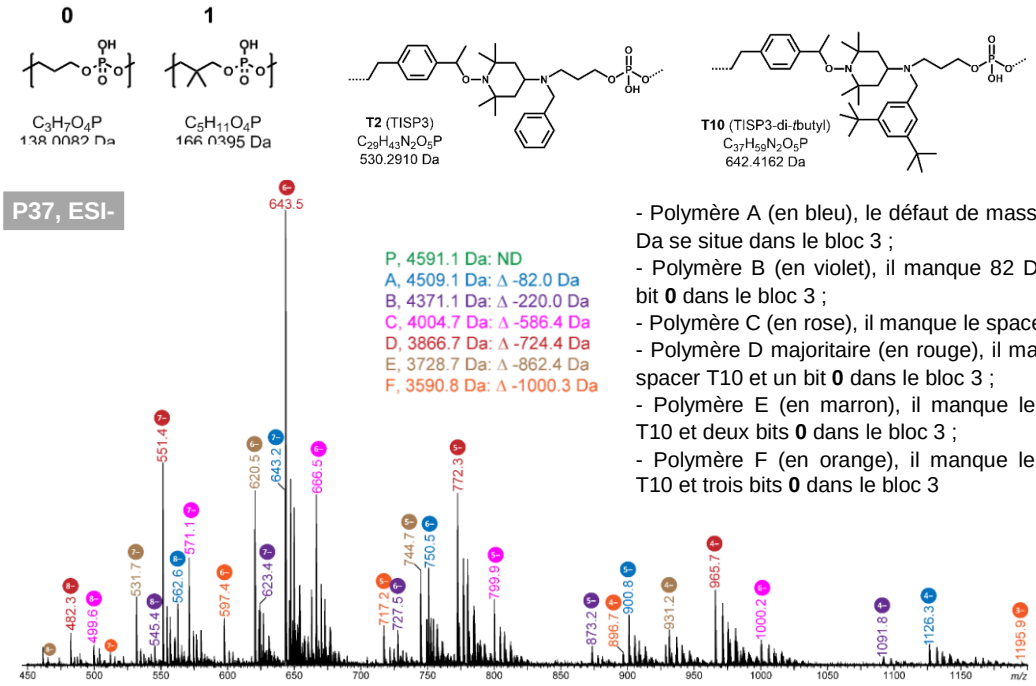


*Figure PE.4. 57 : Spectres MS de l'échantillon **P36**.*
*a) Spectre ESI(–)-MS et b) Spectre MS/MS de l'ion $[P - 6H]^{6-}$ à m/z 784.9 dans l'échantillon **P36**. Les astérisques désignent les échanges H/Na et H/K.*

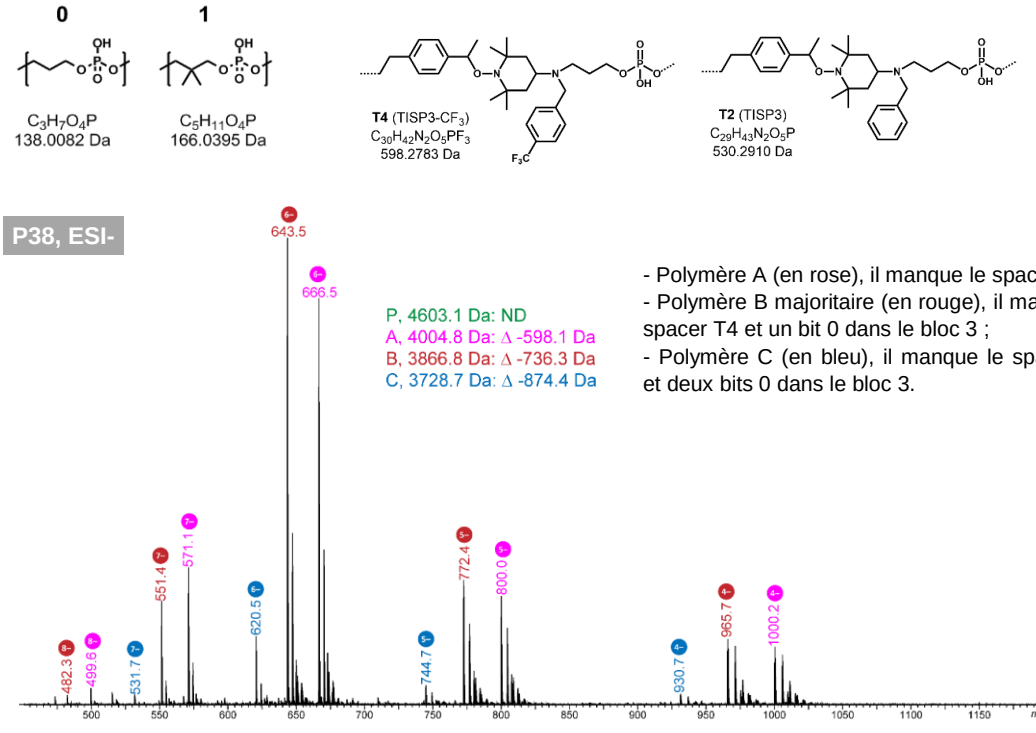
00000000 - TISP3-2 - **11111111** - TISP3-10 - **00000000**



- Polymère A (en bleu), le défaut de masse de 82 Da se situe dans le bloc 3 ;
- Polymère B (en violet), il manque 82 Da et un bit **0** dans le bloc 3 ;
- Polymère C (en rose), il manque le spacer T10 ;
- Polymère D majoritaire (en rouge), il manque le spacer T10 et un bit **0** dans le bloc 3 ;
- Polymère E (en marron), il manque le spacer T10 et deux bits **0** dans le bloc 3 ;
- Polymère F (en orange), il manque le spacer T10 et trois bits **0** dans le bloc 3

*Figure PE.4. 58 : Spectre ESI(–)-MS de l'échantillon **P37**.*
Les impuretés observées dans l'échantillon sont répertoriées en inset.

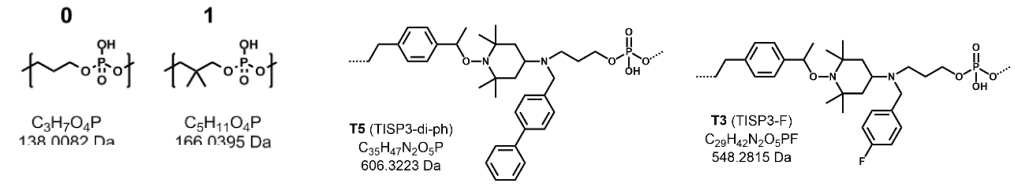
00000000 - TISP3-4 - **11111111** - TISP3-2 - **00000000**



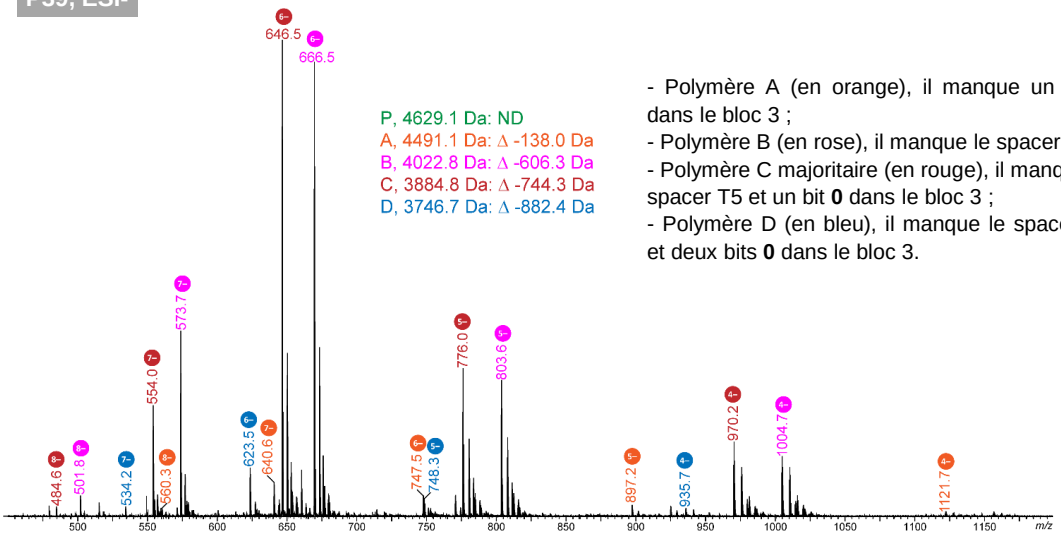
- Polymère A (en rose), il manque le spacer T4 ;
- Polymère B majoritaire (en rouge), il manque le spacer T4 et un bit 0 dans le bloc 3 ;
- Polymère C (en bleu), il manque le spacer T4 et deux bits 0 dans le bloc 3.

*Figure PE.4. 59 : Spectre ESI(–)-MS de l'échantillon **P38**.*
Les impuretés observées dans l'échantillon sont répertoriées en inset.

00000000 - TISP3-5 - **11111111** - TISP3-3 - **00000000**



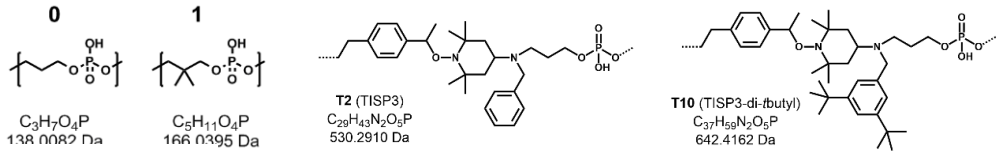
P39, ESI-



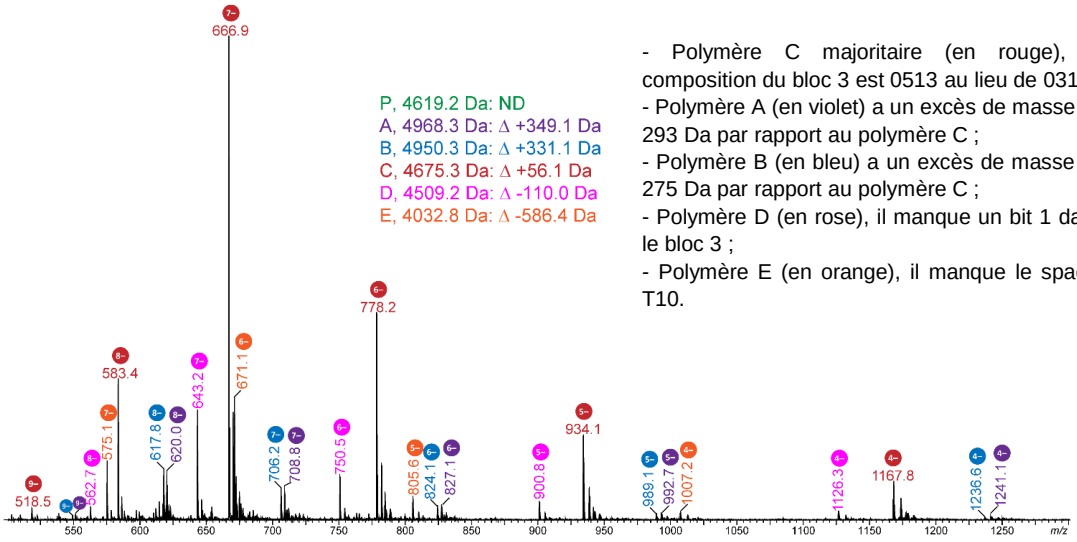
- Polymère A (en orange), il manque un bit **0** dans le bloc 3 ;
- Polymère B (en rose), il manque le spacer T5 ;
- Polymère C majoritaire (en rouge), il manque le spacer T5 et un bit **0** dans le bloc 3 ;
- Polymère D (en bleu), il manque le spacer T5 et deux bits **0** dans le bloc 3.

*Figure PE.4. 60 : Spectre ESI(–)-MS de l'échantillon **P39**.*
Les impuretés observées dans l'échantillon sont répertoriées en inset.

01010000 - TISP3-2 - **01001101** - TISP3-10 - **01000011**



P40, ESI-



- Polymère C majoritaire (en rouge), la composition du bloc 3 est 0513 au lieu de 0315 ;
- Polymère A (en violet) a un excès de masse de 293 Da par rapport au polymère C ;
- Polymère B (en bleu) a un excès de masse de 275 Da par rapport au polymère C ;
- Polymère D (en rose), il manque un bit 1 dans le bloc 3 ;
- Polymère E (en orange), il manque le spacer T10.

*Figure PE.4. 61 : Spectre ESI(–)-MS de l'échantillon **P40**.*
Les impuretés observées dans l'échantillon sont répertoriées en inset.

01010000 - TISP3-4 - **01001101** - TISP3-2 - **01000011**

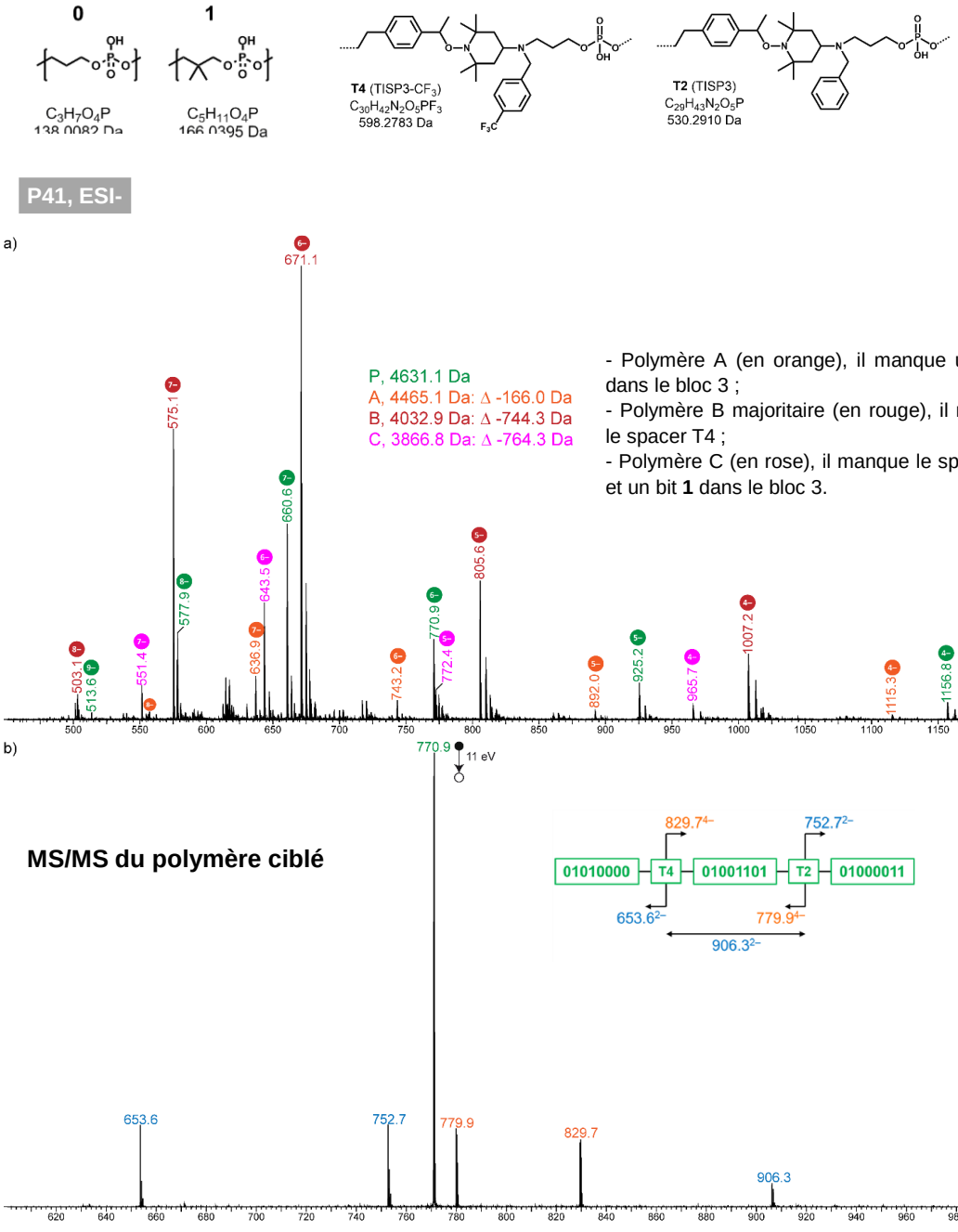


*Figure PE.4. 62 : Spectres MS de l'échantillon **P41**.*
*a) Spectre ESI(–)-MS et b) Spectre MS/MS de l'ion $[P - 6H]^{6-}$ à m/z 770.9 dans l'échantillon **P41**. Les impuretés observées dans l'échantillon sont répertoriées en inset.*

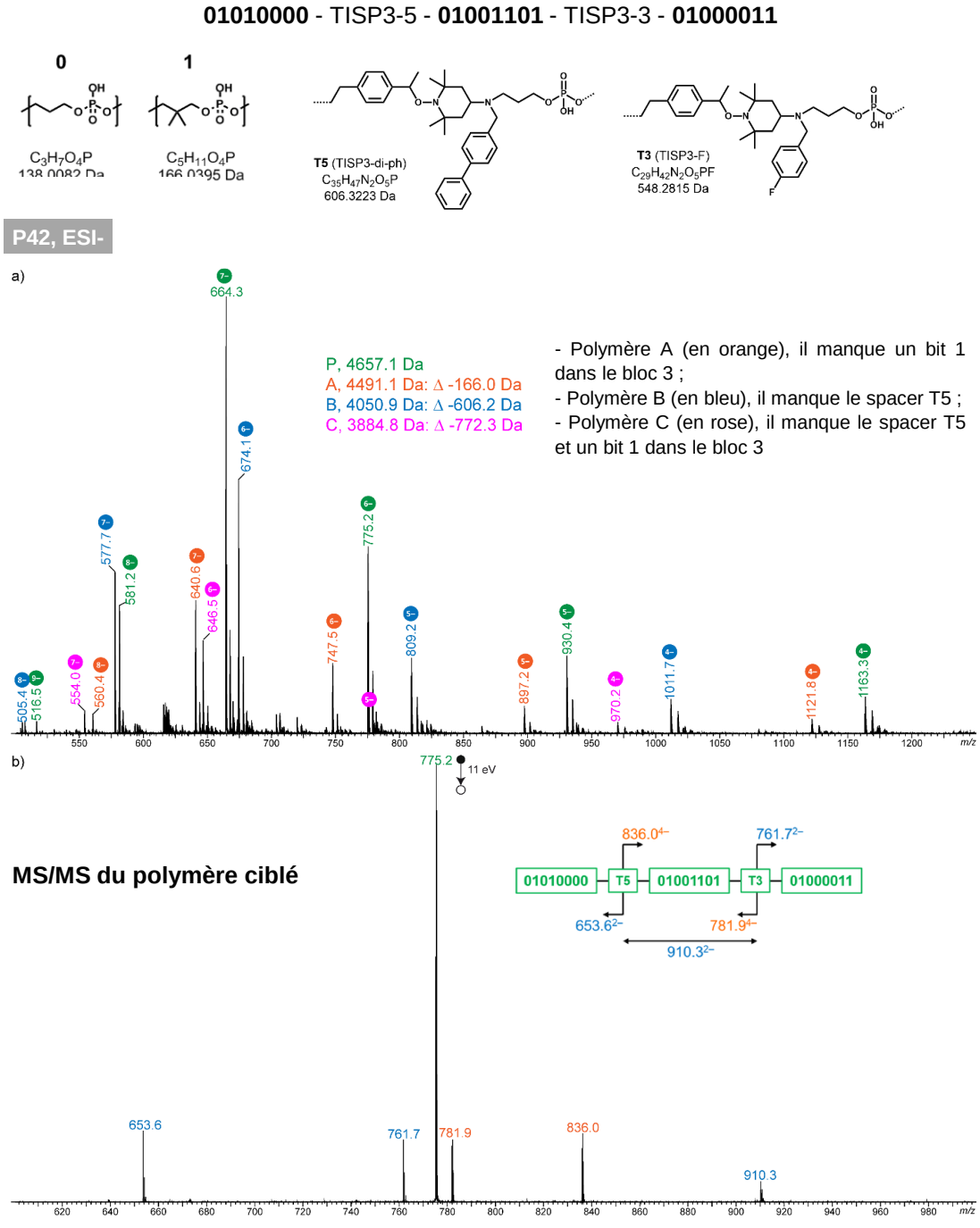
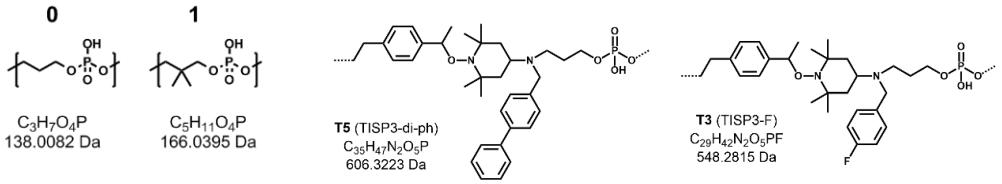


*Figure PE.4. 63 : Spectres MS de l'échantillon **P42**.*
*a) Spectre ESI(–)-MS et b) Spectre MS/MS de l'ion $[P - 6H]^{6-}$ à m/z 775.2 dans l'échantillon **P42**. Les impuretés observées dans l'échantillon sont répertoriées en inset.*

00000000 - TISP3-5 - **11111111** - TISP3-3 - **00000000**



P43, ESI-

- Polymère A majoritaire (en rouge), il manque un bit 0 dans le bloc 3 ;
- Polymère B (en rose), il manque deux bits 0 dans le bloc 3 ;
- Polymère C (en orange), il manque trois bits 0 dans le bloc 3.

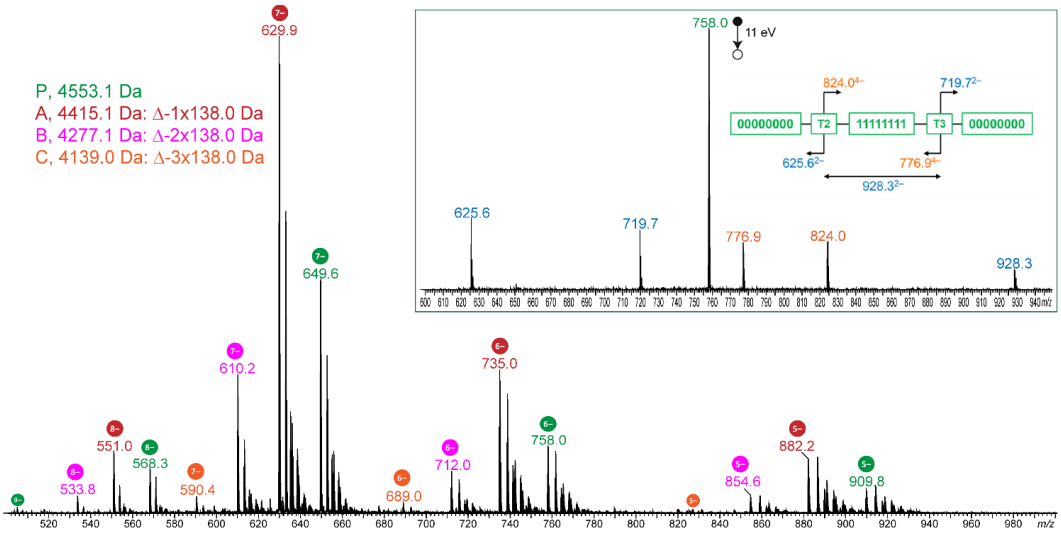


*Figure PE.4. 64 : Spectres MS de l'échantillon **P43**.*
*a) Spectre ESI(–)-MS et Spectre MS/MS de l'ion $[P - 6H]^{6-}$ à m/z 758 dans l'échantillon **P43**. Les impuretés observées dans l'échantillon sont répertoriées en inset.*

5. Spectres UV (efficacité de couplage)

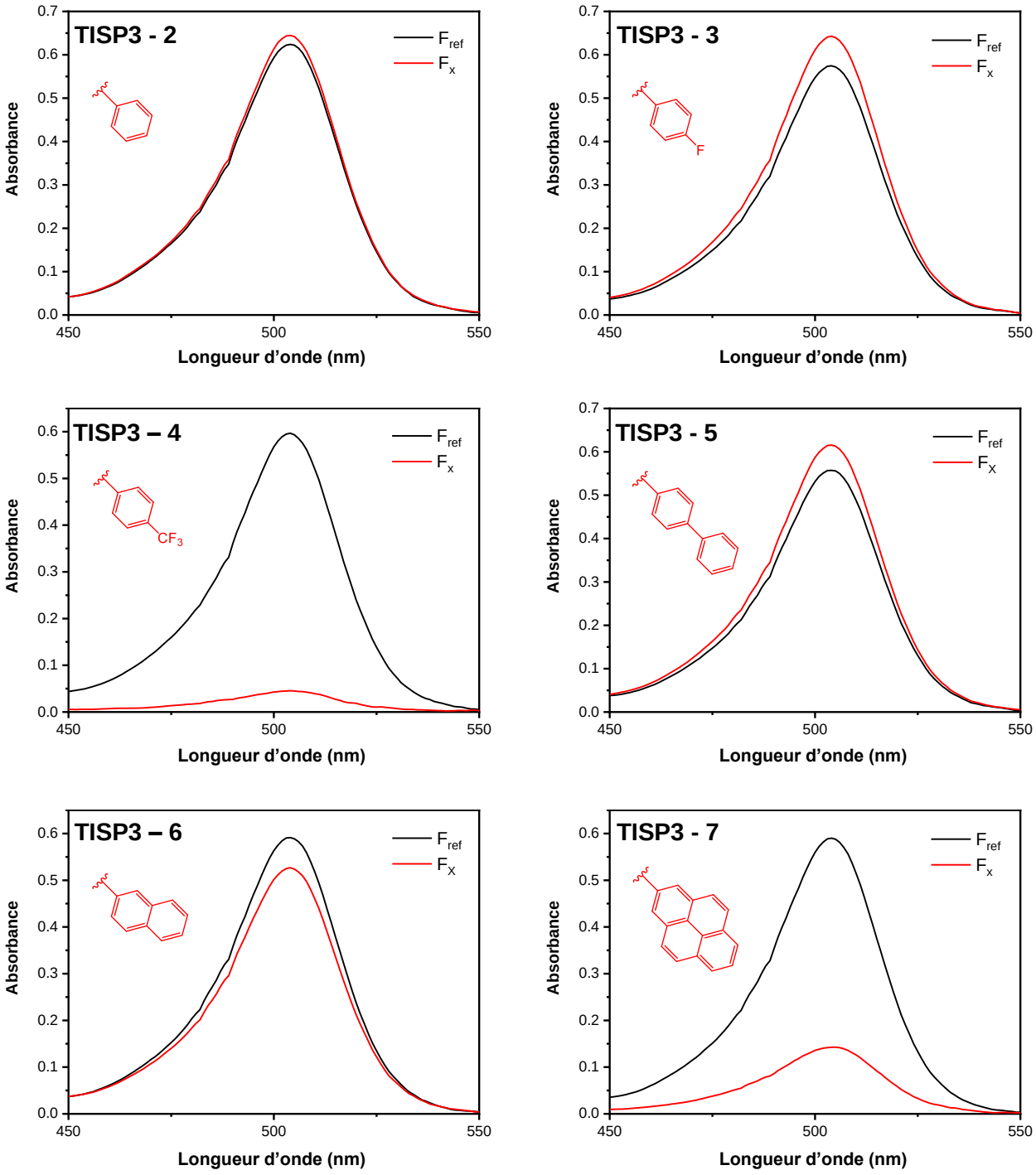


*Figure.PE.5. 1 : Spectres UV enregistrés dans du CH_2Cl_2 pour les fractions de DMT déprotégées lors de la synthèse d'oligomères incorporant **TISP3**.*
*La séquence T1010-x-1010, où x représente un macromonomère de la série **TISP3 - 2-10**. Fx et Fref désignent les fractions de DMT collectées après la liaison du macromonomère **x** et après la liaison du monomère précédent 0, respectivement. **TISP3-2, TISP3-3, TISP3-4, TISP3-5, TISP3-6 et TISP3-7** ont tous été testés à une concentration de 0.1 $mol \cdot l^{-1}$.*

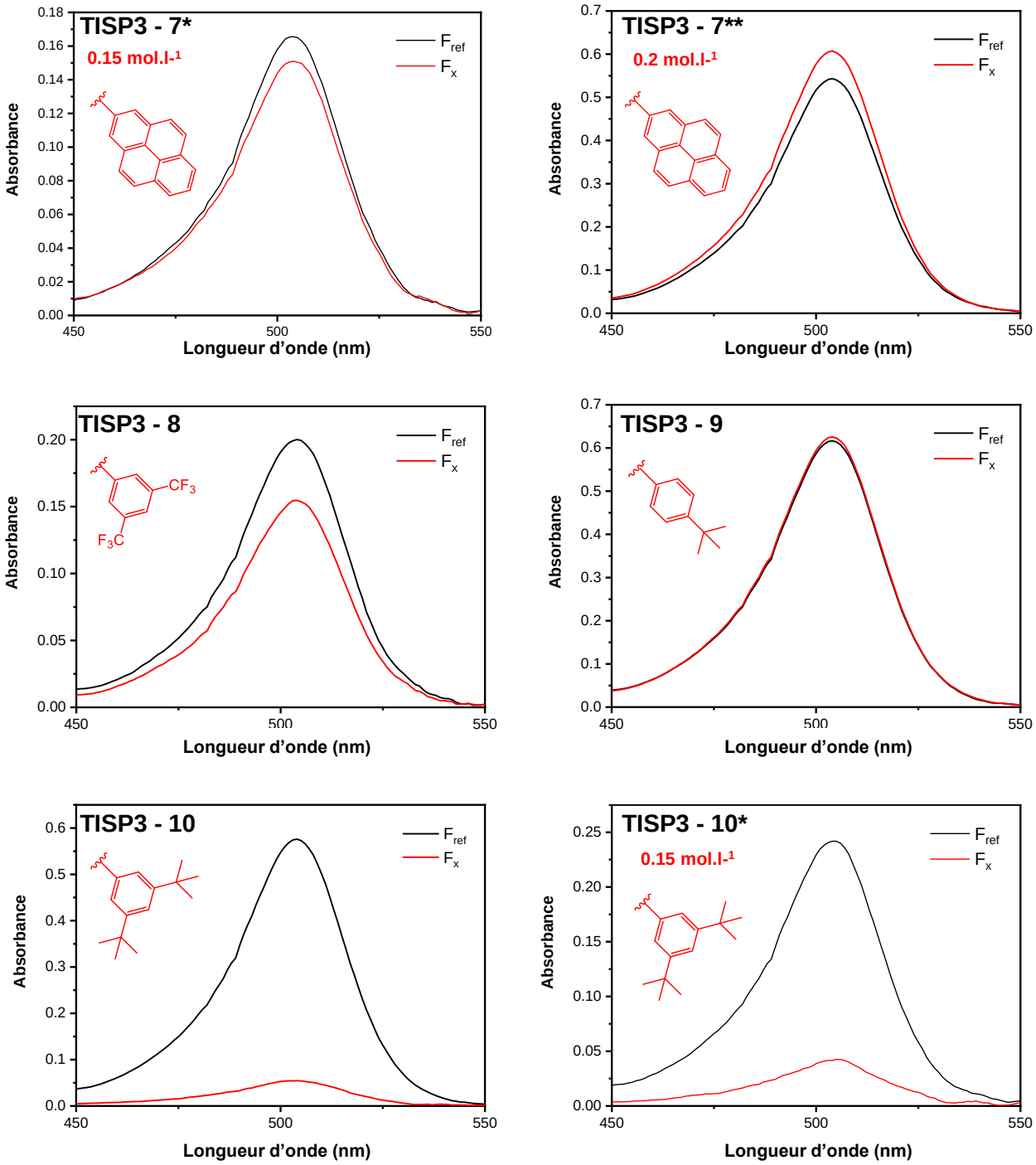


*Figure.PE.5. 2 : Spectres UV enregistrés dans du CH_2Cl_2 pour les fractions de DMT déprotégées lors de la synthèse d'oligomères **TISP3**.*
*La séquence T1010-x-1010, où **x** représente un macromonomère de la série **TISP3 - 2-10**. Fx et Fref désignent les fractions de DMT collectées après la liaison du macromonomère **x** et après la liaison du monomère précédent 0, respectivement. **TISP3-8, TISP3-9** et **TISP3-10** ont été testés à une concentration de 0.1 mol·l^{-1}, **TISP3-7*** et **TISP3-10*** à 0.15 mol·l^{-1}, et **TISP3-7**** à 0.2 mol·l^{-1}.*

6. Chromatogrammes HPLC

En plus de la spectrométrie de masse, l'uniformité des macromolécules numériques contenant les nouveaux Espaceurs ***TISP3*** dans la chaîne polymère a pu être étudiée par HPLC. Cependant, dans le cas des nouveaux support solides universels utilisés, les résultats obtenus étaient moins pertinents, comparativement aux supports chargés de bases nucléiques hautement sensibles aux UV. En effet, avec un bout de chaîne α correspondant à une fonction alcool, les spectres HPLC montraient des signaux larges patatoïdes, difficiles à interpréter. Dépendant des marqueurs utilisés, certains groupes aromatiques comme le ***TISP3*-6** (naphtyle) ou le ***TISP3*-7** (pyrènyle) semblaient présenter des résultats cohérents aux spectres de masses enregistrés, ce qui n'était pas toujours le cas pour les autres dérivés benzyles. De ce fait, pour les recherches effectuées dans cette thèse impliquant le support solide universel, les données HPLC n'ont pas été interprétées. À titre d'exemple, quelques spectres sont donnés pour des échantillons quasi monodisperse discutés dans le Chapitre 3. Par ailleurs, la comparaison des échantillons **P28** et **P29** suggérant une dégradation par la température de clivage est également montrée ci-dessous.

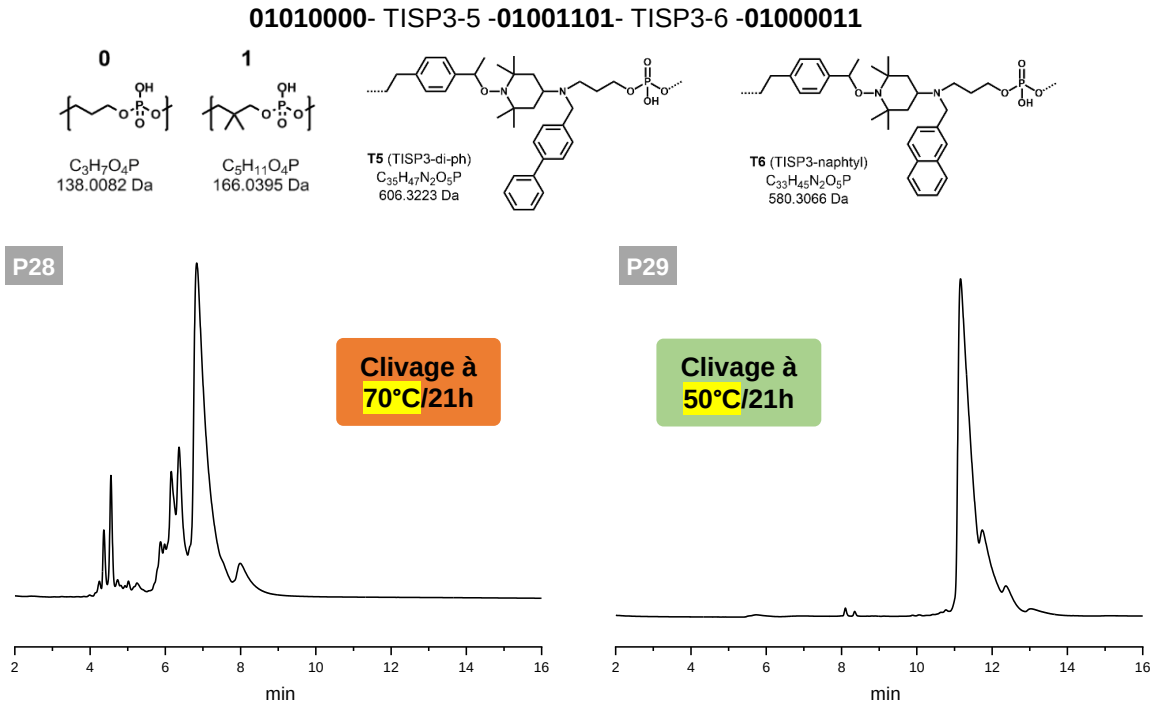


*Figure PE.6. 1 : Spectres HPLC des polymères **P28** et **P29** à différentes températures de clivage.*

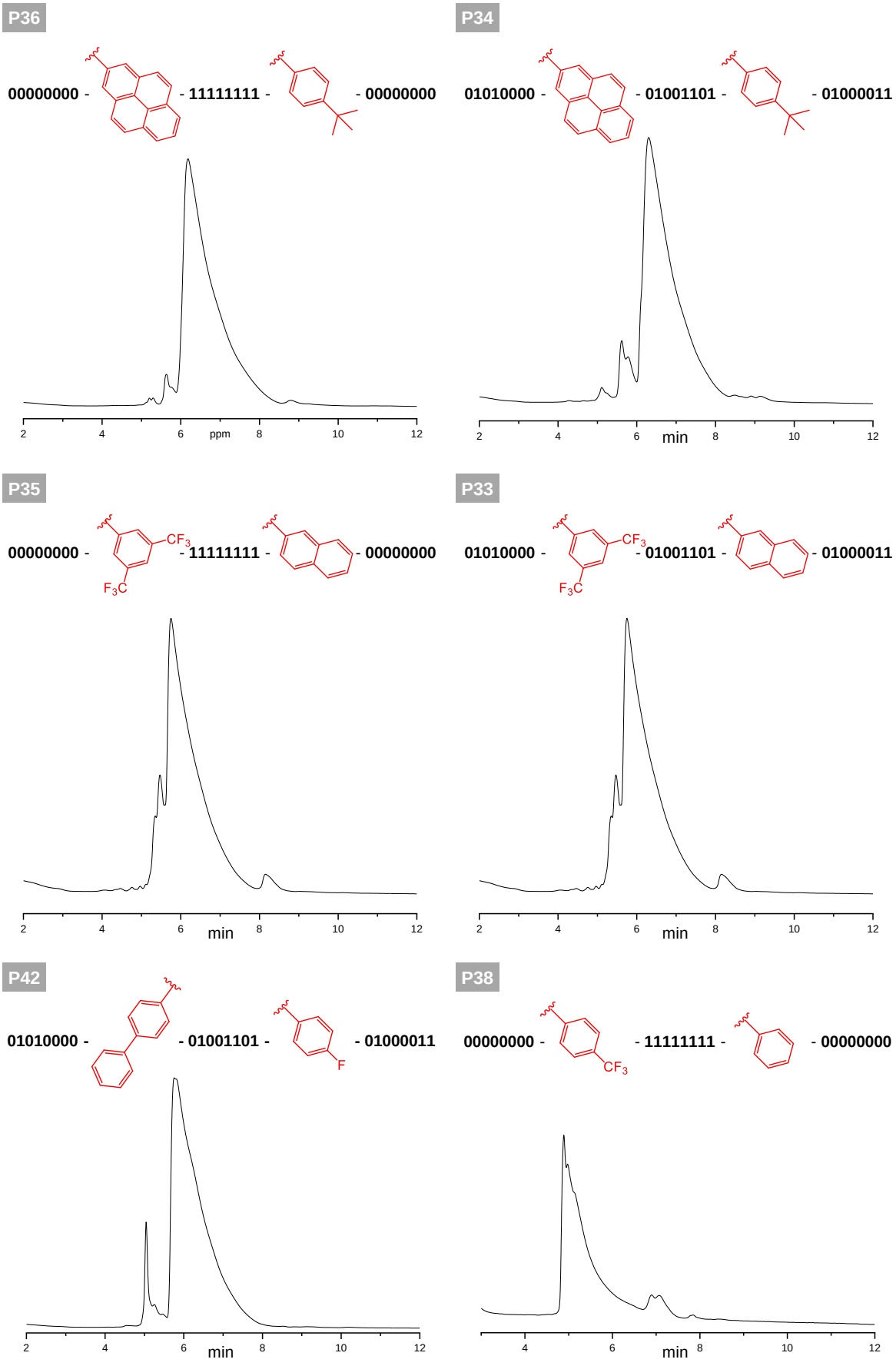


*Figure PE.6. 2 : HPLC des polymères **P33**, **P34**, **P35**, **P36**, **P38** et **P42**.*

Références

[1] D. Reinsel, J. Gantz, J. Rydning, **2018**.
[2] V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church, W. L. Hughes, *Nat. Mater.* **2016**, *15*, 366–370.
[3] N. Jones, *Nature* **2018**, *561*, 163–166.
[4] H. Lee, K. Calvin, D. Dasgupta, G. Krinner, A. Mukherji, P. Thorne, C. Trisos, J. Romero, P. Aldunce, K. Barrett, G. Blanco, W. W. L. Cheung, S. L. Connors, F. Denton, A. Diongue-Niang, D. Dodman, M. Garschagen, O. Geden, B. Hayward, C. Jones, F. Jotzo, T. Krug, R. Lasco, J.-Y. Lee, V. Masson-Delmotte, M. Meinshausen, K. Mintenbeck, A. Mokssit, F. E. L. Otto, M. Pathak, A. Pirani, E. Poloczanska, H.-O. Pörtner, A. Revi, D. C. Roberts, J. Romero, J. Roy, A. C. Ruane, J. Skea, P. R. Shukla, R. Slade, A. Slangen, Y. Sokona, A. A. Sörensson, M. Tignor, D. van Vuuren, Y.-M. Wei, H. Winkler, P. Zhai, Z. Zommers, *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, The Australian National University, **2023**.
[5] M. Gu, Q. Zhang, S. Lamon, *Nat. Rev. Mater.* **2016**, *1*, 1–14.
[6] "Data storage: the DNA revolution," can be found under https://news.cnrs.fr/articles/data-storage-the-dna-revolution, **n.d.**
[7] A. Doricchi, C. M. Platnich, A. Gimpel, F. Horn, M. Earle, G. Lanzavecchia, A. L. Cortajarena, L. M. Liz-Marzán, N. Liu, R. Heckel, R. N. Grass, R. Krahne, U. F. Keyser, D. Garoli, *ACS Nano* **2022**, *16*, 17552–17571.
[8] G. M. Church, Y. Gao, S. Kosuri, *Science* **2012**, *337*, 1628–1628.
[9] W. Kutschera, W. Rom, *Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. At.* **2000**, *164–165*, 12–22.
[10] A. Hoose, R. Vellacott, M. Storch, P. S. Freemont, M. G. Ryadnov, *Nat. Rev. Chem.* **2023**, *7*, 144–161.
[11] M. G. T. A. Rutten, F. W. Vaandrager, J. A. A. W. Elemans, R. J. M. Nolte, *Nat. Rev. Chem.* **2018**, *2*, 365–381.
[12] H. Lu, F. Giordano, Z. Ning, *Genomics Proteomics Bioinformatics* **2016**, *14*, 265–279.
[13] G. Ivády, L. Madar, E. Dzsudzsák, K. Koczok, J. Kappelmayer, V. Krulisova, M. Macek, A. Horváth, I. Balogh, *BMC Genomics* **2018**, *19*, 158.
[14] R. N. Grass, R. Heckel, M. Puddu, D. Paunescu, W. J. Stark, *Angew. Chem. Int. Ed.* **2015**, *54*, 2552–2555.
[15] E. Laurent, R. Szweda, J.-F. Lutz, in *Macromol. Eng.*, John Wiley & Sons, Ltd, **2022**, pp. 1–34.
[16] T. T. Trinh, L. Oswald, D. Chan-Seng, J.-F. Lutz, *Macromol. Rapid Commun.* **2014**, *35*, 141–145.
[17] S. Martens, A. Landuyt, P. Espeel, B. Devreese, P. Dawyndt, F. Du Prez, *Nat. Commun.* **2018**, *9*, 4451.
[18] S. Martens, J. O. Holloway, Filip. E. Du Prez, *Macromol. Rapid Commun.* **2017**, *38*, 1700469.
[19] R. K. Roy, A. Meszynska, C. Laure, L. Charles, C. Verchin, J.-F. Lutz, *Nat. Commun.* **2015**, *6*, 7237.
[20] I. Youssef, S. Samokhvalova, I. Sergent, L. Charles, J.-F. Lutz, *Precis. Chem.* **2023**, DOI 10.1021/prechem.3c00077.
[21] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154.
[22] S. L. Beaucage, R. P. Iyer, *Tetrahedron* **1992**, *48*, 2223–2311.
[23] J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science* **2013**, *341*, DOI 10.1126/science.1238149.
[24] M. H. Caruthers, *Science* **1985**, *230*, 281–285.
[25] A. Al Ouahabi, M. Kotera, L. Charles, J.-F. Lutz, *ACS Macro Lett.* **2015**, *4*, 1077–1080.

[26] E. Laurent, J.-A. Amalian, M. Parmentier, L. Oswald, A. Al Ouahabi, F. Dufour, K. Launay, J.-L. Clément, D. Gigmes, M.-A. Delsuc, L. Charles, J.-F. Lutz, *Macromolecules* **2020**, *53*, 4022–4029.
[27] A. Al Ouahabi, J.-A. Amalian, L. Charles, J.-F. Lutz, *Nat. Commun.* **2017**, *8*, 967.
[28] A. Burel, C. Carapito, J.-F. Lutz, L. Charles, *Macromolecules* **2017**, *50*, 8290–8296.
[29] E. Laurent, J.-A. Amalian, T. Schutz, K. Launay, J.-L. Clément, D. Gigmes, A. Burel, C. Carapito, L. Charles, M.-A. Delsuc, J.-F. Lutz, *Comptes Rendus Chim.* **2021**, *24*, 69–76.
[30] "Du Néolithique au numérique, une histoire de l'information," can be found under http://larevuedesmedias.ina.fr/series/du-neolithique-au-numerique-une-histoire-de-linformation, **n.d.**
[31] V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church, W. L. Hughes, *Nat. Mater.* **2016**, *15*, 366–370.
[32] J. C. K. Lee, Z. Wen, *Nat. Sustain.* **2018**, *1*, 598–605.
[33] T. Cheisson, E. J. Schelter, *Science* **2019**, *363*, 489–493.
[34] B. Deng, X. Wang, D. X. Luong, R. A. Carter, Z. Wang, M. B. Tomson, J. M. Tour, *Sci. Adv.* **2022**, *8*, eabm3132.
[35] N. t. Nassar, X. Du, T. e. Graedel, *J. Ind. Ecol.* **2015**, *19*, 1044–1054.
[36] M. Tyrer, J. P. Sykes, *Significance* **2013**, *10*, 12–16.
[37] J. C. K. Lee, Z. Wen, *J. Ind. Ecol.* **2017**, *21*, 1277–1290.
[38] S. Langkau, M. Erdmann, *J. Ind. Ecol.* **2021**, *25*, 1034–1050.
[39] M. A. B. Siddik, A. Shehabi, L. Marston, *Environ. Res. Lett.* **2021**, *16*, 064017.
[40] "Rankings," can be found under https://cloudscene.com/rankings/leaderboard, **n.d.**
[41] "How a Hard Drive Works," can be found under https://cs.stanford.edu/people/nick/how-hard-drive-works/, **n.d.**
[42] S. M. Yakout, *J. Supercond. Nov. Magn.* **2020**, *33*, 2557–2580.
[43] A. Aoyagi, B. Stipe, R. Wood, S. Campbell, X. Che, *J. Magn. Magn. Mater.* **2022**, *564*, 170146.
[44] Z. Hu, X. Huang, Z. Yang, J. Qiu, Z. Song, J. Zhang, G. Dong, *Light Sci. Appl.* **2021**, *10*, 140.
[45] M. Watanabe, S. Juodkazis, H.-B. Sun, S. Matsuo, H. Misawa, M. Miwa, R. Kaneko, *Appl. Phys. Lett.* **1999**, *74*, 3957–3959.
[46] P. Zijlstra, J. W. M. Chon, M. Gu, *Nature* **2009**, *459*, 410–413.
[47] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.
[48] K. Matange, J. M. Tuck, A. J. Keung, *Nat. Commun.* **2021**, *12*, 1358.
[49] E. Podivinsky, J. L. Love, L. van der Colff, L. Samuel, *Anal. Biochem.* **2009**, *394*, 132–134.
[50] N. V. Ivanova, M. L. Kuzmina, *Mol. Ecol. Resour.* **2013**, *13*, 890–898.
[51] S. Shah, J. G. Elerath, in *Annu. Symp. Reliab. Maintainab. 2004 - RAMS*, **2004**, pp. 163–167.
[52] A. Keller, A. Graefen, M. Ball, M. Matzas, V. Boisguerin, F. Maixner, P. Leidinger, C. Backes, R. Khairat, M. Forster, B. Stade, A. Franke, J. Mayer, J. Spangler, S. McLaughlin, M. Shah, C. Lee, T. T. Harkins, A. Sartori, A. Moreno-Estrada, B. Henn, M. Sikora, O. Semino, J. Chiaroni, S. Rootsi, N. M. Myres, V. M. Cabrera, P. A. Underhill, C. D. Bustamante, E. E. Vigl, M. Samadelli, G. Cipollini, J. Haas, H. Katus, B. D. O'Connor, M. R. J. Carlson, B. Meder, N. Blin, E. Meese, C. M. Pusch, A. Zink, *Nat. Commun.* **2012**, *3*, 698.
[53] J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. García, S. Pääbo, J.-L. Arsuaga, M. Meyer, *Proc. Natl. Acad. Sci.* **2013**, *110*, 15758–15763.

[54] R. P. Feynman, *J. Microelectromechanical Syst.* **1992**, *1*, 60–66.
[55] J. Davis, *Art J.* **1996**, *55*, 70–74.
[56] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, E. Birney, *Nature* **2013**, *494*, 77–80.
[57] L. Organick, S. D. Ang, Y.-J. Chen, R. Lopez, S. Yekhanin, K. Makarychev, M. Z. Racz, G. Kamath, P. Gopalan, B. Nguyen, C. N. Takahashi, S. Newman, H.-Y. Parker, C. Rashtchian, K. Stewart, G. Gupta, R. Carlson, J. Mulligan, D. Carmean, G. Seelig, L. Ceze, K. Strauss, *Nat. Biotechnol.* **2018**, *36*, 242–248.
[58] "Le poids colossal des 'données noires' dans les entreprises," can be found under https://www.bfmtv.com/economie/entreprises/le-poids-colossal-des-donnees-noires-dans-les-entreprises_AN-202301150137.html, **n.d.**
[59] S. Kosuri, G. M. Church, *Nat. Methods* **2014**, *11*, 499–507.
[60] R. E. Franklin, R. G. Gosling, *Acta Crystallogr.* **1953**, *6*, 673–677.
[61] M. H. F. Wilkins, A. R. Stokes, H. R. Wilson, *Nature* **1953**, *171*, 738–740.
[62] M. Meselson, F. W. Stahl, *Proc. Natl. Acad. Sci.* **1958**, *44*, 671–682.
[63] B. Stillman, *Nat. Med.* **2002**, *8*, 211–211.
[64] I. R. Lehman, M. J. Bessman, E. S. Simms, A. Kornberg, *J. Biol. Chem.* **1958**, *233*, 163–170.
[65] "Cells Can Replicate Their DNA Precisely | Learn Science at Scitable," can be found under http://www.nature.com/scitable/topicpage/cells-can-replicate-their-dna-precisely-6524830, **n.d.**
[66] R. Okazaki, T. Okazaki, K. Sakabe, K. Sugimoto, A. Sugino, *Proc. Natl. Acad. Sci.* **1968**, *59*, 598–605.
[67] R. K. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, H. A. Erlich, *Science* **1988**, *239*, 487–491.
[68] K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, H. Erlich, *Cold Spring Harb. Symp. Quant. Biol.* **1986**, *51*, 263–273.
[69] "The Nobel Prize in Chemistry 1993," can be found under https://www.nobelprize.org/prizes/chemistry/1993/mullis/facts/, **n.d.**
[70] F. J. Bollum, *J. Biol. Chem.* **1959**, *234*, 2733–2734.
[71] M. A. Jensen, R. W. Davis, *Biochemistry* **2018**, *57*, 1821–1832.
[72] *DNA Scr.* **n.d.**
[73] "Technology," can be found under https://ansabio.com/technology, **n.d.**
[74] I. Sarac, M. Hollenstein, *ChemBioChem* **2019**, *20*, 860–871.
[75] A. S. Mathews, H. Yang, C. Montemagno, *Org. Biomol. Chem.* **2016**, *14*, 8278–8288.
[76] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, *Nat. Methods* **2009**, *6*, 343–345.
[77] W. P. C. Stemmer, A. Crameri, K. D. Ha, T. M. Brennan, H. L. Heyneker, *Gene* **1995**, *164*, 49–53.
[78] M. G. Mohsen, E. T. Kool, *Acc. Chem. Res.* **2016**, *49*, 2540–2550.
[79] M. M. Ali, F. Li, Z. Zhang, K. Zhang, D.-K. Kang, J. A. Ankrum, X. C. Le, W. Zhao, *Chem. Soc. Rev.* **2014**, *43*, 3324–3341.
[80] D. Verardo, B. Adelizzi, D. A. Rodriguez-Pinzon, N. Moghaddam, E. Thomée, T. Loman, X. Godron, A. Horgan, *Sci. Adv.* **2023**, *9*, eadi0263.
[81] M. A. Cleary, K. Kilian, Y. Wang, J. Bradshaw, G. Cavet, W. Ge, A. Kulkarni, P. J. Paddison, K. Chang, N. Sheth, E. Leproust, E. M. Coffey, J. Burchard, W. R. McCombie, P. Linsley, G. J. Hannon, *Nat. Methods* **2004**, *1*, 241–248.
[82] S. Hecko, A. Schiefer, C. P. S. Badenhorst, M. J. Fink, M. D. Mihovilovic, U. T. Bornscheuer, F. Rudroff, *Chem. Rev.* **2023**, *123*, 2832–2901.
[83] J. Cello, A. V. Paul, E. Wimmer, *Science* **2002**, *297*, 1016–1018.

[84] L. Ceze, J. Nivala, K. Strauss, *Nat. Rev. Genet.* **2019**, *20*, 456–466.
[85] L. C. Meiser, B. H. Nguyen, Y.-J. Chen, J. Nivala, K. Strauss, L. Ceze, R. N. Grass, *Nat. Commun.* **2022**, *13*, 352.
[86] S. Brenner, F. Jacob, M. Meselson, *Nature* **1961**, *190*, 576–581.
[87] P. J. Gomatos, I. Tamm, *Proc. Natl. Acad. Sci.* **1963**, *49*, 707–714.
[88] Y. G. Chen, S. Hur, *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 286–301.
[89] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, in *Mol. Biol. Cell 4th Ed.*, Garland Science, **2002**.
[90] "ribonucleic acid / RNA | Learn Science at Scitable," can be found under http://www.nature.com/scitable/definition/ribonucleic-acid-rna-45, **n.d.**
[91] F. X. Chen, E. R. Smith, A. Shilatifard, *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 464–478.
[92] S. Samokhvalova, J.-F. Lutz, *Angew. Chem. Int. Ed.* **2023**, *62*, e202300014.
[93] P. D. Sun, C. E. Foster, J. C. Boyington, *Curr. Protoc. Protein Sci.* **2004**, *35*, 1711–171189.
[94] C. B. Anfinsen, *Science* **1973**, *181*, 223–230.
[95] L. Pauling, R. B. Corey, H. R. Branson, *Proc. Natl. Acad. Sci.* **1951**, *37*, 205–211.
[96] B. Kuhlman, P. Bradley, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 681–697.
[97] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, in *Mol. Biol. Cell 4th Ed.*, Garland Science, **2002**.
[98] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, in *Mol. Biol. Cell 4th Ed.*, Garland Science, **2002**.
[99] S. L. Rössler, N. M. Grob, S. L. Buchwald, B. L. Pentelute, *Science* **2023**, *379*, 939–945.
[100] C. C. A. Ng, W. M. Tam, H. Yin, Q. Wu, P.-K. So, M. Y.-M. Wong, F. C. M. Lau, Z.-P. Yao, *Nat. Commun.* **2021**, *12*, 4242.
[101] C. C. A. Ng, W. M. Tam, H. Yin, Q. Wu, P.-K. So, M. Y.-M. Wong, F. C. M. Lau, Z.-P. Yao, *Nat. Commun.* **2021**, *12*, 4242.
[102] E. Fischer, *Berichte Dtsch. Chem. Ges.* **1903**, *36*, 2982–2992.
[103] M. Bergmann, L. Zervas, *Berichte Dtsch. Chem. Ges. B Ser.* **1932**, *65*, 1192–1201.
[104] R. B. Merrifield, *Angew. Chem. Int. Ed. Engl.* **1985**, *24*, 799–810.
[105] V. du Vigneaud, C. Ressler, C. J. M. Swan, C. W. Roberts, P. G. Katsoyannis, S. Gordon, *J. Am. Chem. Soc.* **1953**, *75*, 4879–4880.
[106] K. Hofmann, T. Liu, H. Yajima, N. Yanaihara, S. Lande, *J. Am. Chem. Soc.* **1961**, *83*, 2294–2299.
[107] S. Kent, *Nature* **2006**, *441*, 824–824.
[108] B. Gutte, R. B. Merrifield, *J. Biol. Chem.* **1971**, *246*, 1922–1941.
[109] R. A. Houghten, *Proc. Natl. Acad. Sci.* **1985**, *82*, 5131–5135.
[110] B. Merrifield, *Science* **1986**, *232*, 341–347.
[111] E. Atherton, C. J. Logan, R. C. Sheppard, *J. Chem. Soc. Perkin 1* **1981**, 538–546.
[112] C.-D. Chang, J. Meienhofer, *Int. J. Pept. Protein Res.* **1978**, *11*, 246–249.
[113] N. ten Brummelhuis, P. Wilke, H. G. Börner, in *Seq.-Control. Polym.*, John Wiley & Sons, Ltd, **2018**, pp. 117–158.
[114] R. B. Merrifield, *Science* **1965**, *150*, 178–185.
[115] S. L. Pedersen, K. J. Jensen, in *Pept. Synth. Appl.* (Eds.: K.J. Jensen, P. Tofteng Shelton, S.L. Pedersen), Humana Press, Totowa, NJ, **2013**, pp. 215–224.
[116] M. D. Simon, A. J. Mijalis, K. A. Totaro, D. Dunkelmann, A. A. Vinogradov, C. Zhang, Y. Maki, J. M. Wolfe, J. Wilson, A. Loas, B. L. Pentelute, in *Total Chem. Synth. Proteins*, John Wiley & Sons, Ltd, **2021**, pp. 17–57.
[117] G. Sabatino, A. D'Ercole, L. Pacini, M. Zini, A. Ribecai, A. Paio, P. Rovero, A. M. Papini, *Org. Process Res. Dev.* **2021**, *25*, 552–563.

[118] J. M. Collins, in *Microw. Org. Synth.*, John Wiley & Sons, Ltd, **2012**, pp. 897–959.
[119] P. H. Seeberger, W.-C. Haase, *Chem. Rev.* **2000**, *100*, 4349–4394.
[120] O. J. Plante, E. R. Palmacci, P. H. Seeberger, *Science* **2001**, *291*, 1523–1527.
[121] M. Weishaupt, S. Eller, P. H. Seeberger, in *Methods Enzymol.* (Ed.: M. Fukuda), Academic Press, **2010**, pp. 463–484.
[122] J.-F. Lutz, H. G. Börner, *Prog. Polym. Sci.* **2008**, *33*, 1–39.
[123] H. G. Börner, *Prog. Polym. Sci.* **2009**, *34*, 811–851.
[124] S. C. Solleder, R. V. Schneider, K. S. Wetzel, A. C. Boukis, M. A. R. Meier, *Macromol. Rapid Commun.* **2017**, *38*, 1600711.
[125] R. Eritja, *Int. J. Pept. Res. Ther.* **2007**, *13*, 53–68.
[126] B. Lewandowski, G. De Bo, J. W. Ward, M. Papmeyer, S. Kuschel, M. J. Aldegunde, P. M. E. Gramlich, D. Heckmann, S. M. Goldup, D. M. D'Souza, A. E. Fernandes, D. A. Leigh, *Science* **2013**, *339*, 189–193.
[127] A. M. Michelson, A. R. Todd, *J. Chem. Soc. Resumed* **1955**, 2632–2638.
[128] M. Caruthers, R. Wells, *Science* **2011**, *334*, 1511–1511.
[129] H. G. Khorana, G. M. Tener, J. G. Moffatt, E. H. Pol, *Chem Ind* **1956**, 1523.
[130] H. G. Khorana, W. E. Razzell, P. T. Gilham, G. M. Tener, E. H. Pol, *J. Am. Chem. Soc.* **1957**, *79*, 1002–1003.
[131] M. Smith, J. G. Moffatt, H. G. Khorana, *J. Am. Chem. Soc.* **1958**, *80*, 6204–6212.
[132] P. T. Gilham, H. G. Khorana, *J. Am. Chem. Soc.* **1959**, *81*, 4647–4650.
[133] H. Schaller, G. Weimann, B. Lerch, H. G. Khorana, *J. Am. Chem. Soc.* **1963**, *85*, 3821–3827.
[134] M. Smith, D. H. Rammler, I. H. Goldberg, H. G. Khorana, *J. Am. Chem. Soc.* **1962**, *84*, 430–440.
[135] H. G. Khorana, K. L. Agarwal, H. Büchi, M. H. Caruthers, N. K. Gupta, K. Klbppe, A. Kumar, E. Ohtsuka, U. L. RajBhandary, J. H. van de Sande, V. Sgaramella, T. Tebao, H. Weber, T. Yamada, *J. Mol. Biol.* **1972**, *72*, 209–217.
[136] K. L. Agarwal, H. Büchi, M. H. Caruthers, N. Gupta, H. G. Khorana, K. Kleppe, A. Kumar, E. Ohtsuka, U. L. Rajbhandary, J. H. Van De Sande, V. Sgaramella, H. Weber, T. Yamada, *Nature* **1970**, *227*, 27–34.
[137] H. G. Khorana, *Science* **1979**, *203*, 614–625.
[138] M. H. Caruthers, *J. Biol. Chem.* **2013**, *288*, 1420–1427.
[139] M. H. Caruthers, *Proc. Natl. Acad. Sci.* **2014**, *111*, 18098–18099.
[140] R. L. Letsinger, M. J. Kornet, *J. Am. Chem. Soc.* **1963**, *85*, 3045–3046.
[141] R. L. Letsinger, M. J. Kornet, V. Mahadevan, D. M. Jerina, *J. Am. Chem. Soc.* **1964**, *86*, 5163–5165.
[142] R. L. Letsinger, V. Mahadevan, *J. Am. Chem. Soc.* **1965**, *87*, 3526–3527.
[143] R. L. Letsinger, V. Mahadevan, *J. Am. Chem. Soc.* **1966**, *88*, 5319–5324.
[144] M. M. Shemyakin, Yu. A. Ovchinnikov, A. A. Kinyushkin, I. V. Kozhevnikova, *Tetrahedron Lett.* **1965**, *6*, 2323–2327.
[145] H. Hayatsu, H. G. Khorana, *J. Am. Chem. Soc.* **1966**, *88*, 3182–3183.
[146] L. Russell. Melby, D. R. Strobach, *J. Am. Chem. Soc.* **1967**, *89*, 450–453.
[147] R. L. Letsinger, M. H. Caruthers, D. M. Jerina, *Biochemistry* **1967**, *6*, 1379–1388.
[148] G. W. Grams, R. L. Letsinger, *J. Org. Chem.* **1970**, *35*, 868–870.
[149] R. L. Letsinger, K. K. Ogilvie, *J. Am. Chem. Soc.* **1967**, *89*, 4801–4803.
[150] R. L. Letsinger, K. K. Ogilvie, *J. Am. Chem. Soc.* **1969**, *91*, 3350–3355.
[151] R. L. Letsinger, K. K. Ogilvie, P. S. Miller, *J. Am. Chem. Soc.* **1969**, *91*, 3360–3365.
[152] R. L. Letsinger, P. S. Miller, *J. Am. Chem. Soc.* **1969**, *91*, 3356–3359.
[153] R. L. Letsinger, P. S. Miller, G. W. Grams, *Tetrahedron Lett.* **1968**, *9*, 2621–2624.

[154] R. L. Letsinger, M. H. Caruthers, P. S. Miller, K. K. Ogilvie, *J. Am. Chem. Soc.* **1967**, *89*, 7146–7147.
[155] C. B. Reese, R. C. Titmas, L. Yau, *Tetrahedron Lett.* **1978**, *19*, 2727–2730.
[156] S. S. Jones, B. Rayner, C. B. Reese, A. Ubasawa, M. Ubasawa, *Tetrahedron* **1980**, *36*, 3075–3085.
[157] K. G. Devine, C. B. Reese, *Tetrahedron Lett.* **1986**, *27*, 5529–5532.
[158] W. L. Sung, S. A. Narang, *Can. J. Chem.* **1982**, *60*, 111–120.
[159] V. A. Efimov, A. A. Buryakova, I. Y. Dubey, N. N. Polushin, O. G. Chakhmakhcheva, Yu. A. Ovchinnikov, *Nucleic Acids Res.* **1986**, *14*, 6525–6540.
[160] D. M. O'Callaghan, W. J. Donnelly, in *New Nucleic Acid Tech.* (Ed.: J.M. Walker), Humana Press, Totowa, NJ, **1988**, pp. 165–192.
[161] B. E. Kaplan, *Trends Biotechnol.* **1985**, *3*, 253–256.
[162] R. L. Letsinger, J. L. Finnan, G. A. Heavner, W. B. Lunsford, *J. Am. Chem. Soc.* **1975**, *97*, 3278–3279.
[163] R. L. Letsinger, W. B. Lunsford, *J. Am. Chem. Soc.* **1976**, *98*, 3655–3661.
[164] R. C. Pless, R. L. Letsinger, *Nucleic Acids Res.* **1975**, *2*, 773–786.
[165] M. D. Matteucci, M. H. Caruthers, *Tetrahedron Lett.* **1980**, *21*, 719–722.
[166] M. D. Matteucci, M. H. Caruthers, *J. Am. Chem. Soc.* **1981**, *103*, 3185–3191.
[167] S. L. Beaucage, M. H. Caruthers, *Tetrahedron Lett.* **1981**, *22*, 1859–1862.
[168] J. E. Marugg, A. Burik, M. Tromp, G. A. van der Marel, J. H. van Boom, *Tetrahedron Lett.* **1986**, *27*, 2271–2274.
[169] S. L. Beaucage, in *Encycl. Reag. Org. Synth. EROS*, John Wiley & Sons, Ltd, **2003**.
[170] J. Nielsen, M. Taagaard, O. Dahl, J. E. Marugg, J. H. van Boom, *Recl. Trav. Chim. Pays-Bas* **1986**, *105*, 33–34.
[171] N. D. Sinha, J. Biernat, H. Köster, *Tetrahedron Lett.* **1983**, *24*, 5843–5846.
[172] A. F. Sandahl, T. J. D. Nguyen, R. A. Hansen, M. B. Johansen, T. Skrydstrup, K. V. Gothelf, *Nat. Commun.* **2021**, *12*, 2760.
[173] L. J. McBride, M. H. Caruthers, *Tetrahedron Lett.* **1983**, *24*, 245–248.
[174] J. Nielsen, O. Dahl, *Nucleic Acids Res.* **1987**, *15*, 3626.
[175] N. D. Sinha, J. Biernat, J. McManus, H. Köster, *Nucleic Acids Res.* **1984**, *12*, 4539–4557.
[176] J. D. Thorpe, D. O'Reilly, T. Friščić, M. J. Damha, *Chem. – Eur. J.* **2020**, *26*, 8857–8861.
[177] K. W. Knouse, J. N. deGruyter, M. A. Schmidt, B. Zheng, J. C. Vantourout, C. Kingston, S. E. Mercer, I. M. Mcdonald, R. E. Olson, Y. Zhu, C. Hang, J. Zhu, C. Yuan, Q. Wang, P. Park, M. D. Eastgate, P. S. Baran, *Science* **2018**, *361*, 1234–1238.
[178] D. Xu, N. Rivas-Bascón, N. M. Padial, K. W. Knouse, B. Zheng, J. C. Vantourout, M. A. Schmidt, M. D. Eastgate, P. S. Baran, *J. Am. Chem. Soc.* **2020**, *142*, 5785–5792.
[179] S. J. Horvath, J. R. Firca, T. Hunkapiller, M. W. Hunkapiller, L. Hood, in *Methods Enzymol.*, Academic Press, **1987**, pp. 314–326.
[180] E. M. LeProust, B. J. Peck, K. Spirin, H. B. McCuen, B. Moore, E. Namsaraev, M. H. Caruthers, *Nucleic Acids Res.* **2010**, *38*, 2522–2540.
[181] "Polynucleotide Synthesizer Model 280, Solid Phase MicroprocessorController Model 100B," can be found under https://americanhistory.si.edu/collections/search/object/nmah_1451158, **n.d.**
[182] C. Loth, L. Charles, J.-F. Lutz, M. Nerantzaki, *ACS Macro Lett.* **2021**, *10*, 481–485.
[183] N. A. Shabaldin, A. V. Sinitskaya, A. V. Shabaldin, R. A. Mukhamadiyarov, *J. Evol. Biochem. Physiol.* **2023**, *59*, 180–191.
[184] S. Lu, C. R. Fowler, B. Ream, S. M. Waugh, T. M. Russell, J. C. Rohloff, L. Gold, J. P. Cleveland, S. Stoll, *ACS Sens.* **2023**, *8*, 2219–2227.

[185] "ABI 3900 Upgrade - Oligo DNA RNA Synthesizer," can be found under https://www.biolytic.com/t-abi-3900-upgrade-dna-synthesizer.aspx, **n.d.**
[186] *DNA Scr.* **n.d.**
[187] "Oligonucleotide Library Synthesis (OLS) | Agilent," can be found under https://www.agilent.com/en/solutions/genomics-applications-solutions/ols, **n.d.**
[188] S. Vlasenko, *Genet. Eng. Biotechnol. News* **2023**, *43*, 18–19.
[189] P. Hodge, D. C. Sherrington, Eds. , *Polymer-Supported Reactions in Organic Synthesis*, J. Wiley, Chichester ; New York, **1980**.
[190] S. Narang, *Synthesis And Applications Of DNA And RNA*, Elsevier, **2012**.
[191] M. J. Gait, R. C. Sheppard, *J. Am. Chem. Soc.* **1976**, *98*, 8514–8516.
[192] R. Crea, T. Horn, *Nucleic Acids Res.* **1980**, *8*, 2331–2348.
[193] F. Cramer, R. Helbig, H. Hettler, K. H. Scheit, H. Seliger, *Angew. Chem.* **1966**, *78*, 640–641.
[194] F. Cramer, H. Köster, *Angew. Chem. Int. Ed. Engl.* **1968**, *7*, 473–474.
[195] H. Köster, *Tetrahedron Lett.* **1972**, *13*, 1527–1530.
[196] H. Köster, A. Stumpe, A. Wolter, *Tetrahedron Lett.* **1983**, *24*, 747–750.
[197] M. H. Caruthers, *Biochem. Soc. Trans.* **2011**, *39*, 575–580.
[198] G. R. Gough, M. J. Brunden, P. T. Gilham, *Tetrahedron Lett.* **1981**, *22*, 4177–4180.
[199] M. J. Damha, P. A. Giannaris, S. V. Zabarylo, *Nucleic Acids Res.* **1990**, *18*, 3813–3821.
[200] R. T. Pon, S. Yu, *Nucleic Acids Res.* **2004**, *32*, 623–631.
[201] S. Roy, M. Caruthers, *Molecules* **2013**, *18*, 14268–14284.
[202] K. Yamamoto, Y. Fuchi, Y. Ito, Y. Hari, *J. Org. Chem.* **2023**, *88*, 2726–2734.
[203] R. T. Pon, *Curr. Protoc. Nucleic Acid Chem.* **2000**, *00*, 3.1.1-3.1.28.
[204] J. Katzhendler, S. Cohen, E. Rahamim, M. Weisz, I. Ringel, J. Deutsch, *Tetrahedron* **1989**, *45*, 2777–2792.
[205] I. A. Kozlov, M. Dang, K. Sikes, T. Kotseroglou, D. L. Barker, C. Zhao, *Nucleosides Nucleotides Nucleic Acids* **2005**, *24*, 1037–1041.
[206] A. V. Aerschot, P. Herdewijn, H. Vanderhaeghe, *Nucleosides Nucleotides* **1988**, *7*, 75–90.
[207] S. P. Adams, K. S. Kavka, E. J. Wykes, S. B. Holder, G. R. Galluppi, *J. Am. Chem. Soc.* **1983**, *105*, 661–663.
[208] P. Kumar, N. K. Bose, K. C. Gupta, *Tetrahedron Lett.* **1991**, *32*, 967–970.
[209] U. Asseline, T. T. Nguyen, *Tetrahedron Lett.* **1990**, *31*, 81–84.
[210] H. Vu, C. McCollum, C. Lotys, A. Andrus, *Nucleic Acids Symp. Ser.* **1990**, 63–64.
[211] S. A. Surzhikov, E. N. Timofeev, B. K. Chernov, J. B. Golova, A. D. Mirzabekov, *Nucleic Acids Res.* **2000**, *28*, e29-00.
[212] M. P. Reddy, N. B. Hanna, F. Farooqui, *Tetrahedron Lett.* **1994**, *35*, 4311–4314.
[213] B. Mullah, A. Andrus, *Tetrahedron Lett.* **1997**, *38*, 5751–5754.
[214] Q. Zhu, M. O. Delaney, M. M. Greenberg, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1105–1107.
[215] L. C. J. Gillet, J. Alzeer, O. D. Schärer, *Nucleic Acids Res.* **2005**, *33*, 1961–1969.
[216] J. S. Hargreaves, R. Kaiser, P. K. Wolber, *Nucleosides Nucleotides Nucleic Acids* **2015**, *34*, 691–707.
[217] E. Sonveaux, *Bioorganic Chem.* **1986**, *14*, 274–325.
[218] R. P. Iyer, S. L. Beaucage, in *Compr. Nat. Prod. Chem.* (Eds.: S.D. Barton, K. Nakanishi, O. Meth-Cohn), Pergamon, Oxford, **1999**, pp. 105–152.
[219] R. L. Letsinger, J. Fontaine, V. Mahadevan, D. A. Schexnayder, R. E. Leone, *J. Org. Chem.* **1964**, *29*, 2615–2618.
[220] R. P. Iyer, *Curr. Protoc. Nucleic Acid Chem.* **2000**, *00*, DOI 10.1002/0471142700.nc0201s00.

[221] B. S. Sproat, in *Oligonucleotide Synth.* (Ed.: P. Herdewijn), Humana Press, Totowa, NJ, **2005**, pp. 17–31.
[222] W. A. Velema, E. T. Kool, *Nat. Rev. Chem.* **2020**, *4*, 22–37.
[223] J. C. Schulhof, D. Molko, R. Teoule, *Nucleic Acids Res.* **1987**, *15*, 397–416.
[224] H. Köster, N. Hoppe, V. Kohli, M. Kröpelin, H. Kaut, K. Kulikowski, *Nucleic Acids Symp. Ser.* **1980**, 39–60.
[225] H. Büchi, H. G. Khorana, *J. Mol. Biol.* **1972**, *72*, 251–288.
[226] R. Ekins, F. W. Chu, *Trends Biotechnol.* **1999**, *17*, 217–218.
[227] M. J. Heller, *Annu. Rev. Biomed. Eng.* **2002**, *4*, 129–153.
[228] M. Schena, R. A. Heller, T. P. Theriault, K. Konrad, E. Lachenmeier, R. W. Davis, *Trends Biotechnol.* **1998**, *16*, 301–306.
[229] R. A. Hughes, A. D. Ellington, *Cold Spring Harb. Perspect. Biol.* **2017**, *9*, a023812.
[230] K. E. Richmond, M.-H. Li, M. J. Rodesch, M. Patel, A. M. Lowe, C. Kim, L. L. Chu, N. Venkataramaian, S. F. Flickinger, J. Kaysen, P. J. Belshaw, M. R. Sussman, F. Cerrina, *Nucleic Acids Res.* **2004**, *32*, 5011–5018.
[231] S.-H. Shiu, J. O. Borevitz, *Heredity* **2008**, *100*, 141–149.
[232] R. Ekins, F. Chu, J. Micallef, *J. Biolumin. Chemilumin.* **1989**, *4*, 59–78.
[233] S. P. A. Fodor, J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu, D. Solas, *Science* **1991**, *251*, 767–773.
[234] M. C. Pirrung, L. Wang, M. P. Montague-Smith, *Org. Lett.* **2001**, *3*, 1105–1108.
[235] J. Tian, H. Gong, N. Sheng, X. Zhou, E. Gulari, X. Gao, G. Church, *Nature* **2004**, *432*, 1050–1054.
[236] S. P. A. Fodor, R. P. Rava, X. C. Huang, A. C. Pease, C. P. Holmes, C. L. Adams, *Nature* **1993**, *364*, 555–556.
[237] J. M. Rothberg, W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J. A. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth, J. Bustillo, *Nature* **2011**, *475*, 348–352.
[238] P. Cooley, D. Wallace, B. Antohe, *SLAS Technol.* **2002**, *7*, 33–39.
[239] S. Kosuri, N. Eroshenko, E. M. LeProust, M. Super, J. Way, J. B. Li, G. M. Church, *Nat. Biotechnol.* **2010**, *28*, 1295–1299.
[240] E. Schaudy, K. Hölz, J. Lietard, M. M. Somoza, *Nat. Commun.* **2022**, *13*, 3772.
[241] M. Blawat, K. Gaedke, I. Hütter, X.-M. Chen, B. Turczyk, S. Inverso, B. W. Pruitt, G. M. Church, *Procedia Comput. Sci.* **2016**, *80*, 1011–1022.
[242] Y. Erlich, D. Zielinski, *Science* **2017**, *355*, 950–954.
[243] J. Bornholt, R. Lopez, D. M. Carmean, L. Ceze, G. Seelig, K. Strauss, in *Proc. Twenty-First Int. Conf. Archit. Support Program. Lang. Oper. Syst.*, Association For Computing Machinery, New York, NY, USA, **2016**, pp. 637–649.
[244] S. M. H. Tabatabaei Yazdi, Y. Yuan, J. Ma, H. Zhao, O. Milenkovic, *Sci. Rep.* **2015**, *5*, 14138.
[245] J. Chmielecki, M. Meyerson, *Annu. Rev. Med.* **2014**, *65*, 63–79.
[246] S. Mallapaty, *Nature* **2022**, DOI 10.1038/d41586-022-04453-2.
[247] A. Maxmen, *Nature* **2021**, *593*, 21–21.
[248] A. M. Maxam, W. Gilbert, *Proc. Natl. Acad. Sci.* **1977**, *74*, 560–564.
[249] E. J. Boland, A. Pillai, M. W. Odom, P. Jagadeeswaran, *BioTechniques* **1994**, *16*, 1088–92, 1094–5.
[250] F. Sanger, A. R. Coulson, *J. Mol. Biol.* **1975**, *94*, 441–448.

[251] F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci.* **1977**, *74*, 5463–5467.
[252] L. M. Smith, J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. H. Kent, L. E. Hood, *Nature* **1986**, *321*, 674–679.
[253] J. Shendure, S. Balasubramanian, G. M. Church, W. Gilbert, J. Rogers, J. A. Schloss, R. H. Waterston, *Nature* **2017**, *550*, 345–353.
[254] E. D. Hyman, *Anal. Biochem.* **1988**, *174*, 423–436.
[255] *Genomics* **2016**, *107*, 1–8.
[256] M. Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, J. M. Rothberg, *Nature* **2005**, *437*, 376–380.
[257] R. D. Mitra, J. Shendure, J. Olejnik, Edyta-Krzymanska-Olejnik, G. M. Church, *Anal. Biochem.* **2003**, *320*, 55–65.
[258] K. J. McKernan, H. E. Peckham, G. L. Costa, S. F. McLaughlin, Y. Fu, E. F. Tsung, C. R. Clouser, C. Duncan, J. K. Ichikawa, C. C. Lee, Z. Zhang, S. S. Ranade, E. T. Dimalanta, F. C. Hyland, T. D. Sokolsky, L. Zhang, A. Sheridan, H. Fu, C. L. Hendrickson, B. Li, L. Kotler, J. R. Stuart, J. A. Malek, J. M. Manning, A. A. Antipova, D. S. Perez, M. P. Moore, K. C. Hayashibara, M. R. Lyons, R. E. Beaudoin, B. E. Coleman, M. W. Laptewicz, A. E. Sannicandro, M. D. Rhodes, R. K. Gottimukkala, S. Yang, V. Bafna, A. Bashir, A. MacBride, C. Alkan, J. M. Kidd, E. E. Eichler, M. G. Reese, F. M. D. L. Vega, A. P. Blanchard, *Genome Res.* **2009**, *19*, 1527–1541.
[259] J. Shendure, G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra, G. M. Church, *Science* **2005**, *309*, 1728–1732.
[260] D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. Keira Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. J. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. D. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, M. Chiara E. Catenazzi, S. Chang, R. Neil Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. Scott Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschild, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, K. Hoschler, S. Hurwitz, D. V. Ivanov, M. Q. Johnson, T. James, T. A. Huw Jones, G.-D. Kang, T. H. Kerelska, A. D. Kersey, I. Khrebtukova, A. P. Kindwall, Z. Kingsbury, P. I. Kokko-Gonzales, A. Kumar, M. A. Laurent, C. T. Lawley, S. E. Lee, X. Lee, A. K. Liao, J. A. Loch, M. Lok, S. Luo, R. M. Mammen, J. W. Martin, P. G. McCauley, P. McNitt, P. Mehta, K. W. Moon, J. W. Mullens, T. Newington, Z. Ning, B. Ling Ng, S. M. Novo, M. J. O'Neill, M. A. Osborne, A.

Osnowski, O. Ostadan, L. L. Paraschos, L. Pickering, A. C. Pike, A. C. Pike, D. Chris Pinkard, D. P. Pliskin, J. Podhasky, V. J. Quijano, C. Raczy, V. H. Rae, S. R. Rawlings, A. Chiva Rodriguez, P. M. Roe, J. Rogers, M. C. Rogert Bacigalupo, N. Romanov, A. Romieu, R. K. Roth, N. J. Rourke, S. T. Ruediger, E. Rusman, R. M. Sanches-Kuiper, M. R. Schenker, J. M. Seoane, R. J. Shaw, M. K. Shiver, S. W. Short, N. L. Sizto, J. P. Sluis, M. A. Smith, J. Ernest Sohna Sohna, E. J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C. L. Tregidgo, G. Turcatti, S. vandeVondele, Y. Verhovsky, S. M. Virk, S. Wakelin, G. C. Walcott, J. Wang, G. J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J. C. Mullikin, M. E. Hurles, N. J. McCooke, J. S. West, F. L. Oaks, P. L. Lundberg, D. Klenerman, R. Durbin, A. J. Smith, *Nature* **2008**, *456*, 53–59.

[261] M. J. Levene, J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead, W. W. Webb, *Science* **2003**, *299*, 682–686.

[262] A. Rhoads, K. F. Au, *Genomics Proteomics Bioinformatics* **2015**, *13*, 278–289.

[263] D. Deamer, M. Akeson, D. Branton, *Nat. Biotechnol.* **2016**, *34*, 518–524.

[264] L. Song, M. R. Hobaugh, C. Shustak, S. Cheley, H. Bayley, J. E. Gouaux, *Science* **1996**, *274*, 1859–1865.

[265] H. Bayley, P. S. Cremer, *Nature* **2001**, *413*, 226–230.

[266] J. J. Kasianowicz, S. M. Bezrukov, *Nat. Biotechnol.* **2016**, *34*, 481–482.

[267] "MinION," can be found under https://nanoporetech.com/products/minion, **n.d.**

[268] E. Check Hayden, *Nature* **2015**, *521*, 15–16.

[269] E. Kawashima, L. Farinelli, P. Mayer, *Method of Nucleic Acid Sequencing*, **2002**, US20020055100A1.

[270] "Breakthrough Prizes 2022 : Le biophysicien Pascal Mayer lauréat du prestigieux prix en Sciences de la vie - Université de Strasbourg," can be found under https://www.unistra.fr/communiques-presse/detail-des-communiques-et-dossiers/18862-breakthrough-prizes-2022-le-biophysicien-pascal-mayer-laureat-du-prestigieux-prix-en-sciences-de-la-vie, **n.d.**

[271] C. K. Lim, J. W. Yeoh, A. A. Kunartama, W. S. Yew, C. L. Poh, *Nat. Commun.* **2023**, *14*, 3921.

[272] Y. Liu, Y. Ren, J. Li, F. Wang, F. Wang, C. Ma, D. Chen, X. Jiang, C. Fan, H. Zhang, K. Liu, *Sci. Adv.* **2022**, *8*, eabo7415.

[273] A. De, H. Mohammad, Y. Wang, R. Kubendran, A. K. Das, M. P. Anantram, *Sci. Rep.* **2023**, *13*, 6650.

[274] S. Vitak, *Nature* **2021**, DOI 10.1038/d41586-021-00534-w.

[275] H. Lee, D. J. Wiegand, K. Griswold, S. Punthambaker, H. Chun, R. E. Kohman, G. M. Church, *Nat. Commun.* **2020**, *11*, 5246.

[276] R. Heckel, I. Shomorony, K. Ramchandran, D. N. C. Tse, **2017**, DOI 10.48550/arXiv.1705.04732.

[277] D. Paunescu, M. Puddu, J. O. B. Soellner, P. R. Stoessel, R. N. Grass, *Nat. Protoc.* **2013**, *8*, 2440–2448.

[278] W. D. Chen, A. X. Kohll, B. H. Nguyen, J. Koch, R. Heckel, W. J. Stark, L. Ceze, K. Strauss, R. N. Grass, *Adv. Funct. Mater.* **2019**, *29*, 1901672.

[279] E. L. van Dijk, Y. Jaszczyszyn, D. Naquin, C. Thermes, *Trends Genet.* **2018**, *34*, 666–681.

[280] F. Pfeiffer, C. Gröber, M. Blank, K. Händler, M. Beyer, J. L. Schultze, G. Mayer, *Sci. Rep.* **2018**, *8*, 10950.

[281] H. Staudinger, *Berichte Dtsch. Chem. Ges. B Ser.* **1920**, *53*, 1073–1085.

[282] J.-F. Lutz, *Eur. Polym. J.* **2023**, *199*, 112465.

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

[284] M. A. R. Meier, C. Barner-Kowollik, *Adv. Mater.* **2019**, *31*, 1806027.

[285] J.-F. Lutz, in *Seq.-Control. Polym.*, John Wiley & Sons, Ltd, **2017**, pp. 1–26.
[286] G. Decher, M. Eckle, J. Schmitt, B. Struth, *Curr. Opin. Colloid Interface Sci.* **1998**, *3*, 32–39.
[287] R. Szweda, M. Tschopp, O. Felix, G. Decher, J.-F. Lutz, *Angew. Chem. Int. Ed.* **2018**, *57*, 15817–15821.
[288] H. Colquhoun, J.-F. Lutz, *Nat. Chem.* **2014**, *6*, 455–456.
[289] J.-F. Lutz, J.-M. Lehn, E. W. Meijer, K. Matyjaszewski, *Nat. Rev. Mater.* **2016**, *1*, 1–14.
[290] S. Pfeifer, J.-F. Lutz, *J. Am. Chem. Soc.* **2007**, *129*, 9542–9543.
[291] L. Charles, J.-F. Lutz, *Acc. Chem. Res.* **2021**, *54*, 1791–1800.
[292] S. Pfeifer, Z. Zarafshani, N. Badi, J.-F. Lutz, *J. Am. Chem. Soc.* **2009**, *131*, 9195–9197.
[293] A. Meszynska, N. Badi, H. G. Börner, J.-F. Lutz, *Chem. Commun.* **2012**, *48*, 3887–3889.
[294] L. Oswald, A. Al Ouahabi, L. Charles, J.-F. Lutz, *Chem. – Eur. J.* **2016**, *22*, 3462–3469.
[295] L. Oswald, A. Al Ouahabi, C. Laure, J.-A. Amalian, L. Charles, J.-F. Lutz, *J. Polym. Sci. Part Polym. Chem.* **2019**, *57*, 403–410.
[296] A. Al Ouahabi, L. Charles, J.-F. Lutz, *J. Am. Chem. Soc.* **2015**, *137*, 5629–5635.
[297] N. F. König, A. Al Ouahabi, S. Poyer, L. Charles, J.-F. Lutz, *Angew. Chem. Int. Ed.* **2017**, *56*, 7297–7301.
[298] N. F. König, A. Al Ouahabi, L. Oswald, R. Szweda, L. Charles, J.-F. Lutz, *Nat. Commun.* **2019**, *10*, 3774.
[299] N. F. König, S. Telitel, S. Poyer, L. Charles, J.-F. Lutz, *Macromol. Rapid Commun.* **2017**, *38*, 1700651.
[300] R. N. Zuckermann, J. M. Kerr, S. B. H. Kent, W. H. Moos, *J. Am. Chem. Soc.* **1992**, *114*, 10646–10647.
[301] S. Martens, J. O. Holloway, Filip. E. Du Prez, *Macromol. Rapid Commun.* **2017**, *38*, 1700469.
[302] T. T. Trinh, L. Oswald, D. Chan-Seng, L. Charles, J.-F. Lutz, *Chem. – Eur. J.* **2015**, *21*, 11961–11965.
[303] L. Charles, G. Cavallo, V. Monnier, L. Oswald, R. Szweda, J.-F. Lutz, *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1149–1159.
[304] R. Kumar Roy, C. Laure, D. Fischer-Krauser, L. Charles, J.-F. Lutz, *Chem. Commun.* **2015**, *51*, 15677–15680.
[305] G. Cavallo, A. Al Ouahabi, L. Oswald, L. Charles, J.-F. Lutz, *J. Am. Chem. Soc.* **2016**, *138*, 9417–9420.
[306] G. Cavallo, S. Poyer, J.-A. Amalian, F. Dufour, A. Burel, C. Carapito, L. Charles, J.-F. Lutz, *Angew. Chem. Int. Ed.* **2018**, *57*, 6266–6269.
[307] U. S. Gunay, B. 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.
[308] D. Karamessini, B. E. Petit, M. Bouquey, L. Charles, J.-F. Lutz, *Adv. Funct. Mater.* **2017**, *27*, 1604595.
[309] R. L. Kanasty, A. J. Vegas, L. M. Ceo, M. Maier, K. Charisse, J. K. Nair, R. Langer, D. G. Anderson, *Angew. Chem. Int. Ed.* **2016**, *55*, 9529–9533.
[310] T. Mondal, V. Greff, B. É. Petit, L. Charles, J.-F. Lutz, *ACS Macro Lett.* **2019**, *8*, 1002–1005.
[311] T. Mondal, L. Charles, J.-F. Lutz, *Angew. Chem. Int. Ed.* **2020**, *59*, 20390–20393.
[312] P. Cwynar, P. Pasikowski, R. Szweda, *Eur. Polym. J.* **2023**, *182*, 111706.
[313] N. Illy, E. Mongkhoun, *Polym. Chem.* **2022**, *13*, 4592–4614.
[314] P. Espeel, F. E. Du Prez, *Eur. Polym. J.* **2015**, *62*, 247–272.
[315] M. Soete, C. Mertens, R. Aksakal, N. Badi, F. Du Prez, *ACS Macro Lett.* **2021**, *10*, 616–622.
[316] M. Soete, K. De Bruycker, F. Du Prez, *Angew. Chem.* **2022**, *134*, e202116718.

[317] P. Espeel, F. Goethals, F. E. Du Prez, *J. Am. Chem. Soc.* **2011**, *133*, 1678–1681.
[318] 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.
[319] M. Porel, C. A. Alabi, *J. Am. Chem. Soc.* **2014**, *136*, 13162–13165.
[320] S. Martens, J. Van den Begin, A. Madder, F. E. Du Prez, P. Espeel, *J. Am. Chem. Soc.* **2016**, *138*, 14182–14185.
[321] A. Llevot, A. C. Boukis, S. Oelmann, K. Wetzel, M. A. R. Meier, in *Polym. Synth. Based Triple-Bond Build. Blocks* (Eds.: B.Z. Tang, R. Hu), Springer International Publishing, Cham, **2018**, pp. 127–155.
[322] A. C. Boukis, M. A. R. Meier, *Eur. Polym. J.* **2018**, *104*, 32–38.
[323] K. S. Wetzel, M. Frölich, S. C. Solleder, R. Nickisch, P. Treu, M. A. R. Meier, *Commun. Chem.* **2020**, *3*, 1–10.
[324] S. C. Solleder, M. A. R. Meier, *Angew. Chem. Int. Ed.* **2014**, *53*, 711–714.
[325] S. C. Solleder, D. Zengel, K. S. Wetzel, M. A. R. Meier, *Angew. Chem. Int. Ed.* **2016**, *55*, 1204–1207.
[326] J. M. Lee, M. B. Koo, S. W. Lee, H. Lee, J. Kwon, Y. H. Shim, S. Y. Kim, K. T. Kim, *Nat. Commun.* **2020**, *11*, 56.
[327] J. M. Lee, J. Kwon, S. J. Lee, H. Jang, D. Kim, J. Song, K. T. Kim, *Sci. Adv.* **2022**, *8*, eabl8614.
[328] D. A. Tomalia, H. Baker, J. Dewald, M. Hall, G. Kallos, S. Martin, J. Roeck, J. Ryder, P. Smith, *Polym. J.* **1985**, *17*, 117–132.
[329] Z. Huang, Q. Shi, J. Guo, F. Meng, Y. Zhang, Y. Lu, Z. Qian, X. Li, N. Zhou, Z. Zhang, X. Zhu, *Nat. Commun.* **2019**, *10*, 1918.
[330] Q. Shi, H. Yin, R. Song, J. Xu, J. Tan, X. Zhou, J. Cen, Z. Deng, H. Tong, C. Cui, Y. Zhang, X. Li, Z. Zhang, S. Liu, *Nat. Chem.* **2023**, *15*, 257–270.
[331] N. Zydziak, W. Konrad, F. Feist, S. Afonin, S. Weidner, C. Barner-Kowollik, *Nat. Commun.* **2016**, *7*, 13672.
[332] S. C. Leguizamon, T. F. Scott, *Nat. Commun.* **2020**, *11*, 784.
[333] Z. Zhu, C. J. Cardin, Y. Gan, H. M. Colquhoun, *Nat. Chem.* **2010**, *2*, 653–660.
[334] C. E. Arcadia, E. Kennedy, J. Geiser, A. Dombroski, K. Oakley, S.-L. Chen, L. Sprague, M. Ozmen, J. Sello, P. M. Weber, S. Reda, C. Rose, E. Kim, B. M. Rubenstein, J. K. Rosenstein, *Nat. Commun.* **2020**, *11*, 691.
[335] A. C. Boukis, K. Reiter, M. Frölich, D. Hofheinz, M. A. R. Meier, *Nat. Commun.* **2018**, *9*, 1439.
[336] P. Bohn, M. P. Weisel, J. Wolfs, M. A. R. Meier, *Sci. Rep.* **2022**, *12*, 13878.
[337] V. Pardi-Tóth, Á. Kuki, M. Á. Kordován, G. Róth, L. Nagy, M. Zsuga, T. Nagy, S. Kéki, *Sci. Rep.* **2023**, *13*, 16576.
[338] B. J. Cafferty, A. S. Ten, M. J. Fink, S. Morey, D. J. Preston, M. Mrksich, G. M. Whitesides, *ACS Cent. Sci.* **2019**, *5*, 911–916.
[339] L. Charles, in *Seq.-Control. Polym.*, John Wiley & Sons, Ltd, **2018**, pp. 479–504.
[340] S. Poyer, B. E. Petit, S. Telitel, D. Karamessini, J.-F. Lutz, L. Charles, *Int. J. Mass Spectrom.* **2020**, *448*, 116271.
[341] L. Charles, C. Laure, J.-F. Lutz, R. K. Roy, *Macromolecules* **2015**, *48*, 4319–4328.
[342] L. Charles, C. Laure, J.-F. Lutz, R. K. Roy, *Rapid Commun. Mass Spectrom.* **2016**, *30*, 22–28.
[343] M. Frandsen, K. El-Chami, J. Palmfeldt, J. Melgaard Smidt, M. E. T. Langballe, A. Sandahl, K. V. Gothelf, *Angew. Chem. Int. Ed.* **2023**, *62*, e202303170.
[344] K. Ding, Y. Zhang, Z. Huang, B. Liu, Q. Shi, L. Hu, N. Zhou, Z. Zhang, X. Zhu, *Eur. Polym. J.* **2019**, *119*, 421–425.
[345] B. Liu, Q. Shi, L. Hu, Z. Huang, X. Zhu, Z. Zhang, *Polym. Chem.* **2020**, *11*, 1702–1707.

[346] A. ’Bin Sun, S. Li, X. Kou, *Anal. Methods* **2023**, *15*, 868–883.
[347] H. Mutlu, J.-F. Lutz, *Angew. Chem. Int. Ed.* **2014**, *53*, 13010–13019.
[348] G. Montaudo, F. Samperi, M. S. Montaudo, *Prog. Polym. Sci.* **2006**, *31*, 277–357.
[349] B. Zhao, Z. Gao, Y. Zheng, C. Gao, *J. Am. Chem. Soc.* **2019**, *141*, 4541–4546.
[350] A. B. Kanu, P. Dwivedi, M. Tam, L. Matz, H. H. Hill Jr., *J. Mass Spectrom.* **2008**, *43*, 1–22.
[351] J.-A. Amalian, G. Cavallo, A. Al Ouahabi, J.-F. Lutz, L. Charles, *Anal. Chem.* **2019**, *91*, 7266–7272.
[352] Y. Ren, R. Jamagne, D. J. Tetlow, D. A. Leigh, *Nature* **2022**, *612*, 78–82.
[353] C. Cao, L. F. Krapp, A. Al Ouahabi, N. F. König, N. Cirauqui, A. Radenovic, J.-F. Lutz, M. D. Peraro, *Sci. Adv.* **2020**, *6*, eabc2661.
[354] F. Traversi, C. Raillon, S. M. Benameur, K. Liu, S. Khlybov, M. Tosun, D. Krasnozhon, A. Kis, A. Radenovic, *Nat. Nanotechnol.* **2013**, *8*, 939–945.
[355] K. Launay, J.-A. Amalian, E. Laurent, L. Oswald, A. Al Ouahabi, A. Burel, F. Dufour, C. Carapito, J.-L. Clément, J.-F. Lutz, L. Charles, D. Gigmes, *Angew. Chem. Int. Ed.* **2021**, *60*, 917–926.
[356] Y. Zheng, P. Cloutier, D. J. Hunting, L. Sanche, J. R. Wagner, *J. Am. Chem. Soc.* **2005**, *127*, 16592–16598.
[357] W. G. Skene, S. T. Belt, T. J. Connolly, P. Hahn, J. C. Scaiano, *Macromolecules* **1998**, *31*, 9103–9105.
[358] B. Gamoke, D. Neff, J. Simons, *J. Phys. Chem. A* **2009**, *113*, 5677–5684.
[359] I. Roszak, L. Oswald, A. A. Ouahabi, A. Bertin, E. Laurent, O. Felix, I. Carvin-Sergent, L. Charles, J.-F. Lutz, *Polym. Chem.* **2021**, *12*, 5279–5282.
[360] D. Karamessini, S. Poyer, L. Charles, J.-F. Lutz, *Macromol. Rapid Commun.* **2017**, *38*, 1700426.
[361] S. Celasun, D. Remmler, T. Schwaar, M. G. Weller, F. Du Prez, H. G. Börner, *Angew. Chem. Int. Ed.* **2019**, *58*, 1960–1964.
[362] S. Zhang, N. E. Bauer, I. Y. Kanal, W. You, G. R. Hutchison, T. Y. Meyer, *Macromolecules* **2017**, *50*, 151–161.
[363] G.-J. Wu, J. Guan, F.-S. Han, Y.-L. Zhao, *ChemCatChem* **2014**, *6*, 1589–1593.
[364] A. F. Abdel-Magid, K. G. Carson, B. D. Harris, C. A. Maryanoff, R. D. Shah, *J. Org. Chem.* **1996**, *61*, 3849–3862.
[365] T. D. Lee, J. F. W. Keana, *J. Org. Chem.* **1975**, *40*, 3145–3147.
[366] S. X. Cao, V. Feher, T. Ichikawa, B. Jones, S. W. Kaldor, A. A. Kiryanov, Y. Liu, C. Mcbride, S. R. Natala, Z. Nie, J. A. Stafford, B. Lam, *Inhibiteurs De Kinase De Type Polo*, **2009**.
[367] “Glen-Pak™ DNA purification cartridge,” can be found under https://www.glenresearch.com/purification/60-5100.html, **n.d.**
[368] S. Sooriyaarachchi, R. Chofor, M. D. P. Risseeuw, T. Bergfors, J. Pouyez, C. S. Dowd, L. Maes, J. Wouters, T. A. Jones, S. Van Calenbergh, S. L. Mowbray, *ChemMedChem* **2016**, *11*, 2024–2036.
[369] E. D. Funder, A. B. Jensen, T. Tørring, A. L. B. Kodal, A. R. Azcargorta, K. V. Gothelf, *J. Org. Chem.* **2012**, *77*, 3134–3142.
[370] H. Sajiki, T. Ikawa, K. Hattori, K. Hirota, *Chem. Commun.* **2003**, 654–655.
[371] K. Hattori, H. Sajiki, K. Hirota, *Tetrahedron* **2001**, *57*, 2109–2114.
[372] N. Mainolfi, N. Ji, A. F. Kluge, M. M. Weiss, Y. Zhang, X. Zheng, *Irak Degraders and Uses Thereof*, **2020**.
[373] X. Wang, A. C. Burns, J. G. Christensen, J. M. Ketcham, J. D. Lawson, M. A. Marx, C. R. Smith, S. Allen, J. F. Blake, M. J. Chicarelli, J. R. Dahlke, D. Dai, J. B. Fell, J. P. Fischer, M. J. Mejia, B. Newhouse, P. Nguyen, J. M. O’leary, S. Pajk, M. E. Rodriguez,

P. Savechenkov, T. P. Tang, G. P. A. Vigers, Q. Zhao, D. R. Kahn, J. Gaudino, M. C. Hilton, *Kras G12d Inhibitors*, **2021**.
[374] A. T. Tran, V. A. Huynh, E. M. Friz, S. K. Whitney, D. B. Cordes, *Tetrahedron Lett.* **2009**, *50*, 1817–1819.
[375] P. K. Mandal, J. S. McMurray, *J. Org. Chem.* **2007**, *72*, 6599–6601.
[376] W. Long, N. A. Brunelli, S. A. Didas, E. W. Ping, C. W. Jones, *ACS Catal.* **2013**, *3*, 1700–1708.
[377] A. Azhayev, M. Antopolsky, T. Tennilä, H. Mackie, J. Randolph, *Genet. Eng. News* **2005**, *25*.
[378] A. P. Guzaev, M. Manoharan, *J. Am. Chem. Soc.* **2003**, *125*, 2380–2381.
[379] R. K. Kumar, A. P. Guzaev, C. Rentel, V. T. Ravikumar, *Tetrahedron* **2006**, *62*, 4528–4534.
[380] "Glen Report 20.25 - New Universal Support - Glen UnySupport," can be found under https://www.glenresearch.com/reports/gr20-25, **n.d.**
[381] E. Voirin, J.-P. Behr, M. Kotera, *Nat. Protoc.* **2007**, *2*, 1360–1367.
[382] "Expedite 8909 User's Guide," can be found under https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_041567.pdf, **n.d.**
[383] T. Schutz, I. Sergent, G. Obeid, L. Oswald, A. Al Ouahabi, P. N. W. Baxter, J.-L. Clément, D. Gigmes, L. Charles, J.-F. Lutz, *Angew. Chem. Int. Ed.* **2023**, *62*, e202310801.
[384] M. L. Nielsen, J. V. Pustinger, John. Strobel, *J. Chem. Eng. Data* **1964**, *9*, 167–170.
[385] T. Mondal, M. Nerantzaki, K. Flesch, C. Loth, M. Maaloum, Y. Cong, S. S. Sheiko, J.-F. Lutz, *Macromolecules* **2021**, *54*, 3423–3429.
[386] F. E. Hahn, T. Lügger, M. Beinhoff, *Z. Für Naturforschung B* **2004**, *59*, 196–201.
[387] P. Steib, B. Breit, *Chem. – Eur. J.* **2019**, *25*, 3532–3535.

Thibault SCHUTZ

Design et synthèse de poly(phosphodiesters) numériques pour un séquençage facilité

Résumé

L'objectif de cette thèse a été de faciliter le séquençage de poly(phosphodiesters) synthétiques numériques par MS/MS et IMS à l'aide d'un nouveau marqueur clivable. Ces polymères utilisent un alphabet de monomères codants pour 0 et 1 rendant possible le stockage d'information. De tels monomères sont composés d'un groupe phosphoramidite et d'un groupe DMT, permettant d'utiliser la chimie de phosphoramidite automatisée. Afin de faciliter la lecture des polymères numériques, un nouveau design compatible avec cette chimie, incorporant une liaison fragile NO-C et un marqueur de masse a été conçu. Une première approche, motivée par sa simple mise en œuvre et ses bons rendements, a été d'utiliser un alcyne dans la structure du marqueur. Cependant, le séquençage pseudo-MS^3 d'un oligomère contenant cet espaceur a été compromis à cause de sa triple liaison. La réduction de cette dernière n'a pas été encourageante, ce qui a redirigé les travaux de cette thèse vers la base structurelle d'un design antérieur alkyle du laboratoire, quant à elle efficace. Cette nouvelle stratégie a permis d'obtenir une chimiothèque de 12 marqueurs clivables. Leur incorporation en chaîne polymère uniforme, étude d'efficacité de couplage et optimisation sur un support solide universel a été décrite.

Mots clés : Polymère à séquence contrôlée, Macromolécule numérique, Chimie de phosphoramidite automatisée, Alcoxyamine clivable, Séquençage MS/MS

Résumé en anglais

The objective of this thesis was to facilitate the sequencing of synthetic digital poly(phosphodiesters) through MS/MS and IMS using a new cleavable marker. These polymers use a monomer alphabet encoding for 0 and 1, enabling information storage. Such monomers consist of a phosphoramidite group and a DMT group, allowing for automated phosphoramidite chemistry. To enhance the readability of digital polymers, a new design compatible with this chemistry, incorporating a fragile NO-C linkage and a mass tag, was devised. An initial approach, motivated by its simplicity in implementation and high yields, involved an alkyne in the marker structure. However, the pseudo-MS^3 sequencing of an oligomer containing this spacer was compromised due to its triple bond. Reducing this bond did not yield promising results, redirecting the studies of this thesis toward the structural foundation of a previous laboratory design, which proved effective. This new strategy resulted in a library of 12 cleavable markers. Their incorporation into a monodisperse polymer chain, a study of coupling efficiency, and optimization on a universal solid support have been detailed.

Key words: Sequence-controlled polymers, Digital macromolecules, Automated phosphoramidite chemistry, Cleavable alcoxyamine, MS/MS sequencing