HAL
open science

Conception et application d'ADN tétraplexe à topologie contrôlée

Alexandre Devaux

To cite this version:

Alexandre Devaux. Conception et application d'ADN tétraplexe à topologie contrôlée. Biochimie [q-bio.BM]. Université Grenoble Alpes [2020-..], 2020. Français. NNT : 2020GRALV020 . tel-03152509

HAL Id: tel-03152509

https://theses.hal.science/tel-03152509v1

Submitted on 25 Feb 2021

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : **Chimie Biologie**

Arrêté ministériel : 25 mai 2016

Présentée par

Alexandre DEVAUX

Thèse dirigée par **Eric DEFRANCQ**,
et codirigée par **Thomas LAVERGNE**

préparée au sein du **Département de Chimie Moléculaire**
dans **l'École Doctorale Chimie et Sciences du Vivant**

Conception et applications d'ADN tétraplexe à topologie contrôlée

Design and applications of DNA tetraplex with controlled topology

Thèse soutenue publiquement le **1 juillet 2020,**
devant le jury composé de :

Monsieur François MORVAN
Directeur de recherche INSERM, Université de Montpellier (Rapporteur)
Madame Mélanie ETHÈVE-QUELQUEJEU
Professeure, Université Paris Descartes (Rapporteur)
Monsieur Dennis GOMEZ
Chargé de recherche CNRS, IPBS Toulouse (Examinateur)
Monsieur Didier GASPARUTTO
Ingénieur de recherche CEA, Grenoble (Examinateur)
Madame Marie-Paule TEULADE-FICHOU
Directrice de recherche CNRS, Institut Curie, Paris (Présidente de jury)
Monsieur Éric DEFRANCQ
Professeur, Université Grenoble-Alpes (Directeur de thèse)
Monsieur Thomas LAVERGNE
Chargé de recherche, Université Grenoble-Alpes (Co-directeur de thèse)

Sommaire

Abréviations **1**

Chapitre I : Introduction **3**

1. Les acides nucléiques 4
2. L'acide désoxyribonucléique (ADN) et l'acide ribonucléique (ARN) 4
   a) Structures usuelles de l'ADN et de l'ARN 6
   b) Structures non usuelles de l'ADN et de l'ARN 8
3. Les G-quadruplexes (G4) 11
   a) Définition 11
   b) Les différentes topologies des G-quadruplexes ADN 12
   c) Les différentes topologies des G-quadruplexes ARN 14
   d) Localisations et implications biologiques des G-quadruplexes 15
   e) Les G-quadruplexes et leurs ligands spécifiques 21
4. Les i-motifs 23
   a) Facteurs influençant la stabilité et la topologie du i-motif 25
   b) Localisations et implications biologiques des i-motifs 28
   c) Les i-motifs et leurs ligands spécifiques 32
5. Techniques de caractérisation des structures G-quadruplexes et i-motifs 34

Chapitre II : Présentation du projet de thèse **37**

Chapitre III : Synthèse d'un mime ADN i-motif et étude de sa topologie **43**

1. Généralités sur la synthèse peptidique et oligonucléotidique 44
   a) Synthèse peptidique 44
   b) Synthèse oligonucléotidique 45
2. Présentation de la cible i-motif télomérique 46
3. Synthèse du châssis peptidique **1** 51
   a) Synthèse des acides aminés modifiés 52
   b) Synthèse du châssis peptidique **1** 54
4. Synthèse des brins d'ADN 56
   a) Synthèse de l'oligonucléotide 5'-thiol/3'-azoture **6** 56

b) Synthèse de l'oligonucléotide 5'-alcyne/3'-aldehyde **7** 58

5. Synthèse du mime i-motif **F** 59

a) Assemblage de l'intermédiaire **A** 59

b) Assemblage de l'édifice **B** 61

c) Synthèse de l'édifice **C** 63

d) Assemblage de l'édifice **F** 65

6. Bilan intermédiaire 73

7. Etude de la topologie du mime i-motif 73

a) Etude de la topologie i-motif de hTeloC **8** en fonction du pH 74

b) Etude de la topologie i-motif de l'édifice **E** et **F** en fonction du pH 74

c) Etude de stabilité de la topologie i-motif sur l'édifice **F** 75

8. Bilan et perspectives 78

Chapitre IV : Introduction d'une nouvelle ligation chimiosélective pour les acides nucléiques et synthèse d'un conjugué en utilisant quatre ligations orthogonales 79

1. Présentation de différentes ligations chimiosélectives 80

a) Ligation faisant intervenir une fonction cétone ou aldéhyde 81

b) Ligation faisant intervenir une fonction azoture 82

c) Ligation faisant intervenir une fonction thiolate 85

d) Ligation faisant intervenir une fonction oxyde de nitrile ou nitrone 86

e) Ligation faisant intervenir un dérivé tétrazine 87

f) Autres exemples de ligations chimiosélectives 88

2. Présentation de la ligation KAT et son intérêt 90

3. Mise en œuvre et compatibilité de la ligation KAT 93

a) Synthèse du châssis peptidique **10** 94

b) Synthèse du brin d'ADN possédant la fonction acyltrifluoroborate **11** 96

c) Synthèse de l'édifice **G** 98

4. Evaluation de la compatibilité des quatre ligations 100

5. Synthèse des oligonucléotides 101

a) Synthèse de l'oligonucléotide **13** 101

b) Synthèse de l'oligonucléotide **14** 102

c) Synthèse de l'oligonucléotide **15** 103

d) Synthèse de l'oligonucléotide **16** 103

6. Synthèse des châssis peptidiques **17** et **18** 103

a) Synthèse de la Fmoc-azidolysine **19** 104

b) Synthèse des châssis peptidiques **17** et **18** 105

7. Mise au point de la compatibilité entre la chimie KAT et la ligation oxime 114

a) Ligation oxime : formation du composé **I** 117

b) Ligation KAT : formation du composé **J** 119

8. Synthèse de l'édifice **H** 120

a) Synthèse de l'édifice **H1** par ligation oxime 121

b) Voie 1 : synthèse de l'édifice **H2** par réaction KAT 124

c) Voie 1a et 1b : synthèse des édifices **H3** et **H4** par CTC et CuAAC 125

d) Voie 2 : synthèse du composé **H5** par CTC 129

e) Voie 2a et 2b : synthèse des édifices **H3** et **H6** par KAT et CuAAC 131

9. Bilan et perspectives 134

Chapitre V : Synthèse d'édifices G4 contraints d'ADN et d'ARN à une seule topologie et étude d'interaction avec des protéines 138

1. Introduction 139

a) Les protéines favorisant la formation de G4 139

b) Les protéines dégradant les G4 140

c) Les protéines déroulant les G4 : les hélicases 140

2. Présentation des édifices cibles 141

3. Synthèse des édifices G4 et non G4 142

a) Synthèse de l'édifice **Q** 143

b) Synthèse de l'édifice **U** 147

4. Etudes d'interactions G4/Protéines 151

5. Bilan et perspectives 156

Chapitre VI : Conclusions et perspectives 158

Chapitre VII : Partie expérimentale 164

1. Materials and equipment 165

2. Peptides synthesis 165

2.1. General procedure for linear protected peptides synthesis by SPPS 166

2.2. General procedure for cyclisation reaction 166

2.3. General procedure for N-ε Boc or oxyamine deprotection 166

2.4. General procedure for N- ε Dde lysine deprotection 166

2.5. General procedure for coupling chloroacetamide linker on peptide 167

2.6. Experimental procedure 167

3. Oligonucleotides synthesis 177

3.1. General procedure for RNA synthesis 177

3.2. General procedure for DNA synthesis 177

3.3. General procedure for 3'diol DNA and RNA oxidation 178

3.6. General procedure for the addition of the acyl trifluoroborate group 178

3.7. Experimental procedure 179

4. Conjugate synthesis 184

1. General details 184

2. Experimental procedure 185

Références bibliographiques 197

Abréviations

A	Adénosine
AcOEt	Acétate d'éthhyle
AcOH	Acide acétique
ADN	Acide désoxyribonucléique
Alloc	Allyloxycarbonyle
ARN	Acide ribonucléique
ARNm	ARN messager
Bcl-2	B-cell lymphoma 2
Biot	Biotine
Boc	*Tertio*-butyloxycarbonyle
C	Cytidine
CD	Circular Dichroism
CNBP	Cellular Nucleic acid Binding Protein
CPG	Controlled Pore Glass
CTC	Couplage Thiol-Chloroacétamide
CuAAC	Copper Catalyzed Alkyne-Azide
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DCC	*N,N'*-dicyclohexylcarbodiimide
DCM	Dichlorométhane
Dde	1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl
DHX36	DEAH box protein 36
DIEA	*N*,*N*-diisopropyléthylamine
D-loop	Displacement loop
DMAP	4-Dimethylaminopyridine
DMF	*N*,*N*-diméthylformamide
DMTr	Diméthoxytrityle
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
Eei	1-éthoxyéthylidène
eIF4G	Eukaryotic translation Initiation Factor 4 G
ESI-MS	Electrospray Ionisation Mass Spectrometry
Fmoc	9-fluorénylméthyloxycarbonyl
FWJ	Four-Way Junctions
G	Guanosine
G4	G-quadruplexe
GQN1	G quartet nuclease 1
HBTU	*O*-(benzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate
HEPES	Acide 4-(2-hydroxyethyl)-1-piperazineethanesulfonique
HIV	Human Immunodeficiency Viruses
hnRNP	heterogeneous ribonucleoprotein
IEDDA	Inverse Electron Demand Diels-Alder
ISAHC	Imidazole-1-sulfonyl acide hydrochloride
KAHA	α-ketoacid-hydroxylamine amide
KAT	Potassium AcylTrifluoroborate
K_D	Constante d'équilibre de dissociation
MALDI-Tof	Matrix Assisted Laser Desorption Ionisation – Time of Flight
MeCN	Acétonitrile
MeOH	Méthanol
MS	Mass Spectrometry
NCL	nucléoline
NELF	Negative ELongation Factor
NH_4OAc	Acétate d'ammonium

NHC	Carbène N-hétérocyclique
NHE	Nuclease Hypersensitive Element
NHS	N-hydroxysuccinimide
NMM	N-Methylmorpholine
PBS	Phosphate Buffered Saline
PDGFR- *β*	Platelet-Derived Growth Factor Receptor β
PDS	Pyridostatin
PEG	Polyéthylènes glycol
pH_T	pH de transition
PivOM	Pivaloyloxyméthyle
PyBOP	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
RAFT	Regioselectively Addressable Functionalized Template
Rb-1	Retinoblastoma-1
RET	REarranged during Transfection
RHAU	RNA Helicase associated with AU-rich element
RMN	Résonance Magnétique Nucléaire
RP-HPLC	Reverse-Phase High Performance Liquid Chromatography
Rt	Temps de rétention
SEC	Size Exclusion Chromatography
SPAAC	Strain-Promoted Azide Alkyne Cycloaddition
SPANC	Strain-Promoted Alkyne Nitrone Cycloaddition
SPANOC	Strain-Promoted Alkyne Nitrile Oxide Cycloaddition
SPOS	Solid-Phase Oligonucleotide Synthesis
SPPS	Solid-Phase Peptide Synthesis
SPR	Surface Plasmon Resonance
T	Thymidine
TASQ	Template Assembled Synthetic G-quadruplex
tBu	tert-buty
t-BuOH	*tert*-Butanol
TCEP	Tris(2-carboxyethyl)phosphine
TDS	Thermal Difference Spectrum
TEA	Triéthylamine
TEAA	Acétate de triéthylammonium
TERC	Telomerase RNA component
TERRA	Telomeric repeat-containing RNAs
TFA	Acide Trifluoroacétique
TFE	Trifluoroéthanol
THF	Tétrahydrofurane
THPTA	Tris-(hydroxypropyltriazolylmethyl)amine
TIS	Triisopropylsilane
t-loop	Telomere-loop
T_m	Température de dénaturation
TMPyP4	5,10,15,20-Tetrakis-(N-methyl-4-pyridyl)porphyrine
Topo	Topoisomérase
TWJ	Three-Way Junctions
U	Uridine
UPLC-MS	Ultra Performance Liquid-Mass Spectrometry
UTR	Untranslated Transcribed Region
UV	Ultraviolet
VEGF	Vascular Endothelial Growth Factor
WRN	Werner syndrome helicase
WT-1	Wilm's tumor 1
XPB / XPD	Xeroderma Pigmentosum type B /D

Chapitre I

Introduction

1. Les acides nucléiques

C'est au cours du XIXe et début du XXe siècle que les premiers travaux concernant les acides nucléiques sont apparus avec les découvertes de F. Miescher et de R. Altmann notamment[1–3]. En 1929, P. Levene découvre les désoxyriboses et que l'ADN est composé de quatre bases : l'adénine (A), la cytosine (C), la guanine (G), et la thymine (T)[4,5]. Plus tard, E. Chargaff publie des études concernant la composition de différents génomes et observe qu'il y a un pourcentage identique d'adénine et de thymine et un rapport également identique de guanine et de cytosine[6,7]. A l'aide de ces résultats ainsi que des études de M. Wilkins[8] et de R. Franklin[9] qui s'appuient sur des résultats de diffraction de rayons X de cristaux d'ADN, en 1953 les chercheurs J. Watson et F. Crick rédigent la célèbre publication où ils décrivent la structure de la double hélice d'ADN[10]. Cette dernière possède une symétrie de type C2 résultant d'un assemblage de deux chaînes en hélices qui sont en direction opposée. Ils établissent aussi un appariement des bases : A est associée avec T et C est appariée avec G. On retrouve également dans ce modèle que les bases sont espacées de 3,4Å et sont perpendiculaires à l'axe de rotation. Et enfin, le pas complet de la double hélice mesure 34Å et est formé de 10 bases. Ces travaux ont valu à J. Watson, F. Crick et M. Wilkins le prix Nobel de médecine en 1962. Décédée avant l'attribution et ne pouvant pas être décerné à titre posthume, R. Franklin ne figure pas parmi les lauréats.

2. L'acide désoxyribonucléique (ADN) et l'acide ribonucléique (ARN)

Relativement similaire, les brins d'ADN et d'ARN sont tous deux issus d'une polymérisation d'une unité que l'on nomme nucléotide. Ce dernier se compose de trois éléments: la nucléobase, le sucre et le groupement phosphate.

La nucléobase peut présenter un noyau de type purine ou de type pyrimidine. On retrouve dans l'ADN et l'ARN l'adénine (A), la guanine (G) et la cytosine (C). La thymine (T) n'est présente que dans l'ADN alors que l'uracile (U) quant à elle n'est observée que dans l'ARN. Et enfin, les nucléobases possèdent différents sites accepteurs et donneurs de liaisons hydrogènes, indispensables pour l'appariement des bases (**Figure 1**).

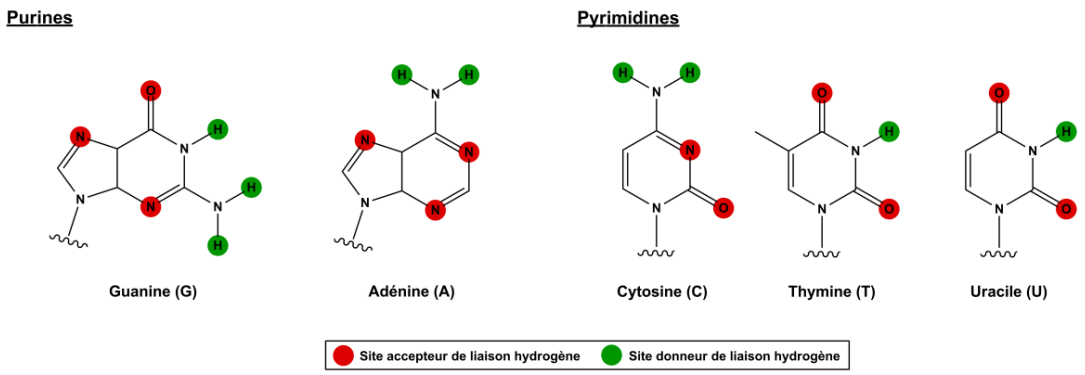


Figure 1 : structures des bases azotées de l'ADN et de l'ARN

Le sucre est un pentose qui est sous la forme 2'-désoxy-D-ribose dans l'ADN et D-ribose dans l'ARN et il est lié à la nucléobase en position C1' par le biais d'une liaison N-glycosidique (**Figure 2a**).

Les différents nucléosides sont reliés entre eux par une liaison phosphodiester. La biosynthèse du polynucléotide (ADN ou ARN) se fait *via* un mécanisme de polymérisation par des polymérases et implique les dérivés 5'-triphosphates des nucléosides (**Figure 2b**).

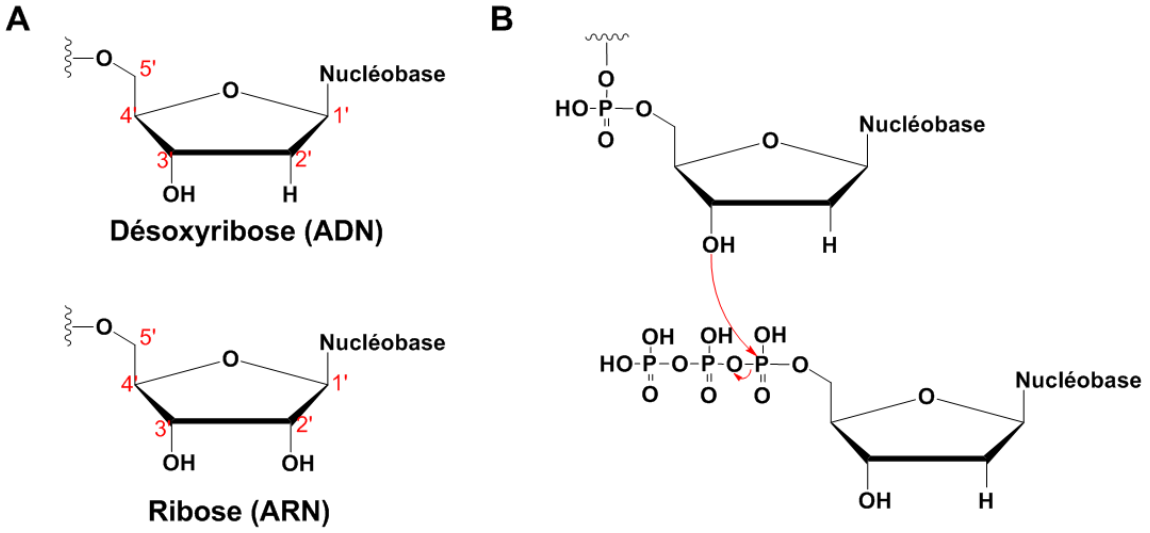


Figure 2 : (a) représentation de la partie saccharidique du nucléotide et (b) mécanisme de formation d'un polynucléotide (ici sous forme ADN)

Différentes conformations peuvent être observées au niveau du nucléotide. En effet, la liaison N-glycosidique peut être orientée soit vers l'extérieur du pentose (conformation *anti*), soit vers l'intérieur du pentose (conformation *syn*) (**Figure 3a**). Concernant le duplexe d'ADN ou d'ARN, la position *anti* est prédominante par rapport à la conformation *syn* pour des raisons d'encombrement stérique. Concernant le pentose, on retrouve deux formes principales, C2'-

endo et C3'-endo qui sont en équilibre entre elles (**Figure 3b**). Ces deux conformations résultent de la position des carbones C2' et C3' selon le plan défini par les atomes C1' et C4'. Si le sucre est un désoxyribose, les deux conformations peuvent être adoptées alors qu'en présence de ribose, l'orientation C2'-endo est généralement défavorisée pour des raisons d'encombrement stérique, des interactions électrostatiques entre l'oxygène 3' et le groupement hydroxyle en 2'[11] et l'effet électro-attracteur de l'hydroxyle en 2'.

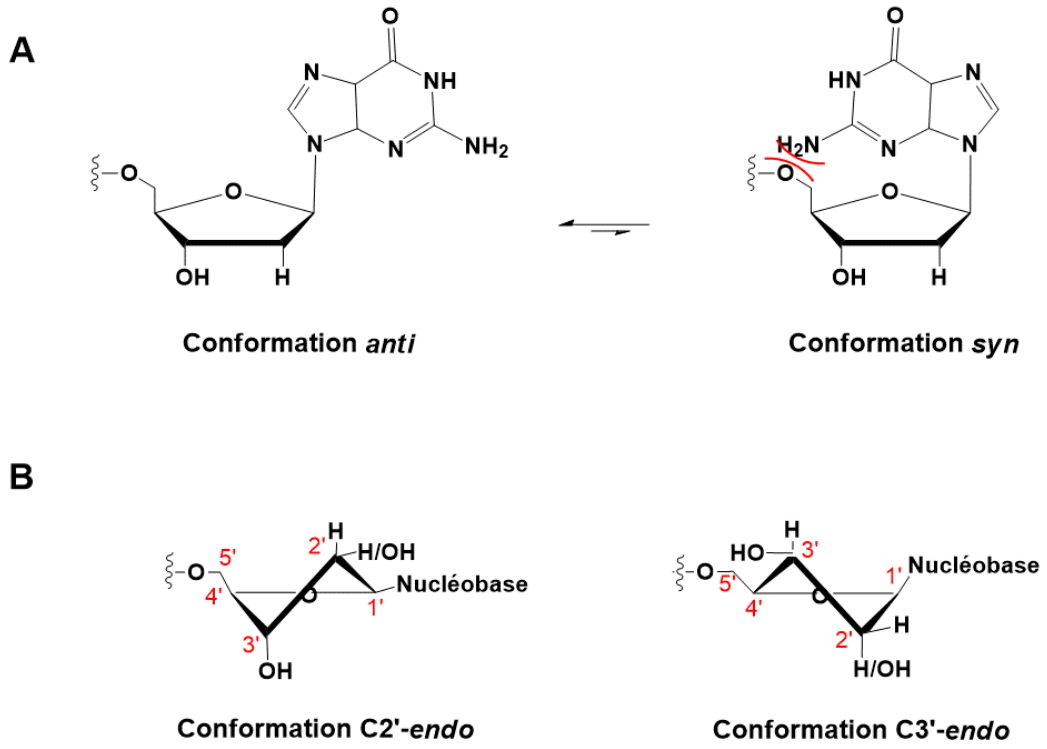


Figure 3 : (a) représentation des conformations *anti* et *syn* du nucléotide avec une guanine et (b) structure du pentose sous sa forme C2'-*endo* et C3'-*endo*

a) Structures usuelles de l'ADN et de l'ARN

L'hélice double brin des acides nucléiques est la structure secondaire la plus connue. On retrouve au sein de cette structure deux polynucléotides complémentaires qui s'apparient par le biais de liaisons hydrogène de type Watson-Crick. Cette association s'effectue entre deux couples de bases, une purine et une pyrimidine, à savoir : l'adénine avec la thymine et la cytosine avec la guanine, créant deux et trois liaisons hydrogènes, respectivement (**Figure 4**). Cette association est la base du modèle double hélice de l'ADN proposé par J. Watson et F. Crick[10].

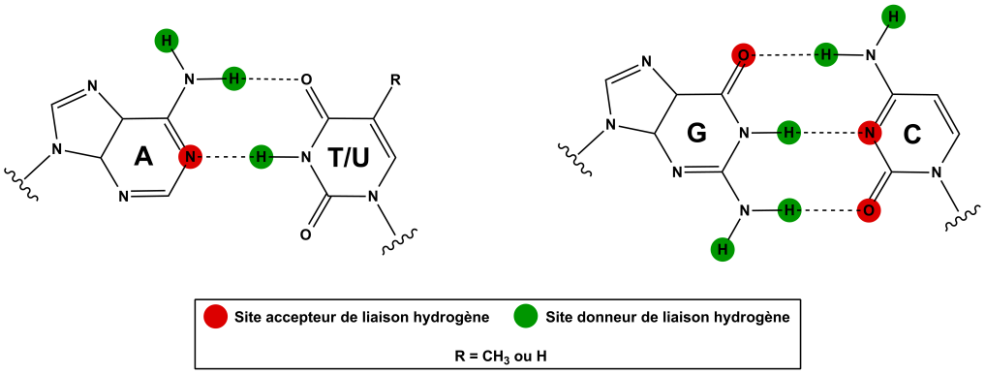


Figure 4 : représentation de l'appariement des bases selon le modèle de Watson-Crick

La topologie de la double hélice d'ADN peut varier et présenter différentes conformations. En effet, les purines et les pyrimidines n'ont pas la même dimension provoquant un appariement asymétrique des bases par rapport à l'axe de rotation. Il en résulte l'apparition d'un petit et d'un grand sillon. Trois conformations de la double hélice d'ADN existent : les hélices A, B et Z (**Figure 5**). Les caractéristiques de ces dernières sont présentées dans le **Tableau 1**. L'ADN B est la forme la plus répandue et connue, alors que les conformations A et Z sont moins prépondérantes. En effet, l'hélice A est observée dans des conditions de faible hydratation et la conformation Z est aperçue dans des séquences riche en bases G-C[12,13].

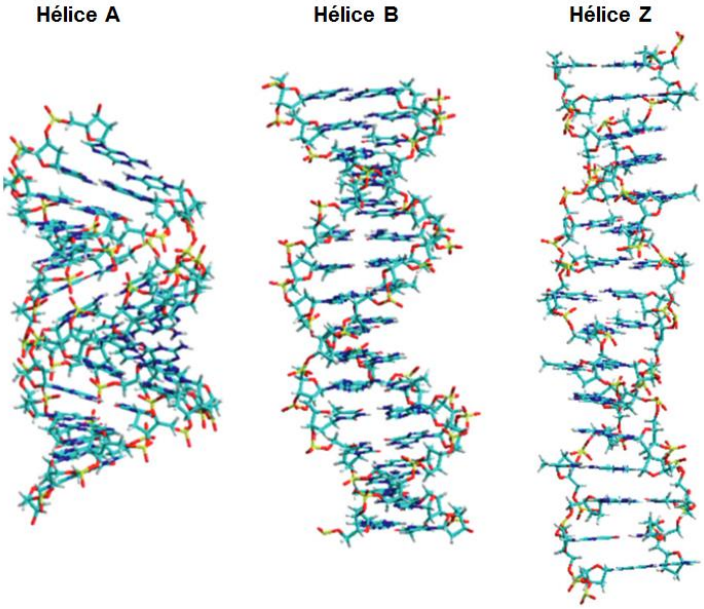


Figure 5 : représentation des conformations A, B et Z de l'ADN. Figure issue et modifiée de Wood B. R. *Chem. Soc. Rev.* **2016**, *45* (7), 1980–1998 [13]

Tableau 1 : caractéristiques des hélices ADN A, B et Z

Paramètres	Hélice A	Hélice B	Hélice Z
Sens	droite	droite	gauche
Pas (Å)	28	34	45
Bases par tour	11	10	12
Distance entres les bases (Å)	2,6	3,4	3,7
Diamètre moyen (Å)	23	20	18
Conformation du désoxyribose	C3'-*endo*	C2'-*endo*	C2'-*endo* : pyrimidine C3'-*endo* : pyrimidine
Conformation de la base	*anti*	*anti*	*anti* : pyrimidine *syn* : purine

Concernant la structure de l'ARN, celle-ci est usuellement simple brin. Toutefois, un repliement intramoléculaire peut également être observé, comme la formation d'une hélice duplexe s'il y a la présence d'une séquence complémentaire. On retrouve ainsi une diversité de structures en ce qui concerne l'ARN et cette hétérogénéité est illustrée à la **Figure 6**[14].

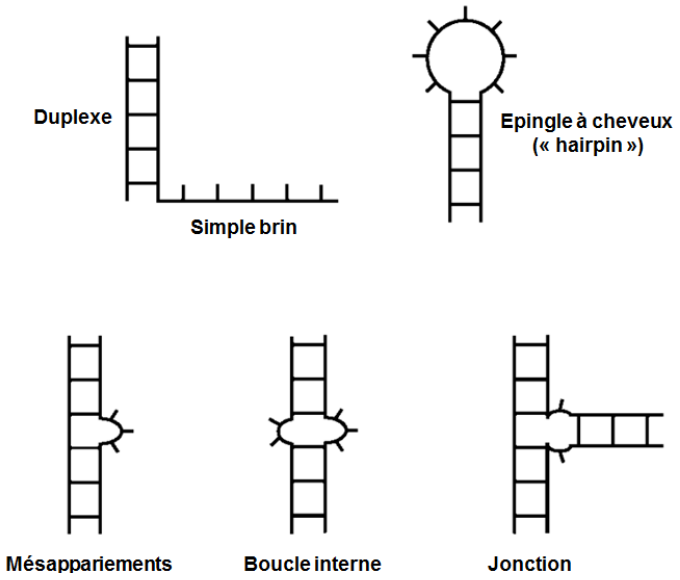


Figure 6 : représentation schématique de différentes conformations possibles de l'ARN. Figure issue et modifiée de Batey *et al*. *Angew Chem Int Ed* **1999**, 38 (16), 2326–2343 [14]

b) Structures non usuelles de l'ADN et de l'ARN

L'environnement ainsi que la nature de la séquence de l'acide nucléique peuvent avoir une influence sur la morphologie et ainsi provoquer la formation de structures non usuelles. On retrouve notamment parmi celles-ci les triplexes, les jonctions à trois ou quatre voies mais également les G-quadruplexes et les i-motifs. Ces deux derniers seront définis et discutés en détail dans la suite du manuscrit.

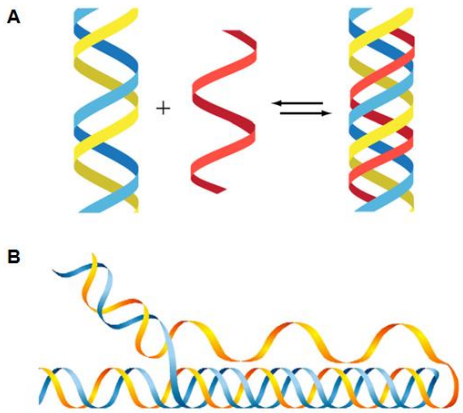


Figure 7 : représentation schématique d'un triplex ADN intermoléculaire (a) et intramoléculaire (b). Figure issue de Jain *et al. Biochimie* **2008**, 90 (8), 1117–1130 [15]

Le triplexe peut se former *via* des interactions intermoléculaire (**Figure 7a**), ou intramoléculaire (**Figure 7b**). La configuration intermoléculaire est due à l'interaction d'un polynucléotide au niveau du grand sillon du duplexe. Dans le cas intramoléculaire, une chaîne polypurine s'apparie avec une chaine polypyrimidine et l'une d'entre elle se replie formant ainsi la structure triplexe. La formation de ce type de morphologie peut s'effectuer avec soit l'ADN, soit l'ARN, soit avec les deux. On y retrouve des interactions de type Hoogsteen ou Hoogsteen inverse entre le troisième brin et l'un des brins du duplex (cas intermoléculaire) ou entre la séquence qui se replie et l'autre brin (cas intramoléculaire)[15]. A noter que les appariements de type Hoogsteen sont des liaisons hydrogène non Watson-Crick (**Figure 8**).

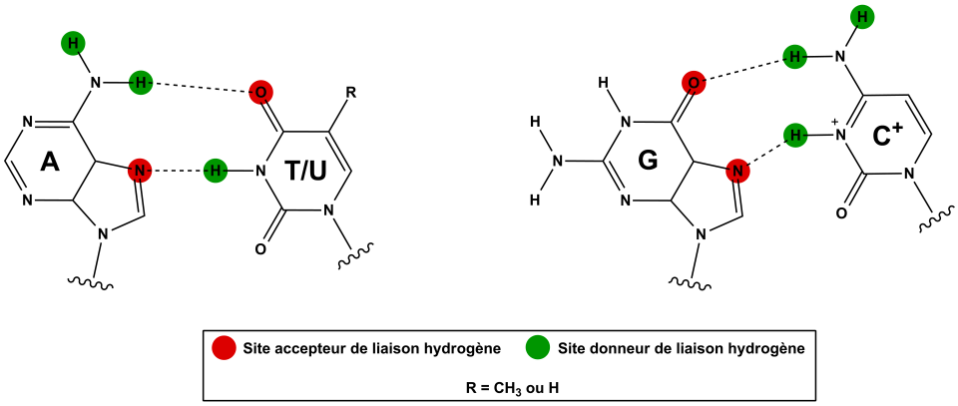


Figure 8 : représentation de l'appariement de type Hoogsteen

Les jonctions à trois voies (« Three-Way Junctions (TWJ) » en anglais) peuvent être vues comme une convergence de trois duplexes en un seul point. On peut y observer un appariement complet ou partiel et dans ce dernier cas, on y retrouve la présence d'une bosse au niveau du point de jonction (**Figure 9**). La présence des nombreux groupements phosphates

provoque un écartement des bras et ce afin de minimiser les répulsions électrostatiques. Dans le cas d'un appariement partiel, on observe que certaines bases sont désappariées au niveau du point de convergence et ceci est dans le but de réduire la tension stérique. De plus, cette conformation peut également varier en présence de cations magnésium vers une structure dite en Y[16].

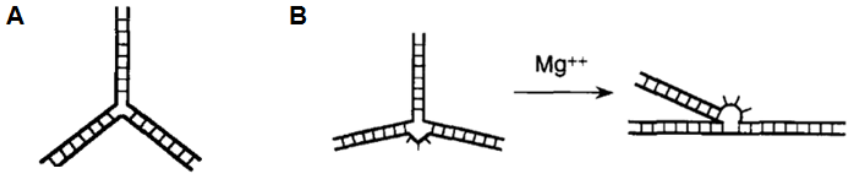


Figure 9 : représentation schématique d'une jonction à trois voies parfaitement appariée (a) et (b) partiellement appariée évoluant vers un repliement (structure en Y) en présence de magnésium. Figure issue et modifiée de Duckett *et al. Biophys. Chem*. **1997**, 68 (1–3), 53–62 [16]

Les jonctions à quatre voies (« Four-Way Junctions (FWJ) » en anglais), également appelés jonctions de Holliday sont le résultat d'une convergence de quatre hélices d'ADN ou d'ARN (**Figure 10**). Ce type de structure est dépendant de la présence de cations divalents tels que le magnésium ou le calcium. En effet en absence de ces derniers, on observe une jonction dite étendue (« extended ») où à l'intersection des quatre hélices le point de jonction est de géométrie carrée. Par contre en présence de cations, on retrouve une forme dite empilée aussi nommée « stacked X structure ». Les jonctions à quatre voies sous forme ADN sont impliquées dans le phénomène de recombinaison homologue, processus essentiel à la vie cellulaire[17].

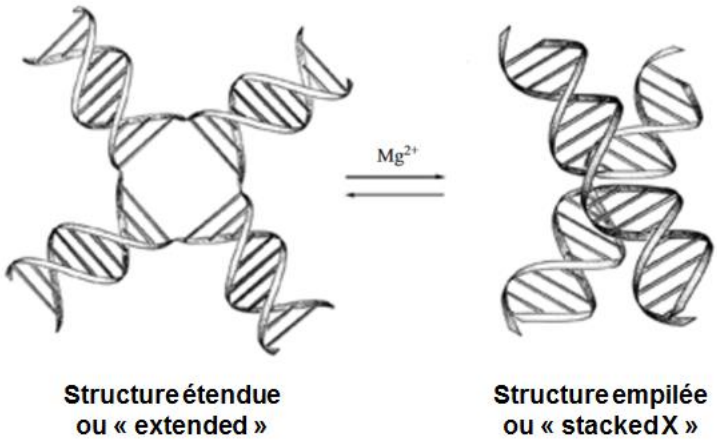


Figure 10 : représentation des structures étendues (ou « extended) et empilées (ou « stacked X ») des jonctions à quatre voies. Figure issue et modifiée de Lilley D. M. J. *Q. Rev. Biophys*. **2000**, 33 (2), 109–159 [17]

3. Les G-quadruplexes (G4)

a) Définition

Les structures G-quadruplexes ou G4 résultent de la capacité de quatre guanines à interagir entre elles dans un même plan par des liaisons hydrogènes selon leurs faces Watson-Crick et Hoogsteen formant ainsi une tétrade de guanine, également appelé G-quartet. Ces derniers peuvent à leur tour interagir entre eux par π–stacking et piéger un cation alcalin tel que Na^+ ou K^+ formant *in fine* une structure G-quadruplexe (**Figure 11**). On retrouve cet auto-assemblage dans des séquences nucléotidiques riches en guanines de type $G_xN_{1-7}G_xN_{1-7}G_xN_{1-7}G_x$ (où G représente la guanine, N peut être n'importe quel nucléotide et x≥3).

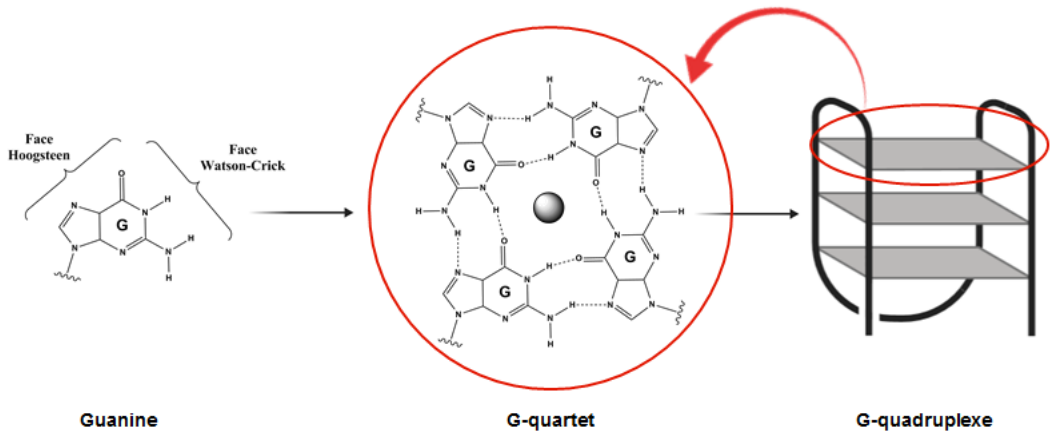


Figure 11 : représentation d'une guanine présentant sa face Hoogsteen et Watson-Crick, d'un G-quartet et d'un G-quadruplexe schématique

Au sein de la structure G4, on y retrouve un nombre de tétrades qui peut varier de deux à quatre usuellement ainsi qu'un cation métallique. Ce dernier participe à la stabilisation en interagissant avec les oxygènes des fonctions carbonyles des guanines. Chaque cation métallique possède un pouvoir stabilisant et il a été observé que le potassium et le sodium stabilisent fortement les G4 ADN en se positionnant, respectivement, entre deux tétrades ou au centre d'une tétrade (**Figure 12**). Le lithium ou encore le césium par exemple possède un très faible pouvoir stabilisant dû à leur rayon atomique ne permettant pas une interaction efficace avec la tétrade[18,19]. A noter que l'ion ammonium possède également un bon pouvoir stabilisant sur la structure G4 et se positionne comme le potassium, entre deux tétrades. Toutefois, il a été observé que l'ion ammonium est trop large pour être intégré au sein des tétrades provoquant une ouverture partielle de ces dernières. Ainsi, la stabilisation entre ce type d'ion et les G4 dépend de la rigidité du G4, elle-même dépendante de la température[19,20].

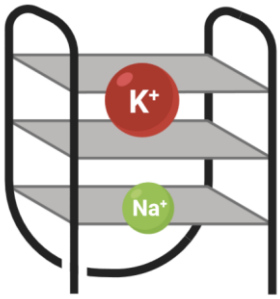


Figure 12 : représentation schématique de la position des cations métalliques potassium et sodium au sein d'un motif G-quadruplexe

b) Les différentes topologies des G-quadruplexes ADN

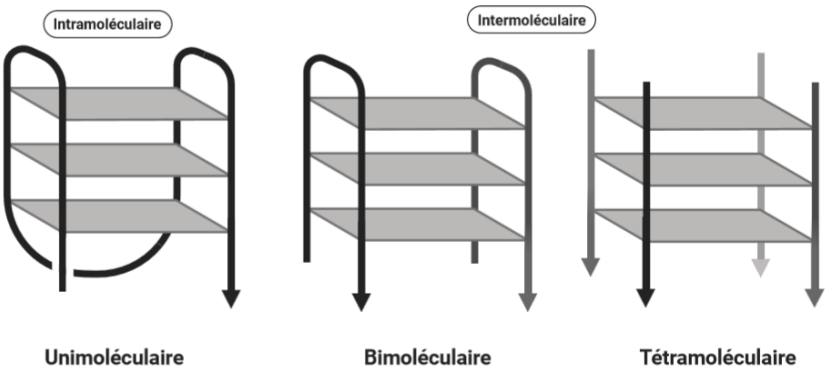


Figure 13 : représentation schématique d'un G4 unimoléculaire, bimoléculaire et tétramoléculaire

Les G-quadruplexes peuvent être formés par un, deux ou quatre brins d'acides nucléiques et dans ce cas, on parle de G4 unimoléculaire (intramoléculaire), bimoléculaire (intermoléculaire) ou enfin tétramoléculaire (intermoléculaire) (**Figure 13**). Une caractéristique importante des G4 est leur variation de topologie et ce peu importe le nombre de brins qui les composent. En effet, selon la direction des brins (parallèle ou antiparallèle) et de l'orientation de la guanine par rapport au sucre (*syn* ou *anti*), on retrouve trois topologies (**Figure 14a**)[19] :

- Dans la topologie **parallèle**, tous les brins sont orientés dans le même sens et toutes les guanines sont positionnées en *anti*.
- Dans la topologie **antiparallèle**, deux brins sont orientés 3'-5' et les deux autres sont en 5'-3'. Dans ce cas-ci, les guanines sont à la fois en *syn* et en *anti*. Il existe deux types de structures antiparallèles, la première dite « chair » et l'autre dite « basket » qui se différencie selon la nature des boucles qui peuvent être latérale, diagonale ou encore « propeller » (**Figure 14b**).

- Dans la topologie **hybride** également appelé « **[3+1]** », trois brins sont dans la même direction et le quatrième est dans le sens opposé à ces derniers et les guanines sont positionnées en *syn* et en *anti*.

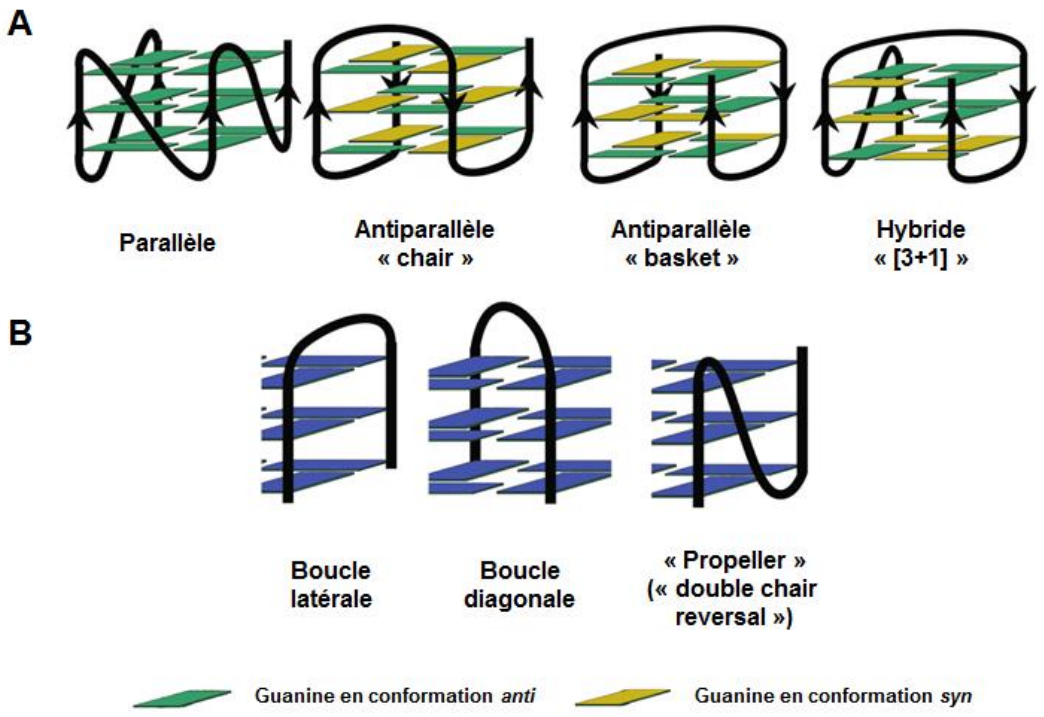


Figure 14 : (a) représentation schématique des différentes conformations d'un G-quadruplex unimoléculaire et (b) des différentes natures de boucles. Figure issue et modifiée de Dolinnaya *et al. Biochem. Mosc*. **2016**, 81 (13), 1602–1649 [19].

De multiples facteurs interviennent dans cette variation de topologie comme par exemple la séquence des brins d'acides nucléiques, la nature et la longueur des boucles, la présence du cation métallique. Ces différentes topologies génèrent des sillons de nature différentes qui peuvent être étroit, moyen ou large selon la conformation des guanines (*syn* ou *anti*)[21,22].

Depuis plusieurs années, différentes études ont démontrées que la formation de structures G4 à partir d'une séquence d'acide nucléique riche en guanine est un processus dynamique faisant appel à différents intermédiaires[23–26] (**Figure 15**). En effet selon le mécanisme proposé[26], la première étape est la formation d'une structure appelée G-hairpin dite 'épingle à cheveux'. Celle-ci est en équilibre avec deux formes, le G4 antiparallèle à deux tétrades et la topologie G-triplexe. Cette dernière est également en équilibre avec deux conformations, soit le G4 antiparallèle, soit la topologie hybride. A noter que la structure G4 antiparallèle à trois tétrades est aussi en équilibre avec sa forme à deux tétrades. La formation

de ce type de structures nécessite la présence de cations métalliques et concernant la structure G-triplexe, il a été observé que des cations monovalents (Na^+ et K^+) et divalents (Mg^{2+} et Ca^{2+}) possédaient un pouvoir stabilisant pour ce type d'édifice[27].

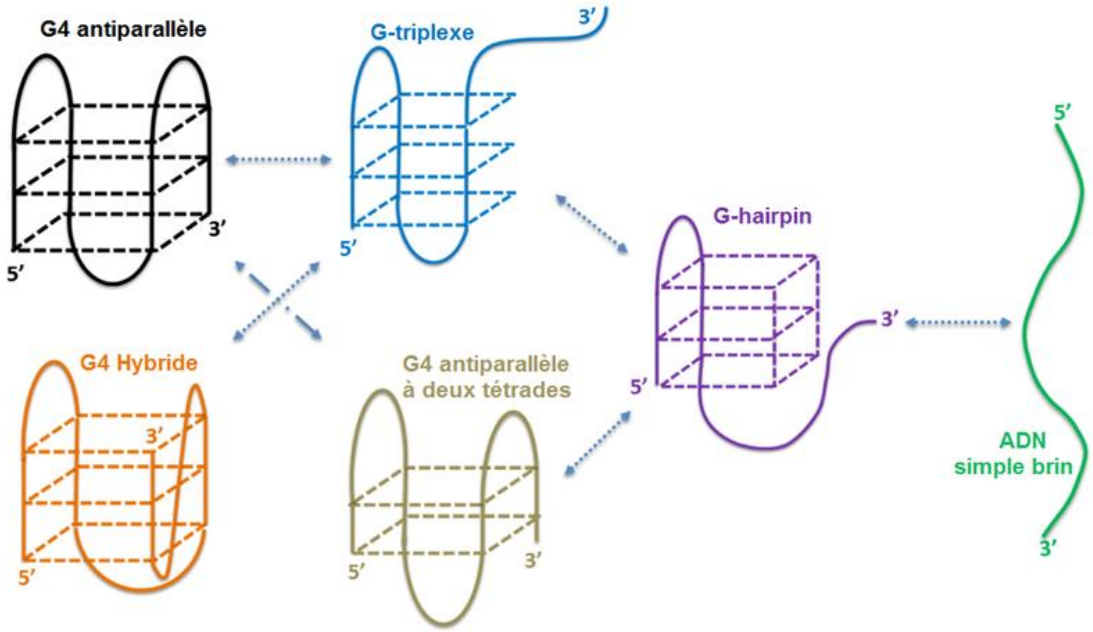


Figure 15 : représentation schématique du mécanisme proposé pour le repliement d'un brin d'ADN en structure G4. Figure issue et modifiée de Hou *et al. Nucleic Acids Res*. **2017**, 45 (19), 11401–11412 [26]

c) Les différentes topologies des G-quadruplexes ARN

Les différences majeures entre l'ARN et l'ADN sont que dans l'ARN, l'uracile remplace la thymine et que le sucre est un ribose, comme vu précédemment. La présence de la fonction hydroxyle supplémentaire sur le pentose va avoir différentes conséquences. Tout d'abord, elle permet la création d'interactions intramoléculaires au sein du G4 ARN causant une augmentation de la stabilité de ce dernier. Ensuite, la présence de ce groupement hydroxyle dans les sillons favorise la capture de molécules d'eau menant également à un gain de stabilité[28]. Et enfin pour des raisons stériques, une orientation *anti* des bases est favorisée limitant la topologie des G4 ARN à une conformation de type parallèle. Ainsi la formation de topologie antiparallèle et hybride de G4 de type ARN ne semble pas être favorisée, confirmée depuis plusieurs années par différentes études[29–31]. Toutefois, en 2014 un aptamère ARN artificiel nommé « Spinach » a été découvert et celui-ci présente une conformation antiparallèle[32]. De plus en 2018, il a été découvert qu'une séquence télomérique ARN possédait une topologie antiparallèle[33] ce qui laisse suggérer une possible extension du répertoire des topologies des séquences ARN. A noter également que dans le cas de structures G4 ARN présentant des

boucles, ces dernières n'affectent pas la topologie qui reste parallèle mais influencent la stabilité du G4[34,35]. En effet, il a été montré que plus les boucles sont longues, plus la stabilité du G4 est amoindrie jusqu'à atteindre un plateau[35]. L'autre paramètre qui différencie les G4 ADN et ARN est leur capacité à interagir avec des cations monovalents ou divalents. En 2016, Guiset Miserachs *et al.* ont montrés que des cations comme le potassium ou le strontium stabilisent les G4 ARN alors que le sodium ne possède cet effet stabilisant[36].

d) Localisations et implications biologiques des G-quadruplexes

Ces dernières années, de nombreux algorithmes ont été développés dans le but de prédire la formation de structures G4 potentielles dans le génome humain[37–41]. Un des premiers, « Quadparser »[37], identifie plus de 370 000 séquences potentielles du génome humain capable de former des G4 *in vivo*. Les critères utilisés dans ces études varient selon les algorithmes. Effectivement, certains ne recherchent que des séquences possédant quatre répétitions de trois guanines, d'autres tiennent compte de la nature et de la longueur des boucles. Ainsi la prédiction de structures G4 au sein du génome est complexe et nécessite la considération de multiples facteurs. Parmi toutes ces séquences, nombreuses d'entre elles peuvent être présentes dans des régions dites 'fonctionnelles', c'est-à-dire là où s'effectuent des processus biologiques. Par conséquent, la formation de structures G4 s'effectue à des moments précis d'évènements cellulaires tels que la transcription, la réplication ou la traduction mais ils peuvent être également aperçus au niveau des télomères[42] (**Figure 16**).

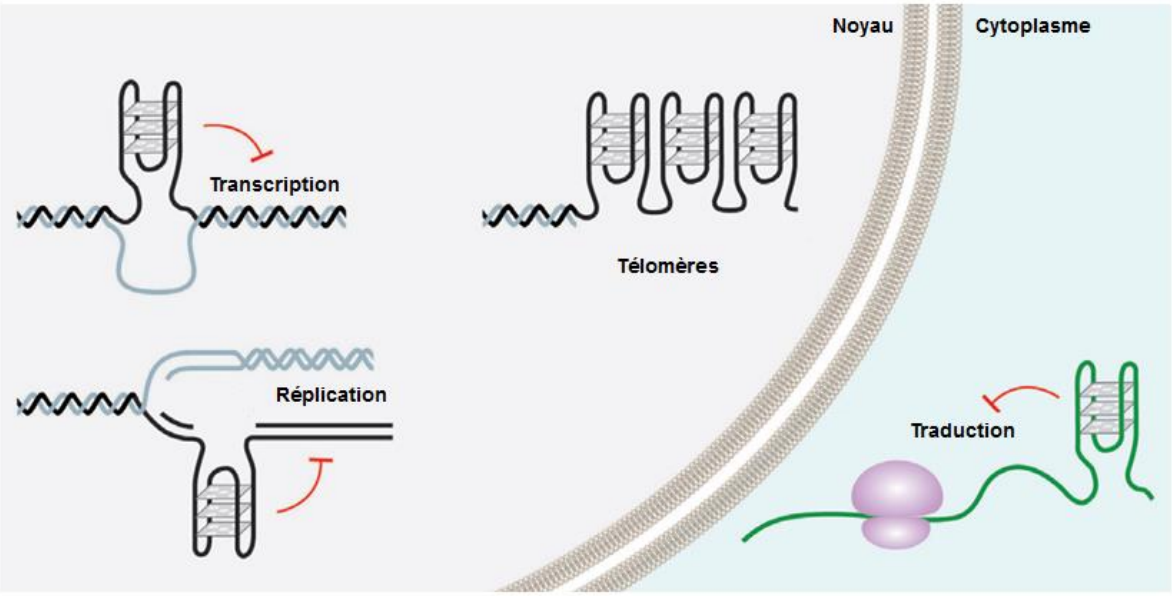


Figure 16 : représentation schématique de l'implication de structures G4 au sein de différents processus cellulaires. Figure issue et modifiée de Rhodes *et al*. *Nucleic Acids Res*. **2015**, 43 (18), 8627–8637 [42]

i. <u>G-quadruplexes télomériques</u>

Dans les organismes eucaryotes, l'ADN est majoritairement retrouvé dans le noyau de la cellule et celui-ci est associé par enroulement à des histones, octamères de protéines en forme de tonneau chargés positivement. Ce phénomène d'enveloppement permet de condenser l'ADN jusqu'à 10 000 fois et ce complexe histone-ADN est connu sous le nom de nucléosome. Ces derniers sont connectés entre eux par des fragments d'ADN et le compactage de nucléosomes forme ce que l'on appelle la chromatine. L'enroulement et l'assemblage de ces unités forme le chromosome où l'on retrouve à l'extrémité de ces derniers, les télomères[43] (**Figure 17**).

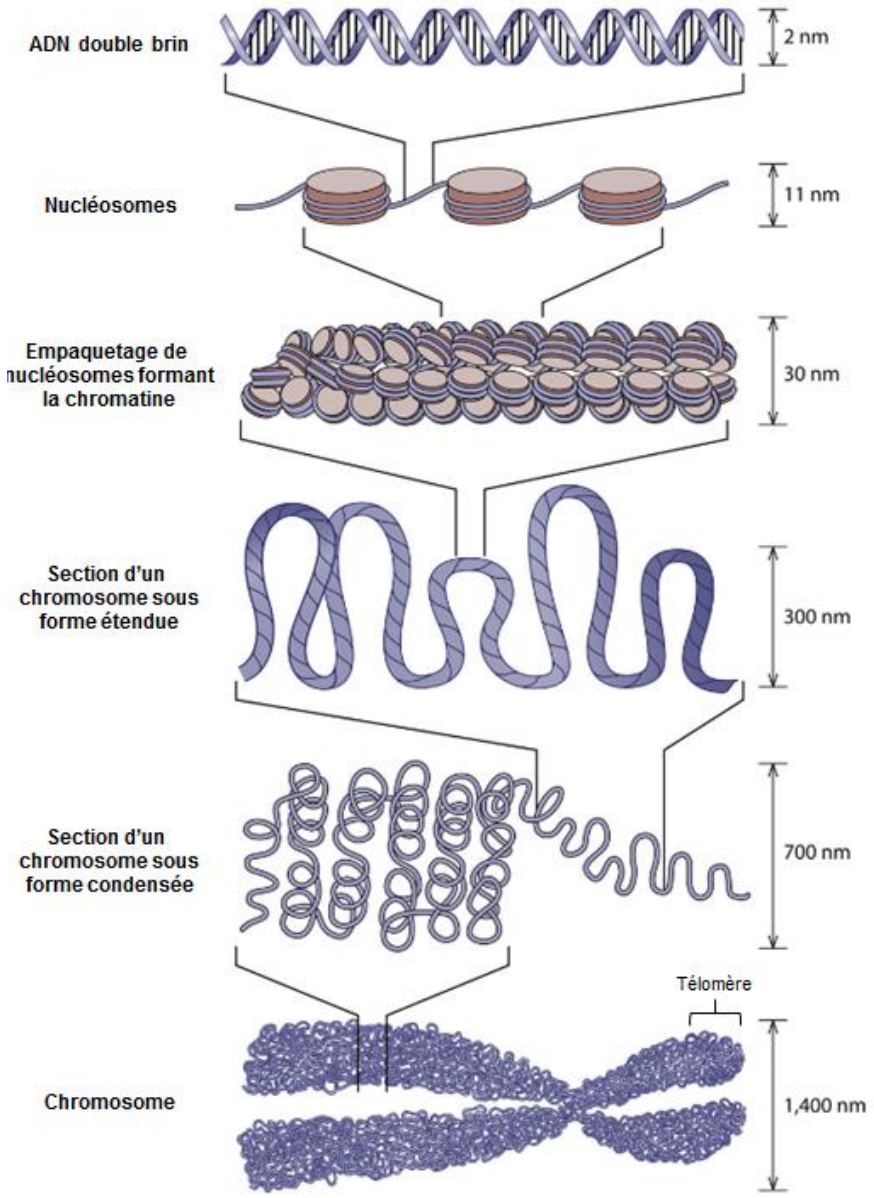


Figure 17 : représentation schématique de la composition d'un chromosome. Figure issue et modifiée de Jansen *et al. Microbiol. Mol. Biol. Rev*. **2011**, 75 (2), 301–320 [43]

Présent au niveau des extrémités des chromosomes eucaryotes, les télomères sont des structures nucléoprotéiques et ont comme rôles de protéger les chromosomes de la dégradation et de la recombinaison. Chez les humains, l'ADN télomérique est majoritairement une séquence double brin, non codante, composé du motif TTAGGG/AATCCC, de longueur variable allant jusqu'à 15 000 paires de bases et possède à son extrémité 3' un simple brin riche en guanines appelé « G-overhang » de longueur variable (150 à 300 bases)[44,45] (**Figure 18**).

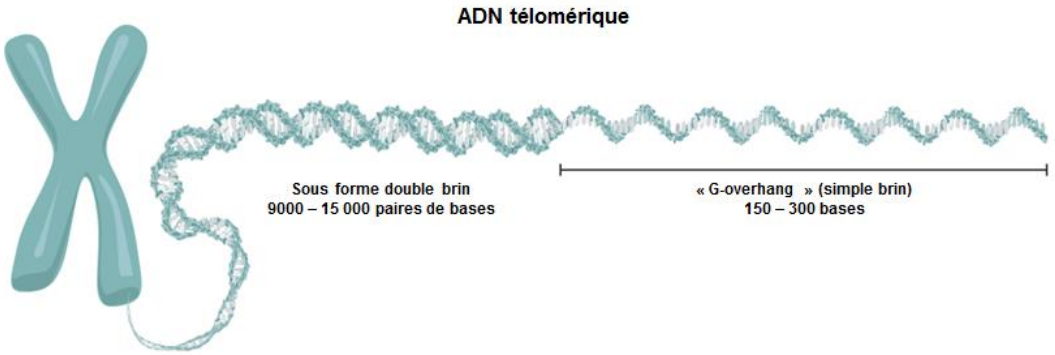


Figure 18 : représentation schématique de l'ADN télomérique. Figure issue et modifiée de Sandin *et al. Curr. Opin. Struct. Biol.* **2014**, 25, 104–110 [44]

Toutefois, la représentation de l'ADN télomérique comme illustré à la **Figure 18** n'est pas linéaire mais possède des boucles. En effet, en 1999 il a été observé pour la première fois par microscopie électronique que les extrémités des chromosomes chez les mammifères sont repliées sur elles-mêmes induisant la formation de deux boucles ou lassos[46] (**Figure 19**). L'une appelée « telomere-loop (T-loop) » composée des deux brins d'ADN riches en G et en C. L'autre nommée « displacement loop (D-loop) » qui est formée suite à l'insertion de l'extension « G-overhang » et qui s'apparie avec une fraction du brin riche en C. A noter que la formation de ces boucles n'est thermodynamiquement pas favorable et qu'on y retrouve l'implication de différentes protéines[45].

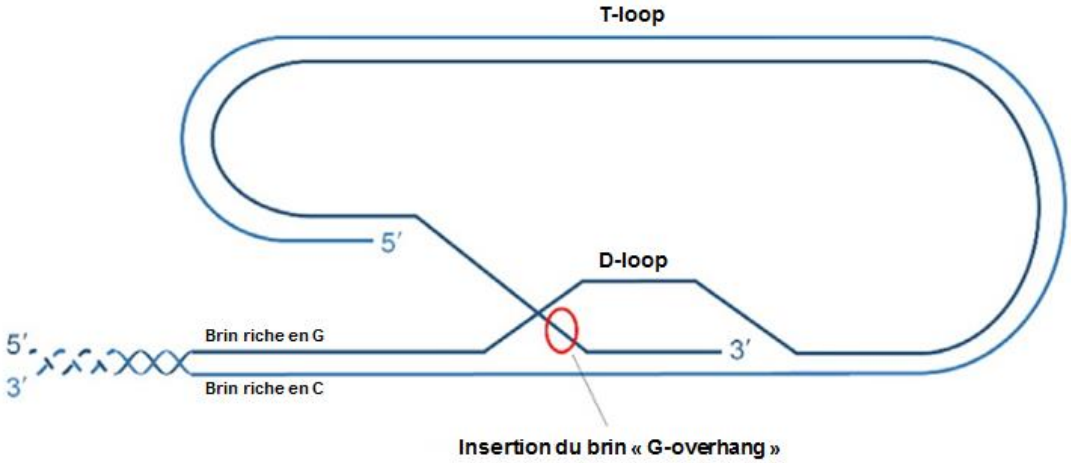


Figure 19 : représentation schématique de la configuration de l'ADN télomérique chez les mammifères. Figure issue et modifiée de Giardini *et al*. *Progress in Molecular Biology and Translational Science*; Elsevier, **2014**; Vol. 125, pp 1–40 [45]

Lors du phénomène de réplication, l'ADN ne peut être répliqué complètement, causant un raccourcissement de l'ADN télomérique à chaque division cellulaire menant *in fine* vers la sénescence réplicative lorsque celui-ci atteint une taille critique (limite de Hayflick). L'un des mécanismes permettant de neutraliser ce raccourcissement des télomères est l'activation d'une enzyme, la télomérase. Découverte par E. Blackburn et C. Greider en 1985 qui leur ont valu le Prix Nobel de médecine en 2009, la télomérase est une ribonucléoprotéine qui maintient la longueur des télomères et est surexprimée dans la plupart des tumeurs humaines et dans les cellules souches mais est peu présente dans les cellules somatiques saines[47]. Toutefois, la région télomérique de par sa richesse en guanine est sujette à la formation de structures G4[48–52] et il a été montré *in vitro* que la présence de G-quadruplexes empêche la télomérase d'effectuer son rôle cellulaire[53]. Ainsi, cibler la formation de structures G4 au sein de l'ADN télomérique est une des stratégies permettant d'induire la sénescence réplicative et/ou l'apoptose des cellules cancéreuses. Un domaine de recherche ayant cet objectif est la conception de petites molécules ou ligands qui permettent la stabilisation de structures G4 télomériques, permettant *in fine* d'inhiber l'activité de la télomérase et causer la mort de cellules cancéreuses. Différents exemples de familles de ligands de G4 seront présentés ci-après (page 20).

ii. <u>G-quadruplexes dans les séquences promotrices</u>

Comme expliqué précédemment, il existe dans le génome humain plusieurs centaines de milliers de séquences capables de former des structures G-quadruplexes. Celles-ci peuvent se former dans les télomères (voir ci-dessus) mais également dans les séquences promotrices (fragment d'ADN sur lequel l'ARN polymérase se fixe afin de réaliser la transcription). Le proto-oncogène *c-myc* est un exemple où une structure G4 parallèle a été observée dans la région riche en G de ce gène[54–56]. D'autres proto-oncogènes montrent également la formation de structures G4 tels que *VEGF*[57], *bcl-2*[58,59] ou encore *c-kit* [60]. La formation de G4 dans ces régions empêche l'interaction entre l'ADN et les facteurs de transcription ce qui peut être vu comme une régulation de la transcription. De plus, de multiples études ont montrées que l'utilisation de ligands spécifiques aux G4 et que l'interaction G4/ligands permet de limiter l'expression de ces gènes[61–63].

La génération de structures G-quadruplexes permet non seulement l'inhibition du processus de transcription mais elle peut permettre également de promouvoir la transcription de certains gènes[64,65]. En effet, lorsque les G4 sont situés en amont du site d'initiation de la transcription, celui-ci induit une inhibition de la transcription (cas précédent) alors que dans

certaines situations, le G4 peut mobiliser des facteurs de transcription de la région simple brin provoquant une initiation de la transcription[65]. Dans le cas où des structures G4 sont formées en aval du site d'initiation de la transcription ceci peut avoir comme conséquence d'empêcher l'ARN polymérase d'effectuer son rôle et ainsi interrompre la transcription[66]. L'impact de la formation de G-quadruplexes au sein de certains gènes sur le processus de transcription est complexe, implique différentes protéines liant des séquences nucléotidiques spécifiques et est de plus en plus étudié ces dernières années. C'est le cas notamment des hélicases, protéines permettant l'ouverture d'acides nucléiques doubles brins (ADN ou ARN). Par exemple, les hélicases XPB et XPD peuvent être mobilisées au niveau de la séquence promotrice et plus précisément là où se forme le G4. Ces dernières déroulent le G4 permettant ainsi la réalisation de la transcription[67]. A noter également qu'il est possible de retrouver des G4 en fin de séquence du gène ce qui laisse supposer que ces derniers peuvent aussi avoir un rôle dans la terminaison de la transcription de certains gènes[68].

iii. G-quadruplexes ARN

Dans les cellules eucaryotes, on retrouve une séquence nommée TERRA, pour « telomeric repeat-containing RNAs (TelRNA) », riche en G, issue de la transcription du brin télomérique riche en C et qui forme des structures G4 ARN *in vitro* et *in cellulo* [69,70]. De plus, il a été montré que la structure G4 de cette séquence présente une topologie de type parallèle[69–71].

Il a également été découvert la formation d'une structure G4 ARN dans la télomérase humaine et plus précisément dans la sous-unité ARN de celle-ci, appelée TERC pour « telomerase RNA component »[72]. De plus, les G-quadruplexes ARN sont fréquemment recensés dans les ARN messagers (ARNm). Des prédictions informatiques ont révélées que des structures G4 ARN potentielles peuvent être présentes dans les régions 5' non traduites de certains gènes (« 5' untranslated region (5'-UTR) ») et sont impliqués dans des processus biologiques tels que la régulation de la traduction[73–75].

iv. Quid des G-quadruplexes *in vivo* ?

Il a été montré précédemment que de multiples études spécifiaient la présence de G-quadruplexes principalement *in vitro* et également *in cellulo*. De plus, différents travaux ont constatés qu'un certain nombre d'hélicases permettent le déroulement de structures G4 télomérique *in vitro* et qu'elles sont nécessaires pour l'intégrité des télomères *in vivo*[76,77]. En

2013, l'équipe de S. Balasubramanian a démontrée l'existence de G4 de type ADN et ARN au sein de cellules vivantes par l'utilisation d'un anticorps spécifique, l'anticorps BG4[78,79]. Récemment, une étude a montré que la spectroscopie RMN peut être utilisée pour investiguer la présence de structures G4 au sein de cellules vivantes et plus précisément dans des cellules d'ovocytes de *Xenopus laevis* (genre d'amphibien)[80]. Tout ceci fournit des preuves circonstancielles de l'existence de ce type de structures. Toutefois, énormément de questions sont encore à résoudre concernant leurs fonctions et leurs mécanismes d'actions *in vivo*.

e) Les G-quadruplexes et leurs ligands spécifiques

Les séquences d'acides nucléiques riches en G pouvant former des structures G4 ne semblent pas être présentes aléatoirement dans le génome et se retrouvent dans les télomères ou encore dans les séquences promotrices de gènes. La formation de structures G4 peut induire différentes réponses cellulaires comme la répression d'un gène ou encore l'inhibition de la télomèrase dans les cellules cancéreuses. Ainsi, concevoir de petites molécules, également appelées ligands, qui interagissent et stabilisent les G4 pourrait aider à découvrir de nouveaux composés possédant des activités (anti-tumorale par exemple).

Les ligands de G-quadruplexes sont de petites molécules ayant la propriété de se lier au G4 avec une grande affinité. Le ligand peut se lier de plusieurs manières au G4. En effet, il peut s'empiler sur les G-quartets externes par interaction de type « π-stacking » mais il peut également s'intercaler dans le G4, se lier au niveau des sillons, des boucles ou effectuer une combinaison de ceux-ci (**Figure 20**).

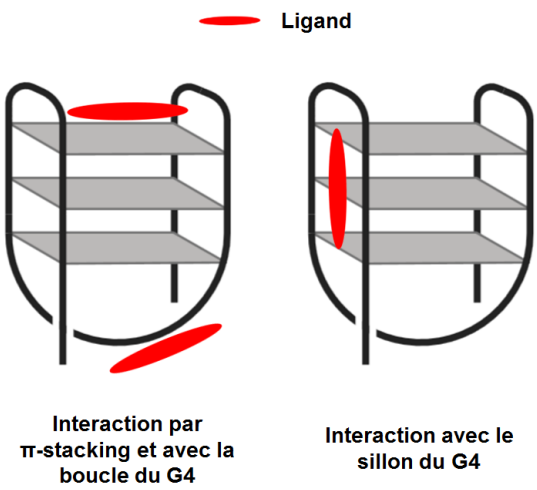
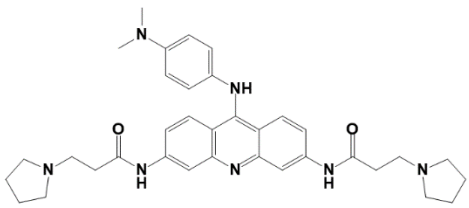


Figure 20 : représentation schématique des différents modes d'interaction entre un ligand et un G-quadruplexe

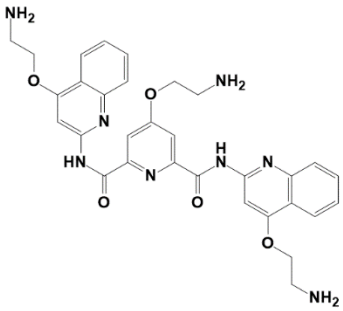
L'interaction G4-ligand peut ainsi s'effectuer par π–stacking, liaisons hydrogènes mais aussi par interactions électrostatiques. Ainsi selon le mode de liaison, on retrouve différentes familles de ligands. Par exemple, dans le but d'interagir de manière optimale par π–stacking avec le G-quartet externe, on retrouve des ligands possédant de larges surfaces aromatiques. Ceci peut conférer une certaine sélectivité du ligand puisque ce dernier étant large ne peut aisément s'intercaler dans l'hélice double brin. L'utilisation de ligands chargés positivement permet l'interaction électrostatique avec l'acide nucléique chargé négativement. Dans ce cas, on n'observe pas de réelle sélectivité pour une structure G4 mais cela contribue à la stabilisation[81–83]. Il existe dans la littérature une pléiade de ligands de G4, il est donc difficile de les énumérer de manière exhaustive, le **Tableau 2** et **Figure 21** présentent quelques ligands de G4 caractéristiques ainsi que leurs activités biologiques.

Tableau 2 : exemples de ligands G4 et leurs activités biologiques

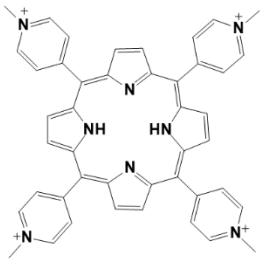
Ligand G4	Activités biologiques
BRACO-19	Inhibition de l'activité de la télomérase[84,85]
Pyridostatin (PDS)	Inhibition de l'activité de la télomérase[86], dommage à l'ADN provoquant un arrêt de croissance de cellules cancéreuse[87]
TMPyP4 (5,10,15,20-Tetrakis-(N-methyl-4-pyridyl)porphyrine)	Favorise la formation de G4 et i-motif[88]. Inhibition de la transcription d'oncogène[89]
Telomestatine	Inhibition de l'activité de la télomerase[90]
PhenDC3	Inhibition de l'activité de la télomerase[91]
PDC 360A	Inhibition de la télomérase[82]



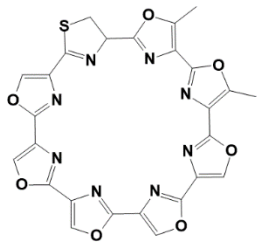
BRACO-19



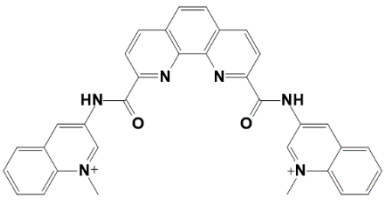
Pyridostatin (PDS)



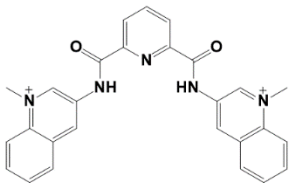
TMPyP4



Telomestatin



PhenDC3



PDC 360A

Figure 21 : représentation de quelques ligands de G4 caractéristiques

4. Les i-motifs

Les i-motifs pour « intercalated motifs », également appelé ADN-i ou i-tétraplexe sont des structures secondaires d'ADN à quatre brins et se forment à partir de séquences d'acides nucléiques riches en cytosines.

En 1992, S. Ahmed et E. Henderson découvrent que la séquence télomérique du brin riche en cytosine des *Tetrahymena* et des *Oxytricha* (genres de ciliés, eucaryotes unicellulaires) se présente sous la forme d'hairpin[92] mais c'est en 1993 que la première structure i-motif a été découverte par Gehring *et al.*[93]. Effectivement, ils montrent qu'un hexamère ADN de séquence d(TC_5) forme à pH acide une structure tétramoléculaire dans laquelle deux duplexes parallèles sont en interactions avec une orientation antiparallèle. L'association se fait par des interactions entre paires de bases semi-protonées $C:C^+$ (**Figure 22**).

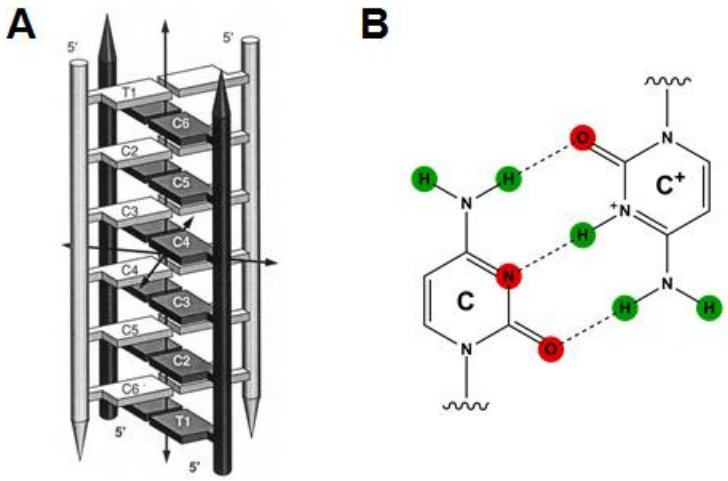


Figure 22 : (a) représentation schématique de la structure i-motif tétramoléculaire de séquence d(TC_5). Figure issue de Gehring *et al. Nature* **1993**, 363 (6429), 561–565 [93]. (b) représentation des interactions au sein de la paire de base $C:C^+$

Suite à cette découverte, différentes structures i-motifs tétramoléculaires ont été caractérisées par RMN et cristallographie[94,95]. Il est notamment observé que la distance entre les paires de bases est de 3,1Å, valeur légèrement inférieure à l'ADN-B (3,4Å) et que l'intercalation des paires de bases mène à une structure possédant quatre sillons, deux larges et deux mineurs[95]. En raison de l'arrangement spatial des paires de bases $C:C^+$, deux topologies principales peuvent être observées concernant les i-motifs, l'une appelée 3'-E (ou forme R), l'autre 5'-E (ou forme S). Lorsque la paire de bases $C:C^+$ la plus externe se situe à l'extrémité 3', on parle de la conformation 3'-E et lorsqu'on la retrouve à l'extrémité 5', on parle de la

topologie 5'-E[96]. La **Figure 23** illustre les deux topologies i-motif 3'-E et 5'-E ainsi que les différentes possibilités d'i-motifs intra- et intermoléculaires. A noter qu'en 1998, Kanaori *et al.* ont montrés que deux structures i-motifs ayant comme séquences d(C_4T) et d(C_3T) sont toutes deux capables d'intervertir lentement leurs conformations 3'-E et 5'-E[97].

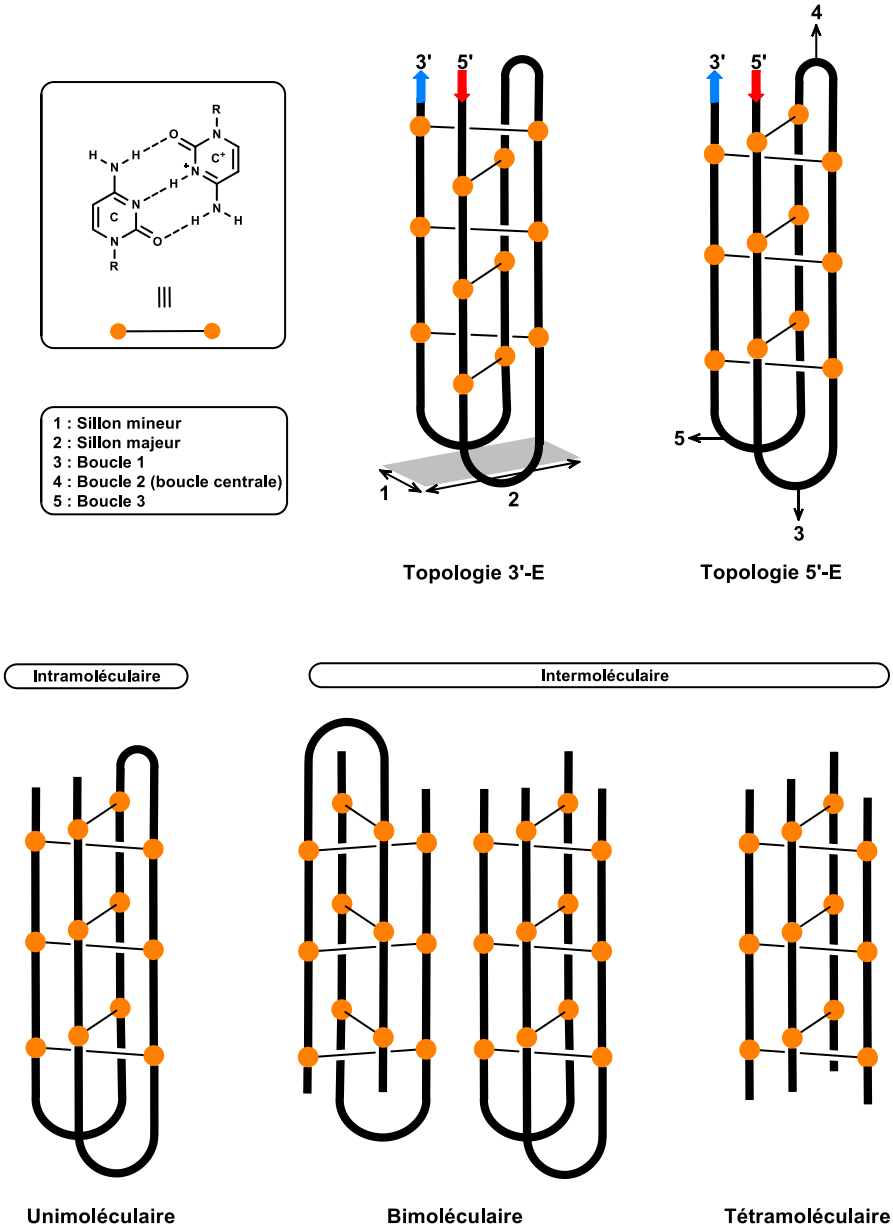


Figure 23 : représentation schématique des deux topologies i-motif 3'-E et 5'-E ainsi que les différentes possibilités d'i-motifs intra- et intermoléculaires

La séquence télomérique humaine d(C_3TAAC_3TAAC_3TAAC_3T) a été résolue par RMN en 2000 par A. Phan, M. Guéron & J-L. Leroy[98]. Celle-ci est composée de six paires de bases C:C^+ intercalées et se présente sous la conformation 5'-E, topologie thermodynamiquement favorable. La configuration 3'-E, topologie cinétiquement favorable, peut également être

observée. L'explication de cette différence provient de la présence d'une interaction T-T supplémentaire au sein de la structure 5'-E[99].

a) Facteurs influençant la stabilité et la topologie du i-motif

De nombreuses études se sont intéressées sur les différents paramètres qui peuvent influencer la stabilité et la topologie des i-motifs. Parmi ces facteurs, on peut notamment retrouver le pH, la nature de la boucle, la force ionique, l'encombrement moléculaire ou encore la nature du pentose.

i. Influence du pH

La formation de structure i-motif nécessite la protonation d'une des deux cytidines formant la paire de base $C:C^+$ ainsi le pH du milieu est un paramètre essentiel. Le pK_a de la cytosine étant de 4,6, il est estimé qu'un pH entre 4 et 7 permet la formation d'une topologie i-motif. A un pH supérieur, les paires de bases $C:C^+$ sont déprotonées, il en résulte une déstructuration menant *in fine* vers la forme simple brin. Lorsque le pH est inférieur à 3, toutes les bases cytosines sont protonées, elles ne peuvent donc plus former de liaisons hydrogènes selon le modèle de la paire de bases $C:C^+$, caractéristique de la structure i-motif[100]. On retrouve ainsi la notion de pH de transition (pH_T) qui désigne le pH à partir duquel on observe la perte de la structure i-motif. Ainsi la valeur du pH_T permet de donner une idée de la stabilité de la topologie i-motif de la séquence étudiée.

ii. Influence de la paire de base $C:C^+$

Les trois liaisons hydrogènes présentes au sein de la paire de base $C:C^+$ jouent un rôle primordial dans la stabilité de la topologie i-motif. Ainsi la présence de modifications au sein de la paire de base $C:C^+$ peut affecter la stabilité de l'édifice. En 2000, L. Lacroix *et al.* ont étudiés la stabilité d'une structure i-motif et l'impact que peut avoir différentes modifications sur la séquence étudiée (5'-TCCTCCT$_4$CCTCCT-3'). Lorsque la cytosine est remplacée par la 5-méthylcytosine, la conformation i-motif est toujours présente et la stabilité n'est pas altérée[101].

En 2014, Wadkins *et al.* se sont intéressés à la séquence du gène *c-myc* et à la modification de la position C5 de la cytosine où l'on pouvait retrouver soit un groupe méthyle ou hydroxyméthyle. Ils ont observés que la méthylation de la cytosine provoque une augmentation de la stabilité contrairement à la présence du groupe hydroxyméthyle qui

déstabilise la topologie[102]. A noter que ces mêmes observations ont été constatées pour la séquence télomérique humaine[103]. Ces quelques exemples illustrent que de légères modifications peuvent affecter la stabilité de la topologie i-motif. Bien entendu, on retrouve dans la littérature d'autres études qui illustrent l'impact de ce paramètre[104–106].

iii. Influence du nombre de cytosines présentes dans la séquence

Récemment il a été observé que plus il y a de cytosine dans la séquence, plus la structure se montre stable[107]. Dans les séquences étudiées, celle possédant le plus grand nombre de cytosines, $C_{10}(T_3C_{10})_3$, montre un pH de transition valant 7,3[107]. Cette tendance a également été constatée avec des séquences oligonucléotidiques dC_n où n est compris entre 10 et 30[108]. Ceci confirmant qu'une topologie i-motif peut tout de même être présente à pH neutre. A noter que d'autres travaux ont aussi démontrés la présence de structures i-motif à pH neutre provenant de certains gènes[109,110].

iv. Influence de la boucle

La composition de la boucle au sein de la structure i-motif est un facteur qui influence sa stabilité. En effet en 2010, Brooks *et al.* se sont intéressés à des séquences riches en C de différents oncogènes et plus précisément sur la composition de ces derniers[111]. La séquence générique de ces oncogènes peut être représentée de la manière suivante : 5'-$C_3XC_3YC_3ZC_3$-3' où X, Y et Z sont les séquences de bases qui séparent les triades C et sont *a fortiori* les boucles qui composent le i-motif intramoléculaire. On peut donc former soit des longues ou des courtes boucles selon la composition de X, Y et Z. Ils ont ainsi différenciés les i-motif selon deux catégories : la classe I et la classe II. La première représente les i-motifs intramoléculaires à boucles courtes et la classe II à boucles longues. Ils ont observés que les i-motifs de la classe I sont moins stables que ceux de la classe II et ce en raison d'un plus faible nombre d'interactions au sein de la boucle.

En 2012, Yang *et al.* ont étudiés l'importance des interactions présentes au sein de la boucle sur la séquence i-motif télomérique humaine. Ils ont montré que lorsqu'on substitue successivement les bases qui composent la boucle par le tétraéthylène glycol, la stabilité décroit progressivement[112].

v. Influence de 'l'ossature phosphate' (« phosphate backbone »)

Au sein du sillon mineur de la structure i-motif, on retrouve notamment les groupements phosphates chargés négativement et la répulsion entre ces derniers doit être contrebalancée pour obtenir un édifice i-motif stable. En 2003, Malliavin *et al.* ont étudié l'effet de la répulsion de ces groupements phosphate sur la stabilité d'un i-motif tétramoléculaire de séquence d(C_4). Cette étude a montré que la topologie 3'-E est plus stable que la 5'-E car elle présente deux interactions supplémentaires favorables au sein de la partie saccharidique[113].

La modification de la partie phosphate peut également induire des variations au sein de la topologie i-motif. Effectivement, l'utilisation de dérivés thiophosphonates permet l'obtention de structures i-motifs intramoléculaire et intermoléculaire[114]. Il a été également comparé l'utilisation de phosphodiester et de thiophosphonate dans de longues séquences (> 16 bases). Il a été observé dans tous les cas la présence de structures i-motifs et ce, même à pH neutre. De plus, il a été constaté que les i-motifs avec les dérivés thiophosphonates présentent une stabilité légèrement plus faible par rapport aux composés phosphodiesters[115].

Une autre étude a montré que l'utilisation d'un dérivé méthylphosphonate ne permet pas la formation d'une configuration i-motif[101]. Ainsi l'ajout d'un groupement légèrement encombrant peut influencer significativement la formation de la structure i-motif.

vi. Influence du dérivé saccharidique

Afin de connaitre l'influence sur la stabilité de la structure en réalisant une modification sur le dérivé saccharidique, l'une des premières études a été de comparer un i-motif sous forme ADN et ARN[116]. Pour rappel, la différence entre les deux est que l'ARN possède une fonction hydroxyle en position C2' du pentose et que la conformation privilégiée est de type C3'-*endo*. La séquence qui a été étudiée est la suivante : 5'-$C_3TC_3T_4C_3TC_3$-3'. Il a été constaté que la séquence ADN présente une stabilité nettement plus importante que la version ARN. Cette différence peut être expliquée par des répulsions existantes entre les groupements 2' OH des pentoses présents dans le sillon mineur causant *in fine* une déstabilisation de l'édifice. La substitution des nucléosides U par des nucléosides dU (c'est-à-dire passer d'un sucre qui possède un groupement 2' OH à un sucre qui n'en contient pas, tout en gardant la base uracile) a permis de confirmer que c'est bel et bien la présence du groupement 2' OH dans cette configuration qui cause l'instabilité de la structure. Cette observation a été confirmée par une

autre étude en utilisant des i-motifs hybrides, c'est-à-dire des séquences mixtes ADN/ARN et celles-ci se sont montrées moins stables que leurs homologues ADN[116,117].

Une autre étude a montré que le remplacement du désoxyribose par l'arabinose (où le groupement 2' OH pointe vers le sillon majeur) ne perturbe pas la formation de structure i-motif[118] alors que la présence d'un groupement 2'-*O*-méthyl la déstabilise fortement[116].

vii. Influence de la force ionique, de l'encombrement moléculaire et de la présence de cations

Il a été constaté que la force ionique de la solution pouvait influencer significativement la stabilité de la structure i-motif. En effet, Mergny *et al.* ont montré qu'à pH ≥ 4,8, l'augmentation de la concentration en NaCl (allant de 0 à 100mM) déstabilise le i-motif. Au-delà de 100mM, aucune augmentation de déstabilisation n'est observée[119]. Cette tendance a également été constatée pour le gène *n-myc*[120].

Le facteur d'encombrement moléculaire (« crowding ») peut aussi influencer la stabilité de l'édifice i-motif. Dans le but d'imiter l'environnement que les acides nucléiques peuvent ressentir au sein de la cellule, des agents moléculaires peuvent être utilisés. Ces derniers possèdent des masses molaires élevées et on retrouve par exemple les polyéthylènes glycol (PEGs). Des conditions d'encombrement favorisent la stabilisation de structures G4 et i-motif par rapport au duplex ou à la structure simple brin[121,122].

Le groupe de Z. Waller a montré que la présence de cations Ag^+ et Cu^+ permettaient la stabilisation des édifices i-motifs, et ce même à pH neutre[123,124]. Ces cations se positionnent comme le proton, au centre de la paire de base C:C. Une récente étude a montré qu'en présence d'ions Mg^{2+} et dans des conditions d'encombrement moléculaire, la structure i-motif se voit stabilisée à pH physiologique[125].

b) Localisations et implications biologiques des i-motifs

Précédemment, il a été vu qu'un nombre important de séquences du génome humain pouvait former des G-quadruplexes. Ainsi en principe, le brin complémentaire de ces séquences riches en G est susceptible de former des structures i-motifs. Tout comme les G4, celles-ci peuvent être retrouvées au niveau des télomères mais également dans les séquences promotrices.

Le groupe de Z. Waller a utilisé l'algorithme « Quadparser » (algorithme utilisé au départ pour déterminer théoriquement le nombre de séquences pouvant former des G4) pour estimer le nombre de séquences potentielles i-motifs[107]. Leurs critères de recherche sont des séquences possédant quatre fragments de cytosines séparés par un certain nombre de nucléotides allant 1 jusqu'à 19. Ils ont ainsi découvert plus de 5000 séquences potentielles dont 637 dans les séquences promotrices de gènes. Le groupe de C. Burrows a également analysé le génome humain mais plus particulièrement les séquences de type dC_n (où n est compris entre 15 et 81) et ils ont trouvés 769 séquences potentielles[108]. Ces dernières sont retrouvées dans les promoteurs de gènes mais également dans les introns. Ces deux études confirment qu'un certain nombre de séquences issues du génome humain peuvent former des structures i-motifs et que celles-ci ne sont pas situées aléatoirement ce qui laisse suggérer qu'elles peuvent être impliquées dans certains processus biologiques.

i. Les i-motifs télomériques

Comme mentionné précédemment, on retrouve dans les télomères une séquence riche en cytosines capable de former une structure i-motif. La séquence télomérique humaine du brin riche en C a été caractérisée par RMN en 2000[98]. Concernant celle-ci, il a été montré que dans des conditions légèrement acides, la formation de structure i-motif était privilégiée par rapport au duplex[126]. De plus, différentes protéines interagissant avec des i-motifs télomériques ont été également identifiées[127,128]. On retrouve par exemple la protéine ASF1/SF2, protéine associée dans l'épissage d'ARNm et impliquée dans la suppression d'introns.

ii. Les i-motifs dans les séquences promotrices

Différents travaux ont montrés qu'au sein du génome humaine, un certain nombre de séquences peuvent potentiellement former des structures i-motifs[107,108]. Parmi ces séquences, on retrouve notamment des gènes tels que *c-myc*, *bcl-2* ou encore *c-kit* et ces structures i-motifs sont impliqués dans la régulation du processus de transcription.

Le proto-oncogène *c-myc* est impliqué dans différents processus biologiques tels que la croissance cellulaire. La dérégulation de ce gène est caractéristique de différents cancers (poumons, pancréas, etc). Au sein du gène *c-myc* on retrouve notamment sept zones appelées « nuclease hypersensitive element (NHE) » que l'on peut définir comme une portion de gène vulnérable aux nucléases. Parmi ces sept NHE, le fragment NHE III_1 contient un brin riche en cytosines capable de former des i-motifs et est également responsable à 90% du contrôle de la

transcription du gène *c-myc* [129,130]. Lors du processus de transcription de ce gène, plusieurs protéines viennent interagir avec cette portion NHE III_1 induisant l'activation de la transcription. On retrouve notamment parmi ces protéines, hnRNP K pour « heterogeneous ribonucleoprotein K » qui est un facteur de transcription interagissant spécifiquement avec le brin riche en C de la région NHE III_1 de *c-myc*. Cependant, il a été émis comme hypothèse que lorsqu'une structure i-motif se forme dans cette région, cela induit une altération de l'interaction entre hnRNP K et le brin d'ADN, affectant ainsi la transcription de ce gène. Dans cette étude, la structure i-motif s'est formée à pH neutre et dans des conditions dites « negative superhelical stress ». A noter que des effets similaires ont été observés avec la formation de G-quadruplexes au sein de la région NHE III_1 du gène *c-myc* [130,131].

L'oncogène *bcl-2* pour « B-cell lymphoma 2 » est impliqué dans la régulation de la mort cellulaire en étant anti-apoptotique, c'est-à-dire empêchant l'apoptose. On retrouve dans ce gène une région riche en C appelée « P1 promoter ». La séquence contient six fragments d'au moins trois cytosines et peut former une structure i-motif de classe II, c'est-à-dire possédant de longues boucles et une haute stabilité (pH_T = 6,6)[111,132]. La protéine WT-1 est connue pour interagir avec la région P1 de *bcl-2* et pour être une protéine répressive de la transcription de ce gène. Le groupe de L. Hurley a émis l'hypothèse que la formation de structures secondaires (G4 et/ou i-motifs) empêche l'interaction entre WT-1 et le brin d'ADN, permettant ainsi l'expression du gène *bcl-2* [111]. Quelques années plus tard, ce même groupe a démontré que la séquence riche en C de la région P1 du gène *bcl-2* est en équilibre entre deux formes, une structure hairpin et i-motif[133]. Lorsque la structure hairpin est stabilisée par un ligand par exemple, il est observé une répression de la transcription alors que si la séquence est sous la forme i-motif, l'expression de la transcription est constatée. A noter que dans ce deuxième cas, l'activation de la transcription est rendue possible grâce à la protéine hnRNP LL, protéine faisant partie de la famille des « heterogenous ribonucleoprotein », qui suite à l'interaction avec le brin riche en C déroule la structure i-motif et induit la transcription du gène. Ceci est un autre exemple où la formation de structures secondaires d'ADN permet l'expression d'un gène.

Une structure i-motif peut également se former à l'extrémité 5' du gène *retinoblastoma* (*Rb-1*). Ce dernier code la protéine du rétinoblastome connue pour exercer un contrôle négatif sur le cycle cellulaire et sa régulation est perturbée dans presque tous les cancers[134,135]. L'édifice i-motif formé au sein de ce gène est de classe I, c'est-à-dire qu'il possède de petites boucles et son pH_T vaut 5,9[111,135]. De plus, il a été constaté que cette séquence pouvait former deux topologies i-motifs intramoléculaires[135].

La séquence promotrice du proto-oncogène *RET* pour « REarranged during Transfection » est également connu pour former des structures G-quadruplexes et i-motif intramoléculaire de classe I avec un pH_T valant 6,4[136]. Un édifice i-motif intramoléculaire de classe I possédant un pH_T de 5,8 a été constaté dans la séquence promotrice du gène *VEGF* pour « Vascular Endothelial Growth Factor »[137].

L'oncogène *c-kit* présente dans sa région promotrice une séquence capable de former une structure i-motif stable dans des conditions légèrement acides (pH 6,8)[138,139]. Il a été démontré que les bases C et G présentes dans les boucles influencent grandement la topologie de l'édifice surtout dans le cas où le pH est proche de 7. En effet, lorsque ces bases sont substituées par T, il a été observé la perte de la structure i-motif à pH 6,8 alors qu'à pH 5, la stabilité de l'édifice n'est que très peu modifiée[139]. Ceci illustrant encore une fois l'impact que peut avoir la nature de la boucle sur la stabilité.

iii. Quid des i-motifs *in vivo* ?

La pertinence biologique des i-motifs a souvent été remise en cause en raison d'un manque de preuves *in vivo* de ce type de structures et du fait qu'un pH acide soit nécessaire pour la formation de ces édifices. Toutefois, très récemment deux études ont démontrées la présence d'i-motifs au sein de cellules humaines vivantes[140,141]. La première décrit l'utilisation de la spectroscopie RMN au sein de cellules vivantes dans le but d'étudier la stabilité des i-motifs. Différentes séquences promotrices humaines ont été pré-structurées et transfectées dans des cellules HeLa. Des signaux RMN caractéristiques des i-motifs ont été détectés jusqu'à 35°C et ceux-ci persistent dans leur environnement. Ceci laisse suggérer que des structures i-motifs peuvent être présentes dans un environnement cellulaire[140]. La seconde étude implique l'utilisation d'un anticorps appelé « iMab » capable de se lier spécifiquement et avec une grande affinité aux séquences riches en C pouvant former des i-motifs. Tout d'abord, il a été démontré qu'iMab peut se lier à différentes topologies bien connues d'i-motifs et n'interagit pas avec d'autres structures ADN telles que le duplex, l'hairpin ou encore le G4. Ensuite par immunofluorescence, il a été mis en évidence la présence de structures i-motifs au sein du noyau cellulaire. Cette étude a également constatée que le nombre de structure i-motifs détecté variait selon le cycle cellulaire ce qui laisse suggérer que la formation d'i-motifs s'effectue à des moments précis du cycle de vie de la cellule et notamment lorsque l'activité transcriptionnelle est élevée[141].

c) Les i-motifs et leurs ligands spécifiques

L'utilisation de ligands interagissant avec des structures i-motifs peut être envisagée dans le but de réguler des processus biologiques tels que l'expression de certains gènes ou l'activité de la télomérase. Dans la littérature, on retrouve de multiples exemples de ligands interagissant avec ce type de structure. En effet, des i-motifs intramoléculaires (séquence télomérique) et intermoléculaires (bimoléculaire et tétratomléculaire) interagissent avec la porphyrine cationique TMPyP4[88,142]. Des dérivées phénanthrolines[143], acridines[144] et berberines[145] peuvent également se lier avec la séquence i-motif télomérique. D'autres petites molécules ont été signalées pour interagir avec des structures i-motifs (séquences télomériques et oncogènes). C'est le cas de la mitoxantrone, la tilorone, la tobramycine, la phénazopyridine, l'amodiaquine, l'harmalol, la quinalizarine et la minocycline tyrothricine[146]. Des ligands G4 bien connus comme BRACO-19, Phen-DC3 ou encore la pyrodistatin (PDS) peuvent également interagir avec les i-motifs[147]. La **Figure 24** illustre les différents ligands i-motifs mentionnés précédemment.

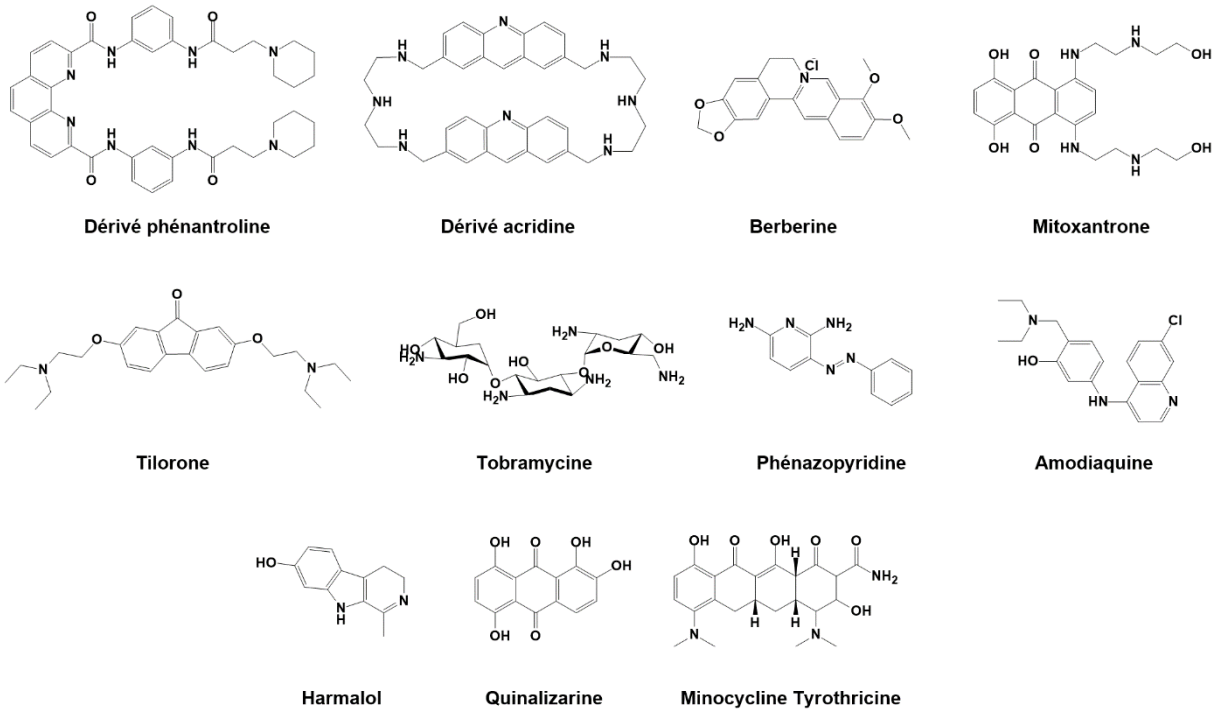


Figure 24 : représentation de ligands interagissant avec des structures i-motifs

Toutefois, l'inconvénient majeur de ces ligands i-motif est la faible sélectivité envers la topologie i-motif *versus* G-quadruplexe. Ainsi, la conception d'un ligand i-motif sélectif nécessite des informations détaillées sur la structure i-motif ciblée afin de pouvoir créer une molécule aux caractéristiques uniques permettant d'interagir avec des zones spécifiques du i-

motif. On retrouve quatre zones où des interactions ligand/i-motif sont envisageables (**Figure 25**) : interaction au niveau des sillons ; intercalation du ligand au sein de la structure ; interaction au niveau des boucles ; interaction du ligand par empilement (« end stacking »).

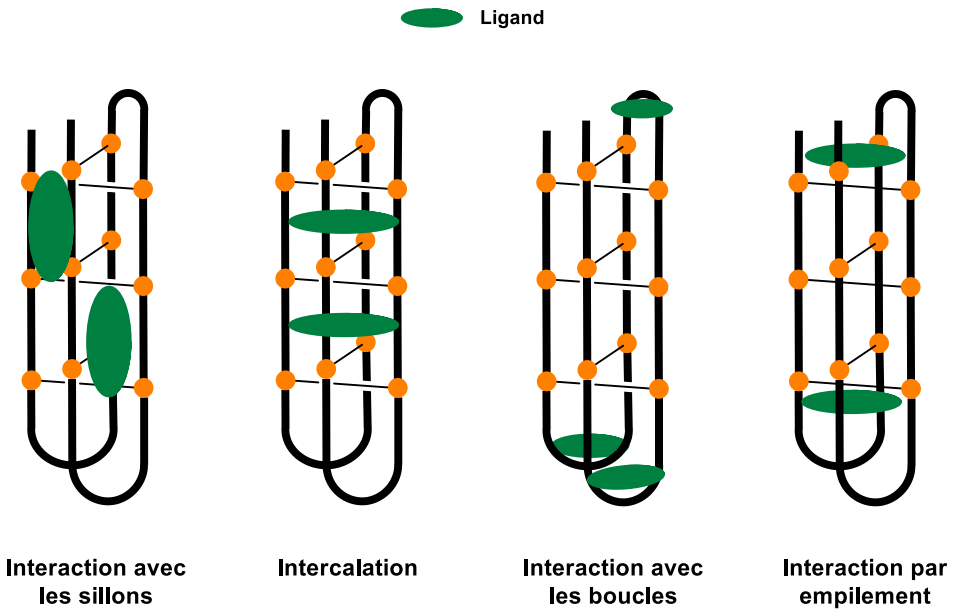


Figure 25 : représentation schématique des différents modes d'interaction entre un ligand et un i-motif

Par l'utilisation de ligand spécifique aux i-motifs, il est possible de réguler des processus biologiques. Un premier exemple a été publié en 2012 avec la formation et la stabilisation de i-motif télomériques par l'utilisation de nanotubes de carbone simple-feuillet et ce à pH physiologique. Il en résulte l'inhibition de l'activité de la télomérase menant *in fine* à la senescence et l'apoptose dans des cellules cancéreuses *in vitro* et *in vivo*[148]. L'expression de certains gènes peut également être contrôlée par le biais de petites molécules. Cela a été effectué par exemple pour l'oncogène *bcl-2* où l'utilisation de ligands permettait soit l'activation ou la répression du gène[133]. Un autre exemple, l'oncogène *PDGFR-β* (« platelet-derived growth factor receptor β ») contient un fragment NHE riche en cytosine capable de former un i-motif de classe II et le ciblage de ce dernier par un ligand de type benzothiophène a été envisagé. La stabilisation du ligand sur la structure i-motif provoque la répression du gène *PDGFR-β*[149]. L'oncogène *c-myc* a également été ciblé par Shu *et al.* en 2018. Ils ont montrés que l'utilisation d'un dérivé acridone permettait la régulation négative du gène *c-myc*[150]. En 2017, Debnath *et al.* ont dévoilés deux ligands peptidomimétiques, l'un spécifique à la reconnaissance d'i-motifs et l'autre spécifique aux G4[151]. Ils ont également démontrés que la stabilisation de l'édifice i-motif régule positivement l'expression du gène *bcl-2* alors que la stabilisation du G4 induit l'effet contraire.

5. Techniques de caractérisation des structures G-quadruplexes et i-motifs

De multiples techniques permettent d'obtenir des informations sur la topologie et la stabilité des structures secondaires d'ADN et d'ARN. On retrouve par exemple la spectroscopie d'absorption UV avec notamment la technique « Thermal Difference Spectra (TDS) », le dichroïsme circulaire (CD), la résonnance magnétique nucléaire (RMN) ou encore la cristallographie aux rayons X.

La technique « Thermal Difference Spectra (TDS) » permet de caractériser la structure d'une séquence oligonucléotidique par le biais du spectre d'absorption UV. De manière générale, les spectres d'absorption de l'acide nucléique sont effectués à basse et à haute température (au-dessus de la température de dénaturation de la séquence étudiée (T_m)) afin d'obtenir le spectre du composé dans sa forme structurée et déstructurée. La différence de ces deux derniers fournit un spectre possédant une forme unique renseignant sur la structure de l'acide nucléique. La **Figure 26** illustre l'allure du spectre pour des acides nucléiques de type duplexes (**Figure 26a**) et G-quadruplexes (**Figure 26b**)[152]. En particulier les spectres TDS des ADN G-quadruplexes sont caractéristiques et montrent une bande négative intense vers 292 nm.

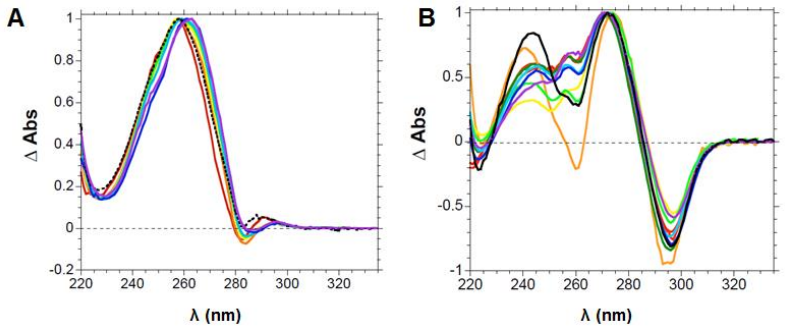


Figure 26 : représentation de l'allure des TDS pour un ensemble d'acides nucléiques de type duplexes (a) et de type G-quadruplexes (b). Figure issue et modifiée de Mergny *et al. Nucleic Acids Res*. **2005**, 33 (16), e138 [152]

La caractérisation de structures secondaires d'ADN et d'ARN peut également être effectuée par dichroïsme circulaire (CD). Cette technique est basée sur la capacité qu'une molécule ou un matériau absorbe différemment la lumière selon sa polarisation (circulaire droite ou circulaire gauche). Ainsi l'une des deux composantes peut être absorbée plus rapidement que l'autre. Le dichroïsme circulaire est couramment utilisé pour des analyses de structures, des changements de conformation et pour la détermination de la température de

dénaturation (T_m) d'acides nucléiques[153]. Cette méthode renseigne sur la topologie adoptée par le G4 ou le i-motif grâce à la présence de signaux caractéristiques. Une conformation parallèle présente des signaux spécifiques, un maximum à 260 nm et un minimum à 240 nm. Une topologie antiparallèle quant à elle est caractérisée par deux maxima à 240 et 290 nm et un minimum à 260 nm. Enfin, la conformation hybride présente deux maxima à 260 et 290 nm et un minimum à 240 nm[21]. La **Figure 27** illustre l'allure des spectres CD pour les trois types conformations. De plus, cette technique est non-destructive et nécessite très peu de quantité de produit (une concentration de quelques µM est suffisante)[21].

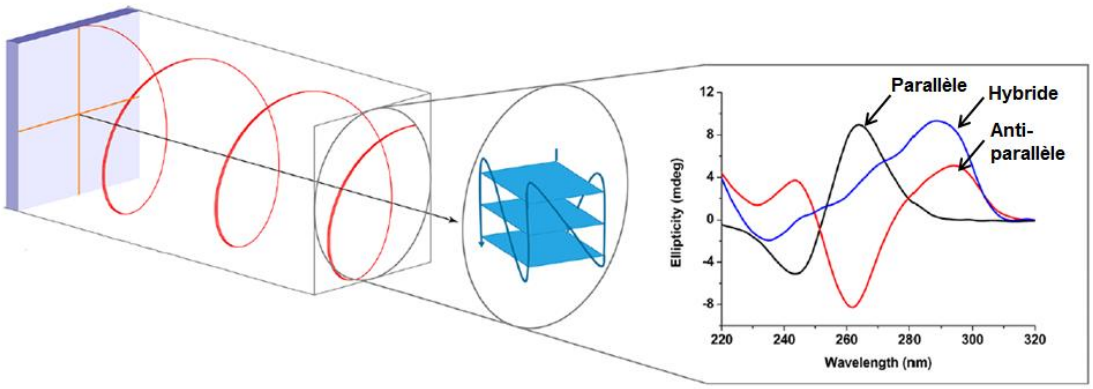


Figure 27 : représentation schématique de la technique dichroïsme circulaire (CD) et de l'allure des spectres CD des trois conformations possibles de G4. Figure issue et modifiée de Carvalho *et al. J. Chem. Educ.* **2017**, 94 (10), 1547–1551 [153]

La résonance magnétique nucléaire (RMN) est également utilisée pour caractériser les G-quadruplexes et les i-motifs. Cette technique est également non-destructive mais nécessite une quantité plus importante de produit (concentration de l'ordre du millimolaire) comparé au dichroïsme circulaire. Les G-quadruplexes et les i-motifs possèdent une signature particulière en RMN. En ce qui concerne les G4, le proton de l'azote N1 des guanines du G-quartet possède un déplacement chimique dans les environs 10-12 ppm. De plus, le proton impliqué dans la formation du G-quartet est moins échangeable et est alors assez fin sur le spectre ^{1}H-RMN et visible dans un solvant aqueux[154]. Concernant les i-motifs, ceux-ci présentent des pics caractéristiques en RMN du proton vers 15 ppm. La caractérisation d'une structure par RMN nécessite la présence d'une espèce stable et majoritaire. En effet, s'il y a la présence de plusieurs conformations en solution, l'interprétation du spectre peut s'avérer complexe. Des modifications de la séquence peuvent également être apportées dans le but de limiter cette hétérogénéité[21]. Plusieurs séquences d'acides nucléiques ont été résolues par RMN comme la séquence télomérique G4[155] et i-motif[98] ou encore l'oncogène *c-myc* [61].

La cristallographie aux rayons X permet également de caractériser de manière précise la topologie de structures G4 ou i-motifs. Par cette méthode, des informations à l'échelle atomique peuvent être obtenues pour une topologie bien précise. Cependant, cette technique ne permet pas déterminer si la séquence étudiée présente un polymorphisme et si celle-ci est l'espèce majoritaire en solution. La séquence télomérique G4 a été caractérisée par cristallographie aux rayons X[156], de même qu'un ensemble d'autres séquences d'acides nucléiques[157].

Chapitre II

Présentation du projet de thèse

Les séquences d'acides nucléiques riches en guanines peuvent adopter une multitude de topologies en solutions. Ce phénomène de polymorphisme est dû à plusieurs facteurs comme la nature de la séquence des brins d'acides nucléiques, la nature et la longueur des boucles ou encore la présence du cation métallique[21,22]. Ce polymorphisme est exacerbé dans la séquence d'ADN télomérique du brin riche en G qui peut présenter de multiples conformations[158] comme illustré dans la **Figure 28**.

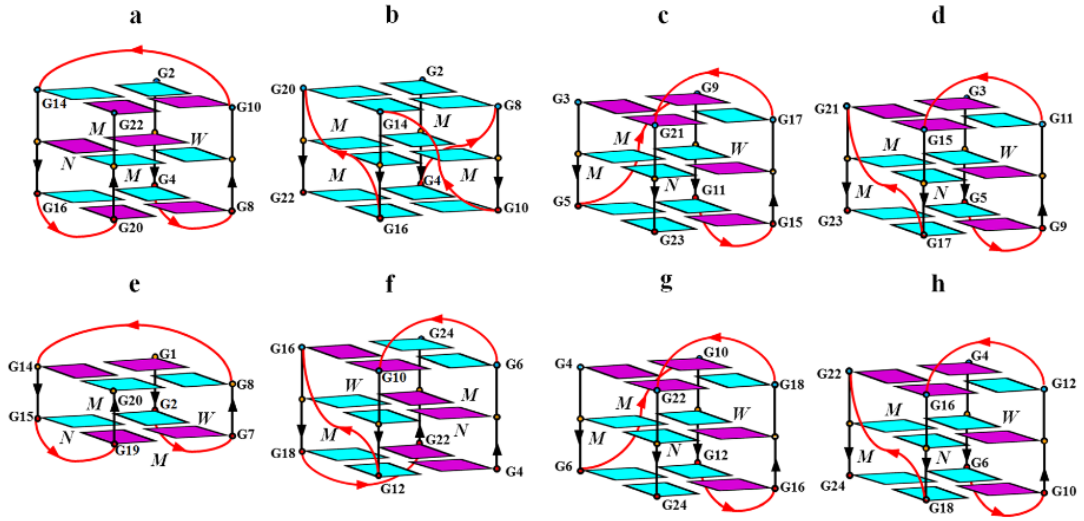


Figure 28 : représentation schématique des différentes conformations possibles de la séquence télomérique humaine. Figure issue de Liu *et al. Biophys. Rev.* **2019**, 11 (3), 389–393 [158]
a) conformation antiparallèle « basket » ; b) conformation parallèle ; c) conformation hybride forme 1 ; d) conformation hybride forme 2 avec la séquence d(TA(GGGTTA)$_3$GGGTT) ; e) conformation antiparallèle à deux tétrades ; f) conformation antiparallèle (2+2) ; g) conformation hybride forme 1 avec la séquence modifiée d(AAA(GGGTTA)$_3$GGGAA) ; h) conformation hybride forme 2 avec la séquence d(TTA(GGGTTA)$_3$GGGTT).

De par leurs implications biologiques, les G-quadruplexes ont depuis ces dernières années suscité un intérêt grandissant et notamment dans le domaine d'identification de protéines interagissant avec les G4 ou encore dans la conception de petites molécules liant les G4 et affectant différents processus cellulaires. Cependant, la plupart de ces ligands ne montrent pas une spécificité suffisante pour une topologie bien particulière de G4. De plus lors de ces études d'interactions *in vitro*, cette caractéristique de polymorphisme est exaltée. Effectivement, il a été montré que de la modification d'un facteur expérimental comme la nature du cation ou la concentration moléculaire peut impacter la conformation du G4[159]. Ainsi, la topologie adoptée par les G4 au sein du milieu cellulaire est difficile à déterminer et est encore à l'étude.

Dans ce contexte, la conception d'édifices G4 possédant une conformation bien définie se montre intéressante et pertinente. Ces dernières années, notre laboratoire a développé un système permettant d'obtenir des édifices tétraplexes d'acides nucléiques avec une topologie bien définie. Ce concept appelé TASQ pour « Template Assembled Synthetic G-Quadruplex » utilise un support peptidique rigide permettant de greffer successivement les séquences d'acides nucléiques et obtenir *in fine* une conformation bien particulière. Ce substrat peptidique est un cyclodécapeptide également appelé « RAFT » pour « Regioselectively Addressable Functionnalised Template » développé dans les années 1990 par Mutter *et al.* [160,161] (**Figure 29a**). Le RAFT se compose de dix acides aminés et possède deux liens glycine-proline coudés induisant à la structure d'être pratiquement rigide. La longueur et la largeur du support sont de 10 Å et 5 Å respectivement. Le RAFT présente deux faces distinctes : une face supérieure où les chaines latérales de quatre lysines distantes de 7-8 Å sont orientées et une face inférieure sur laquelle la chaine latérale d'une lysine est orientée. Ces chaines latérales peuvent être fonctionnalisées afin d'y greffer diverses molécules. Pour le concept TASQ, la face supérieure est utilisée pour lier des oligonucléotides par le biais de réactions chimiosélectives et former une structure tétraplexe. Quant à la face inférieure, la lysine permet de greffer une biotine permettant l'immobilisation de l'édifice lors d'études d'interactions (**Figure 29b**).

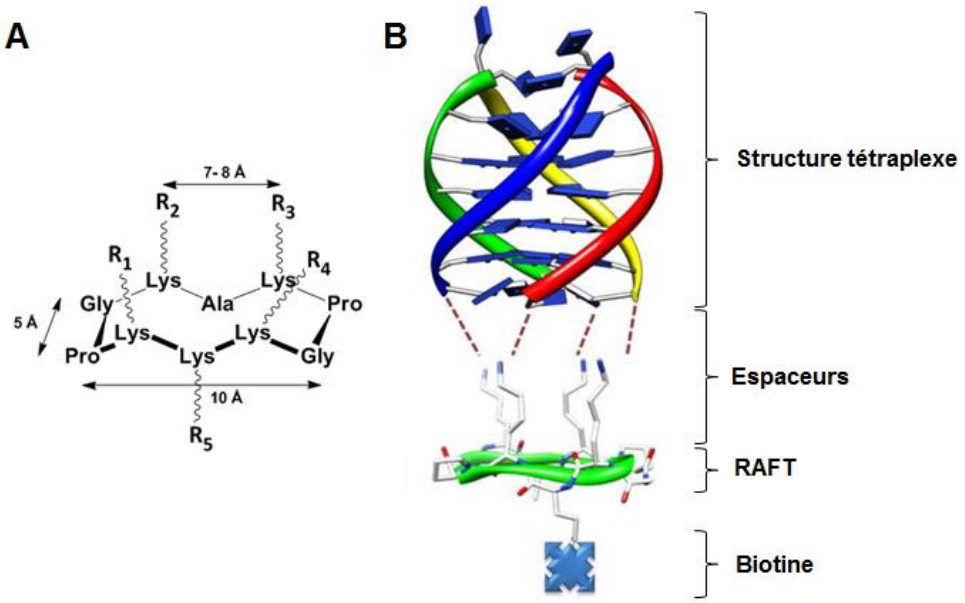


Figure 29 : (a) représentation de la structure RAFT et (b) représentation schématique du concept « Template Assembled Synthetic G-Quadruplex » (TASQ)

Notre laboratoire a pour but depuis de nombreuses années de concevoir des mimes G-quadruplexes et i-motifs possédant une conformation bien définie par l'utilisation du concept TASQ. Ces édifices représentent des cibles biologiques pertinentes étant donné que les séquences immobilisées sont dérivées de la séquence télomérique ou d'oncogènes. Le greffage de ces séquences sur le support peptidique s'effectue à l'aide de différentes ligations chimiosélectives couramment utilisées dans notre laboratoire. La **Figure 30** illustre différents mimes obtenus ces dernières années au sein de notre équipe. On y retrouve des édifices tétramoléculaires ADN, ARN, ADN/ARN et i-motif mais également des composés trimoléculaires et bimoléculaires, tous dérivés soit de la séquence télomérique ou d'origines virale (HIV)[162–167].

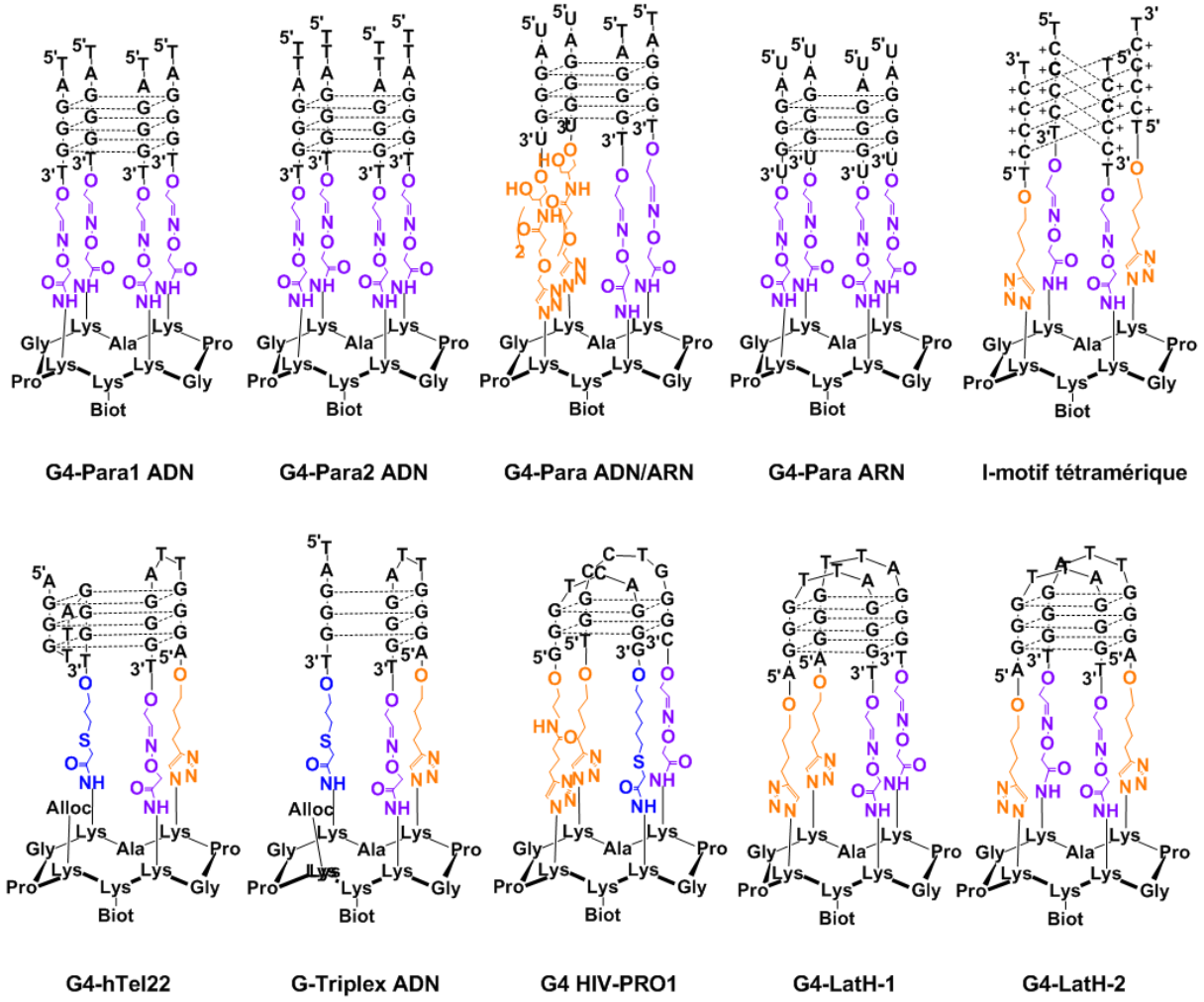


Figure 30 : représentations de différents mimes G4

La conception d'édifices tétraplexes ADN et ARN fait appel à une ou plusieurs ligations chimiosélectives. Au départ, il a été conçu des G4 tétramoléculaires parallèles utilisant qu'une seule ligation chimique[162]. Par la suite, d'autres édifices sollicitant deux ou trois ligations

chimiosélectives ont été synthétisés[163–167]. Parmi celles-ci, on retrouve la ligation oxime, la cycloaddition CuAAC et le couplage thiol-chloroacétamide (CTC). L'utilisation de plusieurs ligations orthogonales permet de diversifier mais également d'envisager l'obtention d'édifices tétraplexes plus complexes. Ces travaux ont permis à la fois de valider le concept TASQ mais aussi d'utiliser ces édifices comme des outils afin de déterminer l'affinité ainsi que la sélectivité d'un ligand pour une topologie bien définie de G4. Par exemple, il a été montré que ligand TmPyP4 possède une bonne affinité pour les G4 (K_D = ~200-300nM) mais n'a pas de réelle sélectivité entre des G4 parallèles ADN et ARN. Le ligand BRACO-19 quant à lui possède une sélectivité plus importante pour les G4 ADN (K_D = 8nM) plutôt que les G4 ARN (K_D = 261nM). Ceci peut être expliqué par le fait que les G4 parallèles ADN et ARN possèdent des sillons différents qui vont influencer sur l'interaction entre le G4 et le ligand[165]. Il a été également montré que les ligands PDS et PhenDC3 possèdent une excellente affinité (K_D < 25nM) pour **G4-HIV-PRO1** et **G4-LatH-1** mais aucune sélectivité n'est observée entre ces deux derniers[166]. De plus, récemment il a été montré que les ligands BRACO-19, PDS et PhenDC3 possèdent une bonne affinité pour l'édifice **G4-hTel22** et **G-Triplex ADN** avec des valeurs de K_D allant de 5,3 à 318nM mais ils ne possèdent pas de réelle sélectivité entre les deux édifices[167].

Toutefois, l'utilisation de plusieurs ligations chimiosélectives permet certes de compléter le panel d'édifices tétraplexes mais ne garantit pas la certitude d'obtenir totalement le produit désiré. En effet pour les composés **G4-HIV-PRO1**, **G4-LatH-1** et **-2**, la dernière étape lors de la synthèse de ces derniers est la fermeture des boucles par la cycloaddition CuAAC qui peut se dérouler soit de manière 'parallèle' (c'est-à-dire que les deux boucles sont parallèles l'une par rapport à l'autre ; produit désiré) ou 'croisée' (c'est-à-dire que les deux boucles s'entrecroisent entre elles lors de la fermeture des boucles sur l'édifice ; produit non désiré). Il a été démontré lors de cette dernière étape qu'une pré-organisation de l'édifice en topologie antiparallèle effectuée avant l'ajout des différents réactifs permet l'obtention du composé désiré. Cependant, il est difficile de confirmer que 100% de la fermeture s'est déroulée de manière 'parallèle'. Deux possibilités sont envisageables pour éviter cette situation. La première est d'utiliser trois ligations chimiosélectives et d'en réutiliser une en cours de synthèse en réalisant une modification sur l'un des substrats. La seconde stratégie est d'utiliser en tout quatre ligations chimiosélectives afin de greffer successivement les oligonucléotides pour obtenir l'édifice désiré. Ces deux possibilités ont été envisagées au cours de ce projet de thèse. En effet, les objectifs ont été de développer de nouveaux mimes G-quadruplexes et i-motif

contraints mais également d'implémenter une quatrième ligation chimiosélective, orthogonale aux trois autres déjà mises en place au sein de notre laboratoire.

Ce travail de thèse est réparti en trois chapitres. Le premier consiste en la synthèse d'un mime i-motif bimoléculaire antiparallèle dérivé de la séquence télomérique (édifice **F**, **Figure 31a**). La synthèse du mime **F** implique l'utilisation de trois ligations chimiosélectives orthogonales. Les détails de la synthèse ainsi que les caractérisations effectuées seront présentés dans le chapitre III. La seconde partie consiste en l'incorporation et l'utilisation d'une nouvelle ligation chimiosélective dans la chimie des oligonucléotides. Ainsi au sein du chapitre IV, il sera développé la méthodologie de cette nouvelle ligation pour les oligonucléotides mais également l'implémentation de celle-ci avec les trois autres chimies couramment utilisées dans notre laboratoire. Il y sera détaillé la synthèse d'un composé 'preuve de concept' (édifice **H**, **Figure 31b**). Enfin le chapitre V exposera les résultats d'études d'interactions entre des mimes G-quadruplexes et des protéines cellulaires. Ces résultats sont issus d'une collaboration établie avec Dr. D. Gomez de l'université de Toulouse (Institut de Pharmacologie et de Biologie Structurale, IPBS).

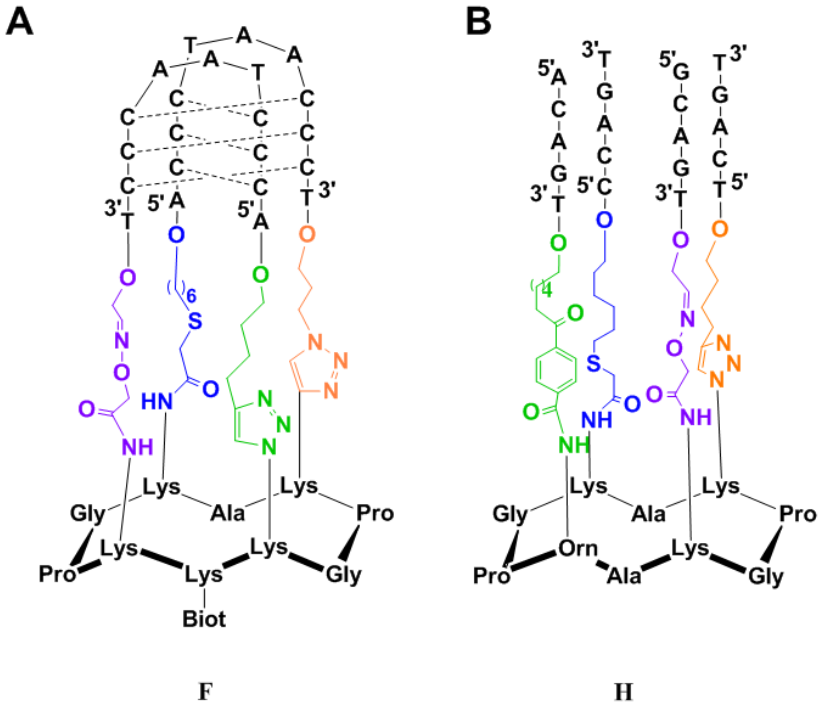


Figure 31 : représentation de l'édifice i-motif bimoléculaire antiparallèle (a) et du composé présentant quatre ligations chimiosélectives (b)

Chapitre III

Synthèse d'un mime ADN i-motif et étude de sa topologie

Au cours de ce chapitre, la synthèse d'un système bioconjugué oligonucléotides-peptide est abordée. Ce dernier représente un mime i-motif possédant la séquence d'ADN télomérique et constitue une cible intéressante d'un point de vue biologique. Dans le but d'obtenir ce type d'édifice, la synthèse d'un châssis peptidique et d'oligonucléotides fonctionnalisés s'effectuent préalablement et indépendamment. On retrouve sur ceux-ci des fonctions chimiques complémentaires afin de les lier ensemble par le biais de ligations chimiosélectives. Ce chapitre expose de manière détaillée la synthèse du support peptidique, des oligonucléotides et des ligations mises en œuvre. Des études concernant la topologie et la stabilité de cette structuration seront également présentées.

1. Généralités sur la synthèse peptidique et oligonucléotidique

a) Synthèse peptidique

La synthèse peptidique s'effectue usuellement sur phase solide (SPPS) de manière manuelle ou automatisée. Cette stratégie, décrite par Merrifield *et al* dans les années 1960[168], permet de réaliser une élongation de la chaîne peptidique en effectuant successivement des étapes de déprotection et de couplage d'acides aminés. La SPPS présente l'énorme avantage d'être exempter de purification entre chaque étape de couplage étant donné que la synthèse se déroule sur un support solide fonctionnalisé et que les différentes espèces réactionnelles peuvent être facilement éliminées par filtration. De plus, différents solvants, agents activateurs ou encore groupements orthogonaux sont compatibles avec la SPPS.

La stratégie de synthèse utilisée au sein de notre laboratoire pour la préparation de châssis peptidique est dite « Fmoc/tBu ». Les acides aminés utilisés possèdent un groupement protecteur Fmoc sur la fonction amine de la chaine principale. Par cette protection, un seul acide aminé est ajouté à la fois et cette élongation s'effectue du C-terminal vers le N-terminal. Entre chaque couplage, le groupement Fmoc est enlevé par une réaction de β–élimination à l'aide d'un mélange DMF/pipéridine (8/2, v/v) permettant ainsi l'ajout du prochain acide aminé. En utilisant cette stratégie de synthèse, il faut s'assurer que les fonctions chimiques présentes sur les chaînes latérales des acides aminées soient protégées et stables au cours des réactions de couplage et de déprotection. De plus, il est possible de suivre les étapes de couplage/déprotection en réalisant un dosage du groupement Fmoc. Enfin, la résine fonctionnalisée utilisée au cours de cette thèse est la 2-chlorotrityl. Celle-ci permet de réaliser l'étape du clivage entre le peptide linéaire et le support solide dans des conditions acides douces

(1% de TFA ou un mélange acide acétique/2,2,2-trifluoroéthanol (TFE)) permettant ainsi de conserver certains groupements protecteurs présents sur le peptide.

b) Synthèse oligonucléotidique

Tout comme les peptides, les oligonucléotides sont également synthétisés en phase solide, on parle alors de SPOS. Au laboratoire, nous utilisons la chimie des phosphoramidites pour l'introduction des nucléotides[169]. La synthèse s'effectue sur support solide et de manière automatisée pour laquelle l'élongation de la chaîne nucléotidique s'effectue de l'extrémité 3' vers l'extrémité 5'. La synthèse oligonucléotidique peut se décomposer en quatre étapes (**Schéma 1**). La première est la détritylation du groupement protecteur présent en 5' de l'oligonucléotide. Celle-ci se réalise à l'aide d'acide trichloroacétique dans le dichlorométhane et un cation de couleur orange est généré, le diméthoxytritylium ($DMTr^+$). Le dosage par spectrophotométrie de ce dernier permet d'évaluer l'efficacité de chaque cycle de synthèse. On retrouve à la fin de cette étape, une fonction hydroxyle en position 5' de l'oligonucléotide. Sur cette position, le prochain nucléotide est ensuite ajouté et ainsi se forme une liaison phosphite triester. Cette seconde étape est effectuée en présence de tétrazole, acide faible ayant un rôle d'activateur. L'avant dernière étape est le coiffage, également appelée « capping » et permet d'éviter que se forme des oligonucléotides présentant des délétions étant donné que l'étape de couplage n'est jamais totale. Pour ce faire, les hydroxyles libres vont être acétyler à l'aide d'anhydride acétique et ainsi ces derniers ne pourront plus réagir lors des cycles suivants et seront facilement éliminés ultérieurement lors de la purification. Et enfin, l'oxydation est la quatrième et dernière étape du cycle de synthèse. Au cours de celle-ci, la liaison phosphite, instable, est transformée en liaison phosphate triester, stable, par oxydation en utilisant une solution de diiode et d'eau. Après synthèse, la séquence d'oligonucléotide est clivée du support solide à l'aide d'une solution d'ammoniaque ou d'un mélange ammoniaque/méthylamine (1/1, v/v). Une étape de purification s'effectue ensuite afin d'éliminer les oligonucléotides de séquences tronquées. Il est important de spécifier que des séquences des oligonucléotides modifiés peuvent être également obtenus. Ceci peut s'effectuer par l'incorporation de nucléotides modifiés en milieu ou en fin de séquence. On peut également utiliser un support solide fonctionnalisé afin d'obtenir *in fine* un oligonucléotide modifié en position 3'.

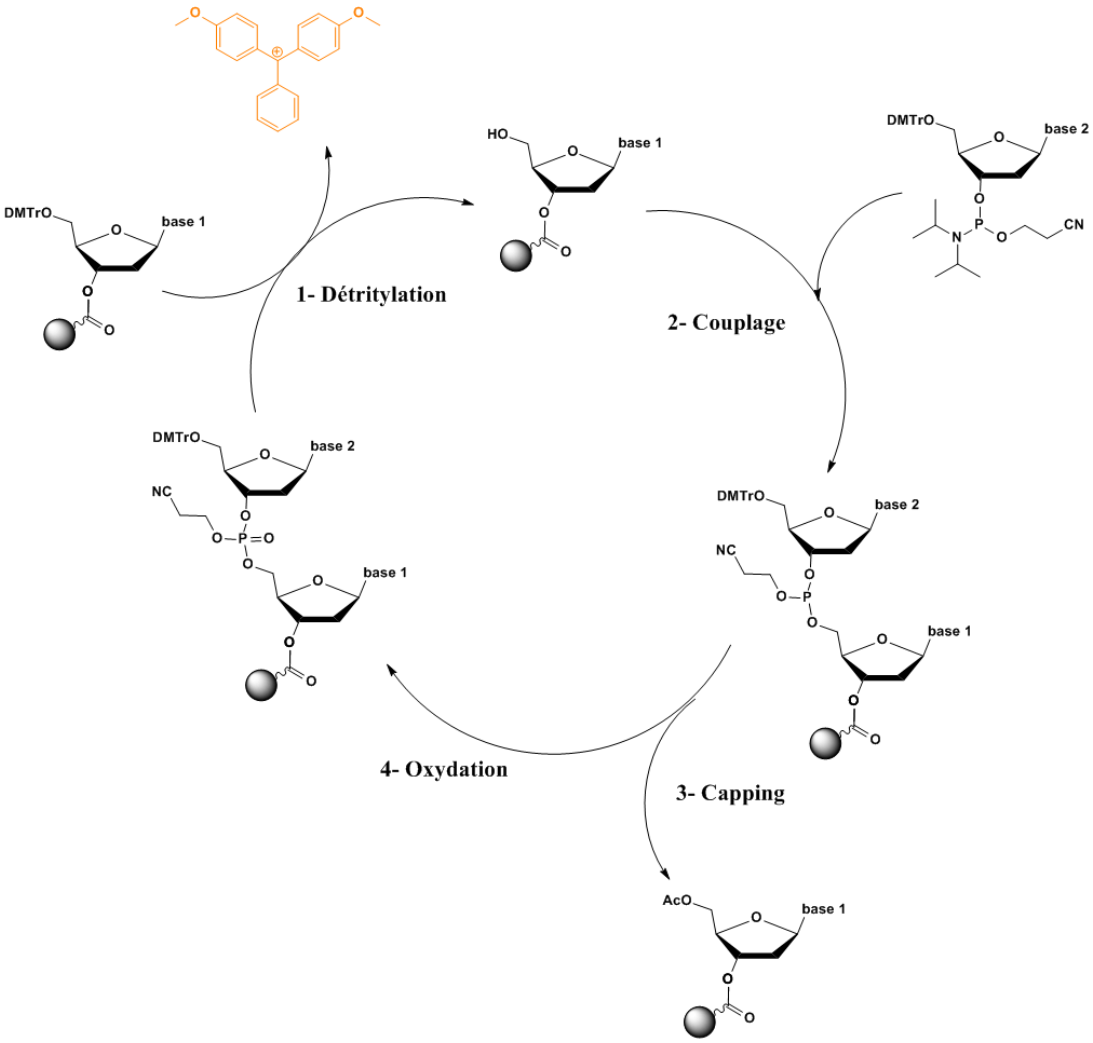


Schéma 1 : cycle de synthèse d'oligonucléotides

2. Présentation de la cible i-motif télomérique

Au sein du génome humain, on retrouve de manière non aléatoire des séquences riches en cytosines et celles-ci peuvent potentiellement former des structures i-motifs[107,108]. Différentes études ont également montrées que ces structures étaient impliquées dans différents processus biologiques[130,131,148] et qu'elles étaient présentes *in vivo* [140,141]. En 2012 au sein de notre laboratoire, un premier modèle i-motif tétramoléculaire a été synthétisé à l'aide du concept TASQ[170]. Ce dernier possède quatre séquences d'ADN $^{5'}$TCCCCT$^{3'}$ immobilisées de manière antiparallèle sur une plateforme peptidique (**Figure 32**). La conception de ce type d'édifice a permis dans un premier temps de confirmer la possibilité d'utiliser la stratégie TASQ et ainsi obtenir une structure i-motif stable sur le cyclodécapeptide. De plus, il a été observé que l'organisation sous forme i-motif était conservée sur une gamme de pH plus étendue que celle d'une séquence naturelle.

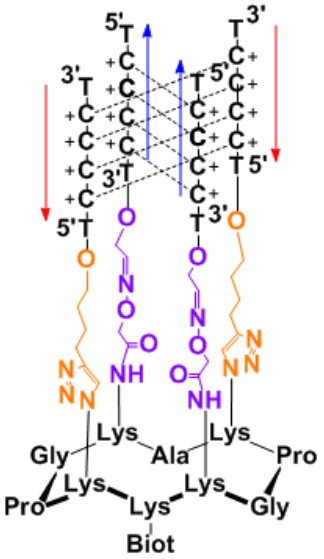


Figure 32 : représentation du modèle i-motif tétramoléculaire utilisant le concept TASQ[170]

Dans le but d'obtenir une cible plus pertinente d'un point de vue biologique (*i.e.* contenant les boucles), la conception d'un édifice biomoléculaire présentant la séquence i-motif télomérique a été élaborée lors de la thèse du Dr. L. Bonnat au sein du laboratoire[171]. La stratégie de synthèse (**Schéma 2**) reposait sur l'utilisation de deux boucles d'ADN présentant leurs extrémités 3' et 5' modifiées afin de les lier indépendamment au support peptidique. Pour cela, trois ligations orthogonales ont été utilisées : la ligation oxime, le couplage thiol-chloroacétamide et la cycloaddition CuAAC.

La première étape a consisté à lier le premier oligonucléotide à l'aide de la ligation oxime et dans un second temps de réaliser la cycloaddition CuAAC afin de fermer la première boucle. Suite à cela, deux voies sont possibles. La première est de greffer le second oligonucléotide grâce au couplage thiol-chloroacétamide suivie d'une réaction de diazo-transfer sur le châssis peptidique (voie a) ou effectuer l'ordre inverse de ces deux réactions (voie b). Dans les deux cas, on retrouve *in fine* l'édifice où la deuxième boucle peut être fermée par une seconde cycloaddition CuAAC. Les raisons de finir par cette réaction sont multiples. En effet, la cycloaddition CuAAC a déjà été utilisée de nombreuses fois dans la fermeture de boucle intramoléculaire mais également cette dernière s'est montrée plus réactive que le couplage thiol-chloroacétamide. Cependant, ce mime i-motif n'a pu être obtenu au final. En effet, la dernière étape, c'est-à-dire la seconde cycloaddition CuAAC permettant de fermer la deuxième boucle, n'a pas permis d'obtenir le produit final. Différentes options ont pourtant été investiguées : réalisation de la cycloaddition dans les conditions réactionnelles usuelles (c'est-à-dire 100 mM d'HEPES à pH 7,4, 6 eq de $CuSO_4$, 30 eq de THPTA et 30 eq d'ascorbate de

sodium), utilisation des conditions classiques de CuAAC mais avec la présence de 500mM de KCl dans le but de réaliser un écrantage de charge afin de limiter la répulsion électrostatique entre les deux brins d'ADN qui pourrait éventuellement éloigner les fonctions alcyne et azoture et enfin explorer l'effet du pH sur la réaction de cycloaddition afin de voir si une pré-organisation de l'édifice permettait la réussite ou non de la réaction. Pour cela un tampon 100mM de PBS a été utilisé à pH 4, 5, 6 ou 7 en présence 6 eq de $CuSO_4$, 30 eq de THPTA et 30 eq d'ascorbate de sodium.

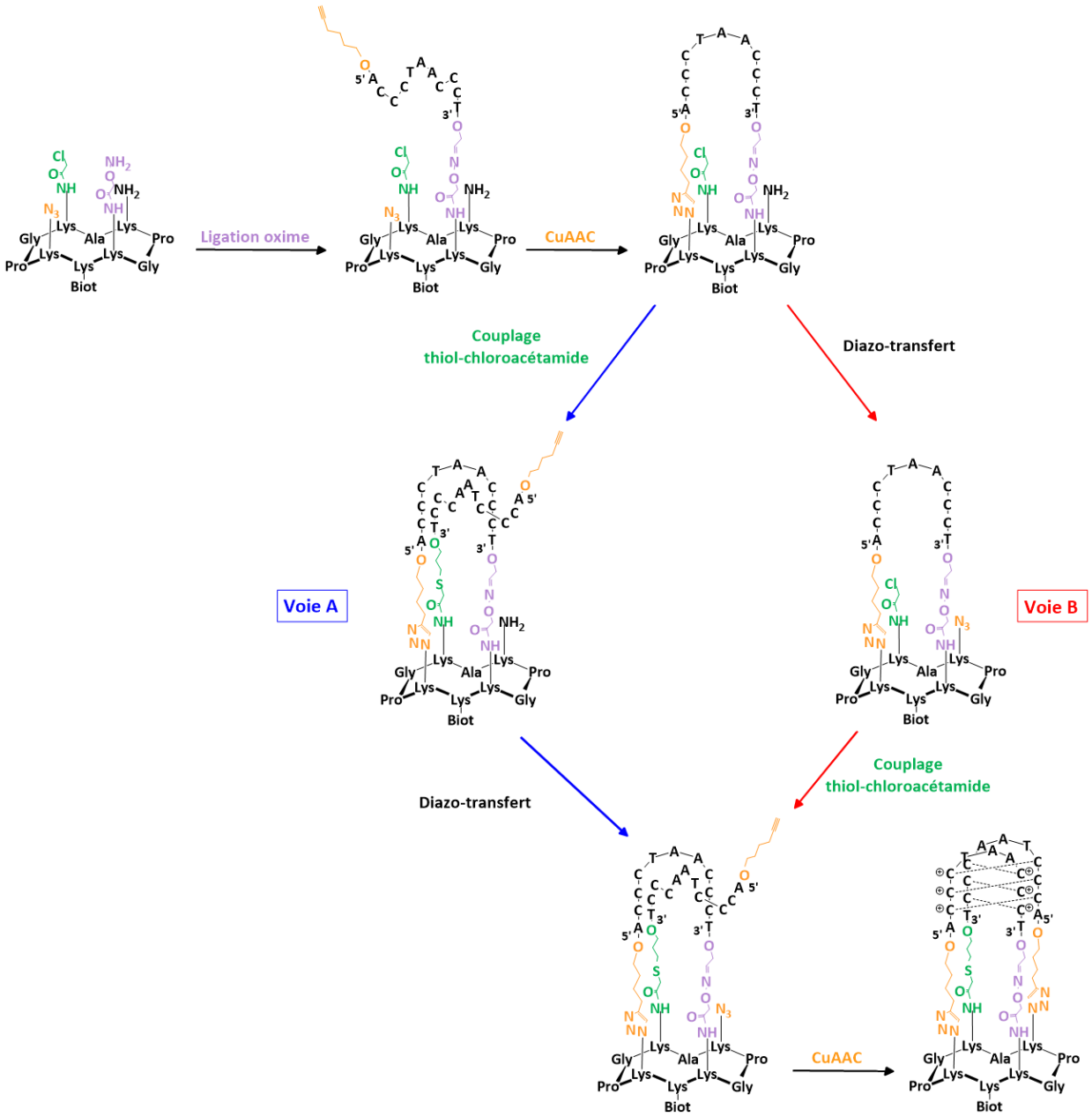


Schéma 2 : représentation de la précédente stratégie de synthèse pour l'assemblage du mime i-motif. Figure issue de Bonnat, L., PhD Thesis, **2017** [171]

Malheureusement, aucune des conditions citées ci-dessus n'a permis d'obtenir l'édifice final. De plus, l'une des difficultés rencontrées au cours de cette dernière réaction est le fait qu'il n'y a aucun changement de masse mais également de charge entre l'intermédiaire et le produit final. Ceci a pour conséquence de rendre la spectrométrie de masse et la chromatographie échangeuse d'ions inutiles afin de confirmer ou infirmer la présence de l'édifice final. Ainsi concernant cette dernière étape, il fallait espérer un changement de temps de rétention lors du suivi par HPLC ce qui n'a pas été le cas.

Suite à cela et dans le cadre cette thèse, une nouvelle stratégie a été élaborée afin d'obtenir un mime i-motif. Comme précédemment, celle-ci repose sur l'utilisation de deux brins d'ADN modifiés aux extrémités 3' et 5'. Dans le plan de synthèse de cet édifice (**Schéma 3**), les trois ligations chimiques mais également la modification sur le châssis peptidique mentionnés auparavant seront de nouveau utilisés.

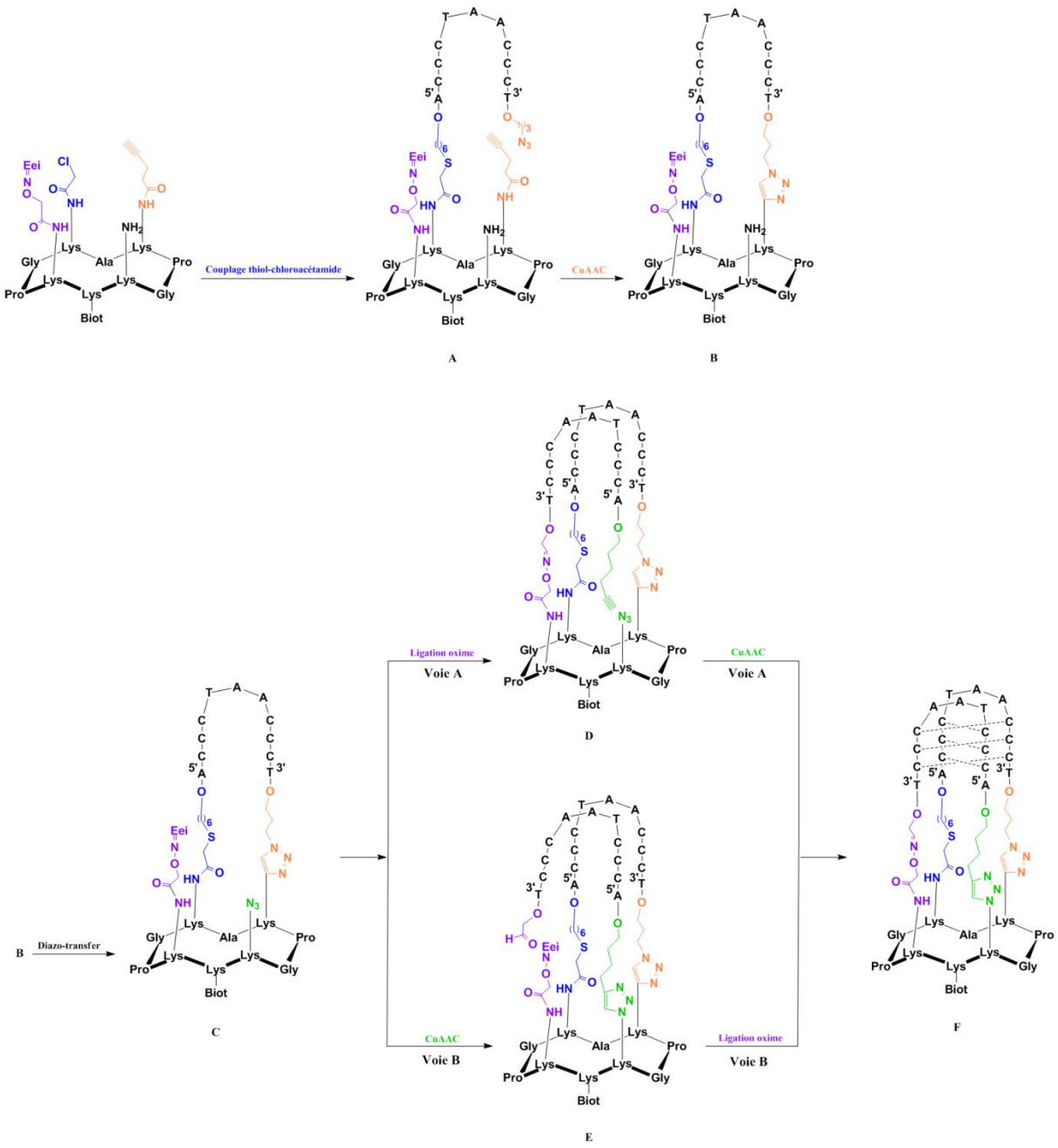


Schéma 3 : plan de synthèse pour le mime i-motif **F**

La première étape consiste à lier le premier oligonucléotide par le couplage thiol-chloroacétamide (édifice **A**). Nous avons choisi d'utiliser cette ligation en premier car c'est la moins réactive des trois et c'est celle qui, usuellement, donne les moins bons rendements. Ainsi, en commençant directement par cette réaction, on s'affranchit de ces désavantages pour la suite. Dans un second temps, il s'en suit une première réaction de cycloaddition CuAAC dans le but de former la première boucle sur le cyclodécapeptide (édifice **B**). Ensuite, une réaction de diazo-transfert est réalisée dans le but de générer une fonction azoture sur le chassis *via* une fonction amine (édifice **C**).

A partir de là, deux possibilités sont envisageables pour former la seconde boucle. Soit greffer le deuxième oligonucléotide par la ligation oxime (édifice **D**) et de finir par la cycloaddition CuAAC (édifice **F**), soit effectuer la seconde réaction de cycloaddition CuAAC (édifice **E**) et de finir par la ligation oxime (édifice **F**). Il est important de préciser qu'au cours de la ligation oxime, une étape de déprotection de la fonction oxyamine doit être réalisée. Ainsi dans la voie B, il sera observé une différence de masse lors de la formation de l'édifice final ce qui n'est pas le cas dans la seconde alternative. Effectivement, dans le cas où l'oximation s'effectue en fin de synthèse, la perte du groupement protecteur de l'oxyamine ainsi que la perte d'eau, intrinsèque à la ligation oxime, sera une preuve de la formation de l'édifice final. Alors que par la voie A, aucun changement de masse n'a lieu lors de la dernière étape, c'est-à-dire lors de la cycloaddition CuAAC ce qui a été problématique auparavant. Ces deux options seront investiguées.

3. Synthèse du châssis peptidique 1

Afin de greffer les deux oligonucléotides sur le châssis peptidique **1**, ce dernier doit présenter à sa surface supérieure les fonctions chimiques adéquates (**Figure 33**). C'est ainsi que l'on retrouve les fonctions chloroacétamide, amine, alcyne et oxyamine protégée. L'incorporation de la fonction chloroacétamide est obtenue au départ d'une lysine commerciale possédant un groupement protecteur DDE. De plus, l'obtention de la fonction amine provient d'une seconde lysine commerciale possédant un groupement protecteur Alloc. Ces deux groupes protecteurs pourront être enlevés sélectivement. En ce qui concerne les fonctions alcyne et oxyamine protégées, les acides aminés correctement fonctionnalisés devront être synthétisées préalablement. Il en va de même pour la lysine possédant la fonction biotine présente sur la face inférieure du châssis peptidique. La synthèse de ces acides aminés modifiés est décrite ci-dessous.

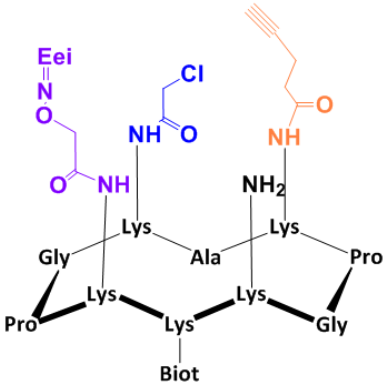


Figure 33 : représentation du châssis peptidique **1**

a) Synthèse des acides aminés modifiés

Pour la réalisation du châssis peptidique **1**, trois « buildings blocks », non commerciaux sont utilisés et doivent donc être synthétisés et introduits lors de la SPPS. Ces trois acides aminés modifiés sont obtenus à partir de l'acide aminé Fmoc-Lys-OH, **2**, issue de la déprotection en milieu acide de l'amine ε de l'acide aminé Fmoc-Lys(Boc)-OH (**Figure 34**).

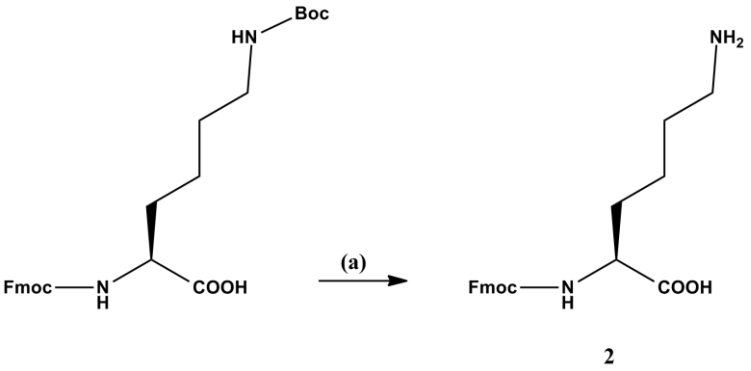


Figure 34 : synthèse de la Fmoc-Lys-OH **2**
(a) HCl (37%)/Dioxane (1/2, v/v), 1h, t.a., 92%

L'acide aminé **3** a été obtenu selon la procédure publiée et décrite par Galibert *et al.*[172]. Afin de préparer la lysine présentant la fonction alcyne **3**, l'acide 4-pentynoique est préalablement activé par l'ajout du groupement N-hydroxysuccinimide (NHS) en présence de N,N'-dicyclohexylcarbodiimide (DCC) (**Figure 35)**. Le greffage de l'ester activé **3a** est ensuite réalisé en milieu basique sur l'acide aminé Fmoc-Lys-OH **2** et donne *in fine* l'acide aminé Fmoc-Lys(*N*- acide 4-pentynoique)-OH **3** avec un rendement de 89 %.

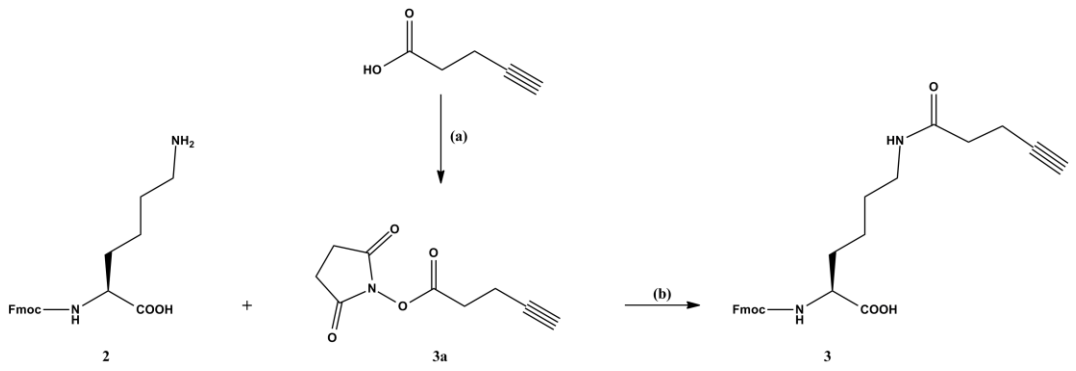


Figure 35 : synthèse de la Fmoc-Lys(*N*- acide 4-pentynoique)-OH **3**
(a) NHS (1eq)/DCC (1eq) dans un mélange AcOEt/Dioxane (1/1, v/v), 5h, t.a.
(b) DMF, DIEA, 2h, t.a., 89%

L'acide aminé **4** portant la biotine a été obtenu selon les procédures publiées et décrites par Youn *et al.*[173]. Il est formé par l'introduction de la partie biotine sur la fonction amine en ε du composé Fmoc-Lys-OH **2** *via* un lien amide (**Figure 36**). Préalablement, la fonction acide carboxylique de la biotine est activée par un groupement NHS, en présence de DCC comme agent de couplage. Une fois activée, la biotine **4a** est introduite sur l'acide aminé Fmoc-Lys-OH **2** en milieu basique et à température ambiante.

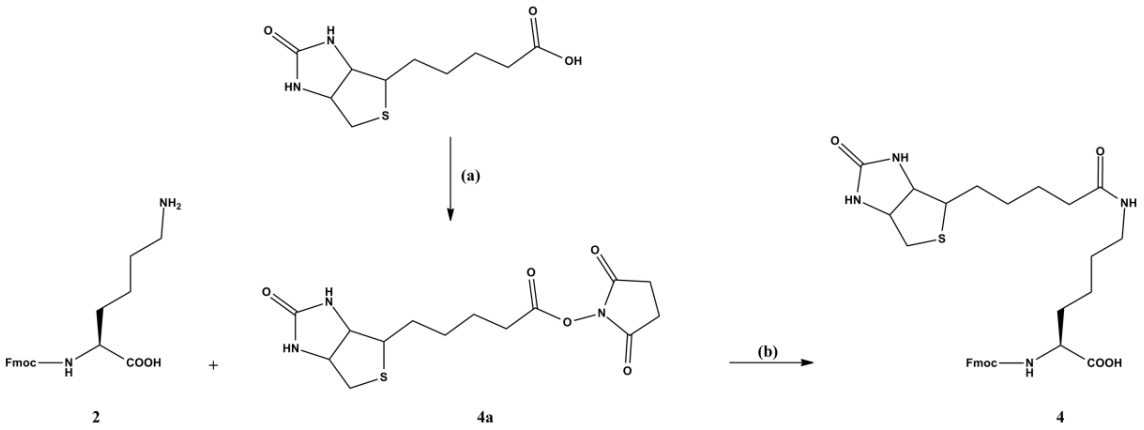


Figure 36 : synthèse de la Fmoc-Lys(biot)-OH **4**
(a) NHS (1eq)/DCC (1eq), DMF anhydre, 30h, t.a., 93% ; (b) DMF, DIEA, 2h, t.a., 89%

L'acide aminé **5** a été obtenu selon les procédures publiées et décrites par Foillard *et al.*[174]. Une première étape consiste à préparer le groupement protecteur (**5a**) par la substitution nucléophile du composé éthyl-N-hydroxyacétimidate sur l'acide iodoacétique. Il s'en suit l'activation de la fonction carboxylique du composé **5a** par une fonction N-hydroxysuccinimide pour former le composé **5b**. Pour finir, ce dernier est ensuite introduit sur la fonction amine ε de l'acide aminé Fmoc-Lys-OH **2** pour donner le composé **5** (**Figure 37**). L'utilisation de ce

groupement protecteur Eei permet le blocage total de la réactivité des oxyamines mais présente également l'avantage d'être éliminé en condition acide douce.

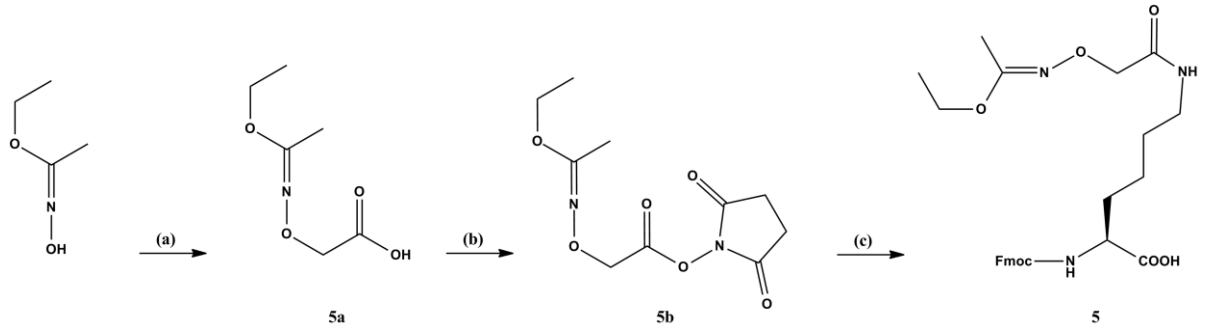


Figure 37 : synthèse de la Fmoc-Lys(Eei-Aoa) **5**
(a) ICH_2COOH, NaOH, H_2O, 5h, 80°C, 53% ; (b) NHS (1eq)/DCC (1eq), DMF, 16h, t.a., 84% ; (c) Fmoc-Lys-OH **2** (1,2 eq), DMF, DIEA, 2h, t.a., 97%

Il est important de spécifier que les acides aminés modifiés utilisés sont totalement compatibles avec les conditions de couplage et de déprotection de la SPPS. De plus, l'utilisation de ces derniers permet de réduire le nombre de fonctionnalisation après la SPPS, ayant comme conséquence directe une augmentation du rendement global.

b) Synthèse du châssis peptidique 1

Le cyclodécapeptide **1** est synthétisé sur support solide en utilisant la stratégie de synthèse au Fmoc/tBu et la résine 2-chlorotrityl (**Schéma 4**). Le premier acide aminé glycine est ajouté manuellement en présence de DCM et DIEA. Ensuite, la chaine peptidique **1a** est allongée à l'aide d'un synthétiseur peptidique. En fin de SPPS, le peptide linéaire **1b** est libéré de la résine en milieu acide faible à l'aide d'une solution DCM/TFE/AcOH (70/20/10). Ces conditions acides douces permettent de conserver le groupement protecteur Eei (voir structure **Figure 37**). Ensuite, le peptide linéaire **1c** est cyclisé en milieu basique (pH 8-9) en présence de PyBOP comme agent de couplage. Il est important de spécifier que la cyclisation s'effectue en milieu dilué (10^{-3}M), dans le but de favoriser la cyclisation intramoléculaire et ainsi éviter une polymérisation du peptide. De plus, le risque de racémisation est évité puisque la cyclisation s'effectue en C-terminal sur une glycine qui ne possède pas de chaîne latérale. Un rendement de 53 % est obtenu pour la formation de **1d**. Après la cyclisation, l'élimination du groupement protecteur DDE est réalisée à l'aide d'une solution d'hydrazine à 2% en volume dans le DMF et d'alcool allylique. La réduction de la double liaison du groupement Alloc par l'hydrazine est évitée grâce à la présence de l'alcool allylique. Par la suite, l'introduction de la fonction chloroacétamide est effectuée à l'aide de l'anhydride chloroacétique en milieu basique.

L'intermédiaire **1f** est obtenu avec un rendement de 49 %. Et enfin, la dernière étape réalisée sur le cyclodécapeptide est la déprotection de l'amine portant le groupement Alloc permettant l'apparition d'une fonction amine sur la chaîne latérale d'une des lysines. Cette déprotection s'effectue en présence de $Pd^0(PPh_3)_4$ et d'un excès de phénylsilane ($PhSiH_3$). Le châssis peptidique final **1** est caractérisé par UPLC-MS ainsi que spectrométrie de masse (**Figure 38**) et est obtenu avec un rendement de 44 %. A noter que lors des purifications, un acide moins fort, l'acide formique, substitue le TFA afin d'éviter la perte du groupement protecteur Eei sur le cyclodécapeptide.

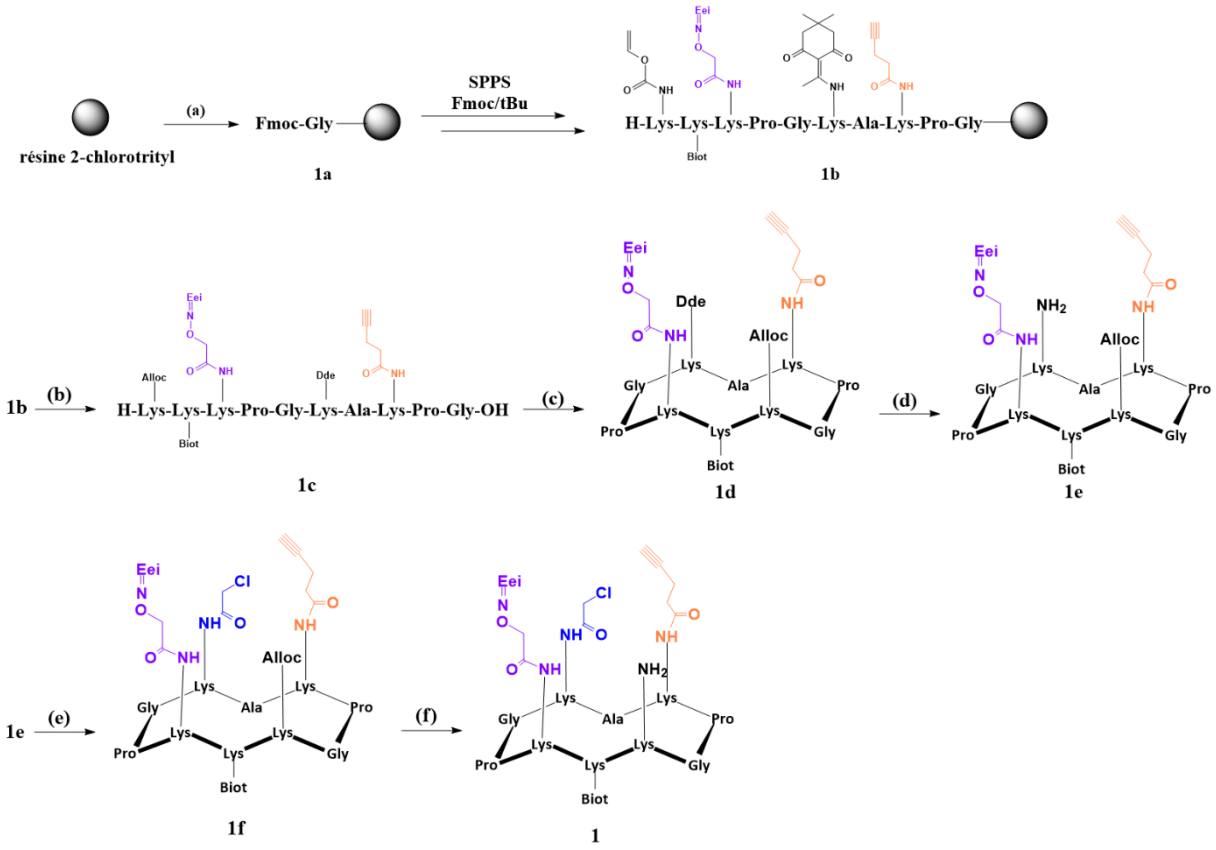


Schéma 4 : synthèse du châssis peptidique **1**
(a) Fmoc-Gly-OH (3eq), DIEA, DCM, 1h, t.a. ; (b) DCM/TFE/AcOH (7/2/1, v/v/v), t.a. ; (c) PyBOP (1,2eq), DIEA, DMF, 2h, t.a., 53% ; (d) 2%vol d'hydrazine dans DMF, alcool allylique (5eq), 20 minutes, t.a. ; (e) anhydride chloroacétique (1,2eq), DIEA, DMF, t.a., 49% ; (f) $Pd^0(PPh_3)_4$ (0,2eq), $PhSiH_3$ (100eq), DCM/DMF anhydre (3/1, v/v), t.a., 44%

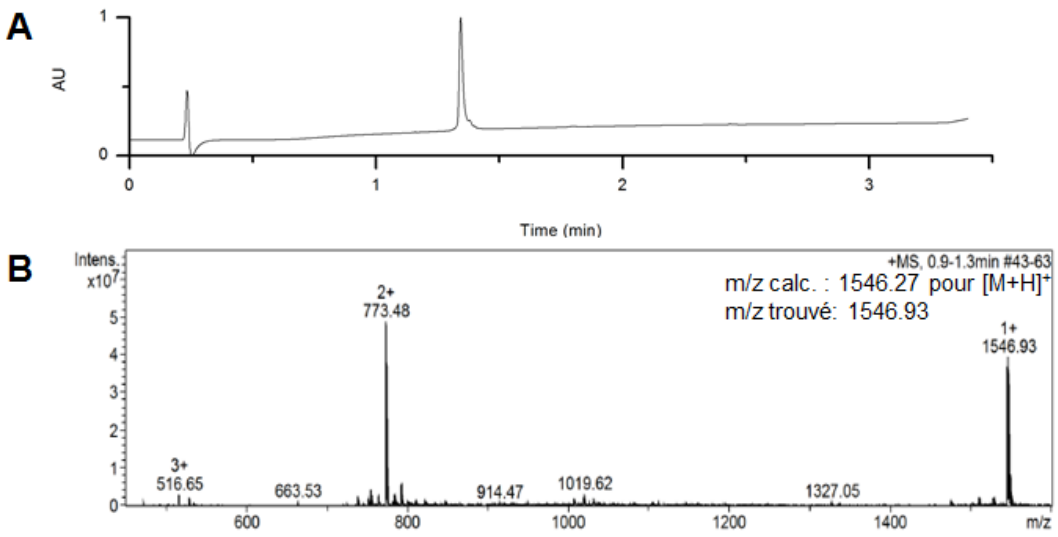


Figure 38 : (a) chromatogramme UPLC et (b) analyse ESI-MS du châssis peptidique **1**

4. Synthèse des brins d'ADN

Ayant comme objectif d'obtenir le mime i-motif **F** avec la présence de deux boucles, la stratégie de synthèse s'est portée sur l'utilisation de deux oligonucléotides modifiés aux extrémités 3' et 5'. Le premier, **6**, possède la séquence suivante : $^{5'}X$-ACCCTAACCCT-$Y^{3'}$ (où *X* représente une fonction thiol et *Y* une fonction azoture). La séquence du second oligonucléotide, **7**, se présente de la manière suivante : $^{5'}W$-ACCCTAACCCT-$Z^{3'}$ (où *W* représente une fonction alcyne et *Z* une fonction aldéhyde). A noter que l'usage de ces différentes fonctionnalisations est connu au sein de notre laboratoire et ces dernières ont été utilisées dans diverses applications[166,175,176].

a) Synthèse de l'oligonucléotide 5'-thiol/3'-azoture 6

La séquence **6** (**Schéma 5**) est obtenue au départ d'une résine « 3'-PT-amino-modifier C3 CPG » permettant d'obtenir une fonction amine en position 3'. De l'autre côté, la fonction thiol est quant à elle incorporée en position 5' sous la forme protégée par un groupement trityl à l'aide d'un phosphoramidite modifié, le « 5'-thiol-modifier C6 ». Suite à l'élongation de la chaîne, la déprotection et l'élimination du support, on obtient ainsi l'intermédiaire **6a**. Ce dernier est alors purifié par RP-HPLC et obtenu avec un rendement de 43 %. Il s'en suit une réaction de diazo-transfer pour introduire la fonction azoture en position 3'. Celle-ci s'effectue à 60°C en présence de sulfate de cuivre, du réactif ISAHC (pour « imidazole-1-sulfuryl azide hydrochloride») dans un mélange tampon $NaHCO_3$ (50mM pH 8.5)/MeOH (3/1, v/v). La réaction est suivie par UPLC-MS jusqu'à complète transformation. Le mélange réactionnel est

ensuite dessalé sur colonne d'exclusion stérique. L'intermédiaire **6b** ainsi obtenu peut subir la détritylation sur la fonction thiol. Cette réaction se déroule en deux temps. Tout d'abord, l'oligonucléotide **6b** est mis en présence de nitrate d'argent dans un mélange tampon TEAA 0.1M à pH 6.5. Après 45 minutes, au mélange réactionnel est ajoutée une solution aqueuse de DTT. Ensuite le brut réactionnel est centrifugé, le surnageant est récupéré pour enfin être dessalé sur colonne d'exclusion stérique. La formation de l'oligonucléotide **6** est considéré comme quantitative et il est caractérisé par RP-HPLC et spectrométrie de masse (**Figure 39**).

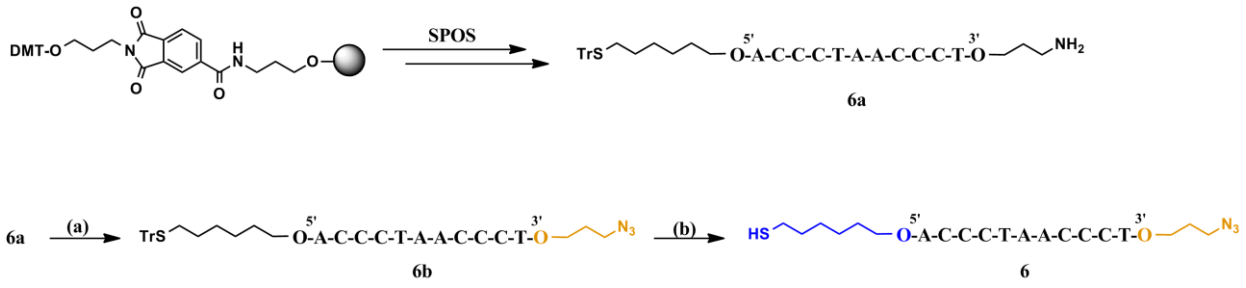


Schéma 5 : Synthèse de l'oligonucléotide **6**
(a) $CuSO_4$ (3eq), ISAHC (30eq), $NaHCO_3$ 50mM pH 8,5/MeOH (3/1, v/v), 3h, 60°C ; (b) 1) $AgNO_3$ 1M, TEAA 0,1M pH 6,5, 45 minutes, t.a. 2) DTT 1M, 15 minutes, t.a.

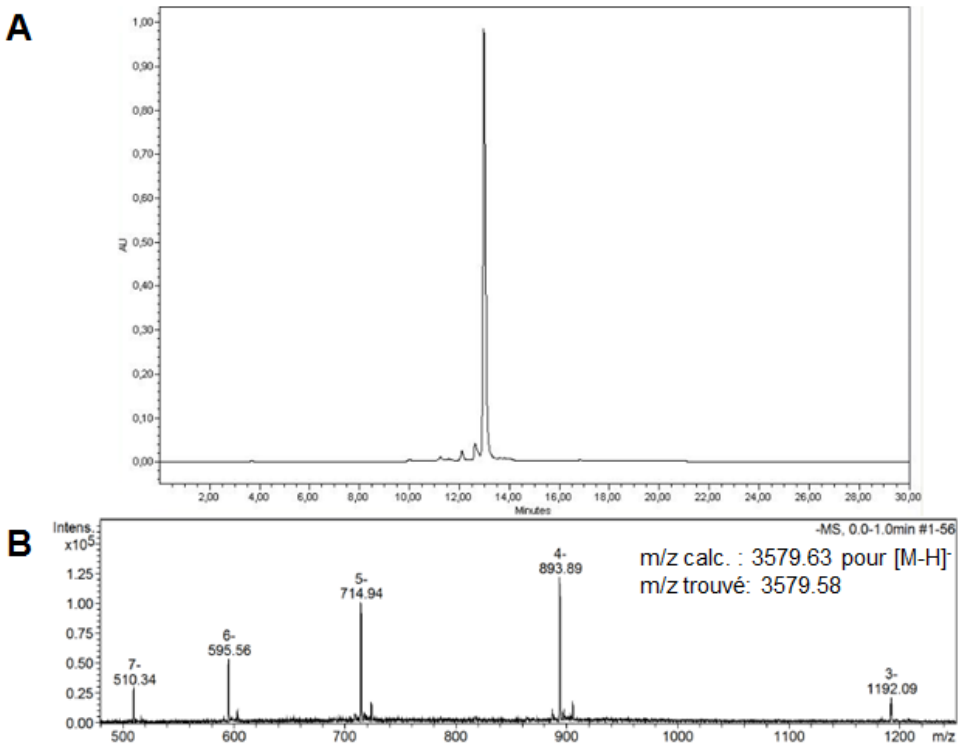


Figure 39 : (a) chromatogramme RP-HPLC et (b) analyse ESI-MS de l'oligonucléotide **6**

b) Synthèse de l'oligonucléotide 5'-alcyne/3'-aldehyde 7

Au départ d'une résine 3'glyceryl CPG et après synthèse supportée, l'oligonucléotide modifié **7a** est obtenu (**Schéma 6**). On retrouve à son extrémité 3' une fonction 1,2 diol et en position 5', une fonction alcyne est incorporée à l'aide d'un phosphoramidite modifié, le « modifier 5' hexynyl ». Après clivage du support dans une solution d'ammoniaque et de méthylamine (1/1, v/v) à 65°C, l'oligonucléotide **7a** est purifié par RP-HPLC et est obtenu avec un rendement de 41 %. Ensuite, la génération de la fonction 3'-aldéhyde est provoquée par le périodate de sodium par coupure oxydative de la fonction diol. Lors de cette réaction, le sel de périodate ainsi que le formaldéhyde sont éliminés sur colonne d'exclusion stérique. On considère la réaction comme quantitative. L'oligonucléotide **7** est caractérisé par RP-HPLC et spectrométrie de masse (**Figure 40**).

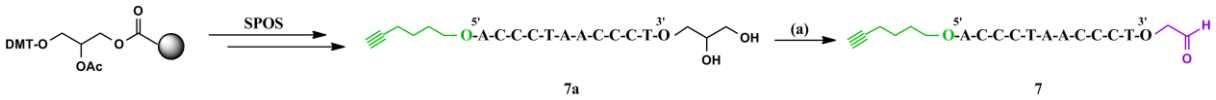


Schéma 6 : synthèse de l'oligonucléotide **7**
(a) $NaIO_4$ (30eq), H_2O, 30 minutes, noir, t.a.

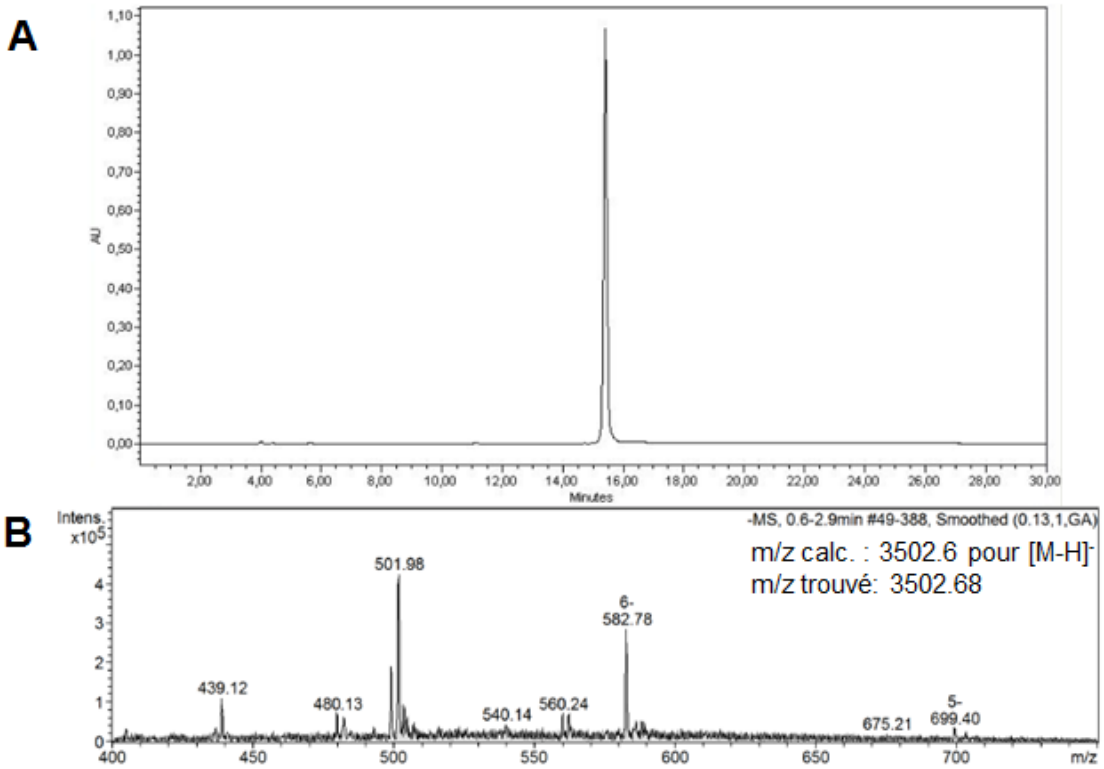


Figure 40 : (a) chromatogramme RP-HPLC et (b) analyse ESI-MS de l'oligonucléotide **7**

5. Synthèse du mime i-motif F

a) Assemblage de l'intermédiaire A

Comme énoncé auparavant dans le plan de synthèse, la première étape consiste à réaliser le couplage thiol-chloroacétamide sur le châssis peptidique **1** à l'aide de la séquence d'ADN **6** (**Schéma 7**).

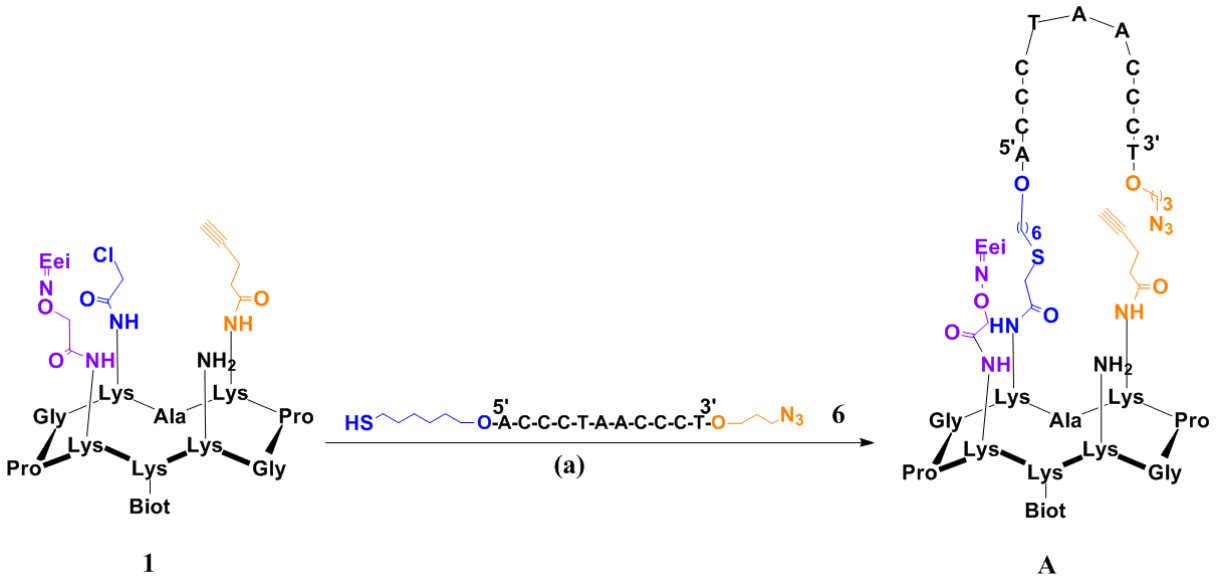


Schéma 7 : synthèse de l'intermédiaire **A**
(a) oligonucléotide **6** (1 eq), TCEP (1eq), DIEA (60eq), KI (60eq), 1h30, 37°C, 53 %

Au cours de cette réaction (**Figure 41**), l'oligonucléotide **6** est mis en présence d'un excès du peptide **1**, de 60eq de DIEA et de KI et de 1eq de TCEP. Ce dernier est un réducteur et permet d'éviter que deux séquences d'ADN **6** s'associent ensemble par l'intermédiaire d'un pont disulfure. Le suivi réactionnel peut s'effectuer soit par RP-HPLC, soit par UPLC-MS. Pour cette réaction, il est plus intéressant de suivre par UPLC-MS car on peut confirmer, par le biais du spectre de masse, la formation ou non d'oligonucléotides **6** sous forme disulfure. Il est important de spécifier que dans notre cas, l'ajout de TCEP ne se fait pas en début de réaction mais au cours de celle-ci. Cet ajout ultérieur s'explique par le fait qu'il a été observé, au cours de la réaction, une réduction du groupement azoture en amine lorsque le TCEP est présent dès le départ. Ainsi utiliser l'UPLC-MS dans le cadre d'un suivi réactionnel permet d'observer si le phénomène de réduction de la fonction azoture a lieu mais permet également d'évaluer l'ampleur de ses conséquences. Dans le but de retarder au maximum l'ajout de TCEP dans le milieu réactionnel, il a été mis en place différents aménagements. En effet, il a été choisi d'effectuer la détritylation la veille du couplage thiol-chloroacétamide entre les deux entités et

ce, afin d'avoir le plus d'oligonucléotide **6** sous forme thiol dès départ de la réaction. De plus, le jour J, la mise en présence des différents réactifs se fait sous atmosphère inerte et avec l'utilisation d'eau désoxygénée. Malgré ces précautions, l'ajout de TCEP est tout de même nécessaire et on observe, malheureusement, le phénomène de réduction de la fonction azoture soit sur l'oligonucléotide **6** non converti, soit sur l'édifice **A**. Le pourcentage de réduction de la fonction azoture se situe en moyenne vers les 10-20 % mais reste variable selon le moment où est ajouté le TCEP. De plus, il a été choisi de stopper la réaction lorsqu'un taux de conversion acceptable est atteint (aux alentours de 70 %) afin de limiter la réduction de la fonction azoture. En fin de réaction, le brut réactionnel est dessalé sur colonne d'exclusion stérique et est ensuite purifié par RP-HPLC. L'intermédiaire **A** est obtenu avec 53 % de rendement. Il est caractérisé par RP-HPLC et spectrométrie de masse (**Figure 42**).

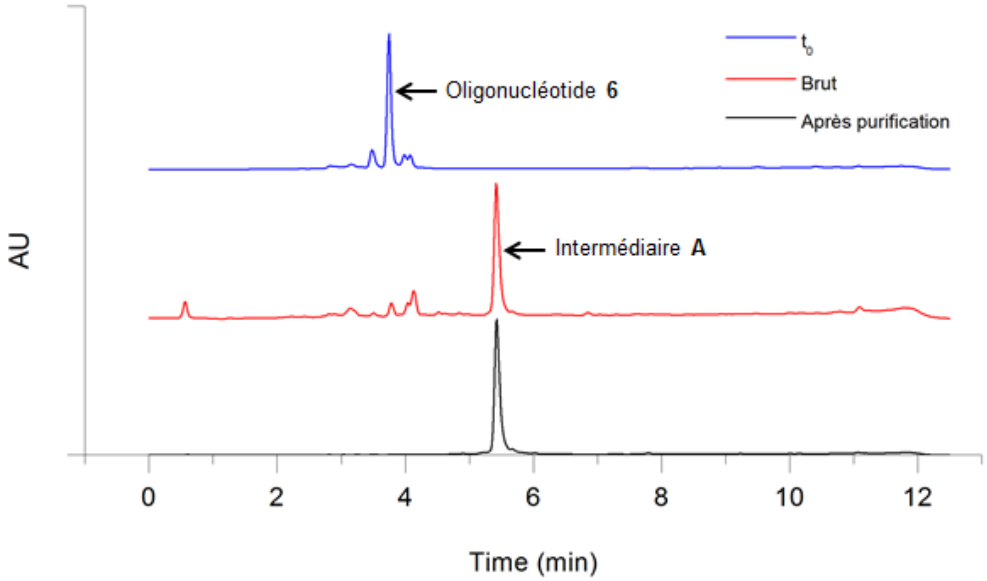


Figure 41 : suivi réactionnel par UPLC du couplage thiol-chloroacétamide entre l'oligonucléotide **6** et le châssis peptidique **1**

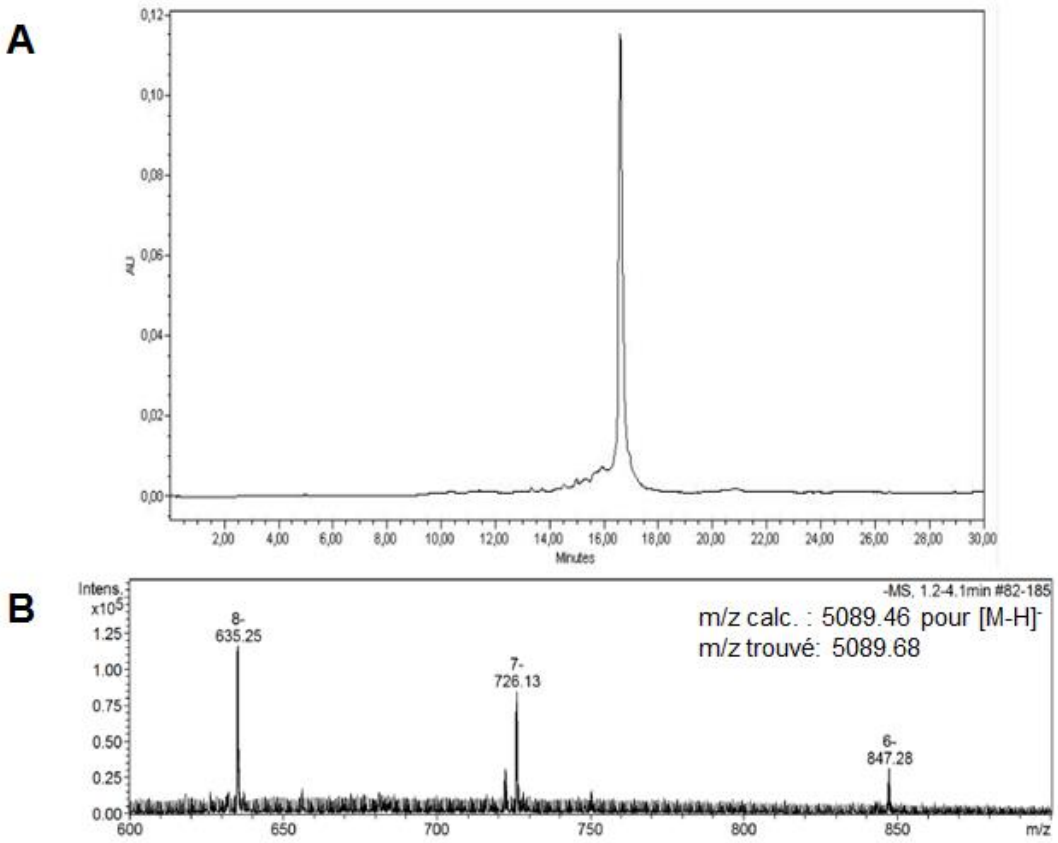


Figure 42 : (a) chromatogramme RP-HPLC et (b) analyse ESI-MS de l'intermédiaire **A**

b) Assemblage de l'édifice B

La formation de l'édifice **B** à partir de l'intermédiaire **A** s'effectue à l'aide d'une cycloaddition CuAAC (**Schéma 8**). Afin d'obtenir cet édifice avec la première boucle, des conditions réactionnelles de CuAAC usuelles ont été utilisées.

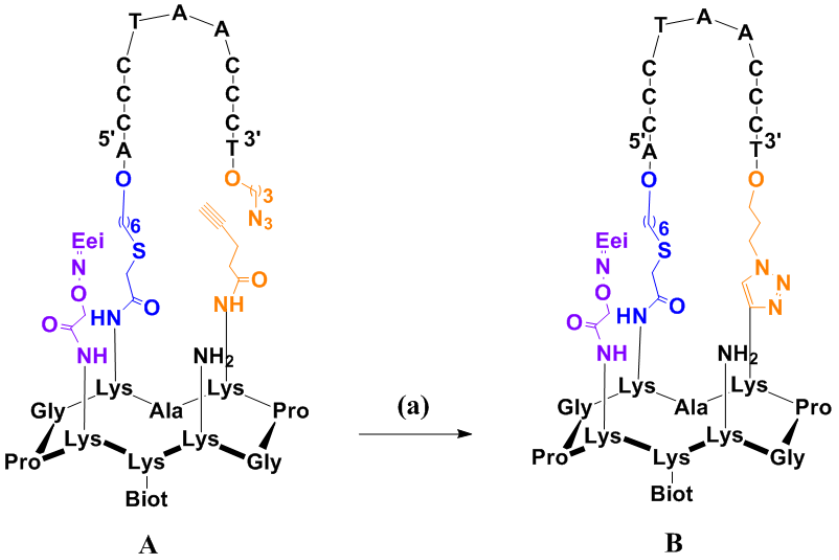


Schéma 8 : synthèse de l'édifice **B**
(a) $CuSO_4$ (3eq), THPTA (30eq), ascorbate de sodium (30eq), 500mM KCl, DMF (10%vol), PBS 50mM pH 7,4, 2h, 37°C, 40%

Le suivi réactionnel peut, comme auparavant, se faire par RP-HPLC ou par UPLC-MS. Malgré qu'aucun changement de masse n'est visible au cours de cette réaction, on observe cependant un changement de temps de rétention net entre l'intermédiaire **A** et le produit **B** (**Figure 43**). On remarque également la présence d'un certain pourcentage de composé **A** qui correspond à une réduction de la fonction azoture (nommé composé **A'**). Ce dernier ne peut, bien évidemment, pas réagir dans ces conditions réactionnelles. De plus, on observe, fort heureusement, un écart de temps de rétention suffisant entre l'édifice **B** et le composé **A'**. En fin de réaction, le brut est dessalé sur colonne d'exclusion stérique et est ensuite purifié par RP-HPLC avec un rendement de 40 %. L'édifice **B** est caractérisé par RP-HPLC et spectrométrie de masse (**Figure 44**)

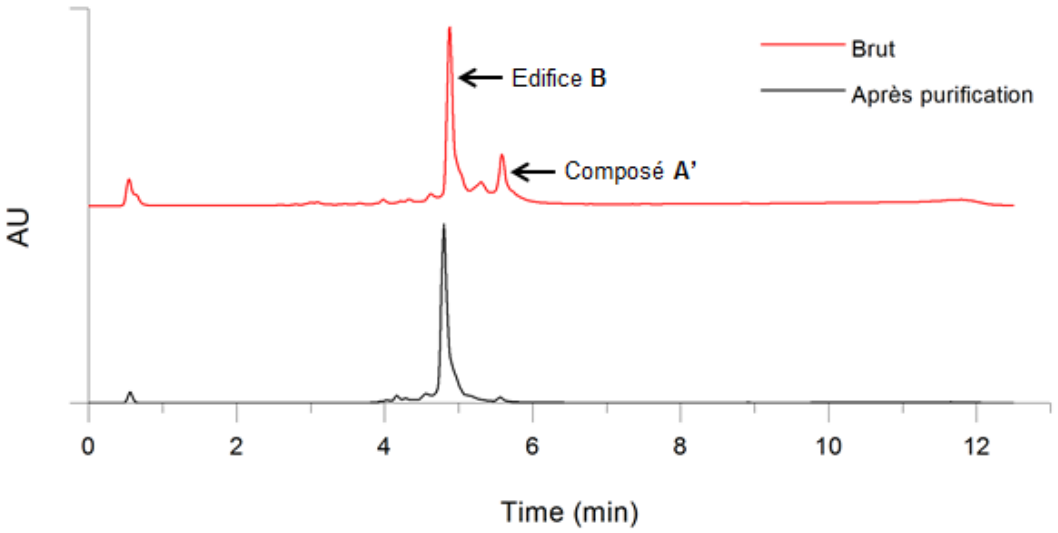


Figure 43 : chromatogrammes de la formation de l'édifice **B** en UPLC du brut réactionnel et après purification

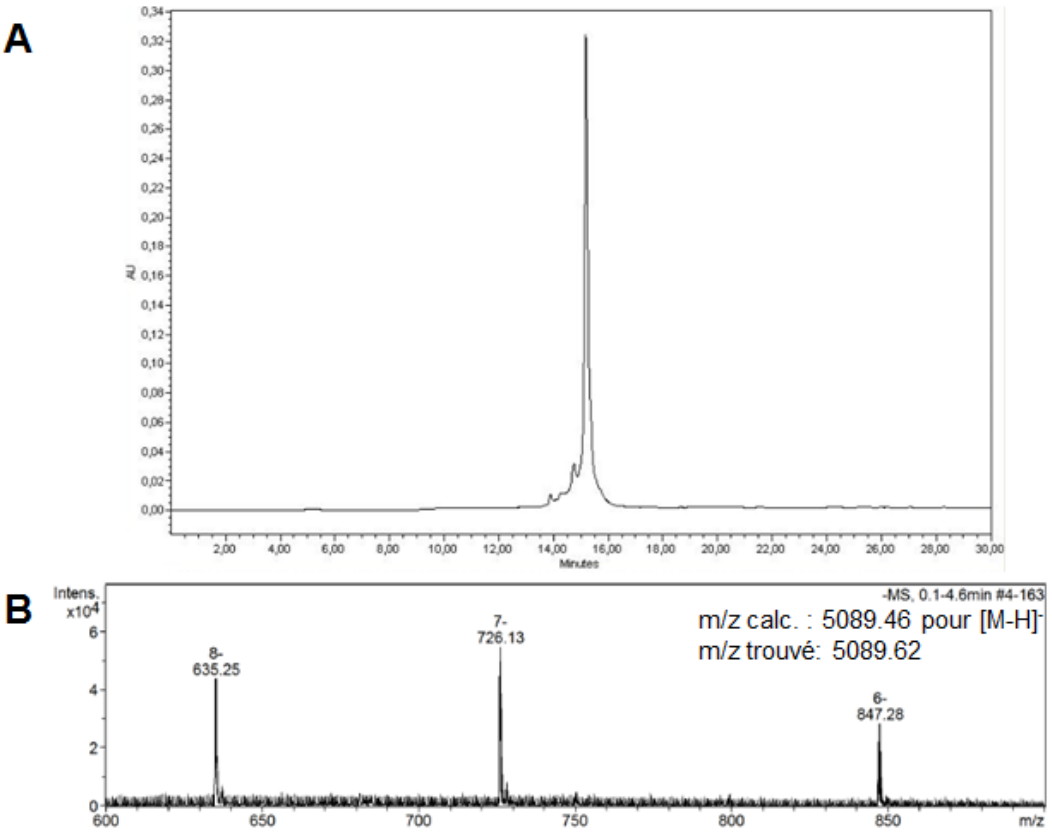


Figure 44 : (a) chromatogramme RP-HPLC et (b) analyse ESI-MS de l'édifice **B**

c) Synthèse de l'édifice C

Deux ligations chimiosélectives ont, jusqu'ici, été utilisées dans l'obtention de l'édifice **B**. Afin d'avoir le composé final **F**, il faut encore faire appel à deux autres ligations. Dans la

stratégie de synthèse établie, il a été choisi d'utiliser la chimie des oximes comme troisième ligation et d'employer de nouveau la cycloaddition CuAAC. Ainsi, il a été privilégié de réaliser une modification sur le châssis peptidique. En effet, on retrouve une fonction amine disponible sur la face supérieure du peptide et cette dernière peut, par diazo-transfert, être transformée en fonction azoture (**Schéma 9**).

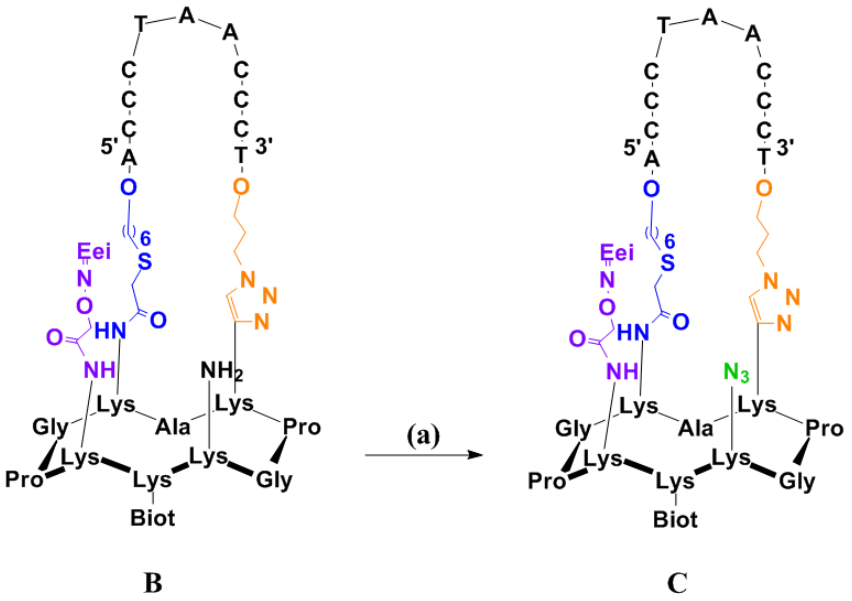


Schéma 9 : synthèse de l'édifice **C**
(a) $CuSO_4$ (3eq), ISAHC (30eq), $NaHCO_3$ 50mM pH 8,5/MeOH (3/1, v/v), 3h, 60°C

Les conditions réactionnelles de diazo-transfert sont identiques à celles utilisées auparavant sur l'oligonucléotide **6**. Le suivi de la réaction s'effectue par UPLC-MS (**Figure 45**). En fin de réaction, le brut réactionnel est dessalé sur colonne d'exclusion stérique et le composé **C** est caractérisé par spectrométrie de masse (**Figure 46**).

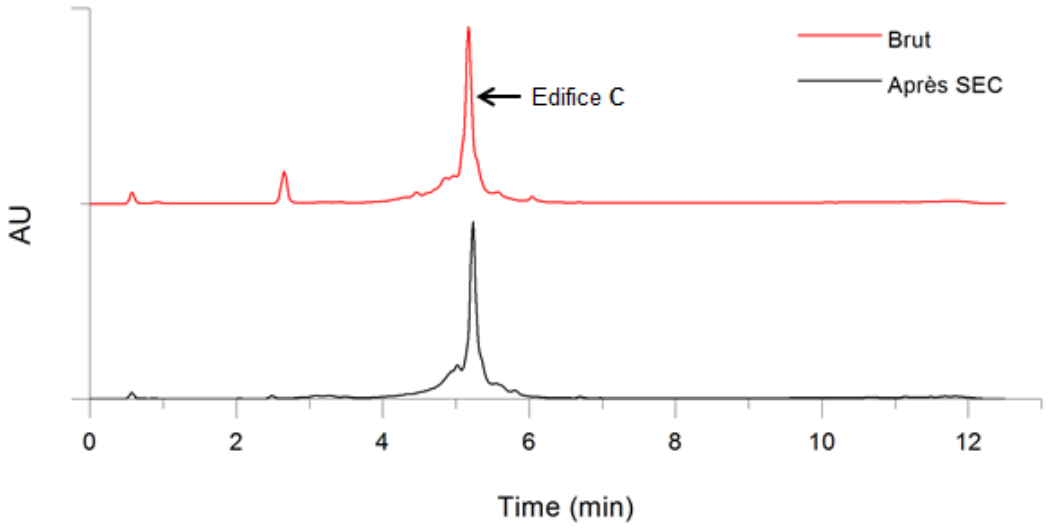


Figure 45 : chromatogrammes de la formation de l'édifice **C** en UPLC du brut réactionnel et après SEC

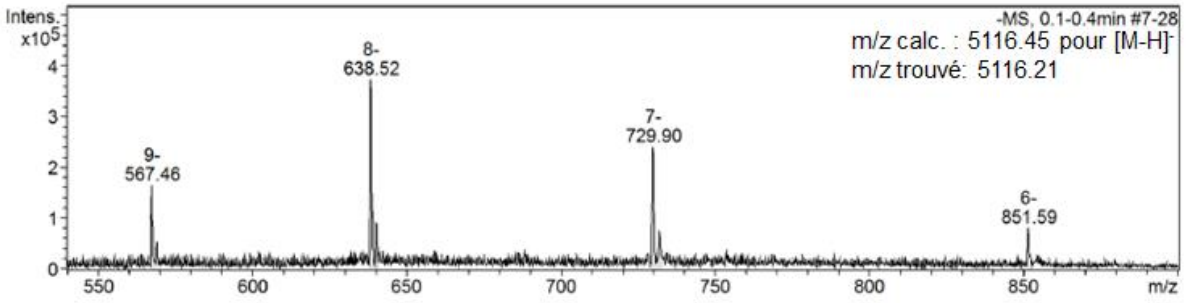


Figure 46 : analyse ESI-MS de l'édifice **C**

d) Assemblage de l'édifice F

A partir du composé **C**, deux voies sont possibles afin d'obtenir l'édifice final **F**. La première stratégie (voie A) consiste à greffer le second oligonucléotide **7** en commençant par la ligation oxime et en finissant par la deuxième réaction CuAAC. La deuxième possibilité (voie B) est de faire l'inverse. (**Schéma 10**). Ces deux voies de synthèse ont été investiguées.

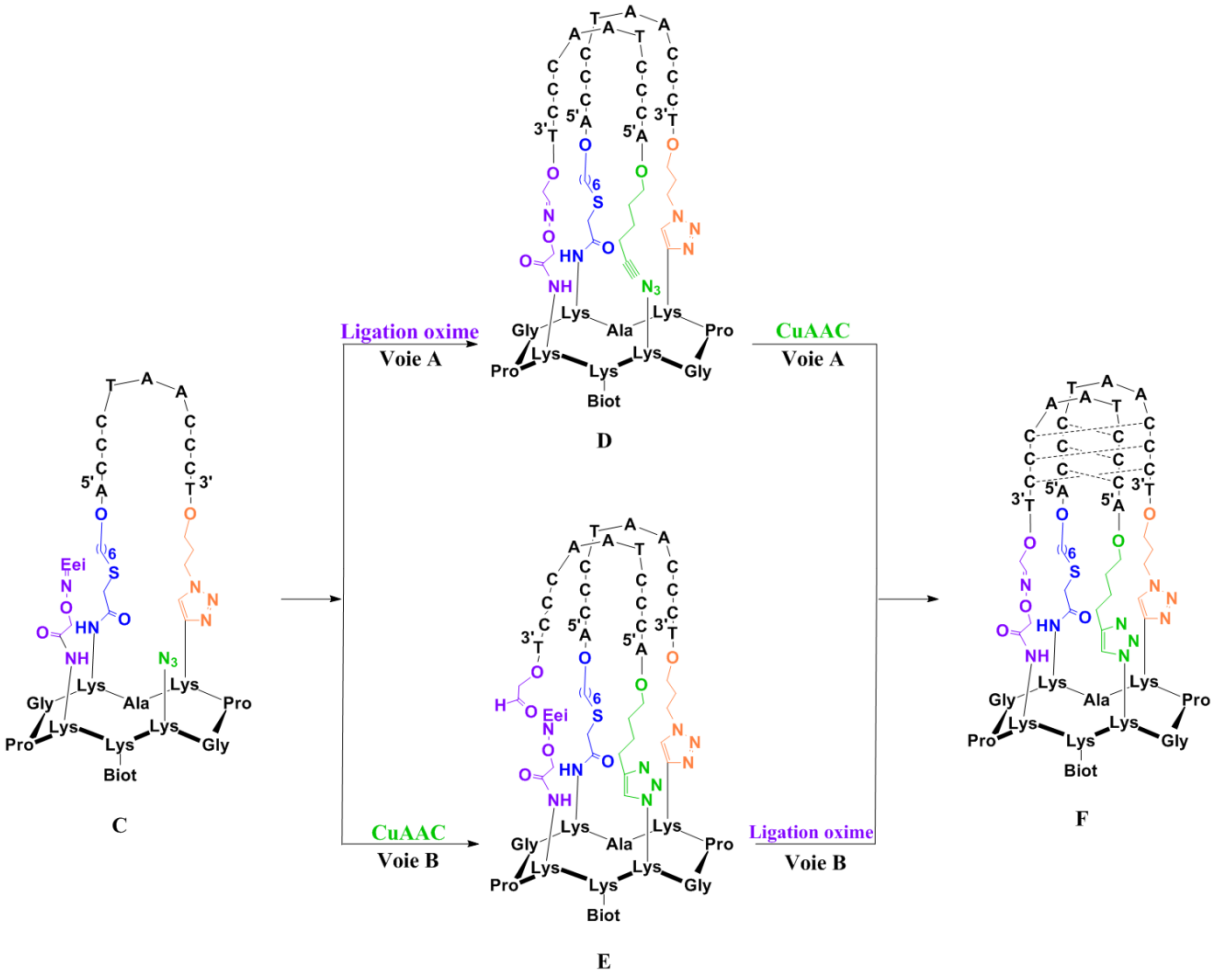


Schéma 10 : représentation des deux voies possibles pour la synthèse du mime i-motif **F**

i. <u>Synthèse de l'édifice **F** : voie A</u>

Dans cette stratégie, la première étape est de solubiliser dans une solution aqueuse de TFA 1% le composé **C** ainsi que l'oligonucléotide **7** afin que se déroule en même temps la déprotection de l'aminooxy et la ligation oxime dans le but de former le composé **D**. Il s'en suit la cycloaddition CuAAC afin d'obtenir *in fine* l'édifice **F** (**Schéma 11**).

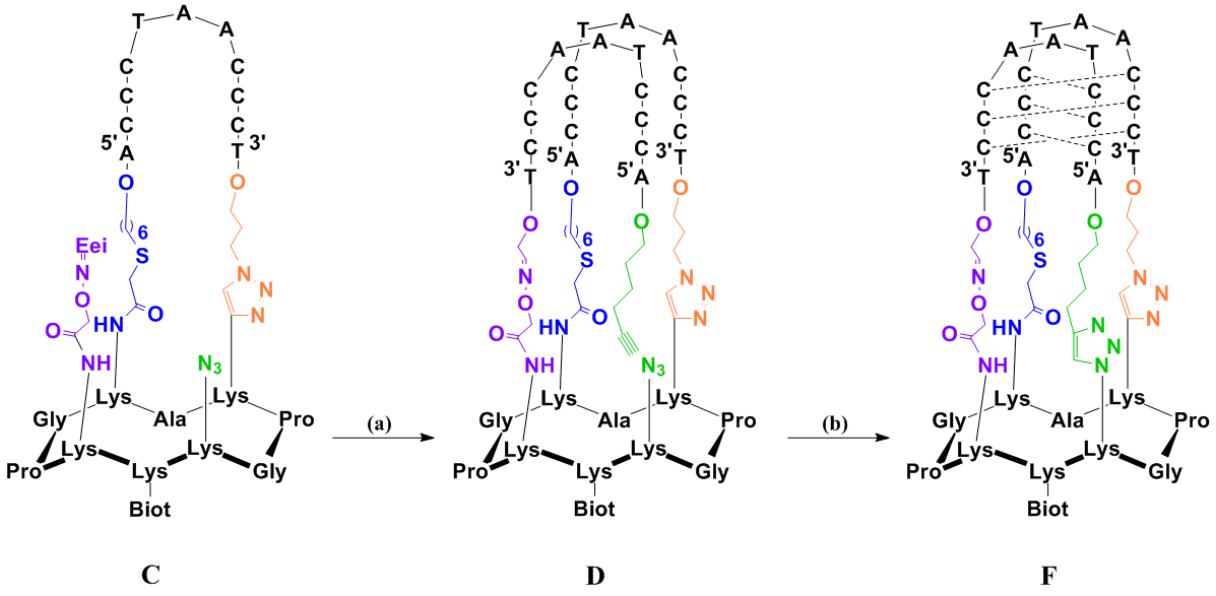


Schéma 11 : synthèse du mime i-motif **F** par la voie A
(a) Oligonucléotide **7** (2eq), TFA 1%vol dans H_2O, quantitatif, t.a. ; (b) $CuSO_4$ (3eq), THPTA (30eq), ascorbate de sodium (30eq), 500mM KCl, DMF (10%vol), PBS 50mM pH 7,4, 1h, 37°C ou 60°C

Lors de la première étape, un suivi réactionnel par UPLC-MS est choisi afin de confirmer par MS la perte du groupement protecteur Eei mais surtout afin de voir s'il y a bien eu la perte d'une molécule d'eau suite à la ligation oxime (**Figure 47**). On observe bien la formation du composé **D** mais on remarque que le pic représentant ce dernier est relativement large et présente un ou plusieurs produit(s) secondaire(s), difficilement interprétable(s) par UPLC-MS. Après purification par RP-HPLC, le produit ainsi obtenu est analysé par UPLC-MS, on retrouve à nouveau un pic relativement large et la présence, encore une fois, d'un ou plusieurs produits secondaires(s), difficilement interprétable par UPLC-MS.

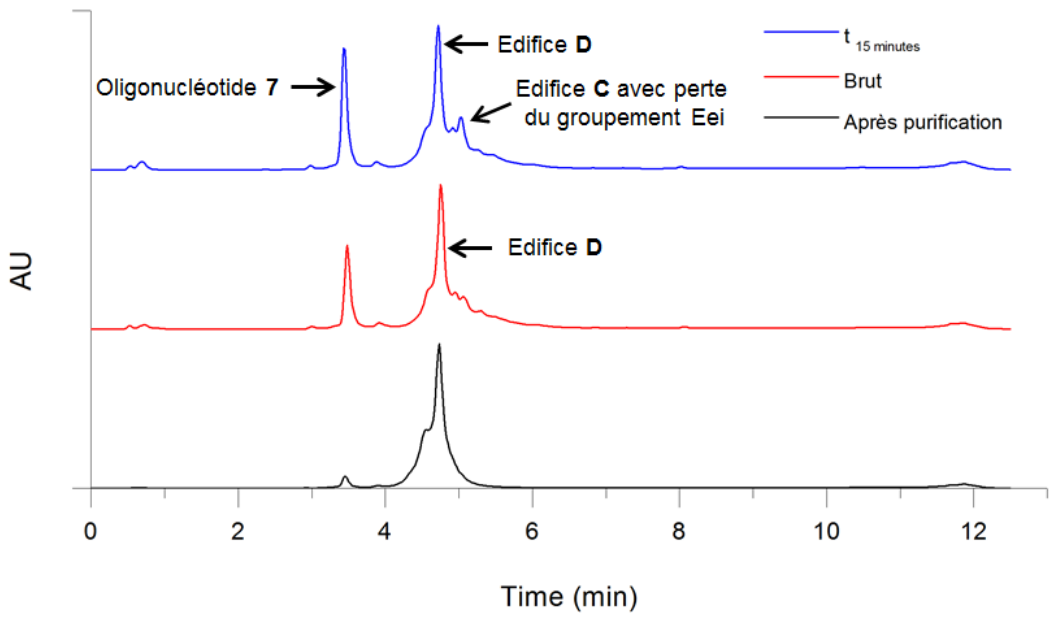


Figure 47 : suivi réactionnel par UPLC de la ligation oxime entre l'oligonucléotide **7** et l'édifice **C**

Malgré cela, la deuxième étape est tout de même réalisée dans l'espoir que la formation du produit final **F** permette d'obtenir une différence suffisante en termes de temps de rétention entre les composés **D** et **F** car aucun changement de masse ou encore de charge ne peut être observé. Différentes tentatives ont été réalisées mais n'ont pas été concluantes (**Figure 48**). En effet, aucun changement de temps de rétention n'a été observé et ce malgré différents tests comme l'utilisation des conditions réactionnelles usuelles de CuAAC, l'augmentation de KCl dans le milieu afin de limiter la répulsion entre les deux brins d'ADN ou encore l'augmentation de la température du mélange réactionnel (passage 37°C à 60°C). En raison de résultats non concluants, cette première voie n'a pas été investiguée plus en profondeur en faveur de la voie B.

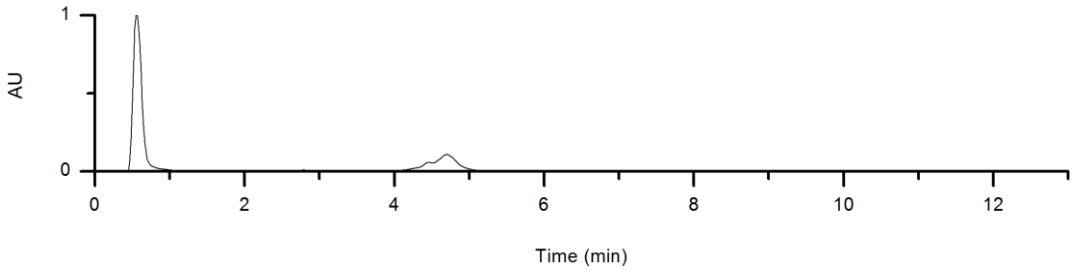


Figure 48 : chromatogramme en UPLC du brut réactionnel de la seconde cycloaddtion CuAAC pour la formation de l'édifice **F**

ii. Synthèse de l'édifice **F** : voie B

La première étape de cette stratégie consiste à réaliser la seconde cycloaddition CuAAC avec le deuxième brin d'ADN **7** afin d'obtenir le composé **E** (**Schéma 12**).

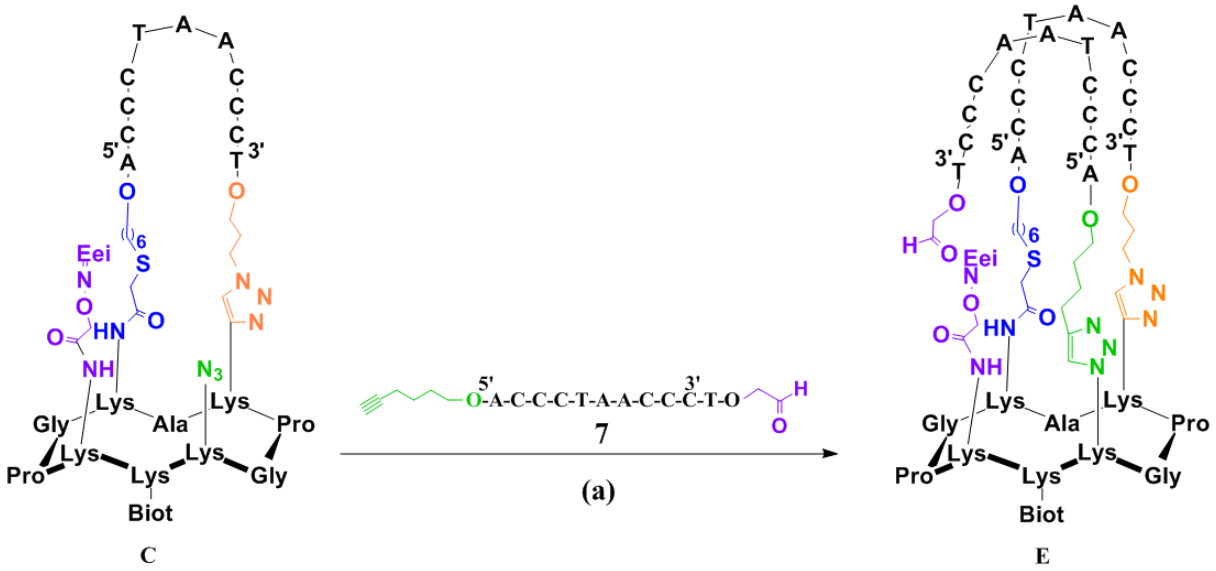


Schéma 12 : synthèse de l'intermédiaire **E**
(a) Oligonucléotide **7** (2eq), $CuSO_4$ (3eq), THPTA (30eq), ascorbate de sodium (30eq), 500mM KCl, DMF (10%vol), PBS 50mM pH 7,4, 1h, 60°C, 48%

Au départ, des conditions usuelles de cycloaddition ont été appliquées, cependant, la réactivité ne s'est pas montrée excellente. En effet, il a été observé l'apparition du produit **E** mais avec un taux de conversion non satisfaisant (aux alentours de 20-30%). Dans le but d'accroitre la formation du produit désiré, il a alors été testé de placer le mélange réactionnel à 60°C tout en gardant les mêmes quantités des autres réactifs. Cette augmentation de température a favorisé l'apparition du composé **E** (**Figure 49**). En fin de réaction, le brut réactionnel a été dessalé sur colonne d'exclusion stérique et ensuite purifié par RP-HPLC avec un rendement de 48%. Le composé **E** ainsi obtenu a été caractérisé par RP-HPLC et spectrométrie de masse (**Figure 50**). A noter qu'il est observé sur le chromatogramme RP-HPLC (**Figure 50a**) la présence de deux pics ce qui laisse suggérer la possible existence de deux structures i-motifs dans les conditions d'analyse.

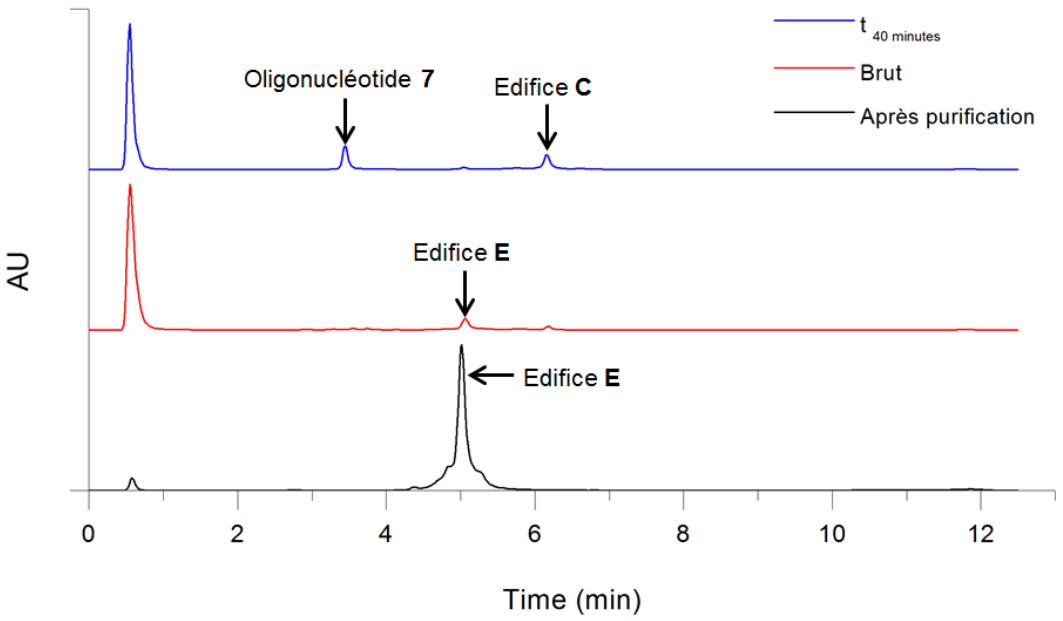


Figure 49 : suivi réactionnel par UPLC de la cycloaddition CuAAC entre l'oligonucléotide **7** et l'édifice **C**

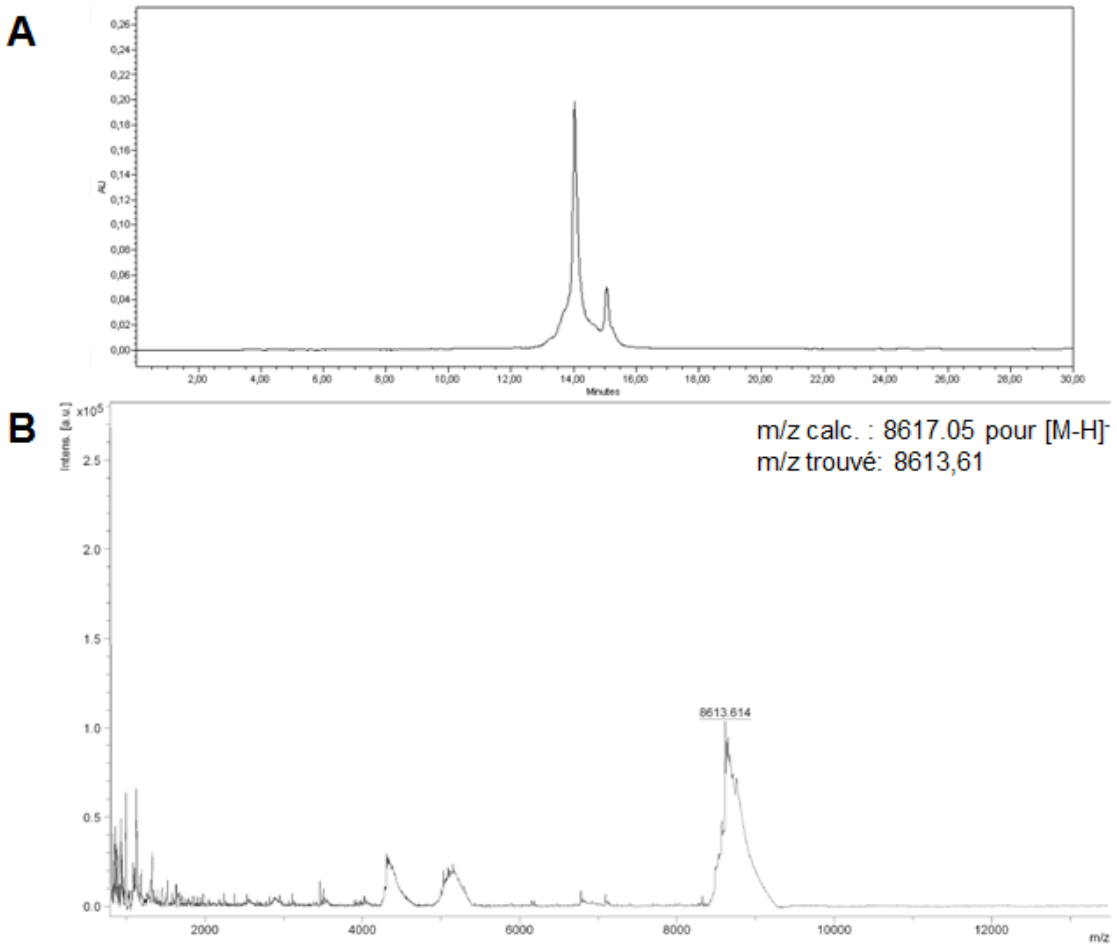


Figure 50 : (a) chromatogramme RP-HPLC et (b) analyse MALDI de l'édifice **E**

La deuxième étape consiste à réaliser à la fois la déprotection de l'oxyamine et la ligation oxime. Pour ce faire, le composé **E** est solubilisé dans une solution aqueuse de TFA 1% (**Schéma 13**).

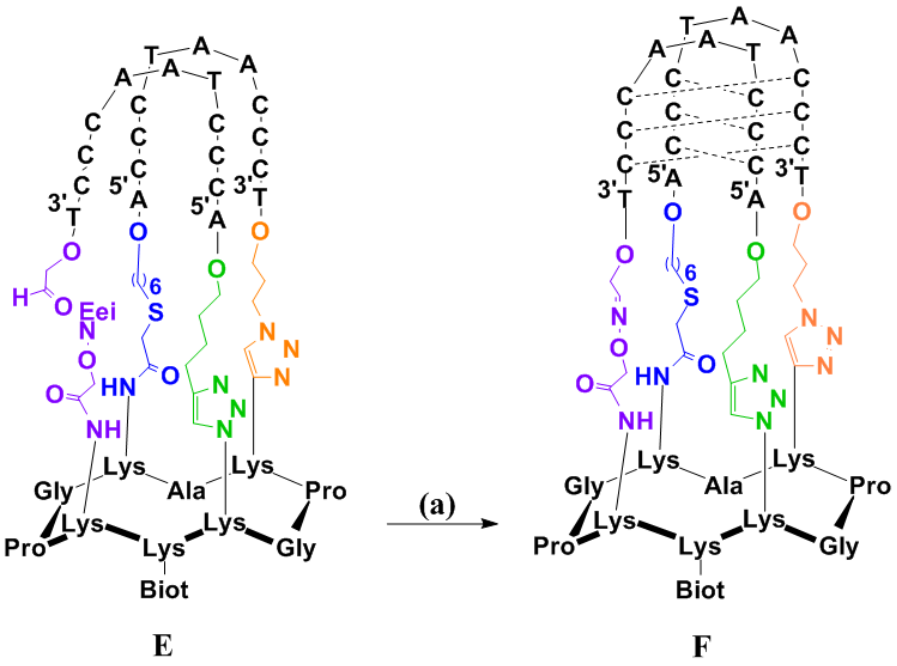


Schéma 13 : synthèse de l'édifice **F**
(a) TFA 1%vol dans H_2O, quantitatif, t.a.

Comme mentionné précédemment, l'un des intérêts de l'utilisation du groupement Eei sur la fonction oxyamine du châssis peptidique est le fait que ce dernier se déprotège facilement en milieu acide. En effet, ici une solution aqueuse de TFA 1% est suffisante pour que la déprotection s'effectue mais permet également de former la deuxième boucle et ainsi obtenir le composé final **F**. Le suivi réactionnel (**Figure 51**) montre que la totalité du composé **E** est déprotégé et on observe aussi la perte d'une molécule d'eau, intrinsèque à la ligation oxime. Le mime i-motif **F** ainsi obtenu est caractérisé par RP-HPLC et spectrométrie de masse (**Figure 52**). A noter qu'on observe sur le chromatogramme RP-HPLC (**Figure 52a**) la présence de deux pics, on pose l'hypothèse qu'il y ait la présence de deux structures i-motifs dans les conditions d'analyse.

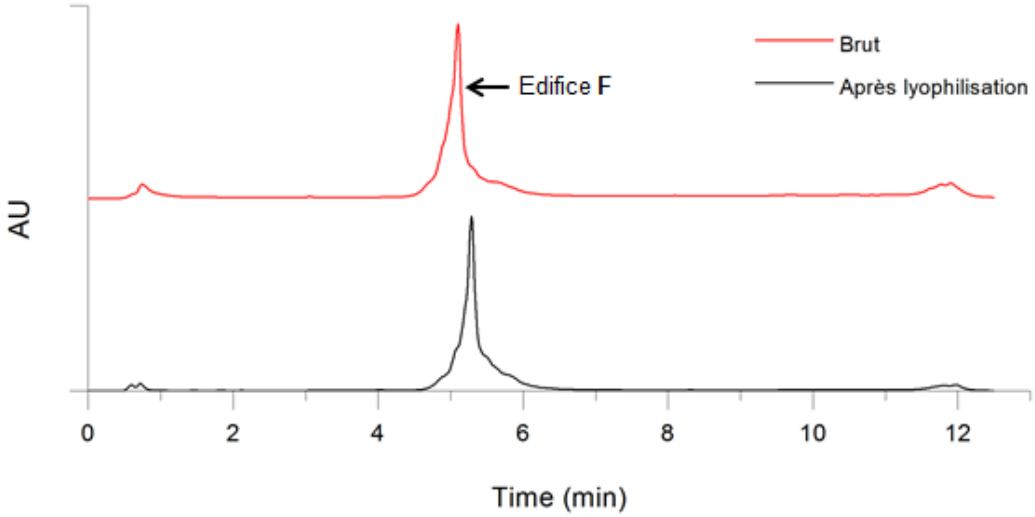


Figure 51 : chromatogrammes de la formation de l'édifice **F** en UPLC du brut réactionnel et après lyophilisation

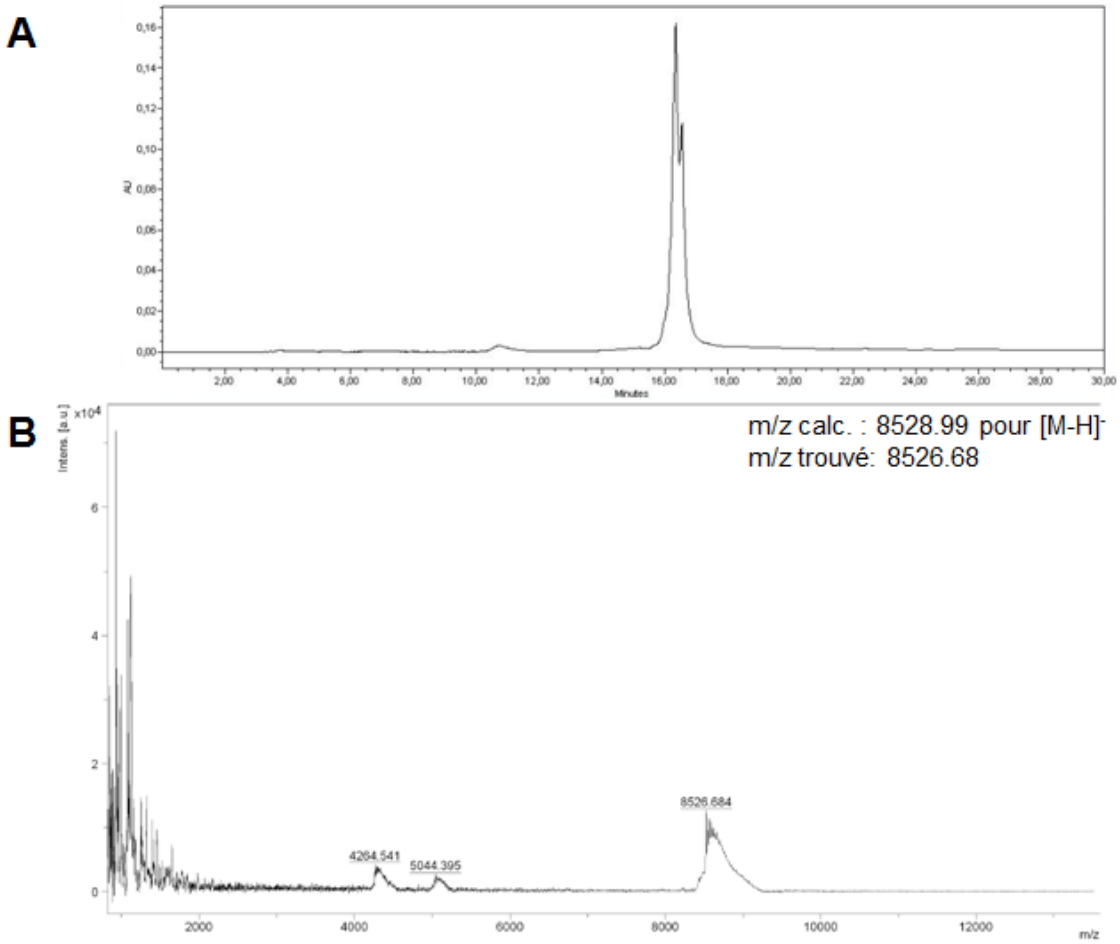


Figure 52 : (a) chromatogramme RP-HPLC et (b) analyse MALDI-Tof de l'édifice **F**

6. Bilan intermédiaire

Au cours de cette partie synthèse, il a été montré qu'une seule voie permettait d'obtenir le mime i-motif **F**. La première étape consiste à réaliser le couplage thiol-chloroacétamide entre l'oligonucléotide **6** et le châssis peptidique **1**. Durant cette première étape, différentes précautions ont été mises en place afin de retarder l'ajout de TCEP dans le milieu réactionnel car ce dernier réduit la fonction azoture présent en position 3' de l'oligonucléotide **6**. Malgré cela, cette réduction reste tout de même présente et fluctue entre 10-20%. Le composé **A** ainsi formé suite à cette première étape peut ensuite subir une réaction de cycloaddition CuAAC afin de former l'édifice **B**. Dans la stratégie de synthèse établie, l'édifice final **F** est obtenu en faisant appel à deux autres ligations pour greffer l'oligonucléotide **7**. Il a été choisi d'employer la chimie des oximes comme troisième ligation et de réutiliser la cycloaddition CuAAC. Pour ce faire, une fonction azoture est générée à la surface supérieure du support à partir d'une fonction amine par le biais d'une réaction de diazo-transfer. A partir du composé **C** ainsi obtenu, deux voies se sont profilées dans le but de former le mime i-motif **F**. Toutefois, une seule voie s'est montrée concluante, la voie B. Celle-ci fait appel à la seconde cycloaddition CuAAC dans un premier temps, et enfin la ligation oxime permet de fermer la deuxième boucle et former *in fine* l'édifice **F**. Au cours de cette dernière réaction, un changement de temps de rétention et de masse permet la confirmation de la formation du composé final ce qui n'a pas été observé lors de l'investigation de la voie A.

7. Etude de la topologie du mime i-motif

Il a été mis au point précédemment une stratégie de synthèse dans le but de former un mime i-motif présentant la séquence télomérique (édifice **F**). Les caractérisations par RP-HPLC, UPLC-MS et par spectrométrie de masse ont permis de confirmer l'obtention de ce dernier. Cependant, la capacité de ce mime à adopter une structure en i-motif stable reste à déterminer. Différentes techniques d'analyses sont disponibles dans le but de vérifier quelle topologie est présente. Afin de caractériser la structuration du mime **F**, nous avons choisi d'utiliser le dichroïsme circulaire (CD), celle-ci présente l'avantage d'être non destructrice et peu de produit est utilisé ce qui est particulièrement avantageux dans notre cas.

Le dichroïsme circulaire permet d'évaluer les structures secondaires qu'adoptent les séquences d'ADN par le biais d'un signal caractéristique. Dans le cas des i-motifs, le spectre CD 'classique', possède un maximum à environ 285 nm et un minimum proche de 260 nm.

Dans le cadre de cette thèse, des études de dichroïsme circulaire ont été réalisées en fonction du pH sur les édifices **E** et **F** mais également sur une séquence d'ADN riche en cytosine (hTeloC, **8**). L'oligonucléotide **8**, de séquence d(C_3TAA)$_3$$C_3$T, a été préparé par SPOS (voir partie expérimentale).

a) Etude de la topologie i-motif de hTeloC 8 en fonction du pH

On peut qualifier la structuration i-motif comme un appariement des cytosines où certaines d'entre elles sont protonées en position N3. Ainsi, il est intéressant de regarder comment le pH influence la structuration i-motif. Nous avons d'abord étudié ceci avec la séquence témoin **8**. Pour cela, hTeloC **8** est solubilisé dans un mélange tampon PBS 10mM où le pH varie de 4 à 7 (**Figure 53**).

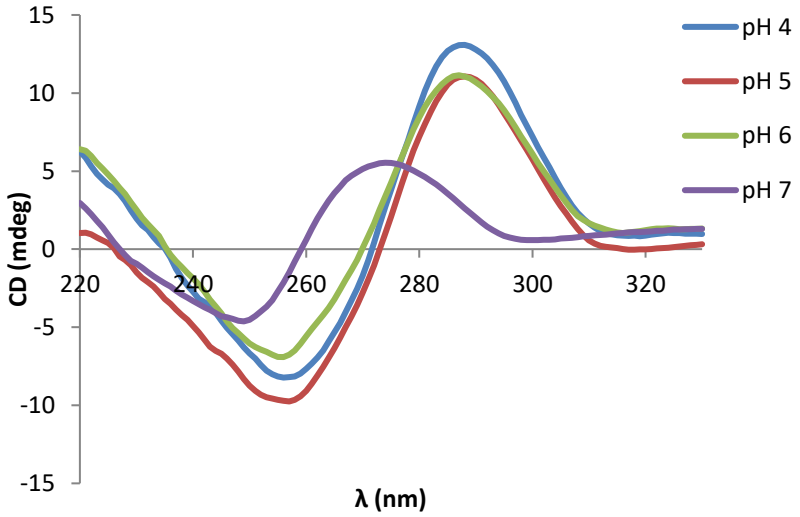


Figure 53 : spectre CD à 20°C de la séquence hTeloC **8**
Condition : 2 µM de **8** dans un tampon PBS 10mM à pH 4, 5, 6 ou 7

On observe la signature caractéristique de la topologie i-motif à pH 4, 5 et 6 pour la séquence **8**, puisque celle-ci présente un maximum proche de 285 nm et un minimum dans les alentours de 260 nm. En revanche, à pH 7, cette signature disparaît étant donné que l'on ne retrouve plus qu'un maximum vers les 275 nm, révélant ainsi la destruction de la topologie i-motif. Ces résultats concordent avec la littérature où l'on retrouve pour la séquence d($^{5'}$TAACCC$^{3'}$)$_4$ un pH de transition (pH$_T$) de 6,5[107].

b) Etude de la topologie i-motif de l'édifice E et F en fonction du pH

Suite aux études effectuées précédemment sur la séquence **8**, il est intéressant d'analyser l'influence du pH sur la topologie des édifices **E** et **F**. En comparant ces deux composés, on

pourra estimer si le fait de greffer les séquences d'ADN sur un support permet d'améliorer la stabilité de la topologie i-motif en fonction du pH mais également l'effet de trois ou quatre attaches. Pour cela, les mêmes conditions expérimentales appliquées pour hTeloC **8** ont été utilisées pour les édifices **E** et **F** (**Figure 54**).

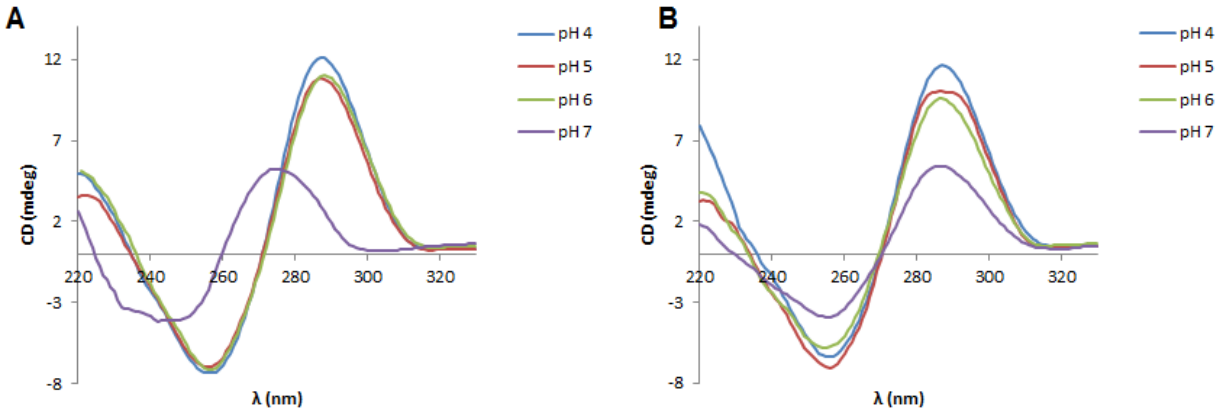


Figure 54 : spectres CD à 20°C de l'édifice **E** (a) et de l'édifice **F** (b)
Conditions : 2 µM de **D ou F** dans un tampon PBS 10mM à pH 4, 5, 6 ou 7

Concernant l'édifice **E**, on observe la même tendance que pour la séquence hTeloC **8**. En effet, on remarque que **E** adopte une topologie i-motif pour les pH 4, 5 et 6 et à pH 7, on perd la signature caractéristique. En ce qui concerne le composé **F**, on voit, comme précédemment à pH 4, 5 et 6, un maximum vers les 285 nm et un minimum dans les alentours de 260 nm, signal caractéristique de la topologie i-motif. A pH 7, cette signature est conservée. Toutefois à ce pH, l'intensité est plus faible, signifiant que la totalité de l'édifice **F** n'adopte pas la structuration i-motif. De plus, de part ces études, on remarque l'impact et l'importance de l'immobilisation des deux séquences d'ADN sur le support peptidique *via* quatre points d'attache du composé **F** puisqu'à pH 7 ce dernier conserve la topologie i-motif contrairement au composé **E** ou à la séquence d'ADN **8**.

c) Etude de stabilité de la topologie i-motif sur l'édifice F

Par dichroïsme circulaire, on peut également évaluer la stabilité de la topologie i-motif en mesurant la température de dénaturation (T_m). Il a été observé précédemment qu'à pH 6, le composé **F** adoptait totalement une topologie i-motif et à pH 7, cette structuration était conservée mais était moins intense. Ainsi, il est intéressant d'étudier la stabilité de cette topologie par la détermination des températures de dénaturation à ces deux pH pour le composé **F**. Pour cela, l'édifice **F** est solubilisé dans un mélange tampon PBS 10mM à pH 6 et 7 et des mesures CD sont réalisées sur différentes gammes de températures (**Figure 55**).

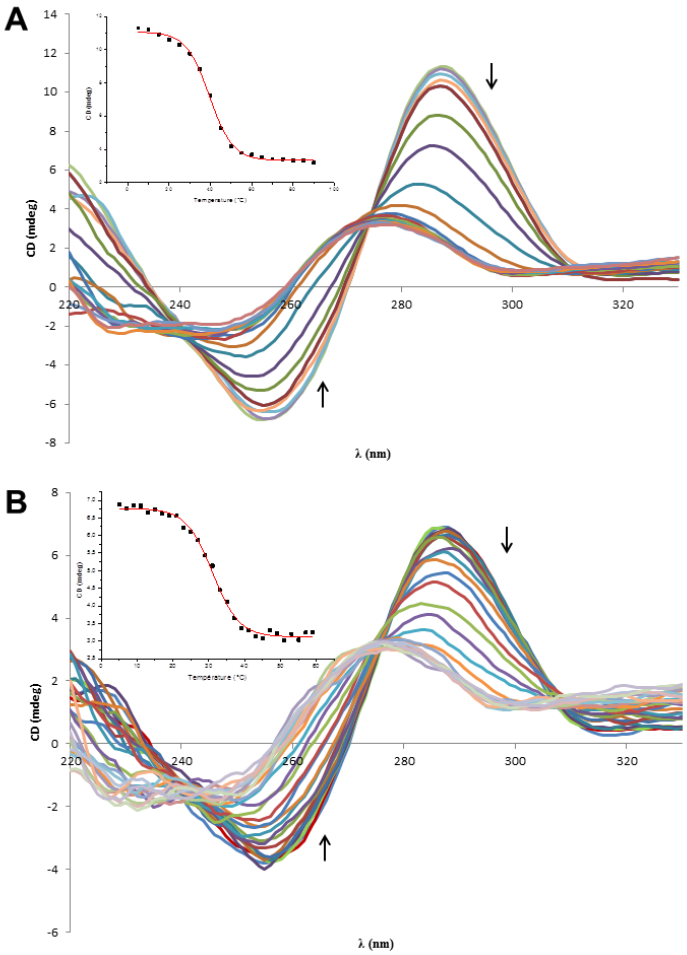


Figure 55 : spectres CD de dénaturation du mime i-motif **F** à pH 6 (a) ou à pH 7 (b)
(a) Condition : 2 µM de **F** dans un tampon PBS 10mM pH 6. Gamme de température : 5°C à 90°C avec un delta T de 5°C.
(b) Condition : 2 µM de **F** dans un tampon PBS 10mM pH 7. Gamme de température : 5°C à 59°C avec un delta T de 2°C.

A pH 6, la température de dénaturation obtenue est de 39,5°C. Celle-ci est similaire à ce qu'on peut trouver dans la littérature concernant la séquence télomérique. En effet, pour la séquence $({}^{5'}TAACCC^{3'})_3$, on retrouve une T_m égale à 41°C à un pH de 5,5. Ceci indiquerait même que la topologie i-motif de l'édifice **F** semble être un peu plus stable étant donné que la T_m a été obtenue à pH 6[107]. Cette légère augmentation semble provenir de la présence du support peptidique et des quatre points d'attache des deux brins d'ADN. A pH 7, on obtient une T_m de 31°C. On observe une nette diminution des T_m entre les deux pH mais reste cependant supérieure à ce que l'on peut retrouver dans la littérature.

Dans le but de s'assurer que le signal obtenu en dichroïsme circulaire à pH 6 n'est pas dû à la formation de structures i-motifs intermoléculaire, c'est-à-dire dû à l'assemblage de brins d'ADN présents sur plusieurs édifices, les températures de dénaturation ont été déterminées à

deux autres concentrations : 0,5µM et 1µM (**Figure 56**). Si les valeurs de T_m obtenues sont similaires, cela indique qu'il n'y a aucune influence de la concentration sur la stabilité de la topologie i-motif, confirmant ainsi une structuration intramoléculaire.

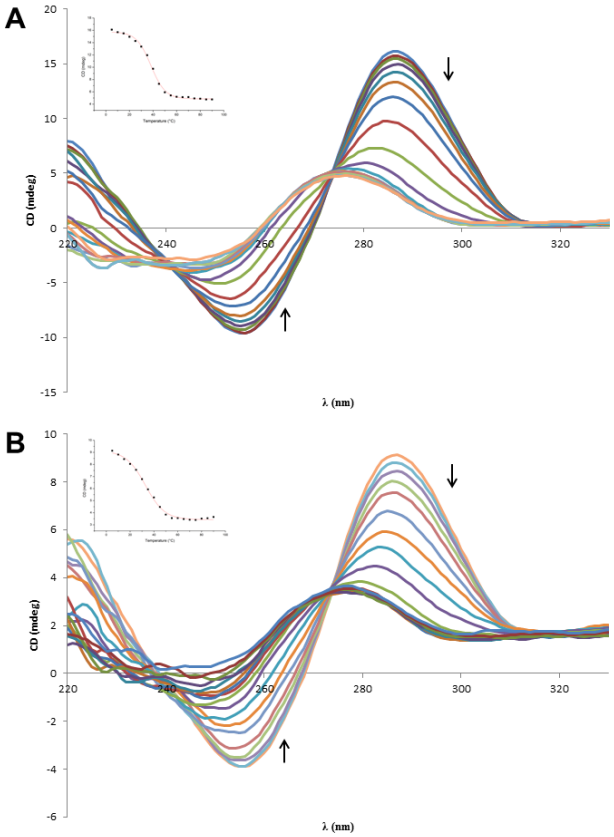


Figure 56 : spectres CD de dénaturation du mime i-motif **F** à 1 µM (a) ou à 0,5 µM (b)
Condition : tampon PBS 10mM pH 6. Gamme de température : 5°C à 90°C avec un delta T de 5°C

Le **Tableau 3** reprend les valeurs de T_m obtenues avec les concentrations 2µM, 1µM et 0,5µM. On remarque que ces valeurs fluctuent légèrement mais sont relativement similaires et n'augmentent pas en proportion de la concentration. On peut ainsi valider que la formation de la topologie i-motif est intramoléculaire en ce qui concerne le mime **F** à pH 6.

Tableau 3 : Températures de dénaturation du mime i-motif **F** obtenues à différentes concentrations dans un tampon PBS 10mM pH 6

	2µM	**1µM**	**0,5µM**
T_m	39,5 °C	38,1°C	40,7°C

8. Bilan et perspectives

Dans ce chapitre, les travaux ont décrits dans un premier temps la synthèse d'un mime i-motif possédant deux séquences d'ADN télomérique. Cela a été permis à l'aide du concept TASQ et l'utilisation conjointe de trois ligations chimiosélectives, à savoir le couplage thiol-chloroacétamide, la cycloaddtion CuAAC et la ligation oxime. De plus, il a été nécessaire de réutiliser une de ces ligations afin de former le mime i-motif final **F**. Ainsi dans le plan de synthèse établi, le choix s'est porté sur la réutilisation de la cycloaddition CuAAC et pour cela, une modification sur le châssis peptidique a été réalisée. A partir de là, deux voies de synthèses se sont profilées mais une seule s'est montrée concluante. En effet, l'utilisation de la seconde cycloaddition CuAAC suivi de la ligation oxime a permis d'obtenir le composé final **F**. Dans un second temps, des études de dichroïsme circulaire ont été effectuées afin de caractériser la topologie de l'édifice **F**. En effet, ces études ont permis de confirmer que ce dernier adopte une structuration i-motif et de plus, cela sur une gamme de pH plus importante que la séquence télomérique naturelle 'libre'. Des températures de dénaturation de 39,5°C et 31°C ont été obtenues respectivement, à pH 6 et 7. Et enfin, il a été confirmé que cette topologie i-motif était bel et bien de type intramoléculaire et non intermoléculaire puisque les études de dénaturation ont données des T_m similaires pour différentes concentrations du composé **F** à pH 6.

Concernant les perspectives, tout d'abord une caractérisation par ^{1}H-RMN permettrait de fournir une nouvelle preuve de l'existence de la structure i-motif du composé **F**. De plus, ce dernier peut être considéré comme une cible attrayante dans le cadre d'études d'interactions. Effectivement, il serait intéressant, par exemple, d'analyser l'interaction entre ce mime possédant les deux brins d'ADN télomérique et un extrait cellulaire contenant un large panel de protéines. Ainsi, il s'effectuerait une 'pêche de protéines' où ces dernières seraient identifiées. *In fine*, un renseignement peut être obtenu concernant quelles sont les protéines qui interagissent avec cette séquence et si celles-ci sont connues pour avoir un rôle biologique spécifique. De plus, des études d'affinité et de sélectivité de ligands peuvent être également imaginées par résonnance plasmonique de surface (SPR). Ces exemples d'études d'interactions sont envisageables puisqu'on retrouve sur le mime i-motif **F** une fonction biotine capable de se lier fortement à la streptavidine, fonction qui peut recouvrir un support solide (exemple : bille magnétique, puce, …).

Chapitre IV

Introduction d'une nouvelle ligation chimiosélective pour les acides nucléiques et synthèse d'un conjugué en utilisant quatre ligations orthogonales

Lors de ce chapitre, il sera abordé l'incorporation et l'utilisation d'une nouvelle ligation chimiosélective aux trois autres couramment utilisées au sein de notre laboratoire. L'utilité d'identifier une nouvelle ligation chimiosélective pour la formation de conjugués d'oligonucléotides ainsi que l'intérêt d'utiliser quatre ligations orthogonales afin de synthétiser des édifices G4 plus complexes, c'est-à-dire possédant des topologies particulières et ainsi obtenir *in fine* des cibles G4 attrayantes d'un point de vue biologique, sera décrit. Ce chapitre présente de manière détaillée l'intégration de cette nouvelle ligation à la conjugaison entre oligonucléotides et peptides. Il sera également abordé la formation d'édifices possédant quatre ligations chimiosélectives.

1. Présentation de différentes ligations chimiosélectives

Les ligations chimiosélectives ont énormément contribué ces dernières années dans différents domaines comme la synthèse organique ou encore la chimie du vivant. On peut définir une ligation chimiosélective comme étant une réaction permettant une association spécifique entre deux molécules présentant plusieurs fonctions réactives. On retrouve également les chimies « Click » qui idéalement se déroulent en présence de fonctions chimiques déprotégées, dans des conditions réactionnelles douces, tolérant un environnement aqueux, en l'absence d'un catalyseur et sans formation de sous-produit(s). Dans le domaine de la synthèse, une ligation chimiosélective dite efficace suit un certain nombre de critères stricts. En effet, l'assemblage de deux composés doit s'effectuer avec une cinétique élevée. Les réactifs ainsi que le produit formé doivent être stables dans les conditions réactionnelles. Et enfin, cette ligation doit fournir un bon rendement avec une stœchiométrie proche du 1/1.

On retrouve également la notion de ligation bio-orthogonale. Elle se définit comme une ligation chimiosélective mais celle-ci fait intervenir au moins une biomolécule. Ainsi, le concept de la chimie bio-orthogonale est de combiner à la fois la chimie organique et la biologie afin de développer des outils permettant de comprendre le monde du vivant. Concernant la ligation bio-orthogonale, d'autres critères peuvent être rajoutés à ceux énumérés précédemment. En effet, les notions de solvant, température mais également de concentration sont à prendre en compte. De plus, généralement les réactions bio-orthogonales se déroulent dans un système aqueux, imposant ainsi aux différents composés mis en jeux d'être solubles mais surtout stables dans ce type de milieu. Il faut également savoir si la/les fonction(s) chimique(s) introduite(s) sur la/les biomolécule(s) ne vont pas altérer une ou plusieurs propriété(s) intrinsèque(s) à celle-ci. De plus, dans le cadre de ligation bio-orthogonale, la

notion de cinétique de réaction est d'autant plus importante. En effet, lorsqu'une biomolécule est mise en jeu, sa concentration dans le milieu est relativement faible ce qui impacte directement la cinétique de la réaction. Ainsi, cette dernière doit être élevée afin qu'un taux de conversion acceptable soit obtenu. Et enfin, l'aspect de sélectivité est également très important. Effectivement, il faut s'assurer qu'il n'y ait pas de de réaction(s) secondaire(s) avec des fonctions chimiques présentes naturellement sur la biomolécule. Tous ces éléments indiquent que la conception et la mise en œuvre d'une chimie bio-orthogonale est un réel challenge.

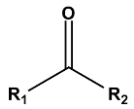
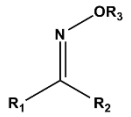
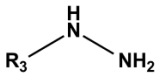
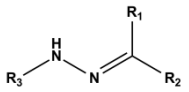
Dans la littérature, on retrouve de nombreuses réactions chimiosélectives et dans cette partie introductive, les réactions les plus utilisées seront présentées.

a) Ligation faisant intervenir une fonction cétone ou aldéhyde

Les composés possédant un groupement carbonyle tels que les cétones ou les aldéhydes sont fréquemment utilisés dans l'écosystème des ligations chimiosélectives. La réaction entre une aldéhyde/cétone et une alcoxyamine ou hydrazine forme, respectivement, un éther d'oxime (**Tableau 4, entrée 1**) ou une hydrazone (**Tableau 4, entrée 2**).

De manière générale, ce type de réaction s'effectue en milieu acide, dans les alentours d'un pH de 4,5 pour que la cinétique de réaction soit optimale. En effet lorsque le milieu est trop acide, le caractère nucléophile de l'alcoxyamine ou de l'hydrazine se trouve amoindrie. De plus, il a été montré que la constante d'équilibre pour les hydrazones est comprise entre 10^4-10^6 M^{-1} et est supérieure à 10^8 M^{-1} pour les oximes[177]. Cependant lorsque cette réaction s'effectue dans des conditions à pH neutre et/ou fait appel à des biomolécules qui sont souvent présentes à de faibles concentrations, cela peut affecter la cinétique de la réaction. Afin de limiter cela, l'une des stratégies a été de développer des catalyseurs permettant une ligation efficace et ce dans des conditions 'biologiques'. On retrouve parmi ceux-ci l'aniline qui joue un rôle de 'catalyseur nucléophile' et augmente jusqu'à 40 fois la vitesse de réaction à pH neutre[177]. Cependant, un désavantage à l'utilisation de ce catalyseur est qu'il faut employer une quantité élevée de celui-ci (10-100mM)[178,179]. Les acides anthraniliques et l'acide 3,5-diaminobenzoique peuvent être également utilisés pour catalyser la formation d'oximes et d'hydrazones à pH neutre. L'avantage de l'utilisation de ces derniers est la faible quantité nécessaire (1mM)[180]. On retrouve dans la littérature de nombreux exemples de bioconjugués utilisant les oximes ou les hydrazones et notamment pour la modification d'oligonucléotides[181–185]. De plus, diverses études ont également montré l'orthogonalité de la ligation oxime avec d'autres types de chimie comme la cycloaddition CuAAC, décrite ci-après[175,186,187].

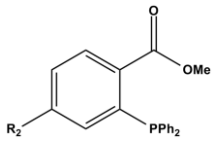
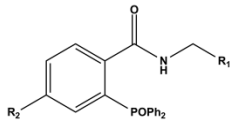
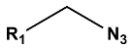
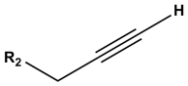
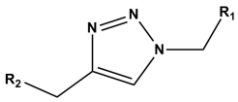
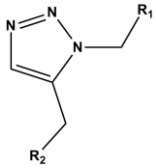
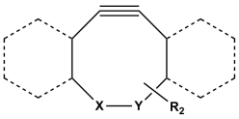
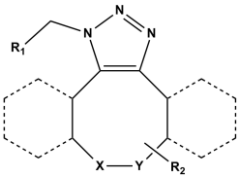
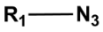
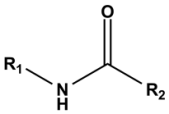
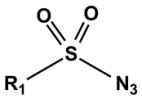
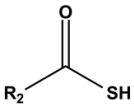
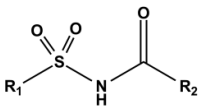
Tableau 4 : représentation des réactions possibles avec une fonction cétone ou aldéhyde

Entrée	Fonction	Partenaire	Produit
1			
2			

b) Ligation faisant intervenir une fonction azoture

Le groupement azoture est l'une des fonctions les plus populaires dans le domaine des réactions chimiosélectives. Sa petite taille mais surtout la capacité à être inerte explique cette notoriété. En effet, cette fonction n'affecte que très légèrement la structure du substrat sur lequel elle est introduite. L'incorporation de la fonction azoture peut s'effectuer de différentes manières, par réaction de diazo-transfert sur une fonction amine ou si le substrat est une protéine, par l'introduction d'un acide aminé non naturel possédant cette fonction azoture.

Tableau 5 : représentation des réactions possibles avec une fonction azoture
[a] L'utilisation d'un catalyseur de Cu(I) forme l'isomère 1,4 ; [b] L'utilisation d'un catalyseur de Ru(II) forme l'isomère 1,5

Entrée	Fonction	Partenaire	Produit
1			
2[a]			
3[b]			
4			
5			
6			

En 2000, Bertozzi *et al.* ont développé une réaction que l'on appelle plus communément la ligation de Staudinger, dérivée de la réduction de Staudinger. Celle-ci s'effectue entre une fonction azoture et un dérivé triphénylphosphine (**Tableau 5, entrée 1**). Au cours de cette réaction, un intermédiaire aza-ylure est formé et celui-ci subit, en présence d'eau, un réarrangement formant *in fine* une liaison covalente amide[188,189]. Malgré le fait que cette ligation se montre sélective et efficace, elle présente certains désavantages. En effet, les fonctions phosphines sont connues pour être oxydable ainsi une possible oxydation peut s'effectuer avant ligation. De plus, la cinétique de la réaction peut être considérée comme relativement faible[189].

Dans le début des années 2000, la cycloaddition 1,3-dipolaire entre les fonctions azoture et alcyne catalysée par le cuivre(I) (également appelée CuAAC pour « Copper catalyzed Azide

Alkyne Cycloaddition ») (**Tableau 5, entrée 2**) a été décrite par Meldal[190] et Sharpless[191]. Depuis, cette réaction est largement utilisée dans des divers domaines comme par exemple dans la synthèse organique[192] ou dans la chimie des biomolécules avec notamment la modification de protéines[193] ou encore en matière de fonctionnalisation d'ADN[194]. En ce qui concerne les oligonucléotides, l'introduction des fonctions chimiques nécessaire à la CuAAC est aisée et bien établie, cette chimie est grandement utilisée et est compatible avec diverses autres ligations[175,176,187,195,196]. Toutefois, l'utilisation du cuivre(I) comme catalyseur pose problème par sa toxicité pour les systèmes vivants, causant la mort de la cellule, une inhibition de la division cellulaire ou encore une dégradation de l'ADN[197]. Cependant, ces effets indésirables peuvent être réduits par l'utilisation de certains ligands comme par exemple le ligand THPTA (pour « tris-hydroxypropyltriazolylméthylamine ») qui limite les effets cytotoxiques du Cu(I)[198–200]. De plus, un contrôle sur la régiosélectivité peut s'effectuer sur le produit formé par l'utilisation de catalyseurs. En effet, l'isomère 1,4 est produit avec l'utilisation d'un catalyseur à base de cuivre (**Tableau 5, entrée 2**) alors que l'isomère 1,5 (**Tableau 5, entrée 3**) est obtenu suite à l'usage d'un catalyseur au ruthénium(II).

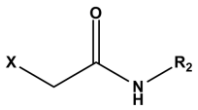
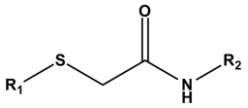
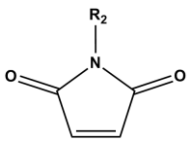
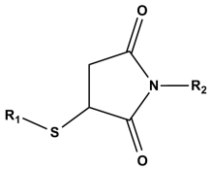
Les problèmes de toxicité liés à l'utilisation du Cu(I) peuvent être évités par l'emploi de ligations bio-orthogonales n'utilisant pas de catalyseur. Dans les années 1950, Blomquist et Liu ont publiés des travaux faisant appel à des composés cycliques contraints, les cyclooctynes[201]. Malgré cette découverte dans les années 1950, ce n'est que dans le début des années 2000 qu'un nouvel intérêt s'est porté sur ces composés[202,203]. Ces derniers, en présence de composés azotures, peuvent réaliser une cycloaddition [3+2] et ce sans l'utilisation d'un catalyseur (**Tableau 5, entrée 4**). Cette réaction est appelée SPAAC pour « Strain-Promoted Azide Alkyne Cycloaddition ». Depuis, on retrouve dans la littérature de nombreux travaux utilisant cette chimie et notamment dans la chimie du vivant[204,205]. Bien que l'utilisation de ces cyclooctynes présente l'énorme avantage d'être biocompatible en n'utilisant pas de catalyseur pouvant être cytotoxique à partir d'un certain seuil, la SPAAC possède tout de même des inconvénients. En effet, en plus d'être non régiosélective (de manière générale), la cinétique de cette ligation peut se montrer basse. De plus, certains dérivés cyclooctynes ne sont pas stables dans le temps[206]. Et enfin, généralement, la synthèse de ces dérivés est relativement longue et peut présenter des rendements faibles. Malgré ces désavantages, il est important de souligner que l'on retrouve énormément d'applications de la SPAAC dans le domaine de la chimie-biologie[207] mais également dans la chimie des oligonucléotides[208,209].

La réaction entre un dérivé azoture et un composé thioacide peut conduire à la formation d'un lien amide (**Tableau 5, entrées 5 et 6**). En 2003, le groupe de L. Williams a publié des travaux mettant en lumière ce type de réaction s'effectuant à pH neutre ou basique[210]. Cette réaction présente les avantages d'avoir une cinétique élevée et de pouvoir se dérouler soit dans un milieu organique ou aqueux. Les réactions ayant recours à l'utilisation de composés azotures aliphatiques nécessite d'être chauffées à 60°C tandis que des composés azotures électro-déficients réagissent à température ambiante[210–212]. Depuis, cette ligation a été appliquée dans divers domaines, par exemple peptidique[213] et récemment dans la sphère des acides nucléiques où il a été également montré que cette chimie était orthogonale à la cycloaddition CuAAC[214].

c) Ligation faisant intervenir une fonction thiolate

La fonction thiolate est également un candidat potentiel pour être utilisé, en tant que nucléophile, dans des réactions chimiosélectives en milieu aqueux. Cette fonction peut réagir avec différents type de composés tels que les halogénoacétamides (**Tableau 6, entrée 1**) et les maléimides (**Tableau 6, entrée 2**). La réactivité relative des α-halogénoacétamides suit la tendance suivante I > Br > Cl > F et l'utilisation de ces derniers permet d'obtenir *in fine* un lien thioéther stable et peu encombrant[215]. On peut retrouver cette chimie dans le domaine des peptides et des protéines mais également au sein des oligonucléotides[216–218]. De plus, ce type de couplage est orthogonale avec diverses chimies comme la ligation oxime et la CuAAC[166,219]. Les composés de type maléimides peuvent être utilisés comme électrophiles et réagir avec une fonction thiolate par le biais d'une addition de Michael. Cependant, l'un des inconvénients est l'hydrolyse en milieu basique des dérivés maléimides[215]. Divers exemples de l'utilisation de cette chimie peuvent être retrouvés dans le domaine peptidique et oligonucléotidique[220].

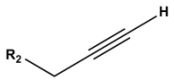
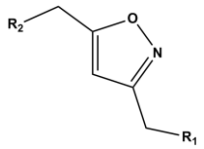
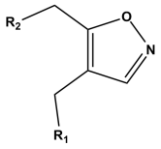
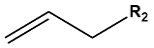
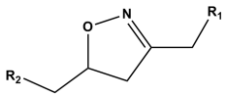
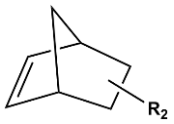
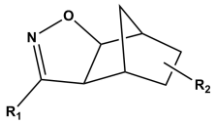
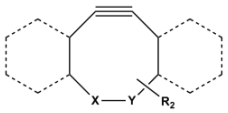
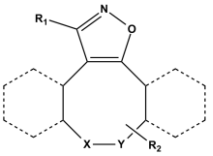
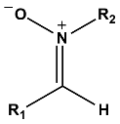
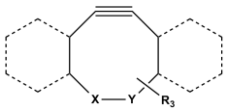
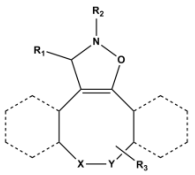
Tableau 6 : représentation des réactions possibles avec une fonction thiolate

Entrée	Fonction	Partenaire	Produit
1	R_1—S^-		
2			

d) Ligation faisant intervenir une fonction oxyde de nitrile ou nitrone

Les nitrones et les oxydes de nitrile sont également des 1,3 dipôles mais sont plus réactifs que les fonctions azotures dans des réactions de cycloaddition. Ceci se présente clairement comme un avantage mais peut également être problématique en termes de biocompatibilité. Effectivement, les nitrones ne peuvent être introduites dans la cellule due à cette instabilité[221] et les oxydes de nitrile doivent être formés *in situ* avant leur utilisation[222,223]. Cependant, la réactivité élevée des oxydes de nitrile est tout de même un avantage étant donné qu'ils peuvent réagir avec des alcynes[224] (**Tableau 7, entrées 1 et 2**), des alcènes[225] (**Tableau 7, entrée 3**) ou encore des norbornenes[226] (**Tableau 7, entrée 4**) et ce sans la présence de catalyseur. Il est important de spécifier que dans le cas d'une réaction entre un oxyde de nitrile et un dérivé alcyne terminal, la régiosélectivité peut être contrôlée par l'utilisation de catalyseurs. En effet, la présence de cuivre(I)[227] ou de carbene N-hétérocyclique (NHC)[228] favorise la formation du cycle 3,5-isoxazole disubstitué (**Tableau 7, entrée 1**) alors que l'utilisation d'un catalyseur de ruthénium(II)[229] contribue à la génération du cycle 3,4 disubstitué (**Tableau 7, entrée 2**). D'autres études ont montrées que les oxydes de nitriles peuvent également réagir avec des dérivées d'alcynes cycliques, réaction qu'on appelle aussi SPANOC pour « Strain-Promoted Alkyne Nitrile Oxide Cycloaddition » (**Tableau 7, entrée 5).** Comparé à la SPAAC, la SPANOC se montre plus réactive et peut être utilisée dans diverses applications dont dans la chimie des oligonucléotides[230]. Tout comme les oxydes de nitrile, les nitrones peuvent également réagir avec des dérivés de cyclooctynes, on parle alors de SPANC pour « Strain-Promoted Alkyne Nitrone Cycloaddition » (**Tableau 7, entrée 6**). Cette ligation montre également une cinétique élevée et peut être utilisée en synthèse organique mais aussi dans des bio-applications[221,231].

Tableau 7 : représentation des réactions possibles avec les fonctions oxydes de nitriles et nitrones
[a] L'utilisation d'un catalyseur de Cu(I) ou de carbene N-hétérocyclique (NHC) forme l'isomère 3,5-isoxazole
[b] L'utilisation d'un catalyseur de Ru(II) forme l'isomère 3,4-isoxazole

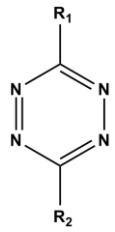
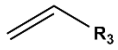
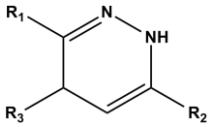
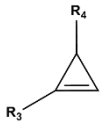
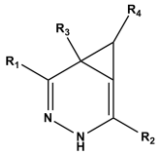
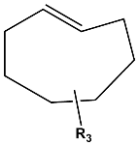
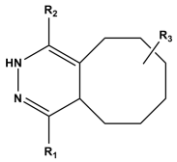
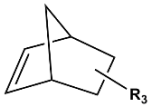
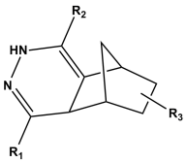
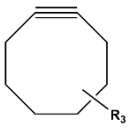
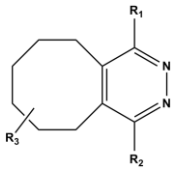
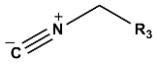
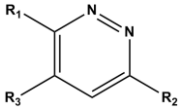
Entrée	Fonction	Partenaire	Produit
1[a]			
2[b]			
3			
4			
5			
6			

e) Ligation faisant intervenir un dérivé tétrazine

La cycloaddition inverse de Diels-Alder, également appelée IEDDA pour « Inverse Electron Demand Diels-Alder » est une alternative aux cycloadditions présentées précédemment. L'IEDDA peut s'effectuer entre un dérivé tétrazine et un composé alcène (**Tableau 8, entrées 1, 2, 3 et 4**), une cyclootyne (**Tableau 8, entrée 5**) ou un dérivé isonitrile (**Tableau 8, entrée 6**). La nature des groupements R présents sur les deux partenaires peuvent affecter la cinétique de la réaction. En effet pour une vitesse de réaction élevée, des groupements électro-attracteurs seront présents sur le partenaire diène (tétrazine) alors que pour le diènophile, on observera plutôt la présence de groupements électro-donneurs. A noter également que plus il y a de N-hétéroatomes au sein de l'hétérocycle, plus la cinétique est élevée[232]. Malgré la formation d'un lien encombrant et d'un manque de régiosélectivité, cette

ligation se montre réactive, sélective et peut être appliquée dans des applications biomédicales[233] mais également dans la chimie des oligonucléotides[232,234].

Tableau 8 : représentation des réactions possibles avec un dérivé tétrazine

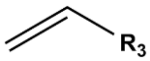
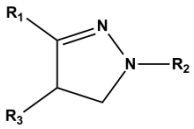
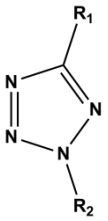
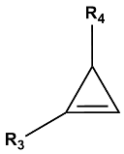
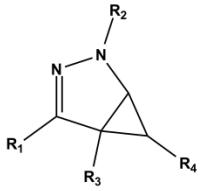
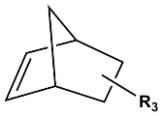
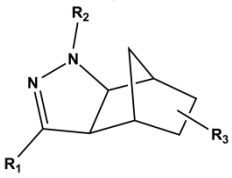
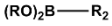
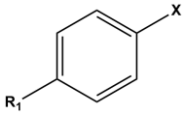
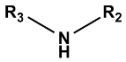
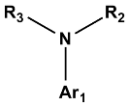
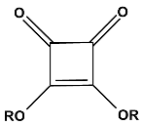
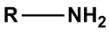
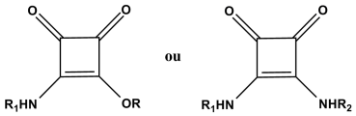
Entrée	Fonction	Partenaire	Produit
1			
2			
3			
4			
5			
6			

f) Autres exemples de ligations chimiosélectives

On retrouve dans la littérature d'autres ligations pouvant être utilisées à l'interface chimie-biologie. Par exemple, des composés de type tétrazoles qui une fois soumis à une irradiation UV forment des 1,3 dipôles nitrilimines et peuvent réagir, par cycloaddition, avec des dérivés d'acrylamides[235] (**Tableau 9, entrée 1**, de cyclopropènes[236] (**Tableau 9, entrée 2**) ou encore de norbornènes[237] (**Tableau 9, entrée 3**) pour former des composés de type pyrazolines. De plus, il a été montré que la réactivité des nitrilimines dépend des substituants présents sur le dérivé tétrazole[238]. Ces dernières années, il y a également eu la possibilité d'utiliser des réactions organométalliques comme le couplage de Suzuki-Miyaura[239] (**Tableau 9, entrée 4**), la réaction Sonogashira[239,240] (**Tableau 9, entrée 5**) ou encore l'amination de

Buchwald-Hartwig[241] (**Tableau 9, entrée 6**) et ce, dans des applications de conjugués d'ADN par exemple. Une alternative est l'usage de fluorosulfonate d'aryle. En 2014, Sharpless *et al.* ont montré une méthode simple et fiable pour la synthèse de ce type de composés[242] et ces derniers peuvent être utilisés dans des réactions de couplage de Suzuki-Miyaura, Sonogashira ou encore d'amination Buchwald-Hartwig[243] (**Tableau 9, entrées 4, 5 et 6**). Et enfin, on retrouve aussi la possibilité d'utiliser des dérivés de squarates pour réaliser des réactions chimiosélectives avec des composés possédant une fonction amine (**Tableau 9, entrée 7**). Cette chimie peut être appliquée dans divers domaines impliquant des biomolécules telles que des peptides, des protéines ou encore des acides nucléiques[244,245].

Tableau 9 : représentation de diverses ligations bio-orthogonales
[a] Réaction faisant intervenir une irradiation UV ; [b] Réaction faisant intervenir un catalyseur de Pd et/ou de Cu ; X = -I, -Br, -Cl ou -OSO_2F

Entrée	Fonction	Partenaire	Produit
1[a]			
2[a]			
3[a]			
4[b]			
5[b]			
6[b]			
7			

2. Présentation de la ligation KAT et son intérêt

Trois ligations chimiosélectives sont couramment utilisées au sein de notre laboratoire pour la fabrication de conjugués d'oligonucléotides fonctionnels, la ligation oxime, la cycloaddition CuAAC et le couplage thiol-chloroacétamide. Ces dernières permettent de synthétiser des édifices de type G-quadruplexe ou i-motif possédant une structuration particulière[162,170,171]. Dans le but d'accéder à de nouvelles structures plus complexes nécessitant une palette plus large de réactivité, deux possibilités sont envisageables. La première est de réutiliser l'une de ces trois réactions en cours de synthèse, comme cela a été effectué pour le

mime i-motif **F** présenté dans le chapitre précédent. La seconde alternative est de trouver une quatrième ligation, à la fois efficace et compatible avec les trois autres. La mise au point de cette deuxième option sera abordée au cours de ce chapitre.

Afin de s'assurer que la ligation choisie soit compatible, différents critères sont à considérer. En effet, il faut que :

- La nouvelle ligation s'effectue en milieu aqueux et que les nouvelles fonctions chimiques mises en jeux soient stables dans cet environnement.
- Les conditions réactionnelles doivent être 'douces' afin qu'aucune dégradation n'ait lieu au cours de la réaction.
- Les nouvelles fonctions chimiques ajoutées doivent être totalement spécifiques entre elles, c'est-à-dire que ces dernières ne doivent pas réagir avec les autres fonctions chimiques déjà présentes sur le composé.
- La réactivité de la nouvelle ligation doit être suffisamment élevée pour obtenir un taux de conversion satisfaisant et la stœchiométrie entre les deux partenaires doit être proche du 1/1.
- Si les nouvelles fonctions chimiques sont protégées au départ, les conditions de déprotection doivent être compatibles avec les fonctions chimiques déjà présentes sur le composé.
- Le nouveau lien effectué par la ligation ne doit pas être trop encombrant et la longueur de celui-ci doit être relativement similaire à ceux des autres ligations utilisées. Ceci est impératif dans le cadre de la conception de mime structurel.

Inspirée de la ligation KAHA pour « α-ketoacid-hydroxylamine amide », développée par Bode *et al.* en 2006[247], la réaction KAT pour « Potassium AcylTrifluoroborate » fait intervenir deux fonctions chimiques, une hydroxylamine et un acyltrifluoroborate et donne *in fine* un lien covalent amide entre les deux composés mis en jeux (**Figure 57a**). En 2012, Bode *et al.* ont publiés les premiers travaux concernant la ligation KAT sur de petites molécules organiques et où différents dérivés acyltrifluoroborates et hydroxylamines ont été testés dans diverses conditions réactionnelles[248]. Cette première étude s'est montrée intéressante étant donné qu'il a été observé une conversion totale (ou presque) avec une cinétique réactionnelle élevée. Cependant, les composés *O*-benzoyle-hydroxylamines utilisés lors de cette première étude ne se sont pas montrés suffisamment stables pour être incorporés dans des biomolécules telles que des peptides. Ainsi, en 2014 Bode *et al.* ont évalué, d'un point de vue stabilité et

réactivité, un panel de dérivés d'hydroxylamines[249]. Le dérivé *O*-diéthylcarbamate, **9**, (**Figure 57b**) s'est montré extrêmement intéressant car celui-ci est tout aussi réactif que son homologue *O*-benzoyle, il est stable vis-à-vis des amines primaires et secondaires dans les conditions de réaction. Et enfin, il est également compatible à différentes conditions de déprotection que l'on peut retrouver dans la chimie peptidique (exemples : élimination du groupement Boc en milieu acide (TFA) ou dans les conditions d'élimination du groupement Fmoc (piperidine/DMF)). Ce dérivé hydroxylamine a pu ainsi être introduit dans une séquence peptidique et différents couplages KAT ont pu être réalisés montrant sa sélectivité et sa réactivité au sein d'une biomolécule. Ainsi, la ligation KAT se présente comme ayant les particularités d'avoir une réactivité unique, d'être stable, de s'effectuer en milieu aqueux, d'être transposable à la chimie peptidique et possède une cinétique de réaction élevée. Ces caractéristiques ont retenu notre attention dans l'éventualité de l'utiliser comme quatrième ligation pour la formation de conjugués d'oligonucléotides.

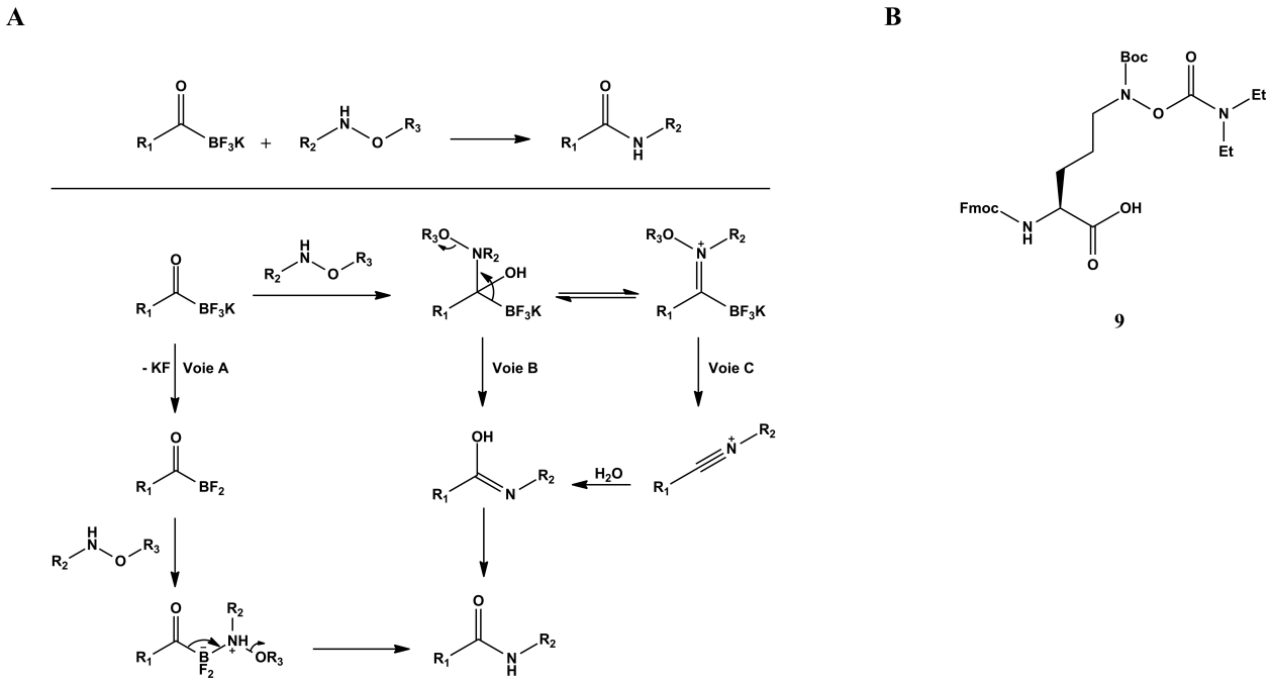
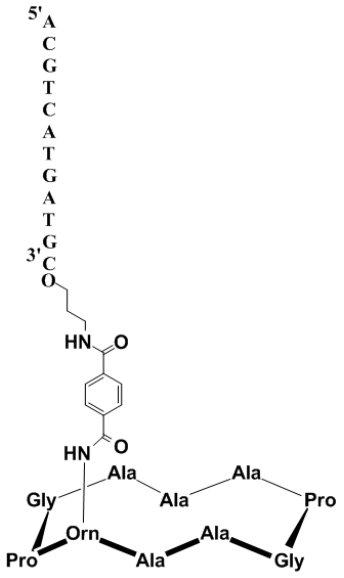


Figure 57 : (a) réaction et mécanisme proposé de la ligation KAT et (b) représentation de l'acide aminé KAT avec la fonction *O*-diéthylcarbamate-hydroxylamine **9**

3. Mise en œuvre et compatibilité de la ligation KAT

Afin de s'assurer que la ligation KAT soit compatible entre des biomolécules telles que des peptides et des oligonucléotides, la synthèse d'un premier conjugué, l'édifice **G**, a été envisagée (**Figure 58**). Nous avons choisi de placer la fonction *O*-diéthylcarbamoyl-hydroxylamine sur le substrat peptidique car on retrouve dans la littérature la synthèse d'un acide aminé modifié présentant cette fonction[249]. De l'autre côté, la fonction acyltrifluoroborate a été implantée sur le substrat oligonucléotidique. Cette fonctionnalisation s'effectue facilement en une seule étape à partir d'un oligonucléotide possédant une fonction amine. Ainsi, il a été envisagé la synthèse d'un cyclodécapeptide présentant la fonction chimique *O*-diéthylcarbamoyl-hydroxylamine **10** et un oligonucléotide possédant la fonction acyltrifluoroborate **11**. Les fonctions chimiques spécifiques à la ligation KAT peuvent être introduites sur les deux partenaires *via* des « building blocks » synthétisés préalablement.



G

Figure 58 : représentation de l'édifice **G**

a) Synthèse du châssis peptidique 10

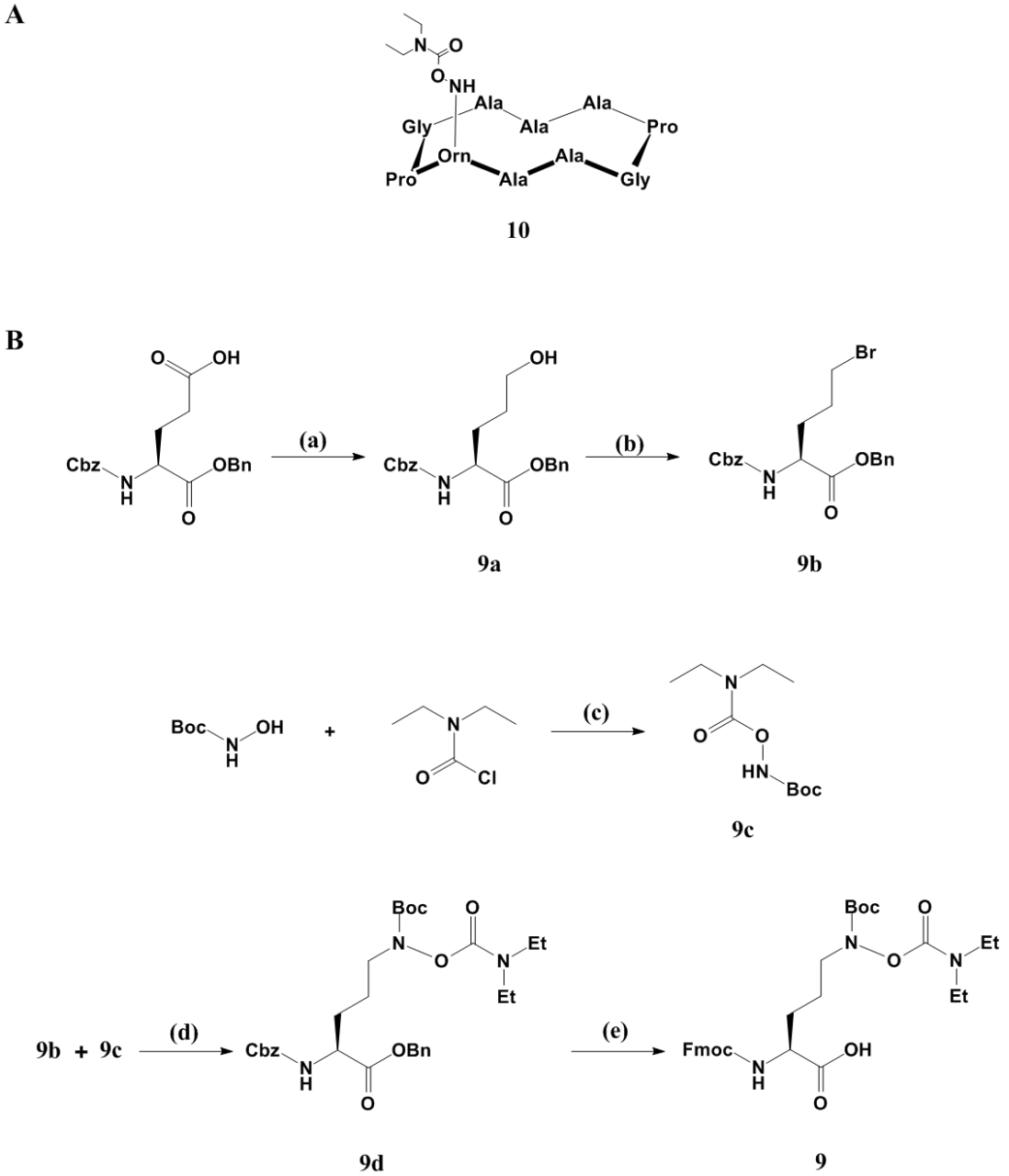


Figure 59 : (A) représentation du châssis peptidique **10**. (B) Synthèse de l'acide aminé KAT **9** (a) 1) $ClCO_2Et$, NMM, THF, -10°C, 2) $NaBH_4$, MeOH, 0°C à t.a. ; (b) PPh_3, CBr_4, DCM, 0°C, 1h, 83% ; (c) NEt_3, DMAP (10%mol), DCM, reflux, 9h, 97% ; (d) K_2CO_3, DMF, 3h, t.a., 82% ; (e) 1) Pd/C (10%w/w), H_2, MeOH, 2h, t.a., 2) Fmoc-OSu, iPr_2NEt, MeCN, H_2O, 2h, t.a., 44%

Le châssis peptidique **10** consiste en un cyclodécapeptide possédant un seul acide aminé modifié (**Figure 59a**). Ce dernier est un dérivé de l'ornithine où l'on retrouve à l'extrémité de la chaîne latérale la fonction *O*-diéthylcarbamoyl-hydroxylamine. La synthèse de cet acide aminé **9** s'effectue en plusieurs étapes et a été obtenu selon les procédures publiées et décrites par Noda *et al.*[249] (**Figure 59b**).

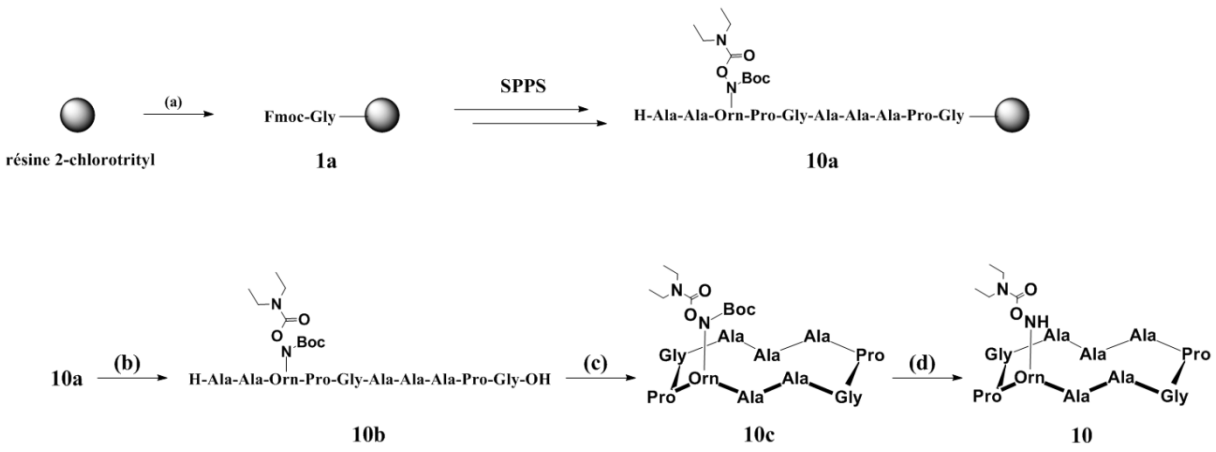


Schéma 14 : synthèse du châssis peptidique **10**
(a) Fmoc-Gly-OH (3eq), DIEA, DCM, 1h, t.a. ; (b) DCM/TFE/AcOH (70/20/10, v/v/v), t.a. ; (c) PyBOP (1,2eq), DIEA, DMF, 2h, t.a., 16% ; (d) TFA/H_2O/TIS (80/16/4, v/v/v), 1h, t.a.

Le châssis peptidique **10** a été synthétisé sur support solide en utilisant la stratégie au Fmoc/tBu à l'aide d'une résine 2-chlorotrityl (**Schéma 14**). Le cyclodécapeptide **10c** a été obtenu après purification par RP-HPLC avec un rendement de 16%. Celui-ci a été caractérisé par UPLC et spectrométrie de masse (**Figure 60**).

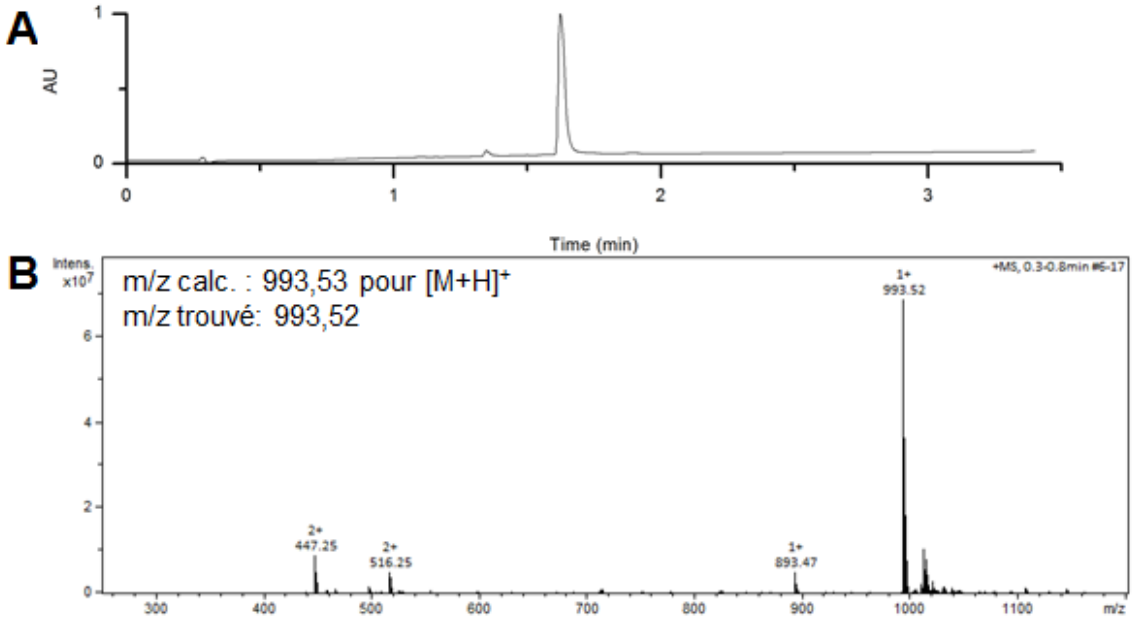


Figure 60 : (a) chromatogramme UPLC et (b) analyse ESI-MS du châssis peptidique **10c**

Et enfin, la dernière étape est l'élimination du groupement protecteur Boc présent sur l'unique acide aminé modifié, ceci est réalisé dans un mélange TFA/H_2O/TIS (80/16/4, v/v/v) pour donner le peptide **10**. Ce dernier a été caractérisé par UPLC-MS ainsi que spectrométrie de masse (**Figure 61**).

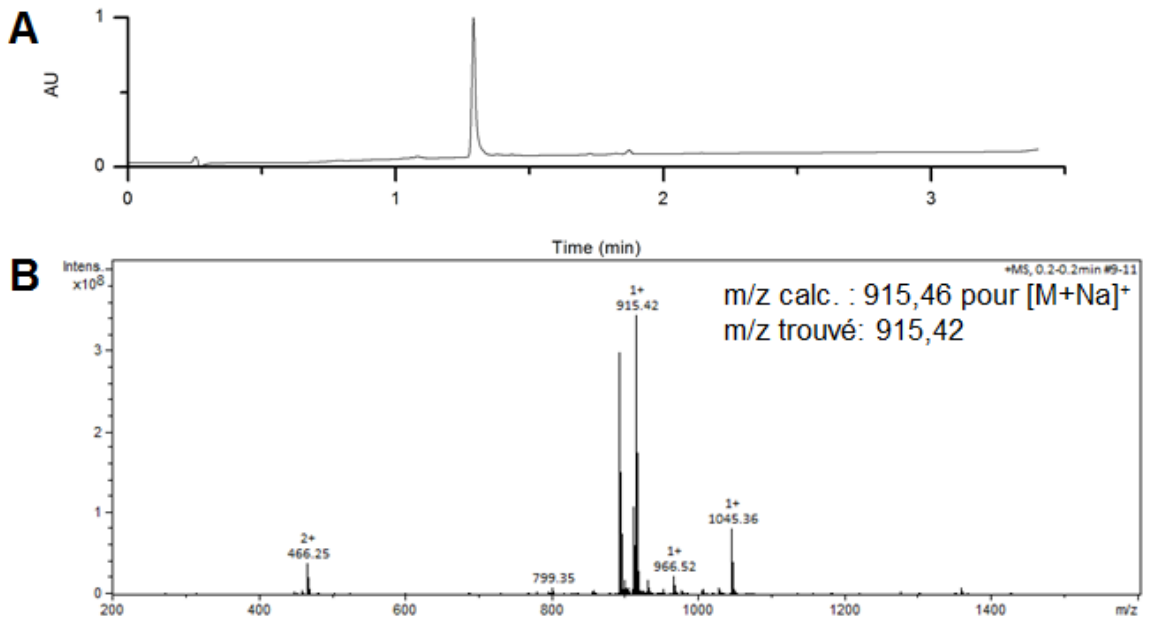


Figure 61 : (a) chromatogramme UPLC et (b) analyse ESI-MS du châssis peptidique **10**

b) Synthèse du brin d'ADN possédant la fonction acyltrifluoroborate 11

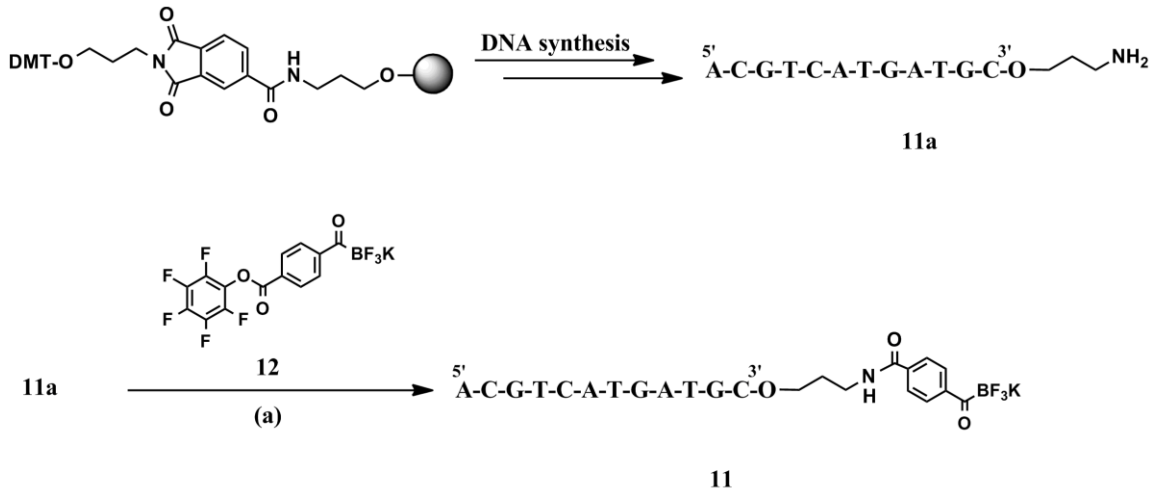


Schéma 15 : synthèse de l'oligonucléotide **11**
(a) Potassium 4-(perfluorophénylcarboxy)benzoyltrifluoroborate **12** (2eq), H_2O/DMF (1/1, v/v), DIEA, 2h, t.a.

La séquence d'ADN **11** (**Schéma 15**) ayant la fonction acyltrifluoroborate est obtenue à partir de l'oligonucléotide **11a**. Ce dernier est synthétisé au départ d'une résine « 3-PT-amino-modifier C3 CPG » permettant d'avoir une fonction amine en position 3'. Après synthèse sur support solide, déprotection et élimination de la résine, l'oligonucléotide **11a** est purifié par RP-HPLC avec un rendement de 44 %.

L'ajout de la partie acyltrifluoroborate sur l'oligonucléotide **11a** a été effectué dans un mélange H_2O/DMF (1/1, v/v) en milieu basique (pH 8-9), à température ambiante et en présence de 2eq de 4-(perfluorophénylcarboxy)benzoyltrifluoroborate de potassium **12** dont la synthèse est présentée ci-après. Un suivi réactionnel par RP-HPLC est réalisé et en fin de réaction, le brut réactionnel est dessalé sur colonne d'exclusion stérique pour enfin être purifié par RP-HPLC avec un rendement de 56 %. L'oligonucléotide **11** a été caractérisé par RP-HPLC et spectrométrie de masse (**Figure 62**).

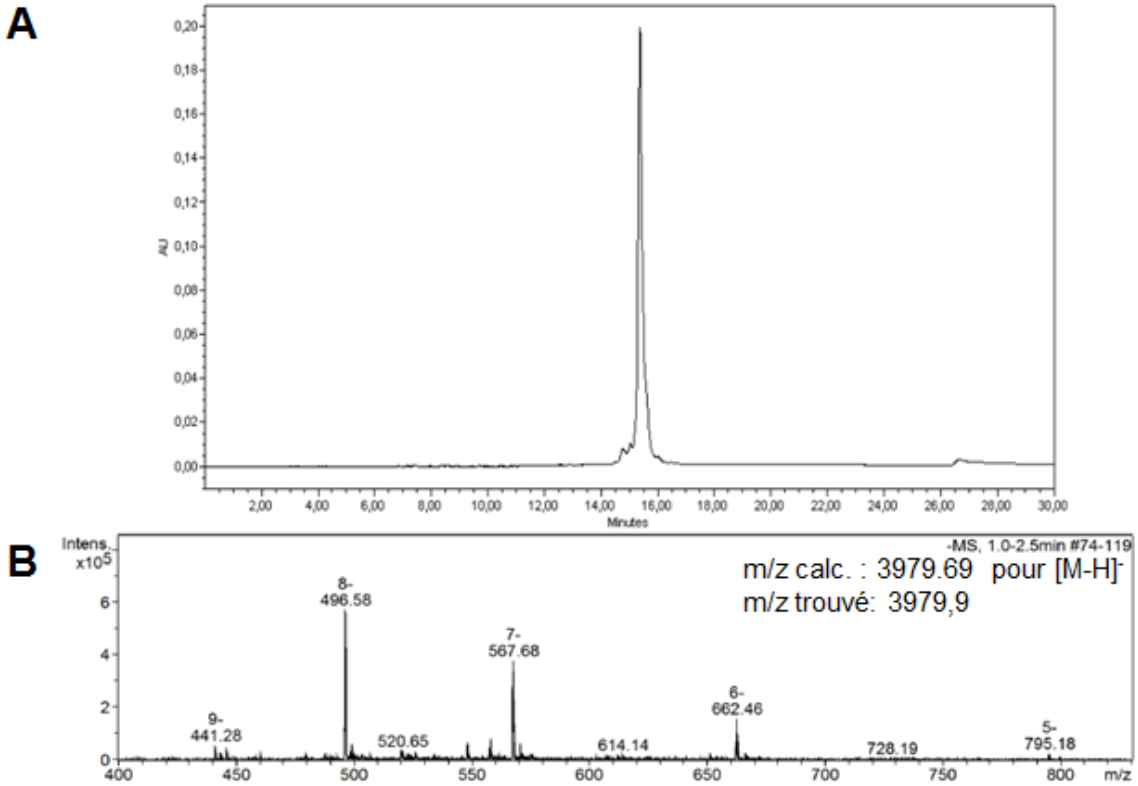


Figure 62 : (a) chromatogramme RP-HPLC et (b) analyse ESI-MS de l'oligonucléotide **11**

La synthèse du « building block » possédant la fonction acyltrifluoroborate (potassium 4-(perfluorophénylcarboxy)benzoyltrifluoroborate **12**) s'effectue en deux étapes et il a été obtenu selon les procédures publiées et décrites par Liu *et al.*[250] (**Schéma 16**).

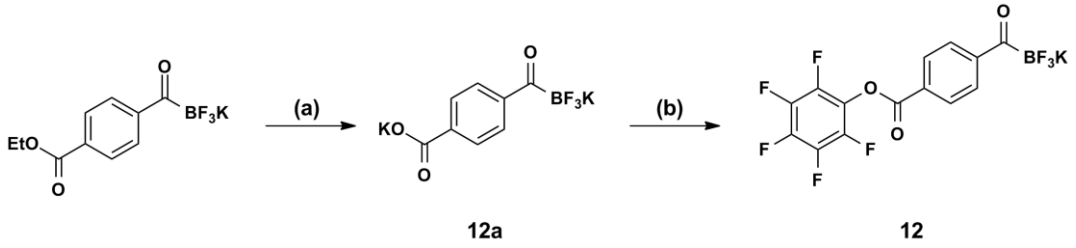


Schéma 16 : synthèse du composé **12**
(a) KOTMS (1eq), MeCN, 16h, t.a., 38% ; (b) $CF_3COOpfPh$ (1,2 eq), DIEA, DMF, 16h, t.a., 21%

c) Synthèse de l'édifice G

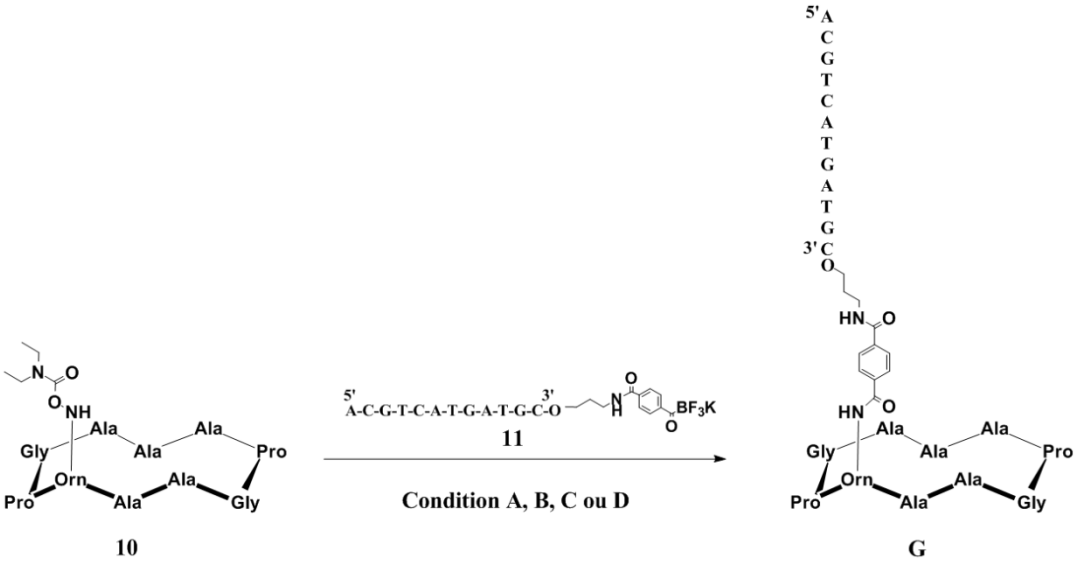


Schéma 17 : synthèse de l'édifice **G**

Pour rappel, l'objectif de la synthèse du composé **G** est de pouvoir confirmer la possibilité de transposer la ligation KAT au couplage entre un oligonucléotide et un peptide. La réaction KAT entre le cyclodécapeptide **10** et l'oligonucléotide **11** pour former l'édifice **G** (**Schéma 17**) a été étudiée dans quatre conditions réactionnelles (A, B, C et D) et suivie par RP-HPLC. Les conditions réactionnelles A et B font références aux conditions testées dans la littérature[248,249]. La condition C est le milieu réactionnel utilisé usuellement pour les ligations oximes mais dont la valeur de pH semble être adéquate pour réaliser la ligation KAT. Enfin, la condition D fait appel à un milieu réactionnel légèrement basique où la ligation KAT devrait s'effectuer mais avec une cinétique plus faible que pour les trois conditions précédentes[248,249]. Le **Tableau 10** reprend ces quatre conditions expérimentales testées avec les quantités engagées des deux partenaires ainsi que le temps nécessaire pour que soit atteint le taux de conversion maximum.

Tableau 10 : Résumé des quatre conditions réactionnelles utilisées pour la ligation KAT entre les composés **10** et **11**

	Conditions réactionnelles	10	11	Conversion maximum
A	t-BuOH/H_2O (1/1, v/v)	2eq	1eq	90 minutes
B	Acide oxalique (100mM)/oxalate de potassium (10mM)	2eq	1eq	90 minutes
C	Tampon NH_4OAc 0,4M pH 4,5	2eq	1eq	90 minutes
D	H_2O/DIEA (pH 8-9)	2eq	1eq	4h

La **Figure 63** illustre les bruts réactionnels dans les différentes conditions utilisées ainsi que l'analyse MALDI du brut réactionnel du composé **G** (condition A).

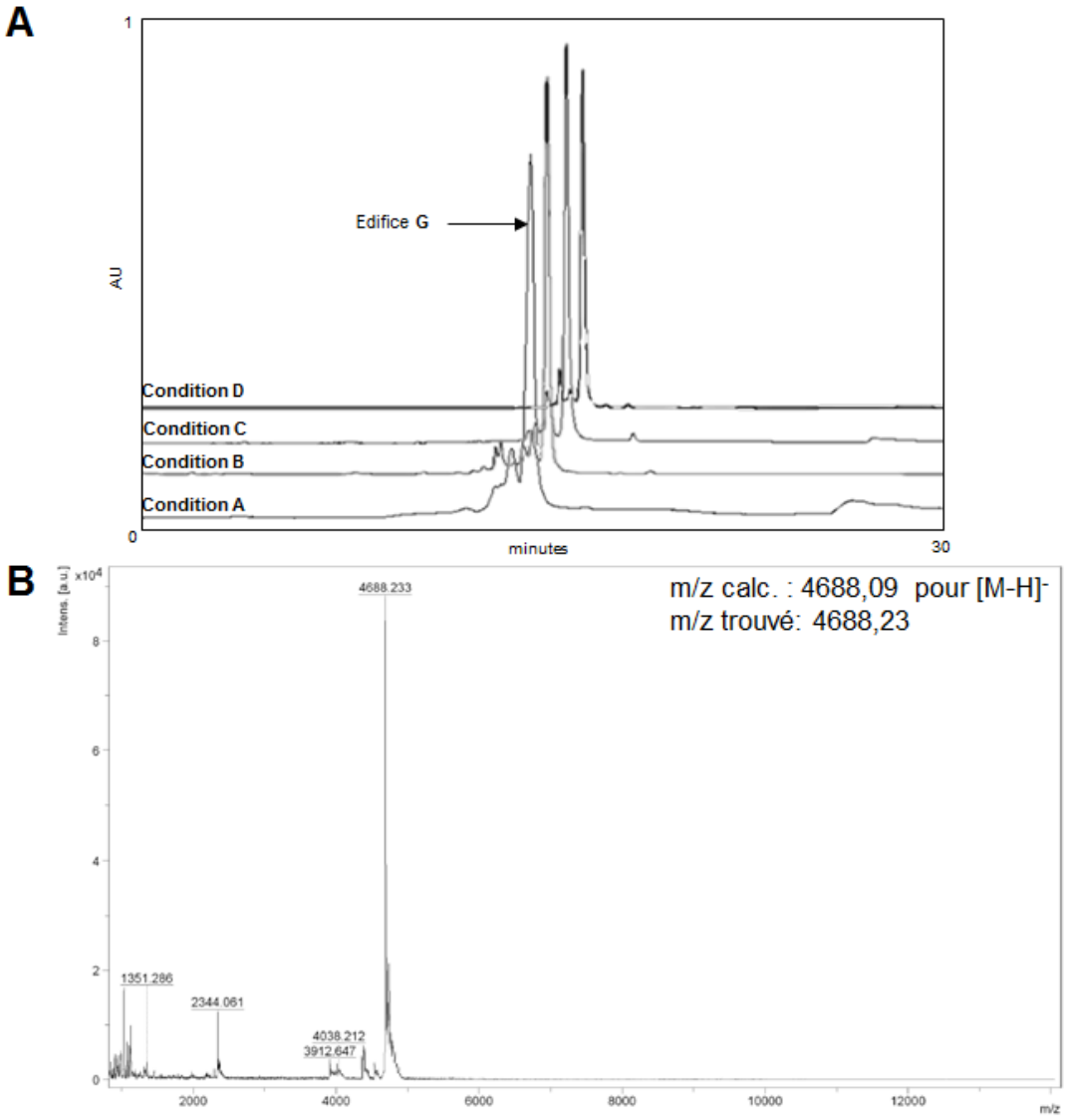


Figure 63 : (a) chromatogrammes RP-HPLC des bruts réactionnels (conditions A, B, C et D) et (b) analyse MALDI du composé **G** (brut ; condition A)

La caractérisation par spectrométrie de masse a permis de confirmer la formation de l'édifice **G** dans les quatre conditions réactionnelles utilisées. On peut également souligner que les conditions A, B et C présentent une cinétique plus importante que la condition D, ce qui est en accord avec la littérature[248,249]. A noter également qu'il a été observé qu'aucune réaction ne se déroulait entre le châssis peptidique **10** et un partenaire oligonucléotidique fonctionnalisé par une fonction aldéhyde. Ces premiers résultats étant satisfaisants dans le contexte de l'assemblage de conjugués peptides/oligonucléotides, nous n'avons pas poursuivi nos efforts par une comparaison plus détaillée de ces conditions qui aurait pu révéler des différences de réactivité dans des conditions stœchiométriques par exemple.

4. Evaluation de la compatibilité des quatre ligations

Précédemment, il a été montré que la ligation KAT était transposable à la conjugaison entre deux biomolécules (peptide et oligonucléotide) et que quatre conditions réactionnelles étaient envisageables pour réaliser cette ligation. Ainsi, la prochaine étape a été de s'assurer de la compatibilité de cette ligation KAT avec les autres ligations utilisées usuellement au sein de notre laboratoire. Pour cela, l'assemblage d'un nouveau conjugué a été envisagé (**Figure 64**). Ce dernier a pour objet d'établir la 'preuve de concept' et consiste en un cyclodécapeptide possédant quatre oligonucléotides modèles greffés par les quatre ligations distinctes. Nous avons choisi d'étudier deux châssis peptidiques, l'un possédant l'acide aminé alanine, l'autre disposant de la lysine-biotine, permettant l'immobilisation du composé sur une surface pour des études d'interactions. Pour cela, la synthèse de deux cyclodécapeptides ainsi que quatre séquences d'ADN modifiées a été envisagée.

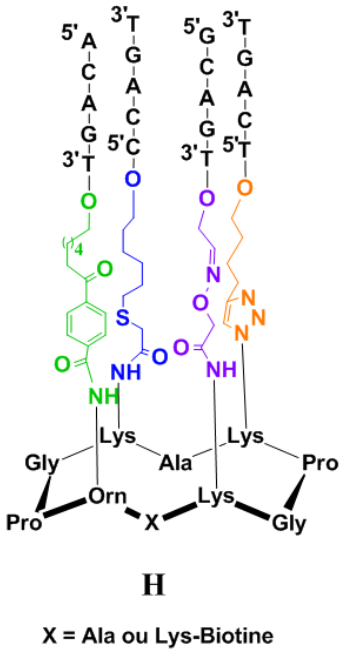


Figure 64 : représentation de l'édifice **H**

5. Synthèse des oligonucléotides

La formation de l'édifice **H** nécessite la synthèse de quatre séquences d'ADN modifiées en position 3' ou 5'. Parmi ces derniers, on retrouve deux oligonucléotides modifiés en position 3' avec soit une fonction aldéhyde (**13**), soit acyltrifluoroborate (**14**) mais également deux séquences modifiées à l'extrémité 5' avec soit une fonction thiol (**15**), soit une fonction alcyne (**16**).

a) Synthèse de l'oligonucléotide 13

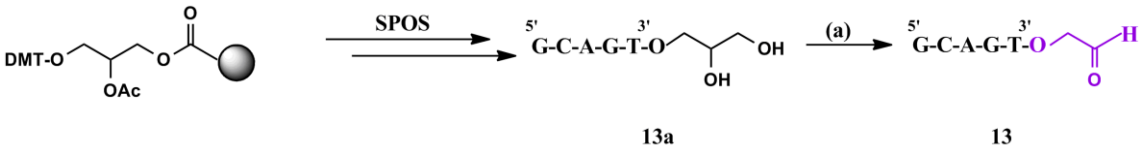


Schéma 18 : synthèse de l'oligonucléotide **13**
(a) $NaIO_4$ (30eq), H_2O, 30 minutes, t.a.

Le protocole pour réaliser la fonctionnalisation de la séquence d'ADN **13** est connu au sein de notre laboratoire[175]. Après synthèse à l'aide d'une résine 3' glyceryl CPG et clivage du support solide, l'oligonucléotide **13a** est purifié par RP-HPLC avec un rendement de 48%. La fonction aldéhyde est générée suite à la coupure oxydative de la fonction diol par le périodate

de sodium (**Schéma 18**). Cette réaction est considérée comme quantitative et l'oligonucléotide **13** est caractérisé par RP-HPLC et spectrométrie de masse (**Figure 65**).

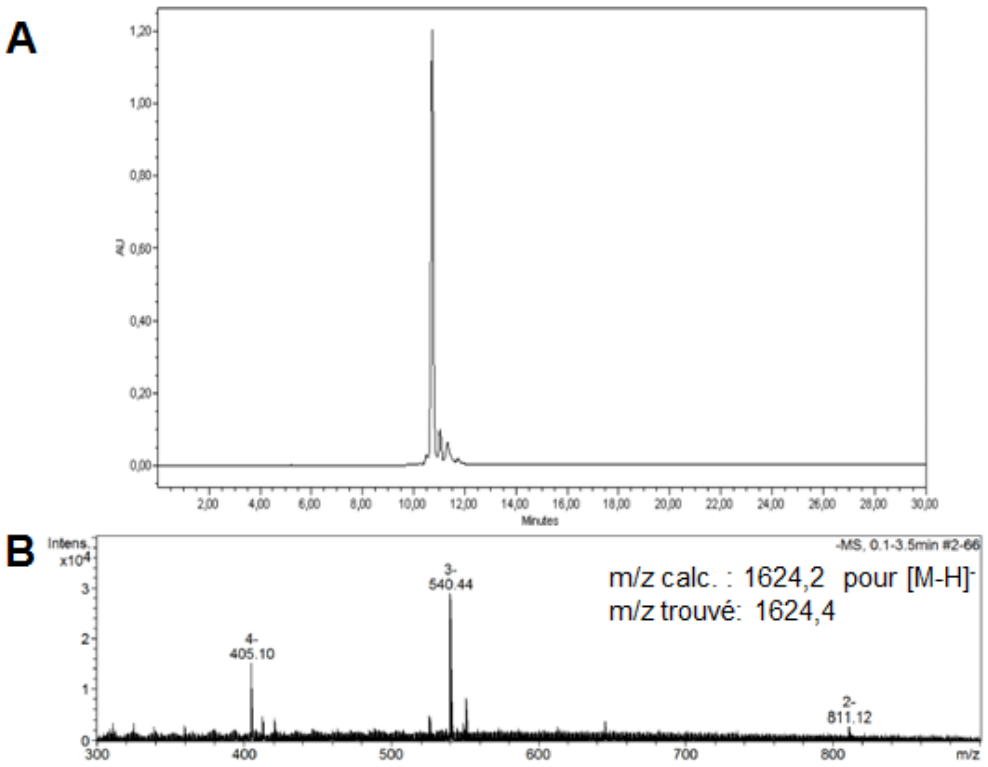


Figure 65 : (a) chromatogramme RP-HPLC et (b) analyse ESI-MS de l'oligonucléotide **13**

b) Synthèse de l'oligonucléotide 14

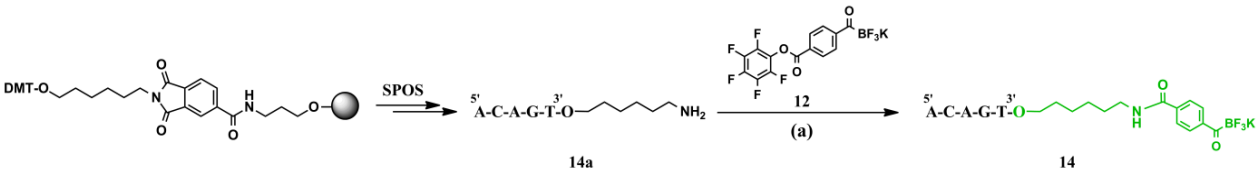


Schéma 19 : synthèse de l'oligonucléotide **14**
(a) Potassium 4-(perfluorophenylcarboxy)benzoyltrifluoroborate **12** (2eq), H_2O/DMF (1/1, v/v), DIEA, 2h, t.a.

L'oligonucléotide **14** (**Schéma 19**) est obtenu par la même stratégie de synthèse que pour l'oligonucléotide **11**. La séquence d'ADN **14** est obtenue, après purification par RP-HPLC, avec un rendement de 55%.

c) Synthèse de l'oligonucléotide 15

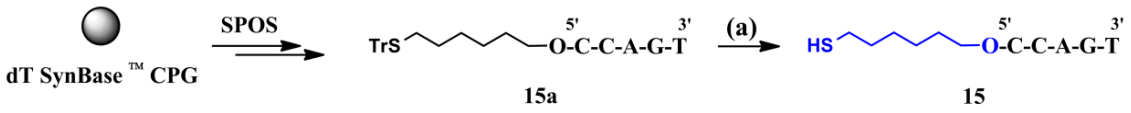


Schéma 20 : synthèse de l'oligonucléotide **15**
(a) 1) $AgNO_3$ 1M, TEAA 0,1M pH 6,5, 45 minutes, t.a. 2) DTT 1M, 15 minutes, t.a.

Le brin d'ADN **15** (**Schéma 20**) est obtenu au départ d'une résine « dT SynBase™ CPG » et d'un phosphoramidite modifié, le « 5'-thiol-modifier C6 ». Après synthèse et détritylation, l'oligonucléotide **15** est obtenu avec un rendement de 43 %.

d) Synthèse de l'oligonucléotide 16

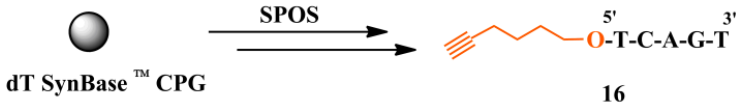


Schéma 21 : synthèse de l'oligonucléotide **16**

Le brin d'ADN **16** (**Schéma 21**) est synthétisé par SPOS à partir de la résine « dT SynBase™ CPG » et de l'incorporation de la fonction alcyne en position 5' par le phosphoramidite modifié, le « modifier 5' hexynyl ». Après clivage du support et purification par RP-HPLC, l'oligonucléotide **16** est obtenu avec un rendement de 41%.

6. Synthèse des châssis peptidiques 17 et 18

L'obtention des cyclodécapeptides **17** et **18** (**Figure 66**) implique l'utilisation d'acides aminés modifiés afin d'introduire les fonctions chimiques adéquates sur le peptide. On y retrouve comme fonctions, une oxyamine protégée, une *O*-diethylcarbamoyl-hydroxylamine, un chloroacétamide et enfin, une fonction azoture. Les deux premières sont obtenues à l'aide des acides aminés synthétisés précédemment (composés **5** et **9**). La fonction chloroacétamide est incorporée comme précédemment par le biais d'une lysine commerciale possédant un groupement protecteur DDE, orthogonal aux autres fonctions. Concernant la fonction azoture, la synthèse d'un acide aminé portant cette fonction est nécessaire et est décrite ci-après. Et enfin, pour le composé **18**, on retrouve une lysine possédant une fonction biotine, cet acide aminé a déjà été synthétisé précédemment (composé **4**). Ces acides aminés modifiés seront introduits lors de la synthèse peptidique et sont totalement compatibles avec les conditions de couplage et de déprotection de la SPPS.

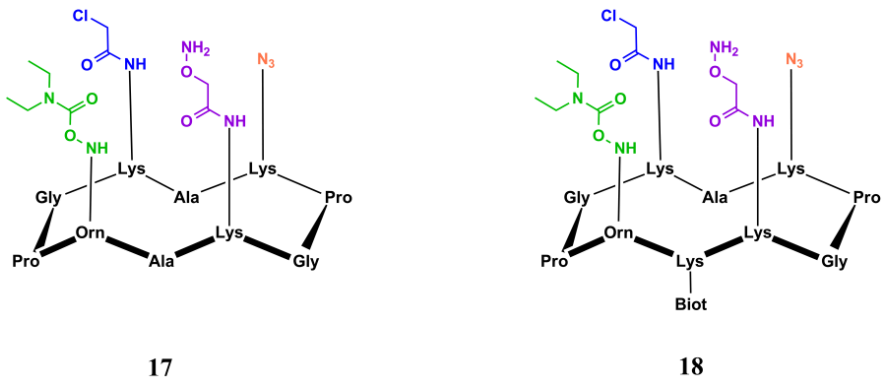


Figure 66 : représentation des cyclodécapeptides **17** et **18**

a) Synthèse de la Fmoc-azidolysine 19

La Fmoc-azidolysine **19** (ou Fmoc-Lys(N_3)-OH) est préparée à partir de la Fmoc-Lys-OH **2** par réaction de diazo-transfer sur la fonction amine présente à l'extrémité de la chaîne latérale. Cette réaction est réalisée en milieu basique, en présence de sulfate de cuivre et d'ISAHC pour « imidazole-1-sulfuryl azide hydrochloride» (**Figure 67**). L'acide aminé **19** a été obtenu selon les procédures publiées et décrites par Byrne *et al.* [251].

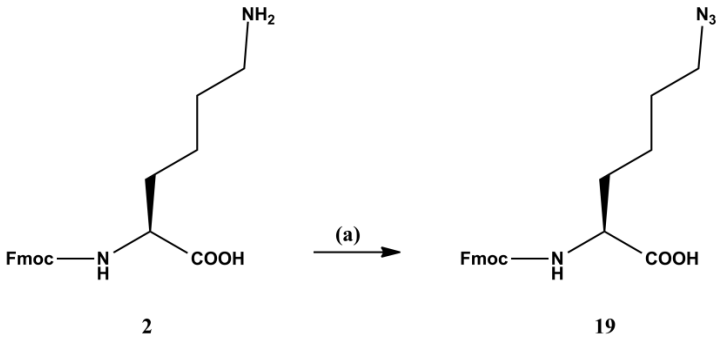


Figure 67 : synthèse de la Fmoc-azidolysine **19**
(a) $CuSO_4$ (0,04eq), $NaHCO_3$ (3,5eq), ISAHC (1,2eq), MeOH/H_2O (1/1, v/v), 16h, t.a., 98%

b) Synthèse des châssis peptidiques 17 et 18

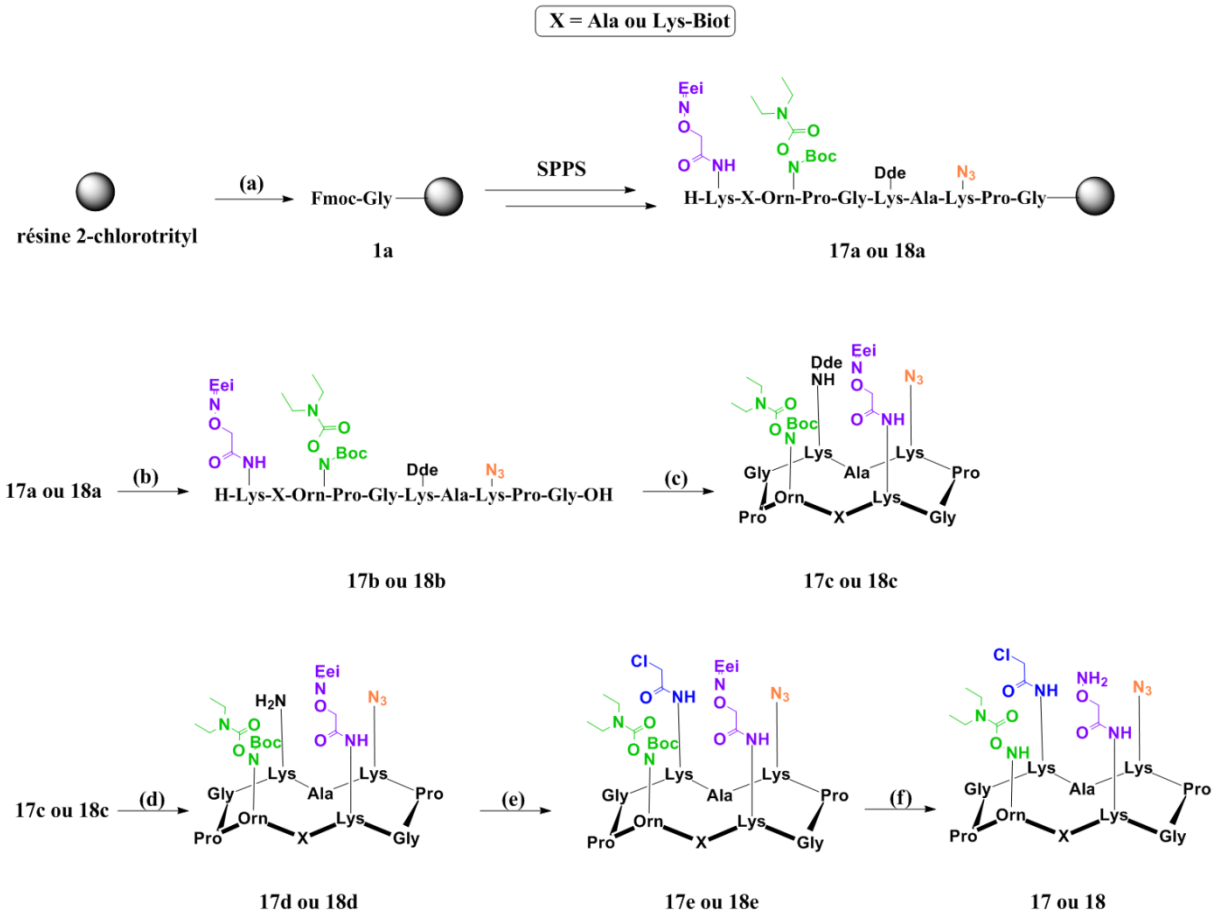


Schéma 22 : synthèse des châssis peptidiques **17** et **18**
(a) Fmoc-Gly-OH (3eq), DIEA, DCM, 1h, t.a. ; (b) DCM/TFE/AcOH (7/2/1, v/v/v), t.a. ; (c) PyBOP (1,2eq), DIEA, DMF, 2h, t.a., 53% pour **17c** et 49% pour **18c** ; (d) 2%vol d'hydrazine dans DMF, 20 minutes, t.a. ; (e) anhydride chloroacétique (1,2eq), DIEA, DMF, t.a., 69% pour **17e** et 61% pour **18e** ; (f) TFA/H_2O/TIS (80/16/4, v/v/v), 1h, t.a.

Les cyclodécapeptides **17** et **18** sont obtenus comme précédemment, c'est-à-dire par synthèse peptidique sur support solide à l'aide d'une résine 2-chlorotrityl (**Schéma 22**). Les cyclodécapeptides **17c** et **18c** ont été obtenus, après purification, avec un rendement de 53% et 49%, respectivement. Après élimination du groupement DDE et introduction de la fonction chloroacétamide, les peptides **17e** et **18e** sont obtenus, après purification, avec un rendement de 69% et 61%, respectivement. Ils sont caractérisés par UPLC-MS et spectrométrie de masse (**Figure 68** et **Figure 69**).

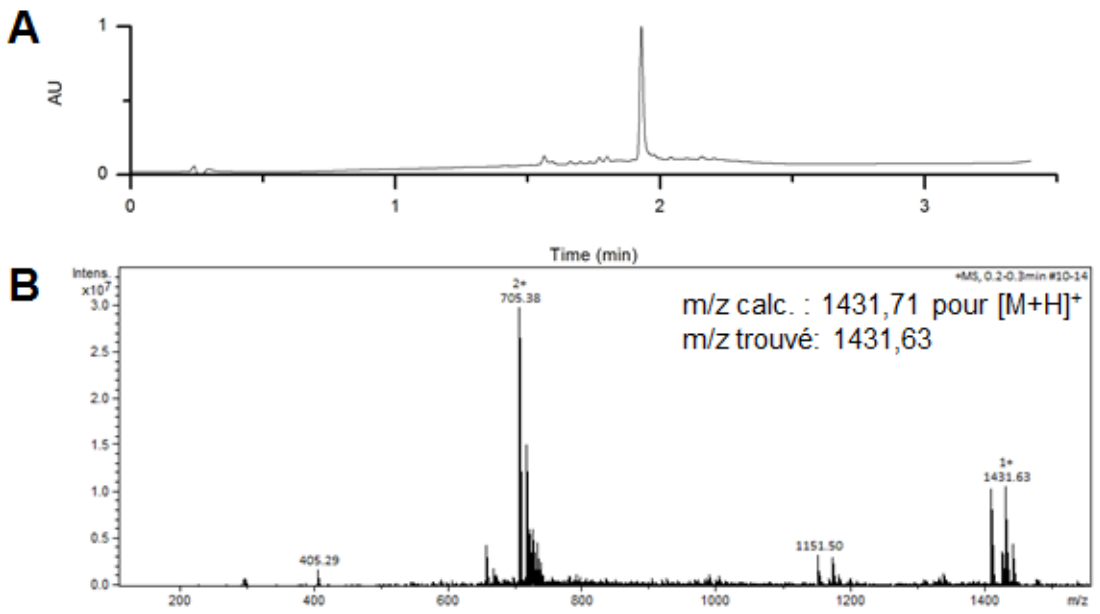


Figure 68 : (a) chromatogramme UPLC et (b) analyse ESI-MS du châssis peptidique **17e**

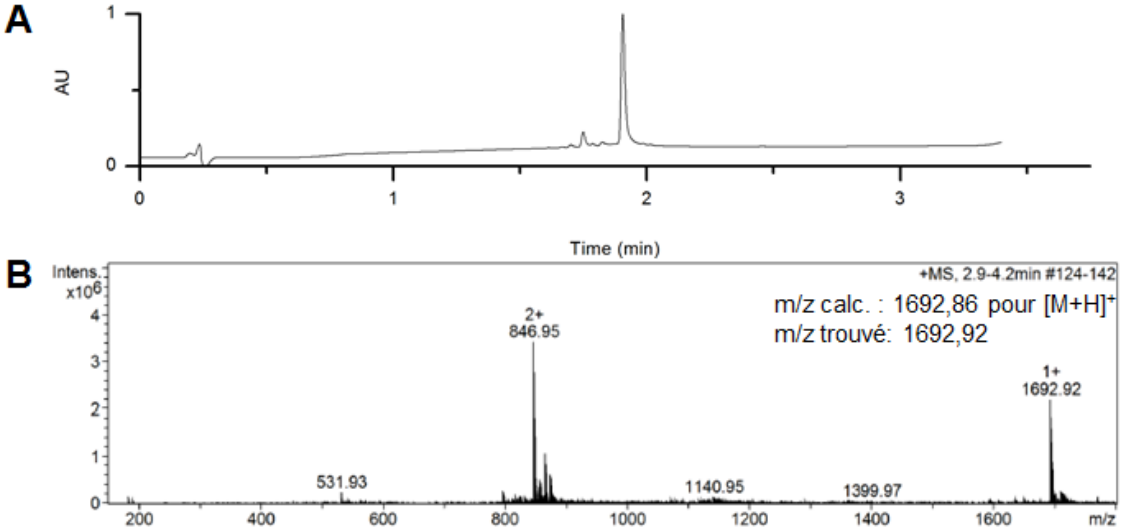


Figure 69 : (a) chromatogramme UPLC et (b) analyse ESI-MS du châssis peptidique **18e**

La dernière étape consiste à éliminer les groupements Eei et Boc présent sur les châssis **17e** et **18e** à température ambiante en présence d'une mélange TFA/H_2O/TIS (80/16/4, v/v/v). La **Figure 70** illustre le suivi de cette étape de déprotection et reprend le spectre ESI-MS du peptide **18**.

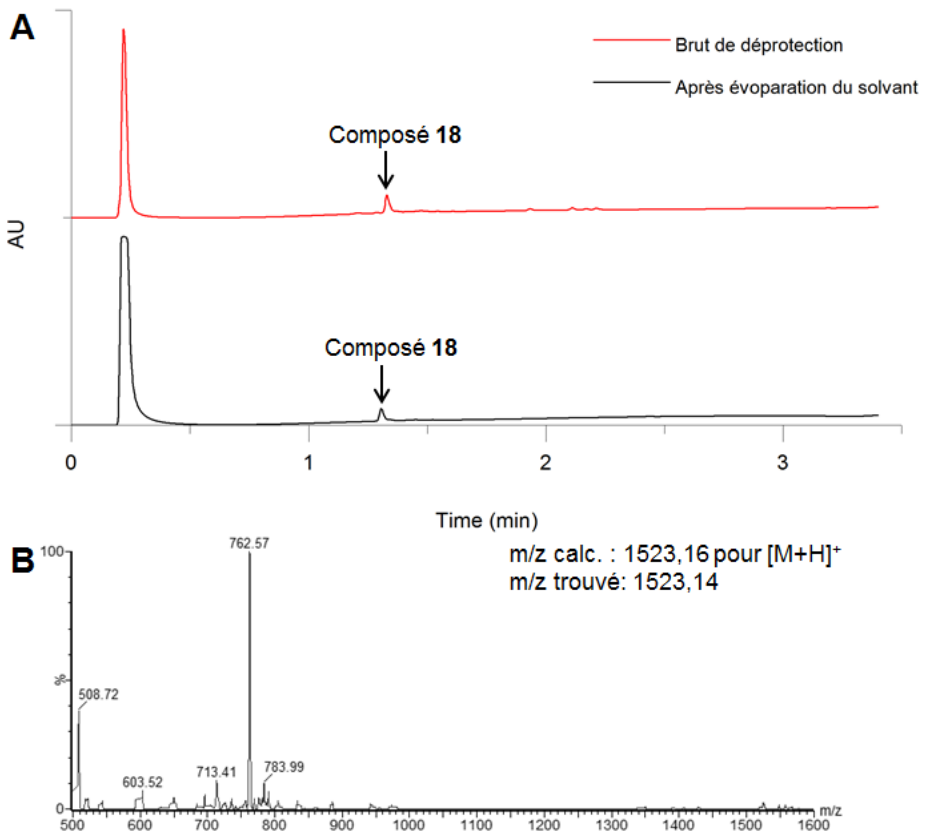


Figure 70 : (a) chromatogramme et (b) spectre UPLC du brut réactionnel de l'étape de déprotection pour le peptide **18e**

Une fois le solvant éliminé, le brut réactionnel a été solubilisé dans un minimum de DCM et est ensuite précipité dans une solution d'éther. Cependant, lors de cette étape de précipitation, une dégradation du peptide totalement déprotégé est observée avec la formation d'un produit secondaire qui se traduit par une perte de masse de 134 g/mol par rapport au peptide attendu. Pour expliquer cette modification, nous avons émis l'hypothèse d'une réaction entre les fonctions déprotégées oxyamine et *O*-diethylcarbamoyl-hydroxylamine. La perte de la fonction *O*-diethylcarbamoyl-hydroxylamine représente une masse de 116 g/mol, il y aurait donc la perte de celle-ci mais également une modification sur le support peptidique. Afin de prévenir cette dégradation, il a été également testé de solubiliser le brut de déprotection du peptide **18** dans un mélange H_2O/CH_3CN (1/1, v/v) et de le lyophiliser. Malheureusement, le produit dégradé est de nouveau observé après lyophilisation. La **Figure 71** reprend les profils UV obtenus après précipitation dans l'éther et après lyophilisation ainsi que le spectre ESI-MS du peptide dégradé. A noter que ceci a également été observé pour le peptide **17**.

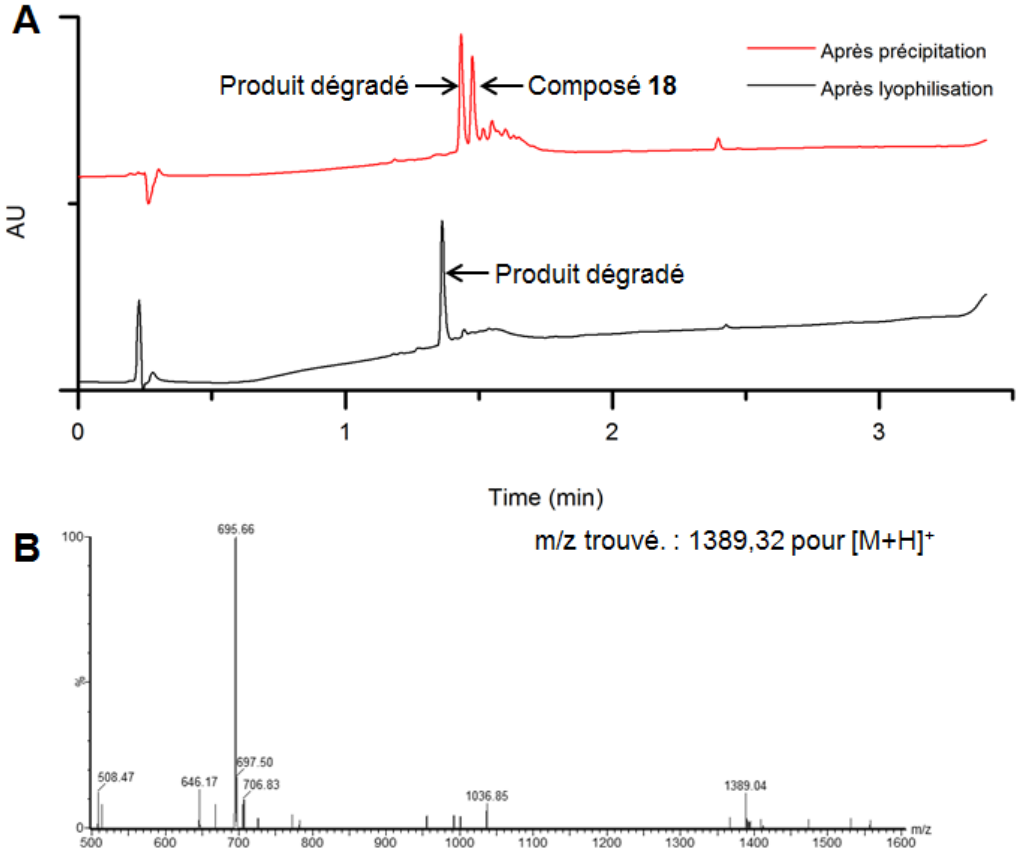


Figure 71 : (a) chromatogrammes après précipitation et lyophilisation du brut réactionnel de l'étape de déprotection du peptide **18e**. (b) Spectre UPLC-MS du produit secondaire formé.

i. <u>Test de ligation oxime entre le brut de déprotection du peptide **18** et l'oligonucléotide **13**</u>

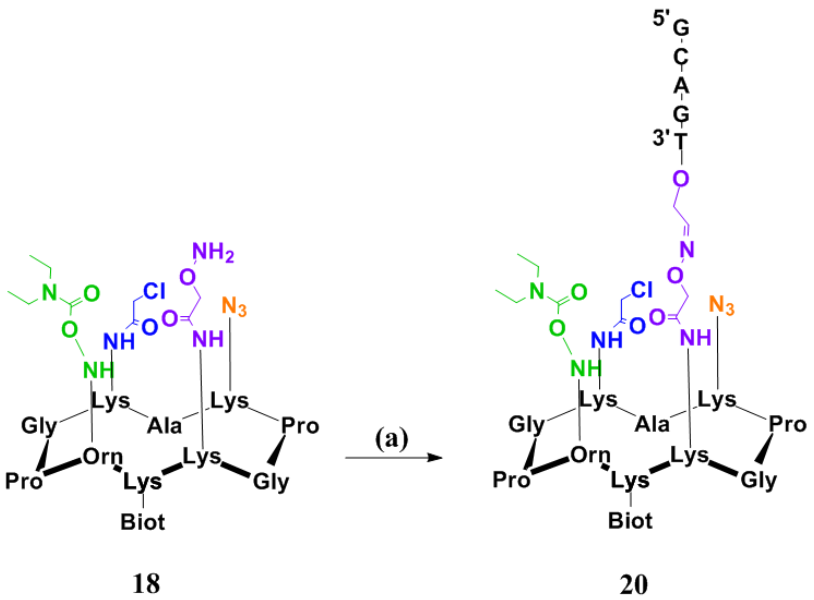


Schéma 23 : synthèse du composé **20**
(a) oligonucléotide **13** (1eq), tampon NH_4OAc 0,4M pH 4,5, t.a.

Etant donné qu'on ne peut obtenir de manière isolée le peptide totalement déprotégé, il a été envisagé de réaliser la ligation oxime entre le brut de déprotection du peptide **18e** et l'oligonucléotide aldéhyde **13** pour former le composé **20** (**Schéma 23**). Pour cela, l'élimination des groupements Eei et Boc s'effectue dans un faible volume (50µL) d'un mélange TFA/H_2O/TIS (80/16/4, v/v/v). Une fois la déprotection finie (**Figure 72**), ce brut est ajouté dans un milieu contenant l'oligonucléotide **13** solubilisé dans un tampon NH_4OAc 0,4M pH 4,5 (à noter que le pH se trouve entre 3 et 4 après cet ajout ce qui ne doit pas compromettre la ligation entre les deux partenaires). Malheureusement, aucune réaction n'a lieu dans ces conditions. En effet, sur la **Figure 73**, on n'observe pas l'apparition du composé **20**, ce qui est confirmé par l'analyse ESI-MS du brut réactionnel (**Figure 73c**) où l'on ne retrouve que la masse de l'oligonucléotide **13**. Ceci démontre que la fonction oxyamine n'est plus présente sur le châssis peptidique dégradé en raison de la réaction secondaire et que la réaction d'oximation ne peut alors s'effectuer.

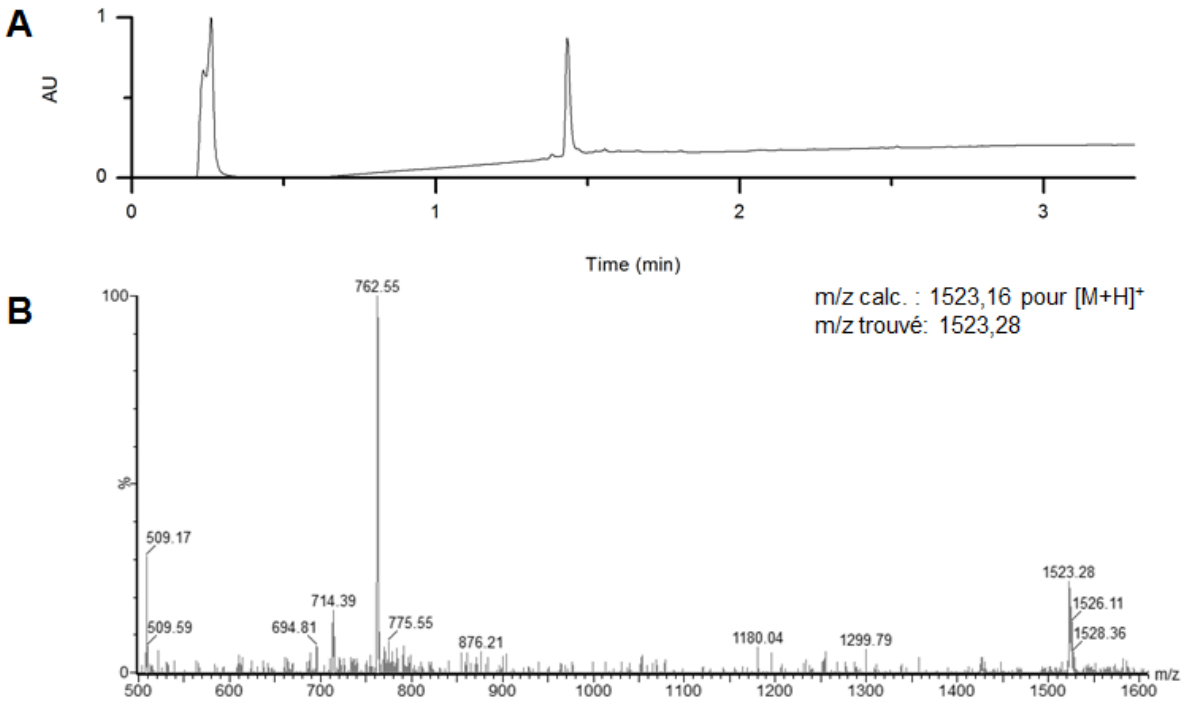


Figure 72 : (a) chromatogramme et (b) spectre UPLC-MS du brut réactionnel de l'étape de déprotection pour le peptide **18e**

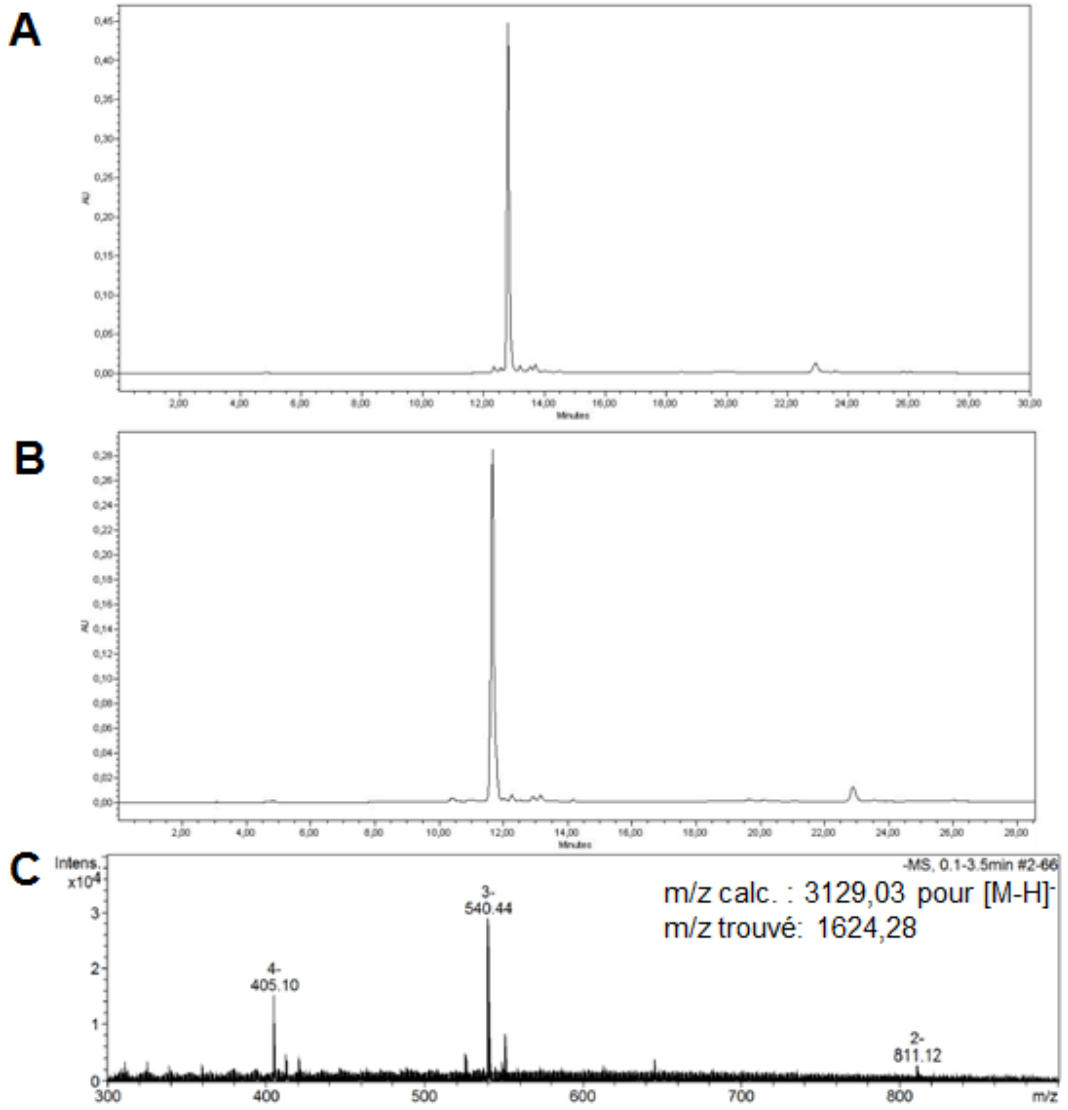


Figure 73 : chromatogrammes RP-HPLC de l'oligonucléotide **13** (a) et du brut réactionnel (b). Spectre ESI-MS du brut réactionnel (c).

Afin de confirmer que la dégradation s'effectue bien entre les fonctions oxyamine et *O*-diéthylcarbamoyl-hydroxylamine, il a été décidé de synthétiser un court peptide reprenant ces deux fonctions chimiques (**Schéma 24**). La synthèse du peptide **21** s'effectue comme précédemment, c'est-à-dire à partir d'une résine 2-chlorotrityl où l'on ajoute successivement les acides aminés par SPPS. Le peptide **21b**, obtenu après clivage de la résine, est caractérisé par UPLC-MS (**Figure 74**) et est obtenu avec un rendement brut de 80%.

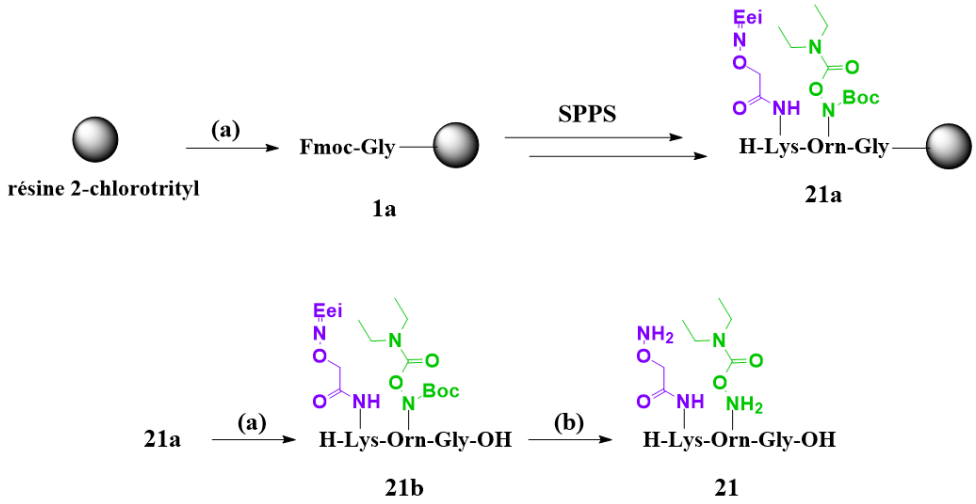


Schéma 24 : synthèse du peptide **21**
(a) DCM/TFE/AcOH (70/20/10, v/v/v), t.a., 80% ; (b) TFA/H_2O/TIS (80/16/4, v/v/v), 1h, t.a.

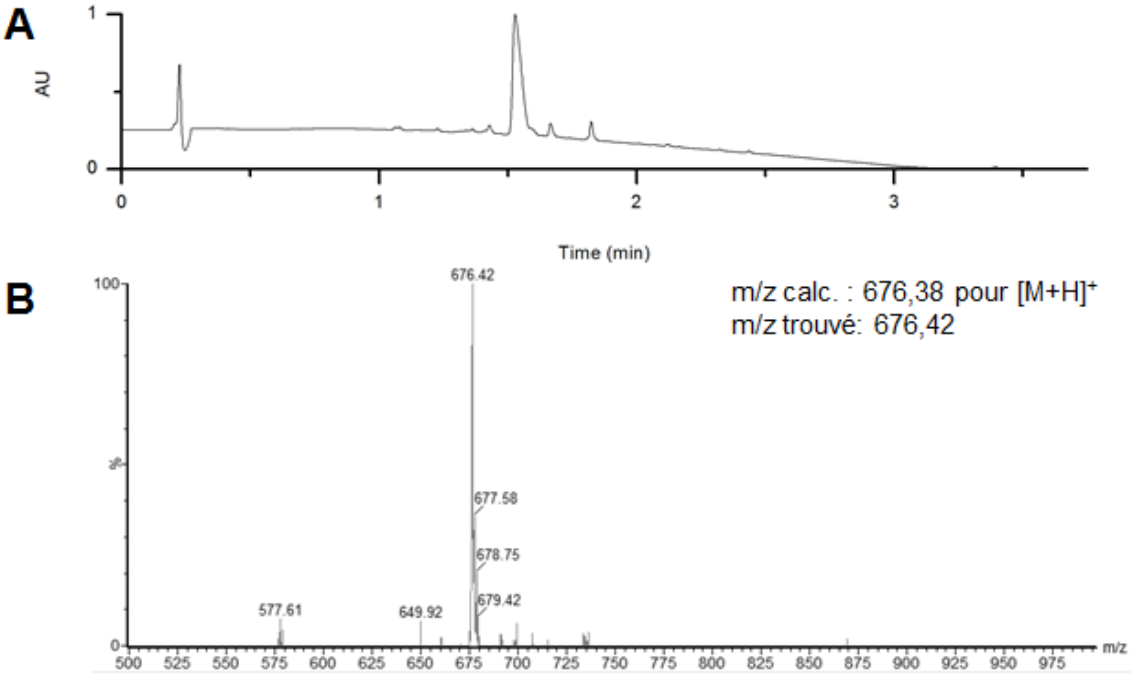


Figure 74 : (a) chromatogramme et (b) spectre UPLC-MS du composé **21b**

Ensuite le peptide **21b** est déprotégé à l'aide d'une solution TFA/H_2O/TIS (80/16/4, v/v/v) pour obtenir le composé **21** (**Figure 75**). Il s'en suit l'étape de précipitation dans l'éther où l'on observe de nouveau la dégradation du produit avec une perte de masse de 134 g/mol (**Figure 76**), ceci confirmant une réaction entre ces deux partenaires. Par manque de temps, une étude plus approfondie sur la nature de cette réaction de dégradation n'a pu être effectuée et nous nous sommes consacré à prévenir son apparition.

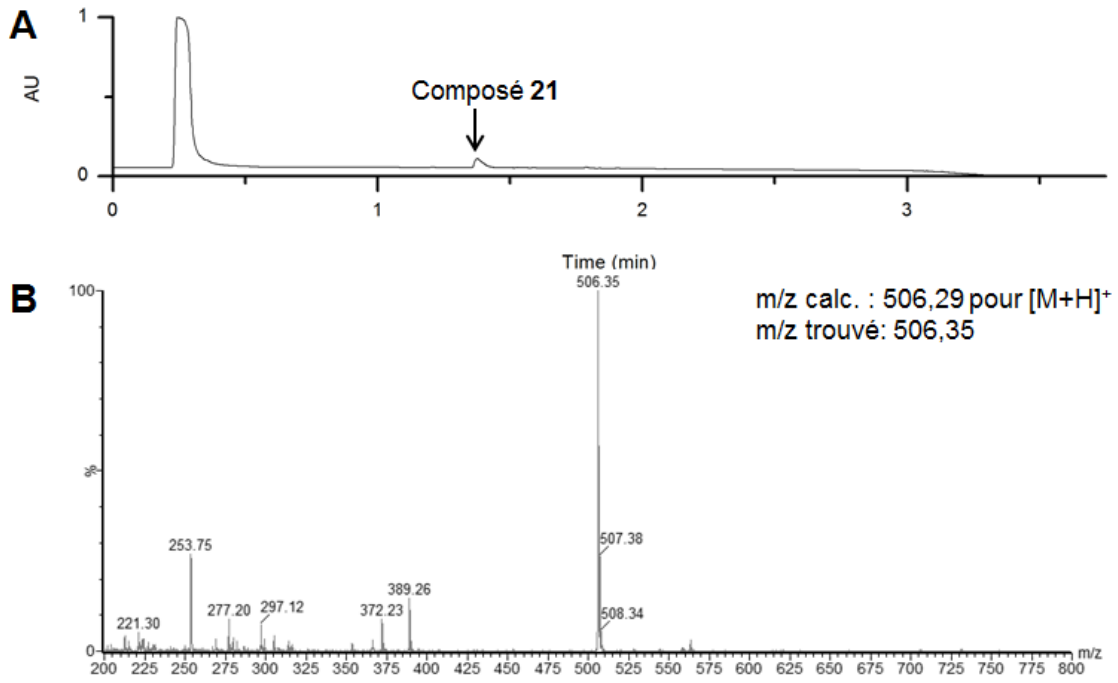


Figure 75 : (a) chromatogramme et (b) spectre UPLC-MS du brut de déprotection du peptide **21b**

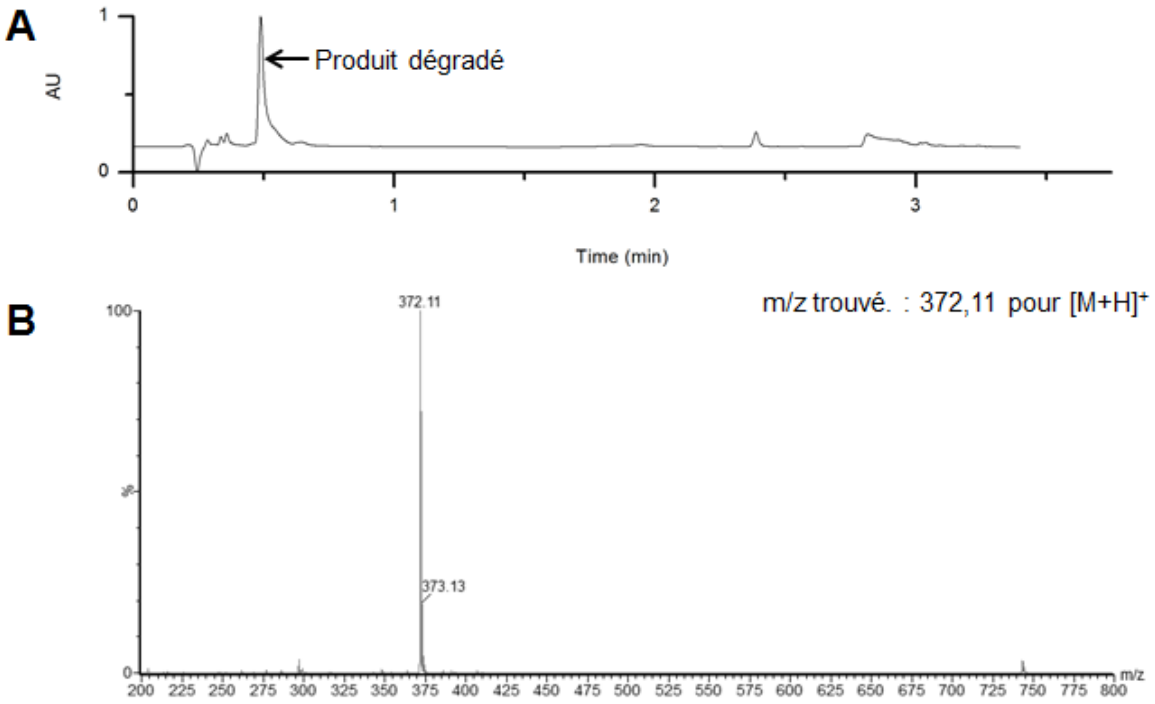


Figure 76 : (a) chromatogramme et (b) spectre UPLC-MS du brut de déprotection du peptide **21b** après précipitation dans l'éther

7. Mise au point de la compatibilité entre la chimie KAT et la ligation oxime

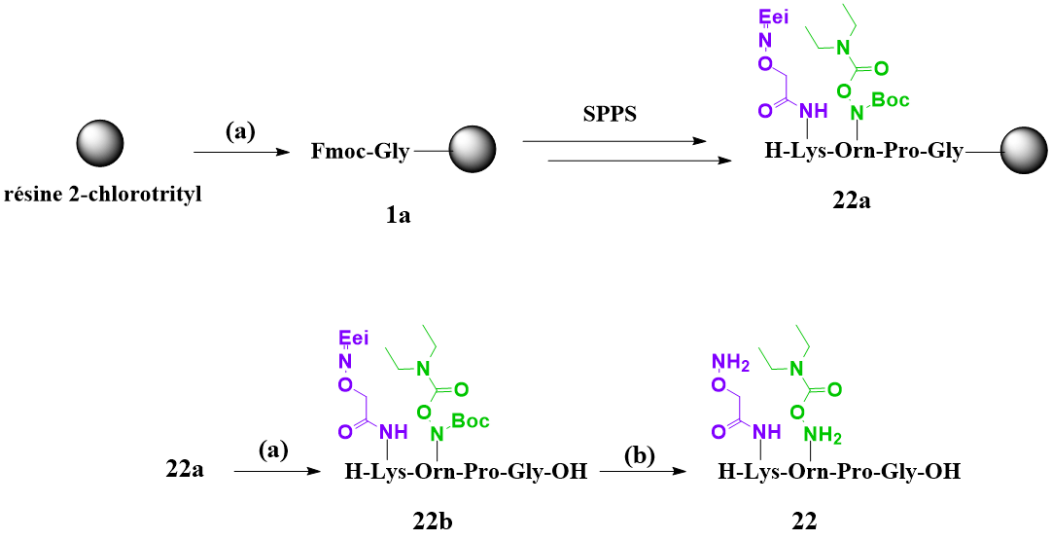


Schéma 25 : synthèse du peptide **22**
(a) DCM/TFE/AcOH (70/20/10, v/v/v), t.a., 76% ; (b) TFA/H_2O/TIS (80/16/4, v/v/v), 1h, t.a.

Précédemment il a été observé que les fonctions oxyamine et *O*-diéthylcarbamoyl-hydroxylamine étaient toujours présentes dans leur milieu de déprotection et ce n'est que lors de l'étape de précipitation que l'on observe une dégradation du peptide. Il est donc nécessaire de trouver des conditions de post-déprotection afin de pouvoir isoler et conserver ces deux fonctions chimiques. Ainsi, un court peptide, **22**, possédant ces deux groupements a été synthétisé et la formation de celui-ci s'effectue de la même manière que précédemment (**Schéma 25**). Après synthèse et clivage, le peptide **22b** est caractérisé par UPLC-MS (**Figure 77**) et est obtenu avec un rendement brut de 76%.

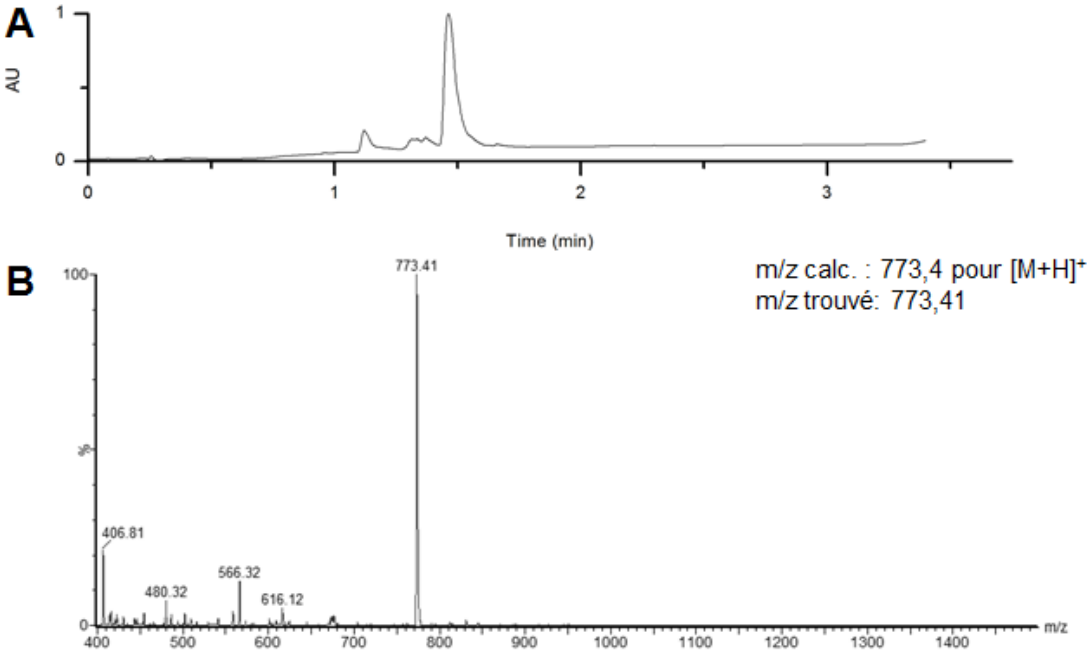


Figure 77 : (a) chromatogramme et (b) spectre UPLC-MS du peptide **22b** (brut)

Les groupements Eei et Boc présents sur le peptide **22b** sont déprotégés à l'aide d'un mélange TFA/H_2O/TIS (80/16/4, v/v/v) pour former le composé **22**. Comme observé auparavant, aucune dégradation ne s'effectue dans ce milieu. Ensuite, il a été testé de neutraliser le mélange réactionnel, essentiellement composé de TFA, par une base forte telle que NaOH (solution 4M). Nous avons alors constaté qu'un passage rapide d'un pH acide à un pH basique permettait de conserver les deux fonctions chimiques sur le peptide **22**. Enfin, afin d'éliminer les sels présents, le mélange réactionnel est élué sur une cartouche Glen-Pak™ et le peptide déprotégé **22** est récupéré à l'aide d'une solution H_2O/CH_3CN (1/1, v/v). Suite à ce traitement de post-déprotection, les fonctions oxyamine et *O*-diethylcarbamoyl-hydroxylamine restent intactes sur le composé **22**. La **Figure 78** illustre l'étape de déprotection effectuée sur le peptide **22b** dans un mélange TFA/H_2O/TIS (80/16/4, v/v/v) ainsi que le traitement de post-déprotection où l'on observe qu'aucune dégradation ne s'effectue sur le peptide **22**. Cependant, il est important de préciser que la solution contenant le peptide **22** ne peut être lyophilisée, ni rester trop longtemps en solution à température ambiante (> deux jours) sous peine de voir la formation du peptide dégradé.

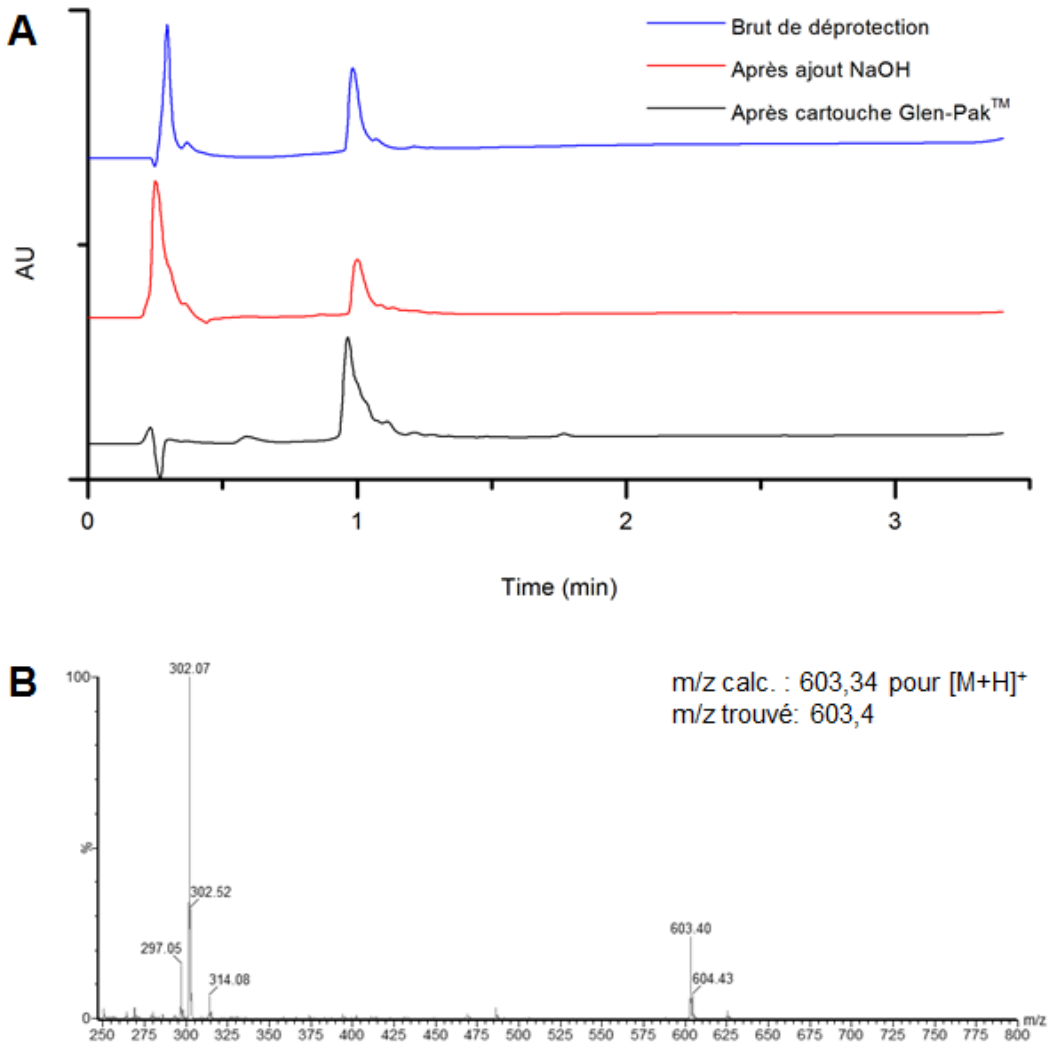


Figure 78 : (a) chromatogrammes UPLC de l'étape de déprotection sur le peptide **22b** ainsi que le traitement de post-déprotection et (b) le spectre UPLC-MS du peptide **22**

a) Ligation oxime : formation du composé I

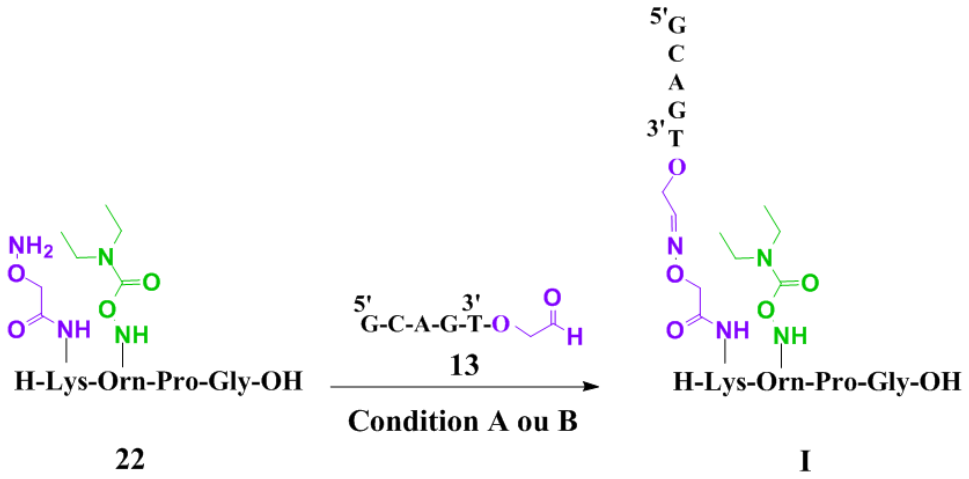


Schéma 26 : synthèse du composé **I**
condition A : oligonucléotide **13** (1eq), tampon NH_4OAc 0,4M pH 4,5, t.a., 51%
condition B : oligonucléotide **13** (1eq), aniline (100eq), tampon NaPhosphate 0,3M pH 7,3, t.a.

Maintenant qu'un traitement de post-déprotection permet de conserver les fonctions oxyamines et *O*-diéthylcarbamoyl-hydroxylamine sur le peptide **22**, il est intéressant de voir si la ligation oxime peut s'effectuer entre ce dernier et l'oligonucléotide **13** afin d'obtenir le composé **I**. Pour cela, deux conditions (A et B) ont été testées (**Schéma 26**). La première (condition A) s'effectue dans un mélange tampon NH_4OAc 0,4M pH 4,5 et à température ambiante (**Figure 79**). La condition B se déroule dans un tampon phosphate salin 0,3M pH 7,3 avec 100eq d'aniline et à température ambiante (**Figure 80**). Dans les deux cas, on observe la formation du produit désiré, confirmé par le spectre ESI-MS (**Figure 79** et **Figure 80**). Le brut réactionnel (condition A) est purifié par RP-HPLC et le composé **I** est obtenu avec un rendement de 51%. En ce qui concerne le brut réactionnel de la condition B, celui-ci n'a pas été purifié par RP-HPLC car il présentait une pureté satisfaisante. Il a donc été dessalé sur colonne d'exclusion stérique. Ces deux conditions se sont montrées concluantes dans l'obtention du composé **I**. De plus, aucune réaction n'est observée de la fonction *O*-diéthylcarbamoyl-hydroxylamine envers les partenaires de la ligation oxime ceci démontrant une orthogonalité dans un système réel. Il est important de préciser qu'il n'est pas possible de déterminer précisément la quantité de peptide **22** obtenu en solution après filtration sur Glen-Pak™ et que cette dernière est initialement estimée pour obtenir une stœchiométrie 1/1 et que la solution de peptide est ajoutée si nécessaire pour obtenir une conversion maximum.

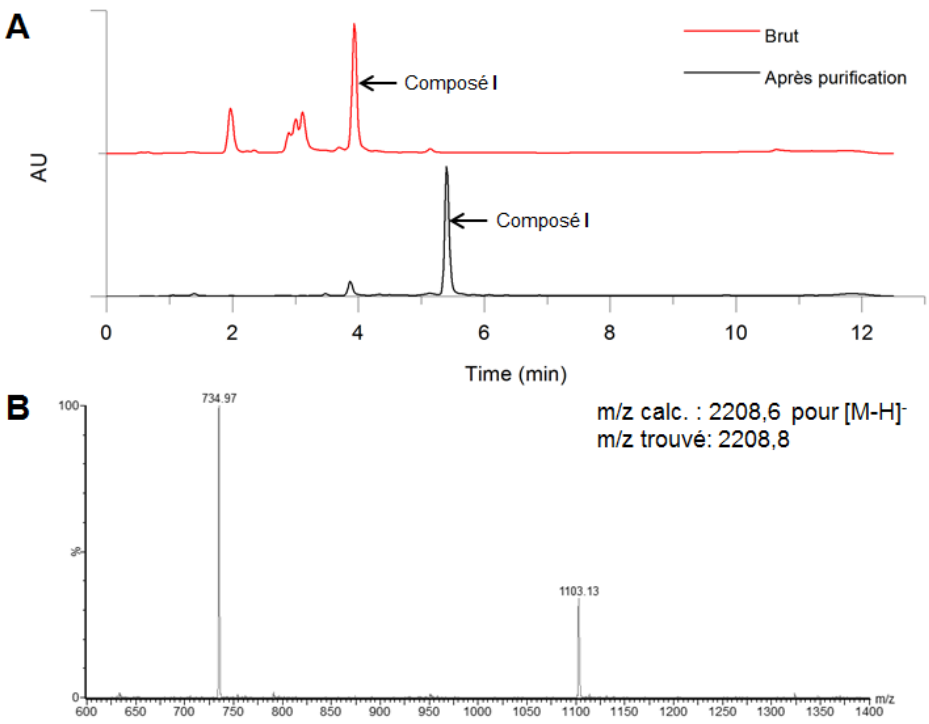


Figure 79 : (a) suivi réactionnel par UPLC de la ligation oxime entre le peptide **22** et l'oligonucléotide **13** (condition A) et (b) spectre UPLC-MS du composé **I**
note : un changement de gradient a été effectué entre l'analyse UPLC du brut réactionnel (gradient III) et celui après purification (gradient I).

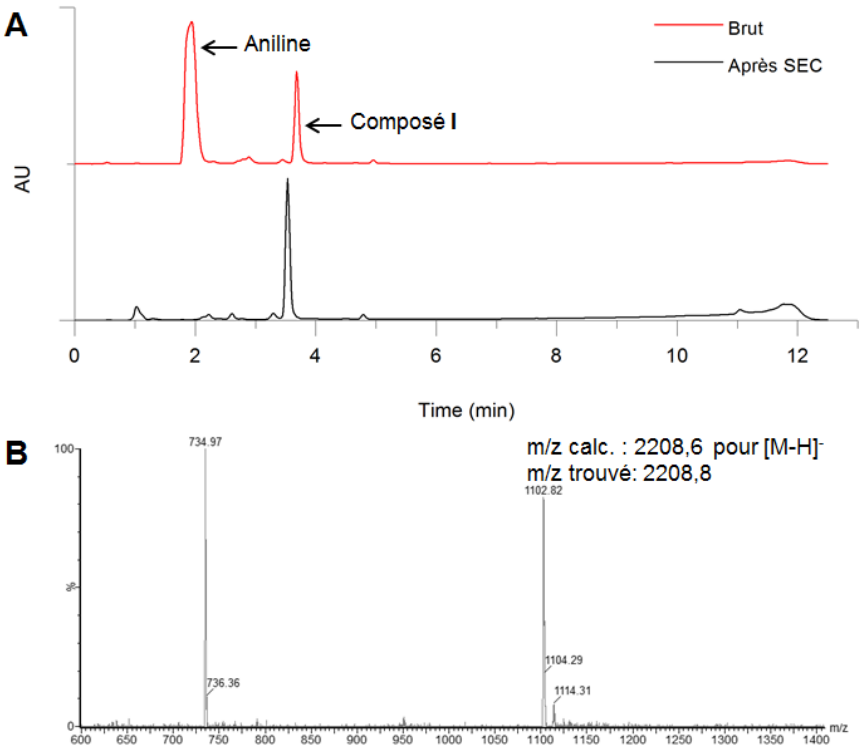


Figure 80 : (a) suivi réactionnel par UPLC de la ligation oxime entre le peptide **22** et l'oligonucléotide **13** (condition B) et (b) spectre UPLC-MS du composé **I**

b) Ligation KAT : formation du composé J

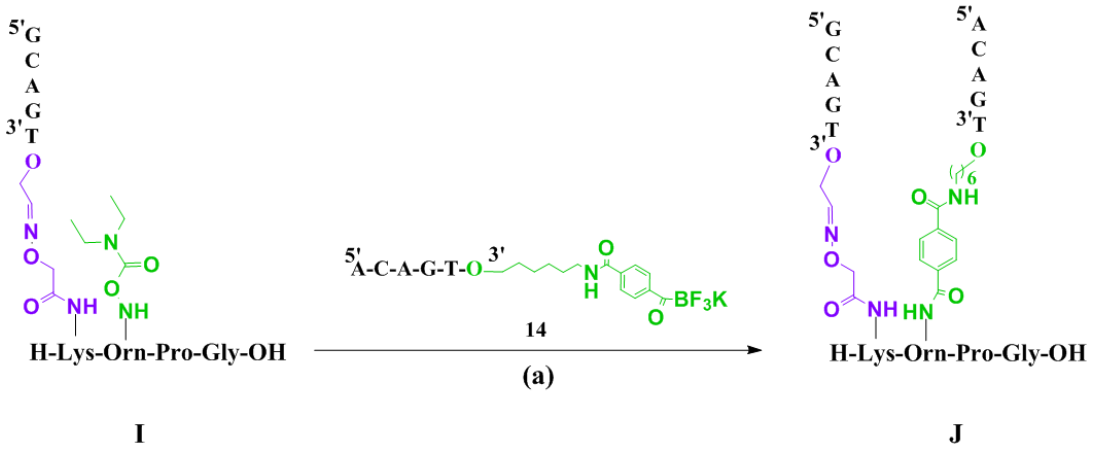


Schéma 27 : synthèse du composé **J**
(a) oligonucléotide **14** (2eq), tampon NH_4OAc 0,4M pH 4,5, t.a., 45 minutes

La formation du composé **J** s'effectue en présence d'un équivalent du composé **I** et de deux équivalents de l'oligonucléotide **14** (**Schéma 27**). Trois conditions réactionnelles ont été utilisées. Les mélanges t-BuOH/H_2O (1/1) et acide oxalique (100mM)/oxalate de potassium (10mM) ont été utilisés dans un but qualificatif et les résultats se sont montrés concluants mais ne sont pas présentées ici. Le meilleur profil a été obtenu avec le mélange tampon NH_4OAc 0,4M pH 4,5 (**Figure 81**). Dans ces conditions, un taux de conversion satisfaisant (> 90%) est obtenu après 45 minutes, le brut réactionnel n'est pas purifié par RP-HPLC étant donné que le composé **J** n'a pas d'application.

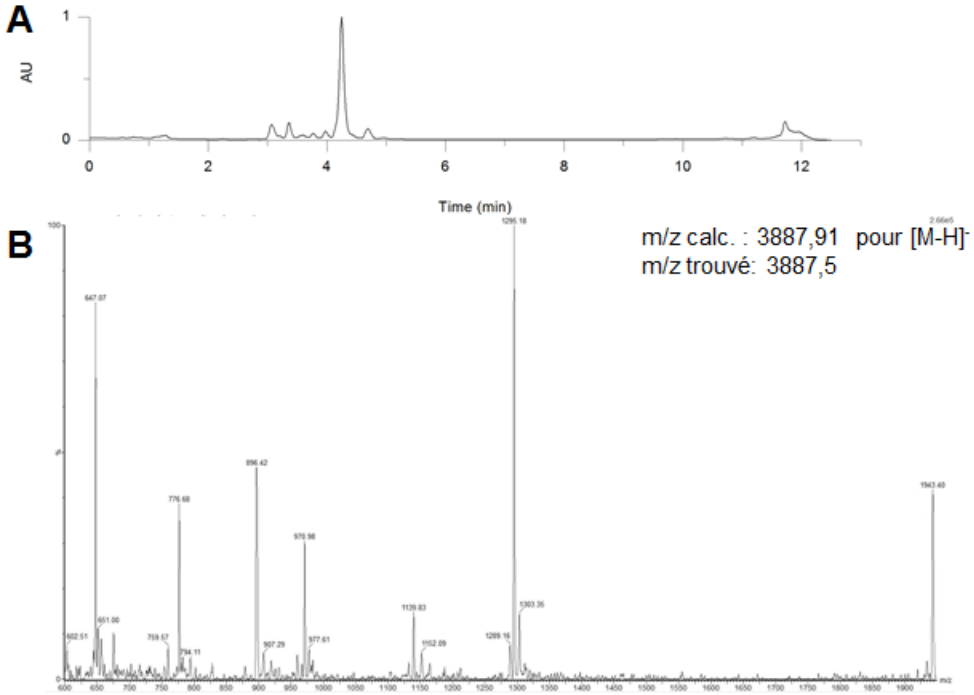


Figure 81 : (a) chromatogramme UPLC du brut réactionnel de la ligation KAT entre le composé **I** et l'oligonucléotide **14** et (b) spectre UPLC-MS du composé **J** obtenu

8. Synthèse de l'édifice H

Il a été démontré précédemment que les ligations oxime et KAT sont compatibles, ainsi la synthèse de l'édifice 'preuve de concept' **H** peut être considérée. Pour obtenir ce dernier, différentes stratégies ont été envisagées étant donné que quatre ligations doivent être réalisées. Cependant, la ligation oxime doit être effectuée obligatoirement en première position dû à la grande réactivité du groupement oxyamine et la possible dégradation du peptide qui peut se dérouler. Une fois la ligation oxime réalisée, trois voies principales sont possibles (**Schéma 28**). En effet, la deuxième ligation peut être la réaction KAT (voie 1), le couplage thiol-chloroacétamide (voie 2) ou la cycloaddition CuAAC (voie 3). Ensuite, pour les trois voies, deux possibilités sont envisageables pour réaliser la troisième et quatrième ligation. Pour la voie 1, le couplage thiol-chloroacétamide et la cycloaddition CuAAC peuvent être réalisé successivement (voie 1a) ou l'inverse (voie 1b). Pour la voie 2, on peut réaliser la ligation KAT suivi de la cycloaddition CuAAC (voie 2a) ou l'inverse (voie 2b). Enfin, pour la voie 3, la ligation KAT et le couplage thiol-chloroacétamide peuvent être envisagées comme troisième et quatrième ligation (voie 3a) ou l'inverse (voie 3b).

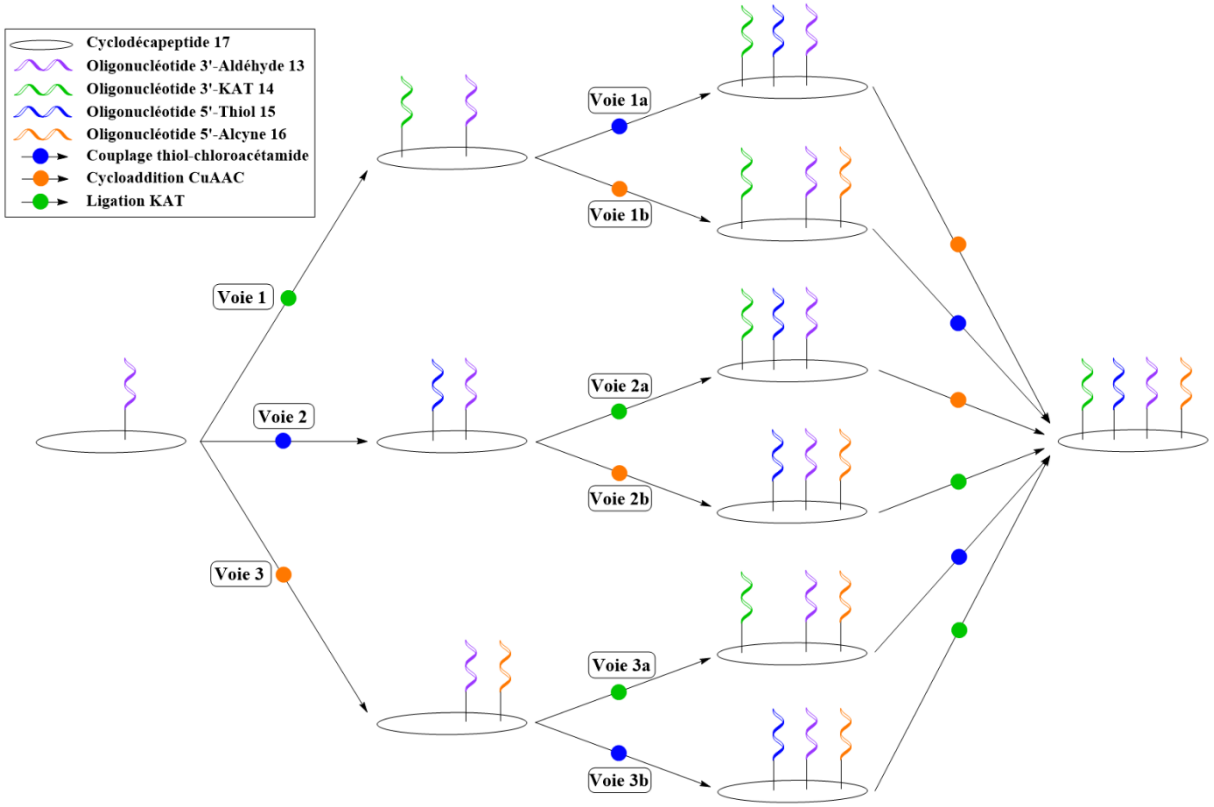


Schéma 28 : illustration des différentes voies envisageables pour synthétiser l'édifice **H**

a) Synthèse de l'édifice H1 par ligation oxime

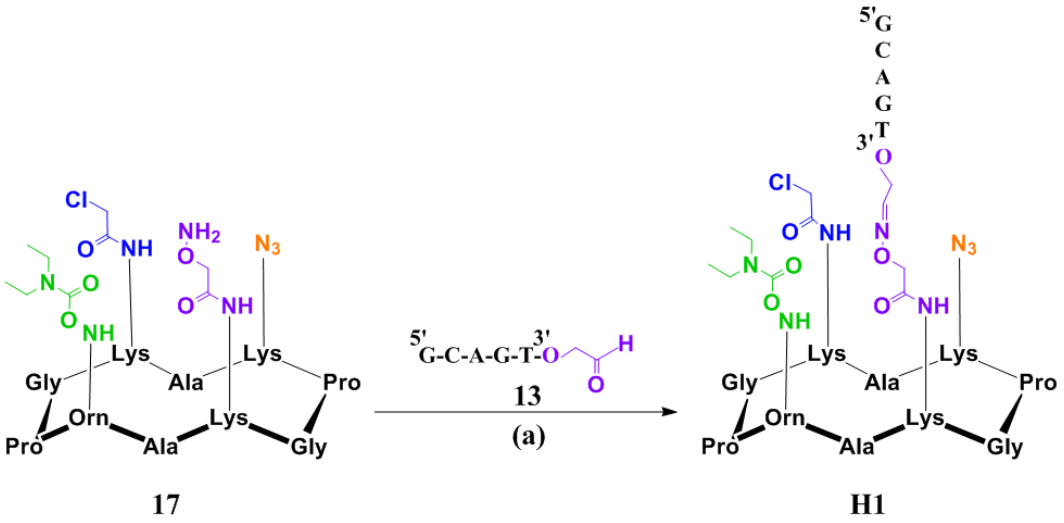


Schéma 29 : synthèse de l'édifice **H1**
(a) oligonucléotide **13** (1eq), tampon NH_4OAc 0,4M pH 4,5, t.a., 62%

Dans le but de former l'édifice **H**, la première ligation est la réaction oxime entre l'oligonucléotide **13** et le châssis peptidique **17** pour former le composé **H1** (**Schéma 29**).

Dans un premier temps, les groupements Boc et Eei du peptide **17e** sont éliminés à l'aide d'un mélange TFA/H_2O/TIS (80/16/4, v/v/v) pour former le composé **17**. Il s'en suit le traitement de post-déprotection dans lequel le TFA est neutralisé par une solution de NaOH 4M jusqu'à atteindre un pH de 9-10. Ensuite, le brut réactionnel est purifié sur une cartouche Glen-Pak™ et le peptide **17** est récupéré à l'aide d'une solution H_2O/CH_3CN (1/1, v/v). Cette étape de déprotection est considérée comme quantitative et le composé **17** est directement utilisé pour effectuer la ligation oxime. La **Figure 82** illustre le suivi réactionnel de l'étape de déprotection pour former le peptide **17**.

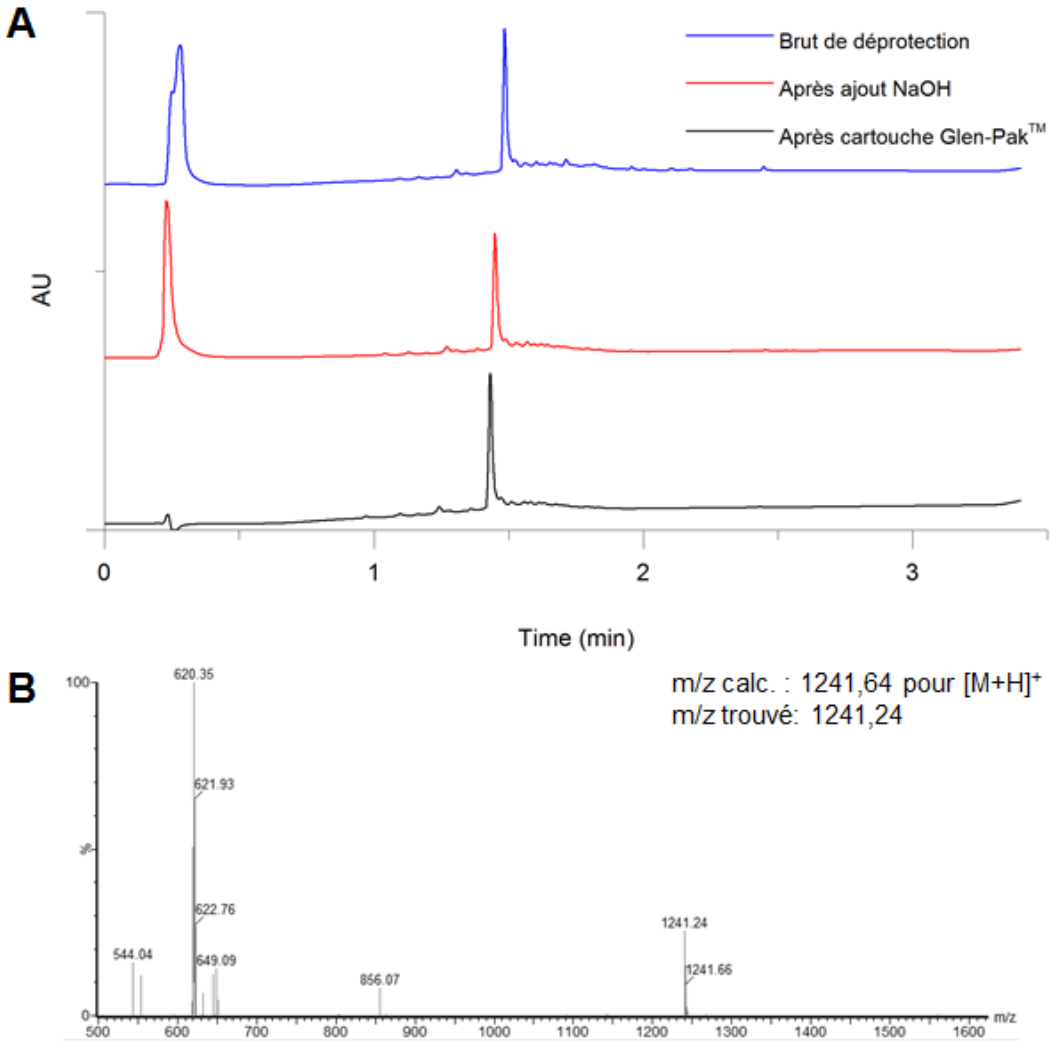


Figure 82 : (a) chromatogrammes UPLC de l'étape de déprotection sur le peptide **17e** ainsi que le traitement de post-déprotection et (b) le spectre UPLC-MS du peptide **17**

Une fois le châssis peptidique **17** obtenu, ce dernier est ajouté à une solution contenant l'oligonucléotide **13** solubilisé dans un mélange tampon NH_4OAc 0,4M pH 4,5. La réaction s'effectue à température ambiante jusqu'à atteindre un maximum de conversion (**Figure 83a**).

En fin de réaction, le brut réactionnel est purifié par RP-HPLC, le composé **H1** est obtenu avec un rendement de 62% et il est caractérisé par UPLC-MS et spectrométrie de masse (**Figure 83**). Il est important de spécifier qu'au cours de la réaction la quantité ajoutée du peptide **17** s'effectue progressivement et jusqu'à atteindre un maximum de conversion. En effet, comme expliqué précédemment pour le peptide **22**, la solution contenant le peptide **17** ne peut être lyophilisée sous peine de dégradation ainsi on ne peut connaitre la quantité exacte de peptide déprotégé récupérée après cartouche Glen-Pak™.

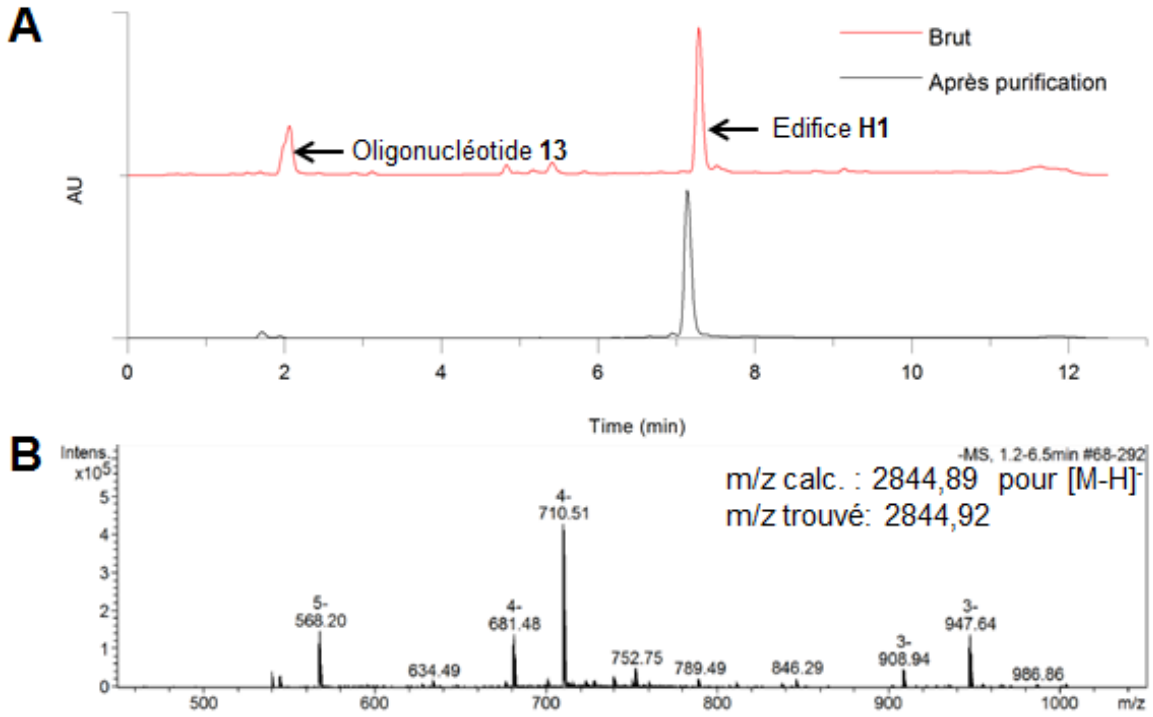


Figure 83 : (a) chromatogrammes de la formation de l'édifice **H1** en UPLC du brut réactionnel et après purification. (b) Analyse ESI-MS de l'édifice **H1**
note : la perte du groupement O-diethylcarbamoyl lors de l'analyse ESI-MS est observée à 908,94 Da et 681,48 Da

b) Voie 1 : synthèse de l'édifice H2 par réaction KAT

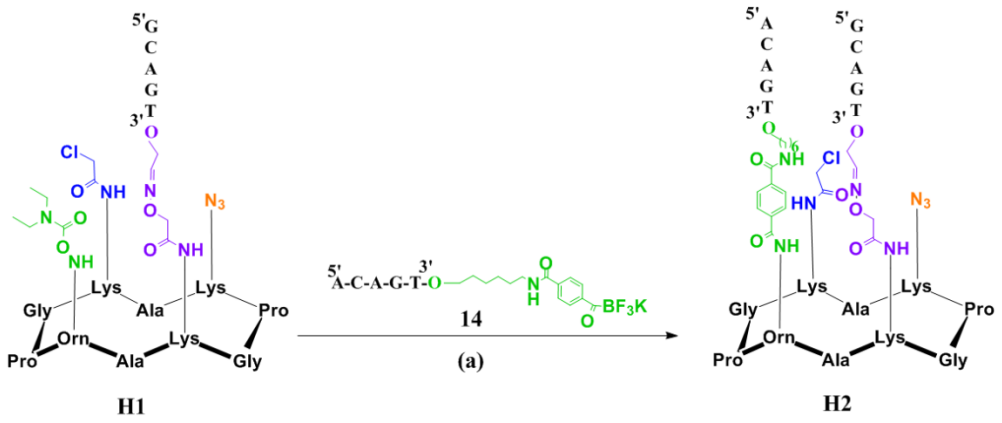


Schéma 30 : synthèse de l'édifice **H2** (voie 1)
(a) oligonucléotide **14** (1,2eq), tampon NH_4OAc 0,4M pH 4,5, t.a., 56%

Par cette stratégie, la seconde ligation réalisée sur l'édifice **H1** est la réaction KAT avec l'oligonucléotide **14** pour former le composé **H2** (**Schéma 30**). Pour ce faire, les deux partenaires sont solubilisés dans le mélange tampon NH_4OAc 0,4M pH 4,5 à température ambiante. En fin de réaction, le brut réactionnel est purifié par RP-HPLC, l'édifice **H2** est obtenu avec un rendement de 56% et il est caractérisé par UPLC-MS et spectrométrie de masse (**Figure 84**).

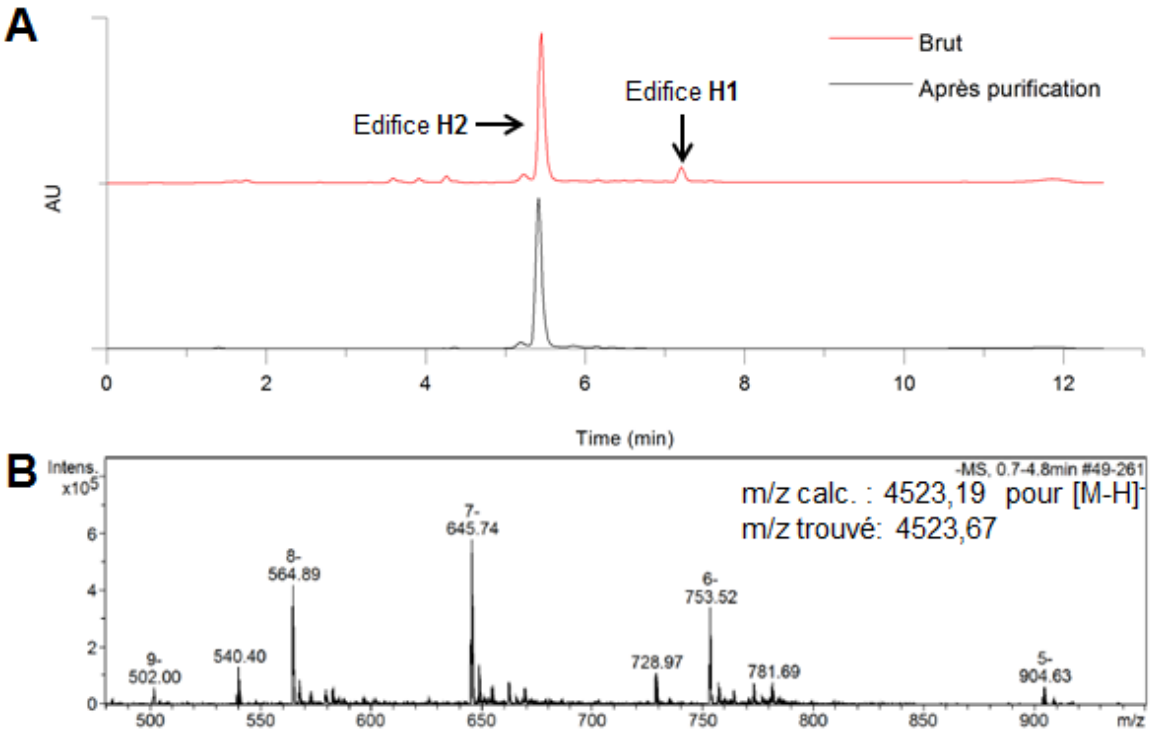


Figure 84 : (a) chromatogrammes de la formation de l'édifice **H2** en UPLC du brut réactionnel et après purification. (b) analyse ESI-MS de l'édifice **H2**

c) Voie 1a et 1b : synthèse des édifices H3 et H4 par CTC et CuAAC

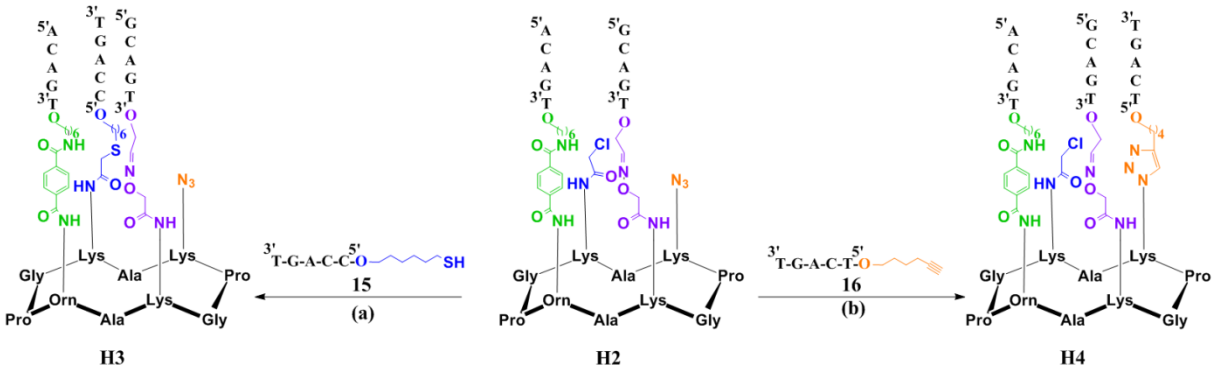


Schéma 31 : synthèse de l'édifice **H3** (voie 1a) et de l'édifice **H4** (voie 1b)
(a) oligonucléotide **15** (2eq), TCEP (1eq), DIEA (60eq), KI (60eq), 400mM KCl, H_2O, 40°C
(b) oligonucléotide **16** (2eq), $CuSO_4$ (3eq), THPTA (30eq), ascorbate de sodium (30eq), 500mM KCl, DMF (10%vol), tampon PBS 50mM pH 7,4, 1h, 37°C, 69%

A partir de l'édifice **H2**, deux possibilités sont envisageables, soit d'effectuer le couplage thiol-chloroacétamide avec l'oligonucléotide **15** et former le composé **H3** (voie 1a), soit de réaliser la cycloaddition CuAAC avec l'oligonucléotide **16** pour obtenir l'édifice **H4** (voie 1b) (**Schéma 31**).

i. <u>Synthèse de l'édifice **H3** (voie 1a) par CTC</u>

Dans le but de former l'édifice **H3**, le composé **H2** et l'oligonucléotide **15** sont solubilisés dans l'eau et mis en présence de 60eq de DIEA et de KI et de 1eq de TCEP et le mélange réactionnel est placé à 40°C. De plus, le suivi de la réaction s'effectue par UPLC-MS afin de s'assurer, par le biais de la spectrométrie de masse, que l'oligonucléotide **15** est bien sous forme thiol et non disulfure. Il est observé au cours du suivi que la réaction entre les deux partenaires s'effectue difficilement. Les réactifs tels que DIEA et KI sont rajoutés au milieu afin d'être sûr qu'ils soient présents en large excès. Il en est de même pour le TCEP où 1eq est rajouté afin d'éviter la formation de pont disulfure entre oligonucléotides **15**. Après 4h de réaction, on observe dans le brut réactionnel la présence du produit désiré mais où la fonction azoture présent sur le support peptidique a été réduite (composé nommé **H3'**) (**Figure 85**). Il est important de spécifier que différentes précautions ont été mise en place avant et pendant la réaction afin d'éviter cette réduction. En effet, l'ajout de TCEP dans le milieu réactionnel ne s'effectue pas en début de réaction mais au cours de celle-ci, lorsqu'on observe la présence d'oligonucléotide **15** sous forme disulfure. De plus, la réaction de détritylation de l'oligonucléotide **15** s'effectue la veille du couplage chloro-acétamide afin d'avoir ce dernier sous forme thiol en début de réaction. Et enfin, la mise en place de la réaction se fait sous

atmosphère inerte et avec l'utilisation d'eau désoxygénée. Toutefois, l'ajout du TCEP au mélange réactionnel est nécessaire en cours de réaction et la réduction de la fonction azoture est, malheureusement, observée. Face à cette problématique, la voie 1a a été abandonnée.

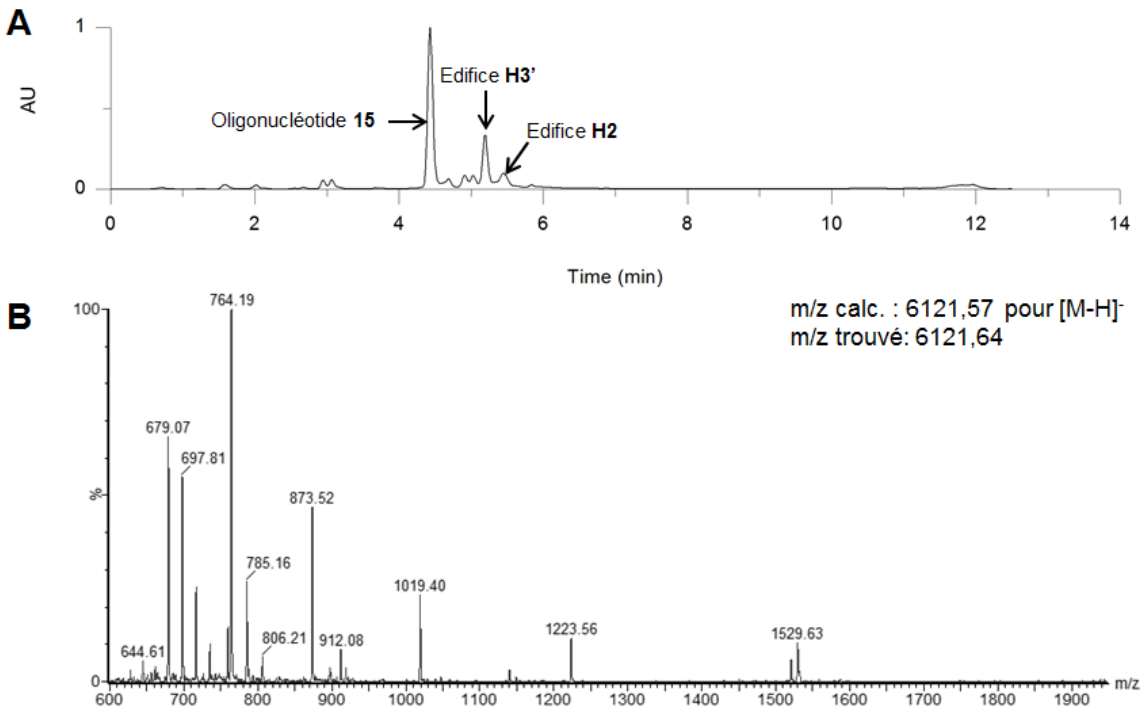


Figure 85 : (a) chromatogramme en UPLC du brut réactionnel pour la formation du composé **H3** et (b) le spectre UPLC-MS du composé **H3'**

ii. <u>Synthèse de l'édifice **H4** (voie 1b) par CuAAC</u>

Le couplage entre l'édifice **H2** et l'oligonucléotide **16** s'effectue dans des conditions usuelles de cycloaddition CuAAC. En fin de réaction, le brut réactionnel est purifié par RP-HPLC, l'édifice **H4** est obtenu avec un rendement de 69% et il est caractérisé par UPLC-MS et spectrométrie de masse (**Figure 86**).

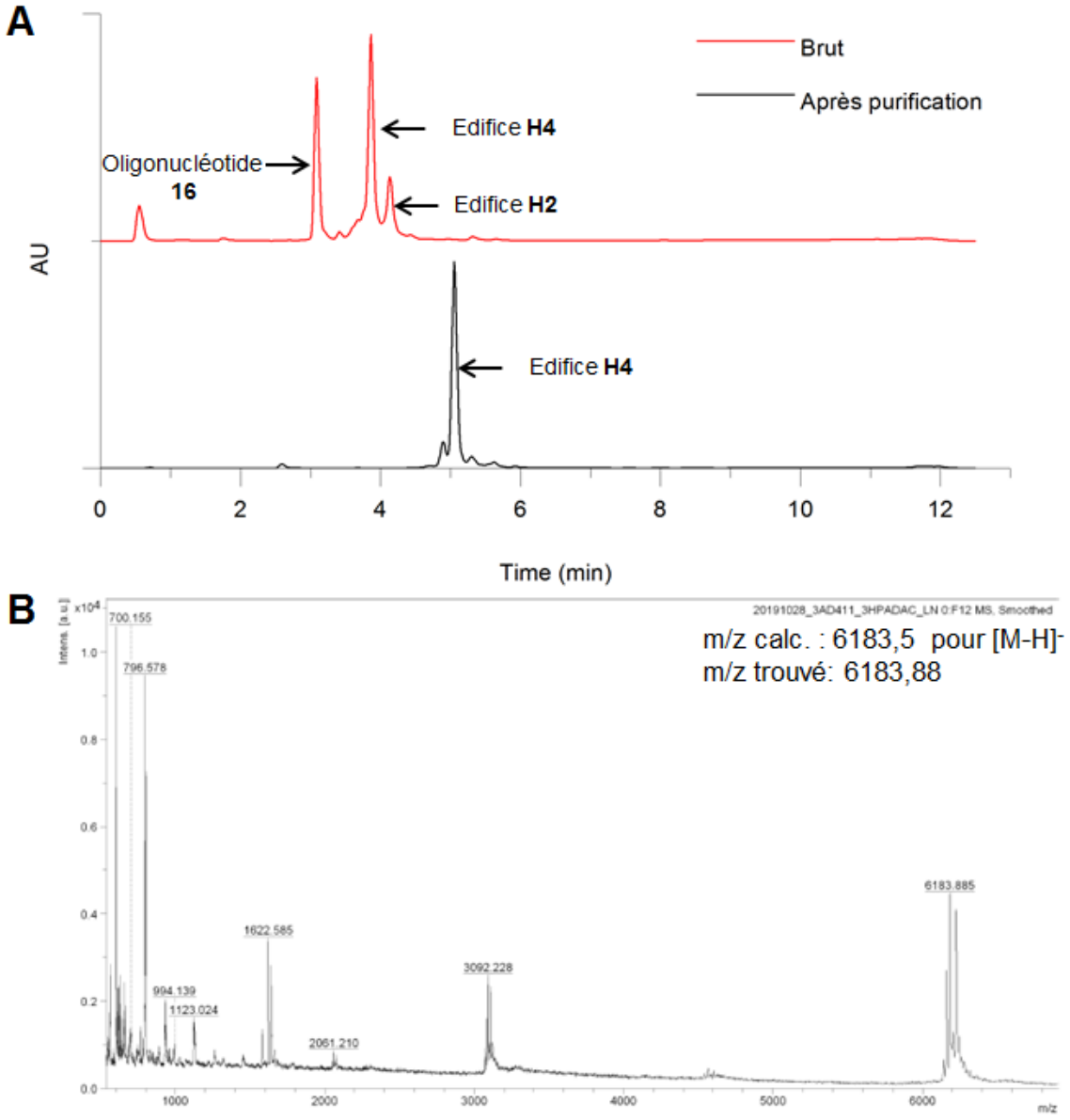


Figure 86 : (a) chromatogrammes de la formation de l'édifice **H4** en UPLC du brut réactionnel et après purification. Note : un changement de gradient a été effectué entre l'analyse UPLC du brut réactionnel (gradient II) et celui après purification (gradient I). (b) Analyse MALDI de l'édifice **H4**

iii. Synthèse de l'édifice **H** (voie 1b) par CTC

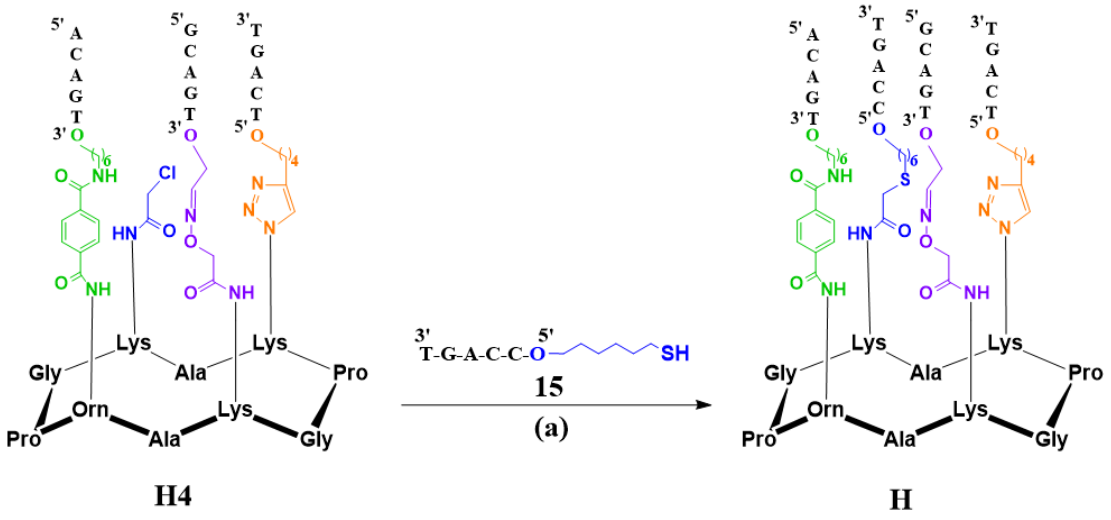


Schéma 32 : synthèse de l'édifice **H** (voie 1b)
(a) oligonucléotide **15** (2eq), TCEP (1eq), DIEA (60eq), KI (60eq), 400mM KCl, H_2O, 40°C

La dernière étape de la voie 1b est de réaliser le couplage thiol-chloroacétamide entre l'oligonucléotide **15** et le composé **H4** pour former l'édifice final **H** (**Schéma 32**). Pour ce faire, des conditions usuelles de couplage thiol-chloroacétamide ont été utilisées (comme pour la synthèse du composé **H3**) et l'ajout de TCEP dans le milieu peut s'effectuer dès le départ de la réaction étant donné qu'aucune fonction pouvant être réduite n'est présente. Lors du suivi réactionnel, on observe que la réaction s'effectue difficilement et ce malgré la présence des différents réactifs en large excès. De plus, on constate la présence de deux produits secondaires (**H4'** et **H4''**) de nature inconnue possédant une masse de 6147 g/mol et 6378 g/mol, respectivement. Une hypothèse à la formation de ces derniers est la présence de cuivre résiduel dans le milieu réactionnel et ce malgré les précautions réalisées après la cycloaddition CuAAC (piégeage du cuivre par une solution aqueuse d'EDTA et dessalage sur colonne d'exclusion stérique). Toutefois, on remarque la légère formation de l'édifice **H** (**Figure 87**). Etant donné le faible taux de conversion ainsi que la proximité en terme de temps de rétention des composés **H4**, **H4'**, **H4''** et **H**, cette voie n'a pas été investiguée plus en profondeur.

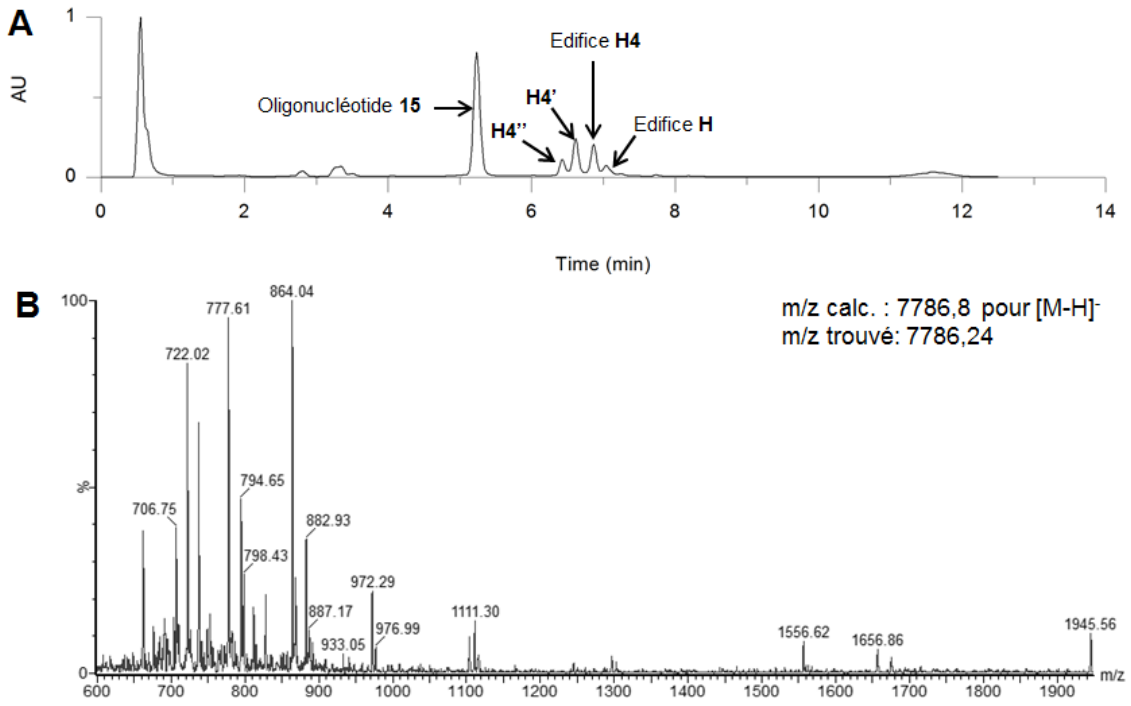


Figure 87 : (a) chromatogramme en UPLC du brut réactionnel pour la formation du composé **H** et (b) le spectre UPLC-MS du composé **H**

d) Voie 2 : synthèse du composé H5 par CTC

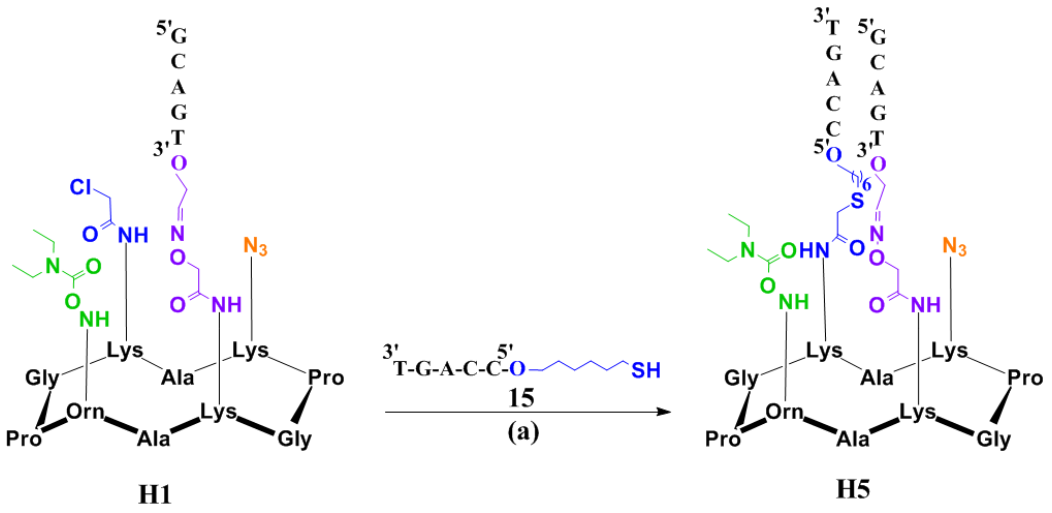


Schéma 33 : synthèse de l'édifice **H5** (voie 2)
(a) oligonucléotide **15** (2eq), TCEP (2eq), DIEA (60eq), KI (60eq), 400mM KCl, H_2O, 1h30, 40°C, 47%

Malgré que par la voie 1b, il a été observé la formation du produit final désiré, l'édifice **H**, cette stratégie ne s'est finalement pas montrée convaincante et différents problèmes ont été constatés autour du couplage thiol-chloroacétamide. En effet, sa faible réactivité ainsi que la réduction du groupement azoture a limité la formation de l'édifice **H**. Par la voie 2, ce couplage

est réalisé en deuxième position pour former le composé **H5** (**Schéma 33**). Etant donné la présence du groupement azoture sur la surface supérieure du châssis peptidique, les mêmes précautions appliquées précédemment ont été utilisées. Il en va de même concernant les conditions réactionnelles. Après 1h30, il a été observé sur le brut réactionnel la formation du composé **H5** sans réduction du groupement azoture et avec un taux de conversion acceptable (75%). Après purification par RP-HPLC, l'édifice **H5** a été obtenu avec un rendement de 47% et a été caractérisé par UPLC-MS et spectrométrie de masse (**Figure 88**).

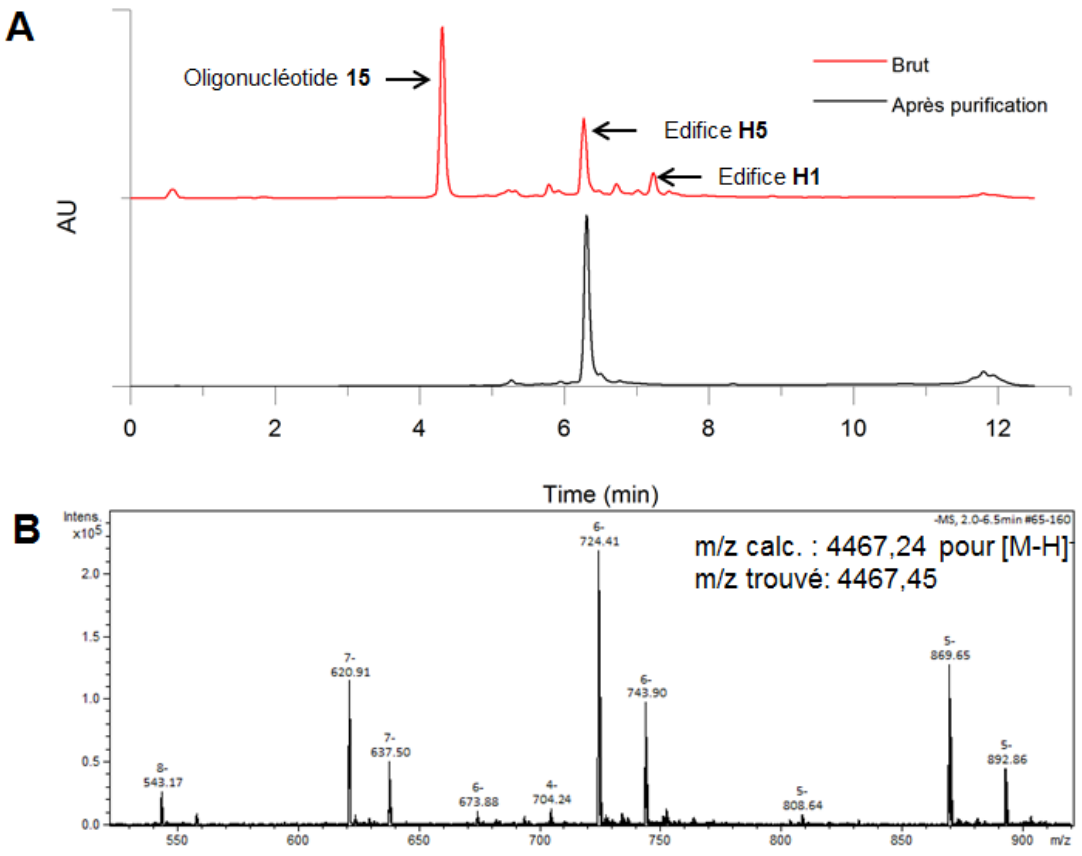


Figure 88 : (a) chromatogrammes de la formation de l'édifice **H5** en UPLC du brut réactionnel et après purification. (b) Analyse ESI-MS de l'édifice **H5**.
Note : la perte du groupement *O*-diethylcarbamoyl lors de l'analyse ESI-MS est observée à 869,65 Da, 724,41 Da, 620,91 Da et 543,17 Da

e) Voie 2a et 2b : synthèse des édifices H3 et H6 par KAT et CuAAC

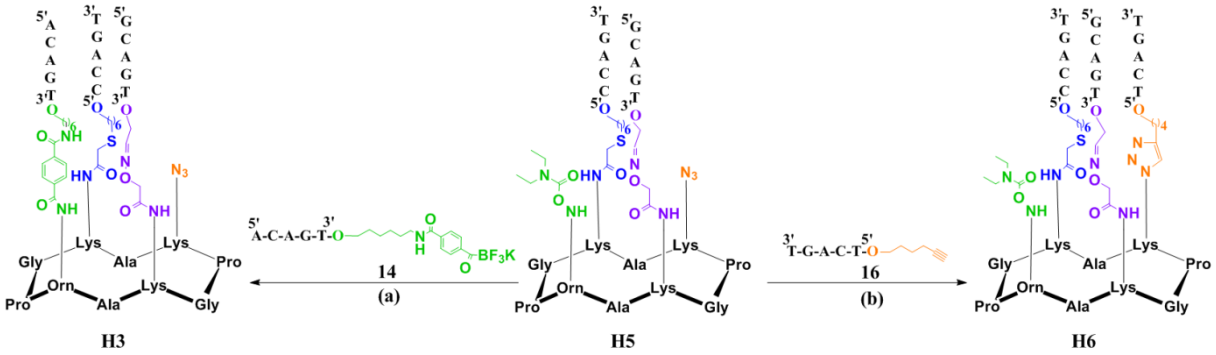


Schéma 34 : synthèse de l'édifice **H3** par la voie 2a et de l'édifice **H6** par la voie 2b
(a) oligonucléotide **14** (1,2eq), tampon NH_4OAc 0,4M pH 4,5, t.a., 57%
(b) oligonucléotide **16** (2eq), $CuSO_4$ (3eq), THPTA (30eq), ascorbate de sodium (30eq), 400mM KCl, DMF (10%vol), tampon PBS 50mM pH 7,4, 37°C

A partir de l'édifice **H5**, deux possibilités sont envisageables, soit d'effectuer la ligation KAT avec l'oligonucléotide **14** et former le composé **H3** (voie 2a), soit de réaliser la cycloaddition CuAAC avec l'oligonucléotide **16** pour obtenir l'édifice **H6** (voie 2b) (**Schéma 34**).

i. <u>Synthèse de l'édifice **H3** (voie 2a) par réaction KAT</u>

La formation du composé **H3** s'effectue en présence d'un équivalent de l'édifice **H5** et de 1,2 équivalents de l'oligonucléotide **14** dans un mélange tampon NH_4OAc 0,4M pH 4,5 à température ambiante. La réaction est suivie par UPLC-MS et le brut réactionnel est purifié par RP-HPLC, le composé **H3** est obtenu avec un rendement de 57% et il est caractérisé par UPLC-MS et spectrométrie de masse (**Figure 89**). A noter qu'on observe la formation d'un produit secondaire (15%), le composé **H5'**. Ce dernier représente l'édifice de départ **H5** mais où le groupement *O*-diéthylcarbamoyle est perdu, montrant une certaine fragilité de ce dernier.

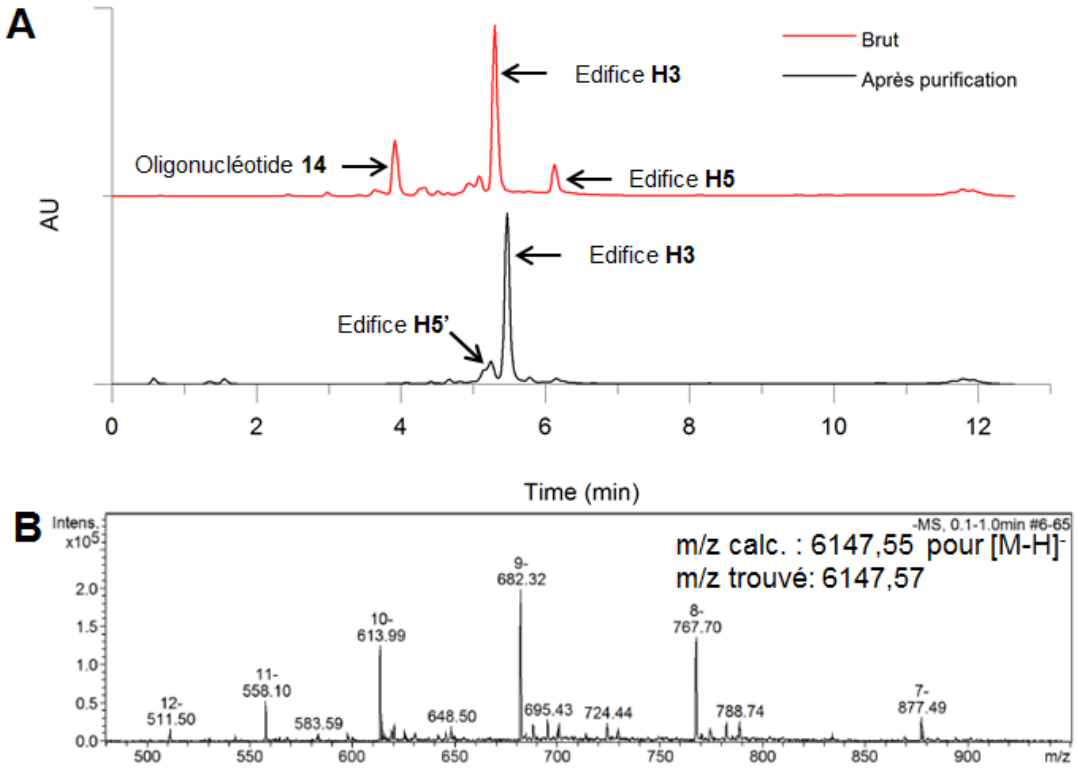


Figure 89 : (a) chromatogrammes de la formation de l'édifice **H3** en UPLC du brut réactionnel et après purification. (b) Analyse ESI-MS de l'édifice **H3**

ii. <u>Synthèse de l'édifice **H** (voie 2a) par CuAAC</u>

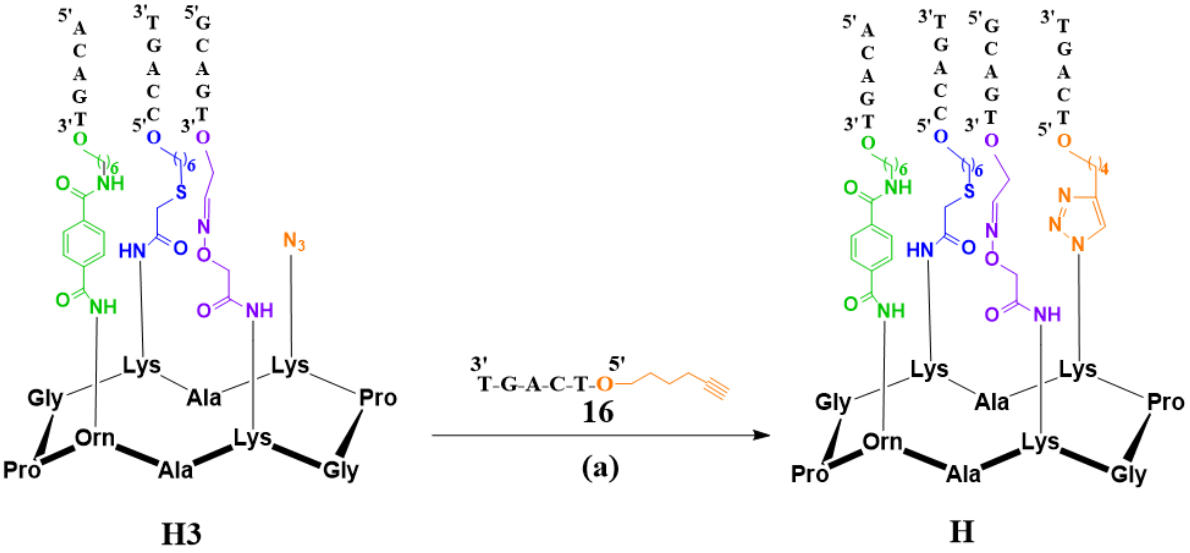


Schéma 35 : synthèse de l'édifice **H** par la voie 2a
(a) oligonucléotide **16** (1,2eq), $CuSO_4$ (10eq), THPTA (30eq), ascorbate de sodium (100eq), 400mM KCl, DMF (10%vol), tampon PBS 50mM pH 7,4, 1h, 60°C, 48%

Par la voie 2a, l'ultime étape est de réaliser la cycloaddition CuAAC entre le composé **H3** et l'oligonucléotide **16** (**Schéma 35**). Pour ce faire, les conditions réactionnelles utilisées ont été modifiées par rapport aux conditions usuelles. En effet, la quantité de certains réactifs ont été augmentées afin qu'ils soient présents en large excès dans le milieu et le mélange réactionnel a été placé à 60°C pour favoriser la cycloaddition CuAAC. On observe sur le brut réactionnel, la formation de l'édifice final **H**. Après purification par RP-HPLC, ce dernier est obtenu avec un rendement de 48% et est caractérisé par UPLC-MS et spectrométrie de masse (**Figure 90**).

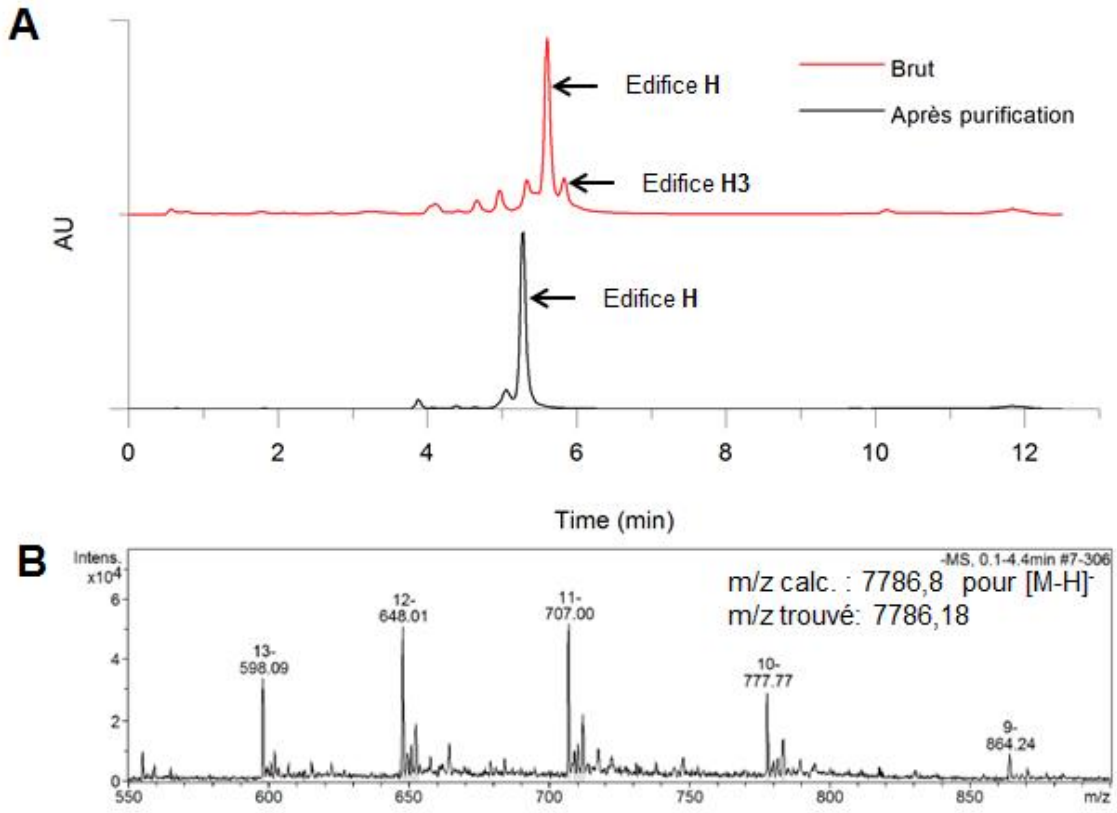


Figure 90 : (a) chromatogrammes de la formation de l'édifice **H** en UPLC du brut réactionnel et après purification. (b) Analyse ESI-MS de l'édifice **H**

iii. <u>Synthèse de l'édifice **H6** (voie 2b) par CuAAC</u>

La seconde alternative à partir de l'édifice **H5** est de réaliser la cycloaddition CuAAC pour former le composé **H6** (**Schéma 34**). Celle-ci a été effectuée dans les conditions usuelles de CuAAC. Toutefois, on remarque, sur le brut réactionnel, que le groupement *O*-diethylcarbamoyl présent sur le composé **H5** est perdu lors de la réaction rendant ce composé, **H5'**, inutilisable pour la suite (**Figure 91**). A noter qu'il avait déjà été constaté précédemment (synthèse de l'édifice **H3** par la voie 2a) que ce groupement possédait une certaine fragilité.

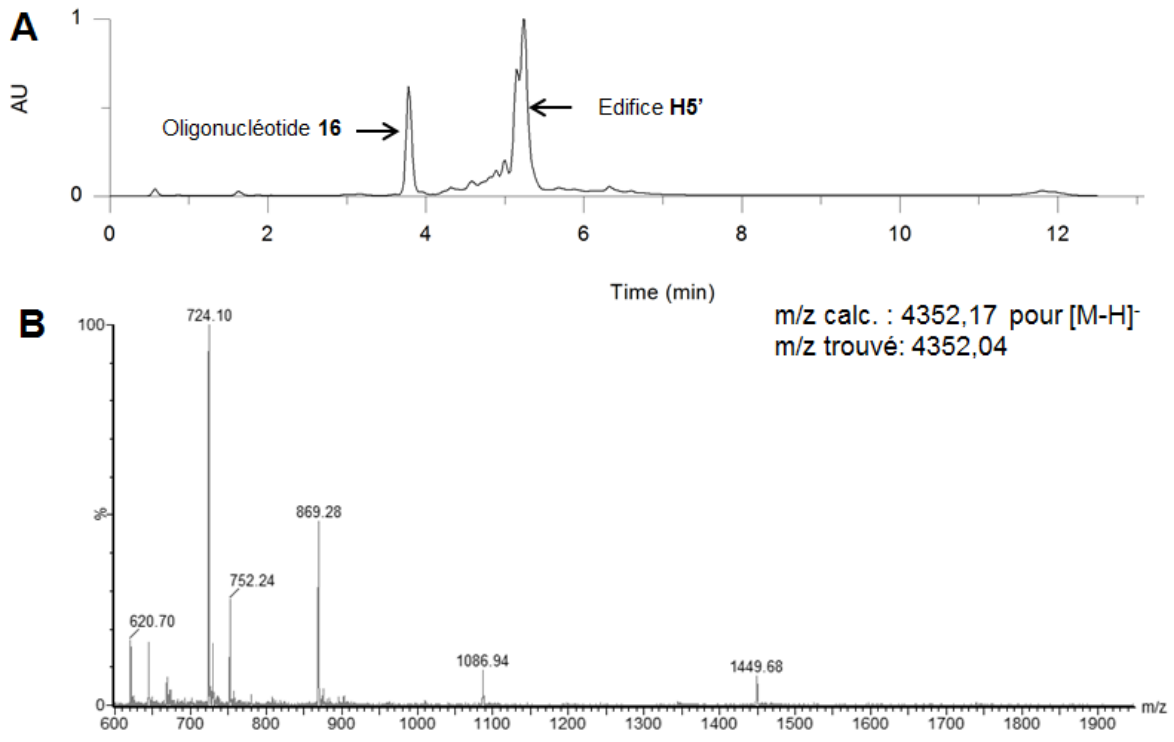


Figure 91 : (a) chromatogramme en UPLC du brut réactionnel pour la formation du composé **H6** et (b) le spectre UPLC-MS du composé **H5'**

Il est important de spécifier que par manque de temps, la stratégie de synthèse de l'édifice **H** par la voie 3 n'a pas pu être investiguée. Toutefois, cette voie est peut-être la moins intéressante des trois étant donné que le couplage thiol-chloroacétamide et la ligation KAT s'effectue en troisième et quatrième position.

9. Bilan et perspectives

Les travaux décrits dans ce chapitre ont tout d'abord montrés que la ligation KAT pouvait être transposée à la synthèse de conjugués d'oligonucléotides. A notre connaissance, c'est le premier exemple de composé faisant intervenir cette ligation entre ces deux types de partenaires.

Dans un second temps, il a été observé une interférence entre les fonctions oxyamine et *O*-diéthylcarbamoyle. Effectivement, lorsque celles-ci sont déprotégées, dans un milieu concentré qui n'est pas à pH acide fort, il est constaté une réaction entre ces deux fonctions provoquant une perte de masse de 134 g/mol sur le composé étudié. Cette réaction de dégradation n'a pas été étudiée plus en profondeur, on sait que suite à cette dégradation, la fonction oxyamine n'est plus présente sur le châssis peptidique étant donné qu'aucune réaction n'est observée lorsqu'un partenaire aldéhyde est mis en présence. Toutefois, un traitement de

post-déprotection adapté, neutralisant l'acidité du milieu, permet de conserver de manière intacte ces deux fonctions pour qu'elles soient utilisées dans leurs ligations respectives (synthèse des édifices **I** et **J**).

Suite à cela, il a été possible d'envisager la synthèse de l'édifice **H**, composé prouvant qu'il est possible de greffer quatre oligonucléotides séquentiellement par quatre ligations différentes. Pour cela, différentes voies de synthèse sont envisageables mais toutes ont comme première ligation la réaction oxime étant donné la grande réactivité du groupement oxyamine et la possible réaction de dégradation du peptide. Sur les trois voies principales de synthèse, deux ont pu être investiguées.

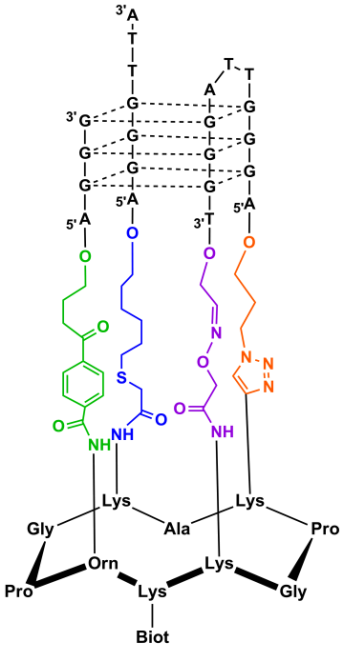
La voie 1 met en jeu la réaction KAT comme seconde ligation et le produit **H2** est obtenu sans contrainte. Le couplage thiol-chloroacétamide et la cycloaddition CuAAC peuvent être utilisées comme troisième et quatrième ligation (voie 1a) ou on peut également effectuer l'ordre inverse de ces deux réactions (voie 1b). Dans la voie 1a, il a été constaté lors du couplage thiol-chloroacétamide une réduction du groupement azoture présent sur le châssis peptidique dû à la présence de TCEP. Malgré différents aménagements effectués pour retarder l'ajout de TCEP, ce dernier doit être tout de même présent pour éviter la formation d'oligonucléotides sous forme disulfure causant malheureusement la réduction de la fonction azoture. Ainsi la voie 1a ne s'est pas montrée concluante. La voie 1b fait intervenir dans un premier temps, la cycloaddition CuAAC pour former le composé **H4** pour ensuite réaliser le couplage thiol-chloroacétamide et donner l'édifice final **H**. Le composé **H4** a été obtenu dans des conditions usuelles de cycloaddition CuAAC mais lors de la dernière étape, le couplage thiol-chloroacétamide ne s'est pas montré efficace et il a été observé la formation du composé final **H** en faible proportion ainsi que deux produits secondaires, **H4'** et **H4''**. Etant donné le taux de conversion faible et la proximité en termes de temps de rétention des différents composés, cette voie n'a pas été étudiée plus en profondeur.

Par la voie 2, le couplage thiol-chloroacétamide est utilisé comme deuxième ligation pour former l'édifice **H5**. Celui-ci a été obtenu, cependant il a été observé une perte du groupement *O*-diéthylcarbamoyle sur ce composé (composé **H5'**). A partir de là deux possibilités sont possibles, soit réaliser la ligation KAT en troisième position pour former le composé **H3** et finir avec la cycloaddition CuAAC pour obtenir l'édifice final **H** (voie 2a), soit d'effectuer l'ordre inverse de ces deux réactions (voie 2b). La voie 2a s'est montrée concluante pour obtenir l'édifice **H** alors que par la voie 2b, lors de la réaction de cycloaddition CuAAC,

la perte du groupement *O*-diéthylcarbamoyle a de nouveau été observée rendant inutilisable le composé pour réaliser la dernière étape. En ce qui concerne la voie 3, elle n'a pas pu être étudiée par manque de temps.

Par ces études, il en ressort différents points importants. Le premier est que la réactivité du couplage thiol-chloroacétamide ne s'est pas montrée efficace lorsque celui-ci était utilisé en troisième ou quatrième position. Ainsi, positionner ce couplage en début de stratégie permet de s'affranchir de ce désavantage pour la suite de la synthèse. De plus lors de ce couplage, il peut se dérouler une réduction du groupement azoture présent sur le support peptidique rendant inutilisable le composé pour réaliser une cycloaddition CuAAC. Une solution à ce problème serait d'inverser les fonctions chimiques des deux partenaires, c'est-à-dire introduire la fonction azoture sur l'oligonucléotide et avoir la fonction alcyne sur le châssis peptidique. Un autre point important à souligner, c'est la fragilité du groupement *O*-diethylcarbamoyl. En effet lors de ces études, il a été observé à plusieurs reprises la perte de ce groupement provoquant l'impossibilité de réaliser la ligation KAT, ainsi positionner cette ligation en fin de synthèse peut être risqué.

En ce qui concerne les perspectives, il aurait été intéressant d'investiguer la voie de synthèse de l'édifice **H** numéro 3. En effet, celle-ci aurait pu donner ou confirmer différentes informations pouvant être utiles dans la synthèse d'édifices utilisant quatre ligations différentes. Une autre perspective serait de réaliser certaines ligations en « one-pot ». Par exemple, effectuer dans un premier temps la ligation oxime et de réaliser ensuite la ligation KAT puisque ces deux dernières peuvent se dérouler dans le même milieu réactionnel. De plus, il serait également intéressant d'entreprendre la synthèse d'édifices G4 utilisant quatre ligations orthogonales, comme l'édifice **K** par exemple (**Figure 92**) qui représente un G4 télomérique en topologie hybride. Effectivement, utiliser quatre ligations permet d'ouvrir le champ des possibilités concernant la synthèse d'édifices G4 et leur variété de topologie. Ainsi différentes cibles ayant des structures bien particulières peuvent être envisagées et être considérées comme des cibles attrayantes dans le cadre d'études d'interactions (par exemple : études d'affinité et sélectivité de ligands de G4, études d'interaction entre G4 et protéines, etc.)



K

Figure 92 : représentation de l'édifice **K**

Chapitre V

Synthèse d'édifices G4 contraints d'ADN et d'ARN à une seule topologie et étude d'interaction avec des protéines

Dans ce chapitre, la synthèse d'un panel d'édifices G-quadruplexes et non G-quadruplexes de type ADN et ARN sera décrite. Ces différents composés ont pour but d'être utilisés dans des études d'identification et de caractérisation de protéines en interactions avec ces édifices. En particulier, l'objectif est de tirer profit de la stabilisation des structures G4 par le châssis moléculaire et du contrôle topologique afin d'identifier des protéines affines et spécifiques de certaines structures. La genèse de ces études provient d'une collaboration établie avec le laboratoire du Dr. D. Gomez à l'université de Toulouse (Institut de Pharmacologie et de Biologie Structurale, IPBS). Les résultats préliminaires des études d'interactions entre protéines et édifices seront également présentés.

1. Introduction

Comme mentionné dans l'introduction générale de ce manuscrit, les G-quadruplexes sont retrouvés de manière non-aléatoire au sein du génome humain et de multiples études confirment leurs implications lors de différents évènements cellulaires tels que la réplication d'ADN[252], la réparation des dommages d'ADN[253], la transcription[254] ou encore la traduction[255]. Bien entendu, les G4 ne peuvent réguler à eux seuls ces processus et de nombreuses protéines sont également impliquées. Les protéines interagissant avec les G4 peuvent être divisées en trois groupes selon le lien qu'il existe entre ces deux entités. Tout d'abord, on retrouve les protéines qui favorisent la formation ou la stabilisation des G4. Il y a également les protéines qui 'dégradent' les G4 et enfin, il existe des protéines de 'déroulement' de G4 (« unwinding proteins » en anglais). Ces processus, qui contribuent à la dynamique de la régulation des G4, sont encore largement méconnus.

a) Les protéines favorisant la formation de G4

Au sein de ce premier groupe, la nucléoline (NCL) est la protéine la plus connue et la plus étudiée. Celle-ci est une phosphoprotéine nucléolaire qui est impliquée notamment dans la transcription et possède également comme rôle d'aider à la formation de structures complexes d'acides nucléiques[256]. La nucléoline peut, par exemple, promouvoir et se lier à des structures G4 parallèles (*in vitro*) du gène *c-myc* avec une haute affinité et sélectivité. La stabilisation de ce genre de structures par la nucléoline permet l'inhibition de l'activité du gène *c-myc*[257]. On retrouve également dans ce premier groupe de protéines les ADN topoisomérase (Topo) qui sont des enzymes contrôlant la topologie de l'ADN et sont impliquées dans la réplication ou encore dans la transcription. Certaines de ces protéines peuvent se lier et favoriser la formation de structures G4 parallèles et antiparallèles[258].

b) Les protéines dégradant les G4

Les protéines permettant la dégradation des structures G4 de type ADN ou ARN sont un peu moins étudiées et la plupart d'entre elles sont des nucléases. La GQN1 pour « G quartet nuclease 1 » est une nucléase dont l'expression est élevée dans les lymphocytes B qui dégrade seulement les G4 de type ADN mais pas les duplex ni les simples brins d'ADN, ni les G4 de type ARN[259]. Un autre exemple de ce type de protéine est la protéine Mre11 qui possède une haute affinité pour les G4 de type ADN et est impliquée dans le processus de maintenance des télomères[260,261].

c) Les protéines déroulant les G4 : les hélicases

Les hélicases peuvent être définis comme étant des moteurs moléculaires puisqu'elles utilisent l'énergie d'hydrolyse de l'ATP pour catalyser l'ouverture d'acides nucléiques (ADN ou ARN) appariés sous forme double brins. On retrouve communément dans la littérature cinq familles d'hélicases qui sont discriminées selon le substrat sur lequel elles agissent (ADN et/ou ARN), mais également sur la direction de leurs 'mouvements', c'est-à-dire si elles se déplacent sur le substrat de la position 5' vers l'extrémité 3', du 3' vers le 5' ou les deux. Toutefois, toutes les hélicases nécessitent la présence d'une queue simple brin soit à l'extrémité 3', soit en 5' afin que celles-ci puissent s'associer à leur substrat[262,263]. Elles sont impliquées dans de nombreux processus biologiques comme la réplication, la réparation, la transcription ou encore la maintenance des télomères[264,265]. C'est le cas par exemple des hélicases XPB et XPD qui peuvent être mobilisées au niveau de certaines séquences promotrices et plus précisément là où se forme des structures G4. Ces protéines vont ainsi dérouler les structures G4 permettant ainsi la réalisation de la transcription du gène concerné[67]. Si ces processus, dépendant d'hélicases, sont déficients, cela peut entrainer comme conséquence une réplication ou une transcription défectueuse mais également une mutagenèse, une sénescence cellulaire ou encore une cancérogenèse. Ainsi les hélicases, de par leurs actions sur l'ADN ou l'ARN et leurs interactions avec des protéines, permettent de préserver une homéostasie du génome. Par conséquent, l'étude et la compréhension des liens, des interactions entre ces différents partenaires sont devenus importants[262].

2. Présentation des édifices cibles

Dans le cadre de cette thèse, il a été établi une collaboration avec le laboratoire du Dr. Dennis Gomez à l'université de Toulouse (Institut de Pharmacologie et de Biologie Structurale, IPBS). A travers ce partenariat, notre laboratoire a pour objectif de synthétiser et fournir un panel d'édifices G4 à topologie contrôlée dans le but de pouvoir réaliser des études d'identification et de caractérisation de protéines en interactions avec des G4. Pour cela, un ensemble d'édifices G4 ADN et ARN dérivés de la séquence télomérique ont été choisis (**Figure 93**). On y retrouve trois édifices tétramoléculaires parallèles, de type ADN et ARN (**L**, **M** et **T**). Trois composés bimoléculaires antiparallèles présentant des boucles sont également visés (**N**, **O** et **Q**). Et enfin, on retrouve également des édifices non G4. Parmi ceux-ci, les composés **R** et **S** présentent la séquence télomérique mais ne possédent qu'une seule boucle et les édifices **P** et **U** sont aussi ciblés et serviront de contrôle lors de ces études d'interactions.

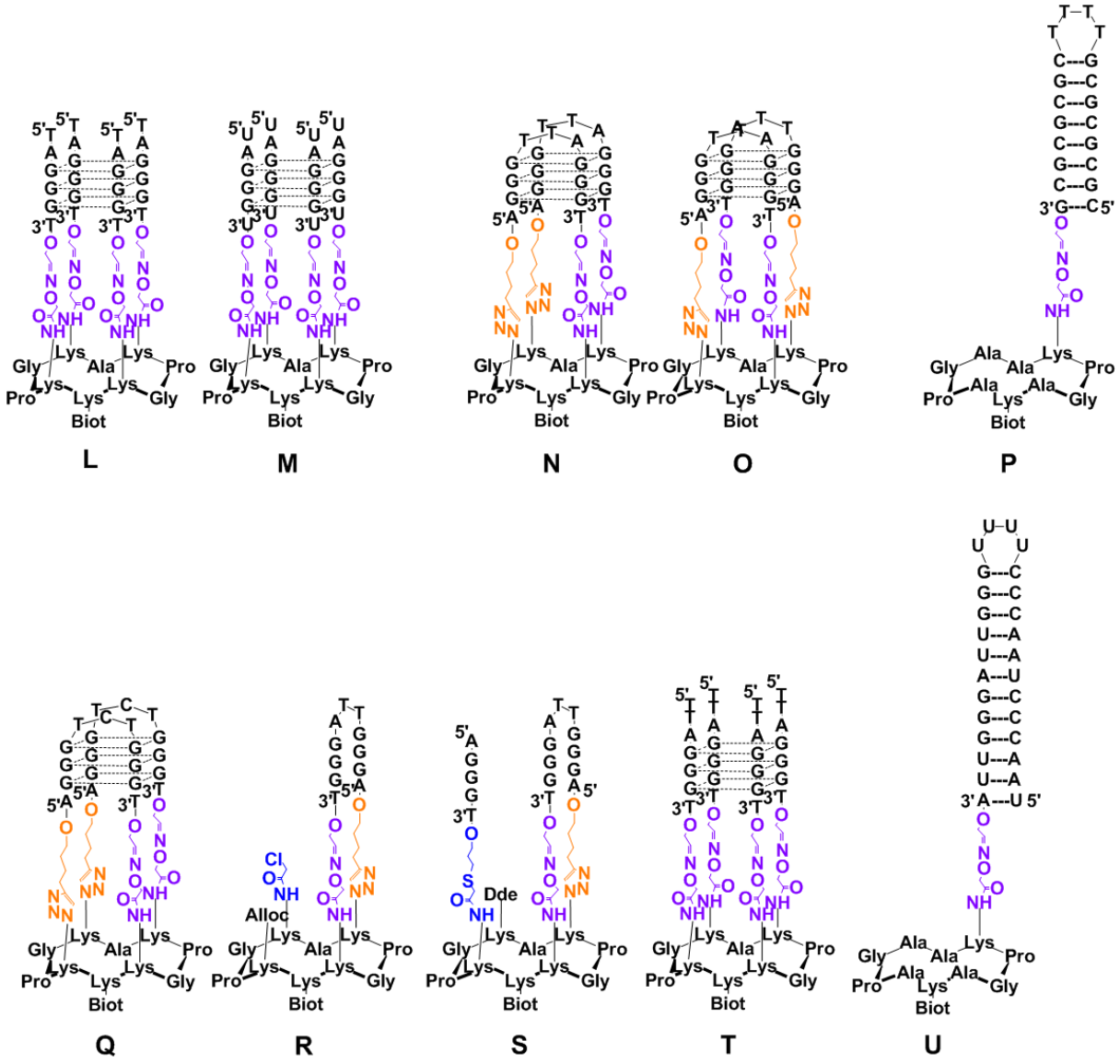


Figure 93 : représentation des édifices G4 et non G4 ADN et ARN utilisés pour les études d'interactions

3. Synthèse des édifices G4 et non G4

La synthèse de la plupart de ces édifices a été mise au point précédemment au sein de notre laboratoire[162,164,165] (édifices **L**, **M**, **N**, **O**, **P**, **R**, **S** et **T**), toutefois ces derniers ont été resynthétisés pour leurs utilisations dans le cadre de la collaboration avec l'IPBS mais la description de l'obtention de ces composés ne sera pas mentionnée. En ce qui concerne les édifices **Q** et **U**, ceux-ci sont nouveaux et la synthèse est décrite ci-après.

a) Synthèse de l'édifice Q

i. Synthèse du cyclodécapeptide **23**

La synthèse du cyclodécapeptide **23** (**Figure 94**) utilisé pour la synthèse de l'édifice **Q** a été mise au point auparavant au sein de notre laboratoire[171] ainsi la description détaillée de ce composé ne sera pas mentionnée.

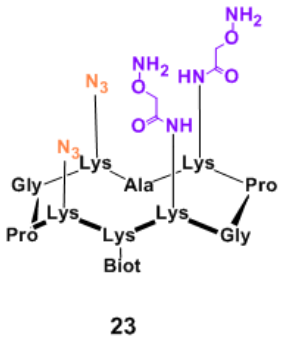


Figure 94 : représentation du cyclodécapeptide **23**

ii. Synthèse de l'oligonucléotide **24**

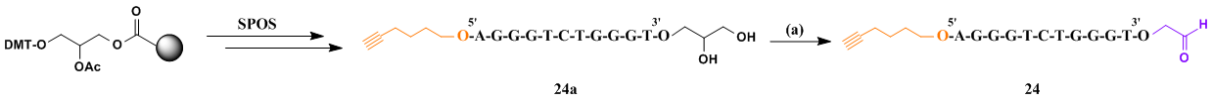


Schéma 36 : synthèse de l'oligonucléotide **24**
(a) $NaIO_4$ (30eq), 30 min, noir, t.a.

En ce qui concerne le partenaire oligonucléotidique mis en jeu **24**, celui-ci est bifonctionnalisé avec une fonction aldéhyde en position 3' et une fonction alcyne à l'extrémité 5' (**Schéma 36**). La stratégie de synthèse de ce composé s'effectue comme pour l'oligonucléotide **7** (page 56).

iii. Synthèse de l'édifice **Q** par ligation oxime et CuAAC

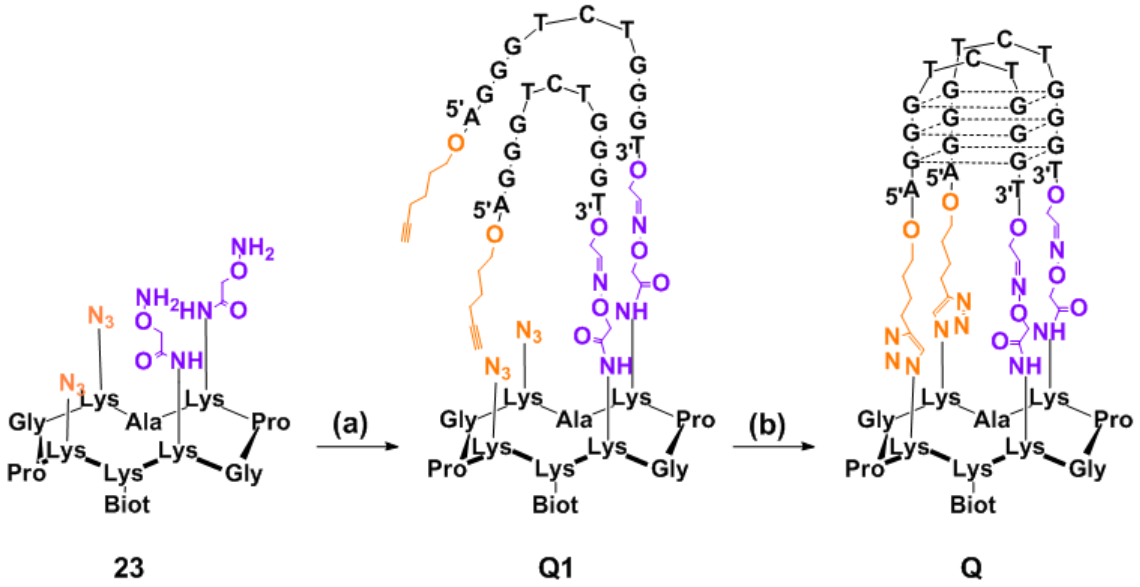


Schéma 37 : synthèse de l'édifice **Q**
(a) oligonucléotide **24** (2,4eq), tampon NH_4OAc 0,4M pH 4,5, t.a., 32%
(b) $CuSO_4$ (3eq), THPTA (30eq), ascorbate de sodium (30eq), 0,1M NaCl, DMF (10%vol), HEPES 100mM pH 7,4, 1h30, 37°C, 60%

La synthèse de l'édifice **Q** s'effectue en deux étapes (**Schéma 37**). La première consiste à réaliser la ligation oxime entre le cyclodécapeptide **23** et deux oligonucléotides **24**. Des conditions réactionnelles usuelles de ligation oxime ont été utilisées pour obtenir l'intermédiaire **Q1**. En fin de réaction, le brut réactionnel est purifié par RP-HPLC. Le composé **Q1** est obtenu avec un rendement de 32% et est caractérisé par RP-HPLC et spectrométrie de masse (**Figure 95**).

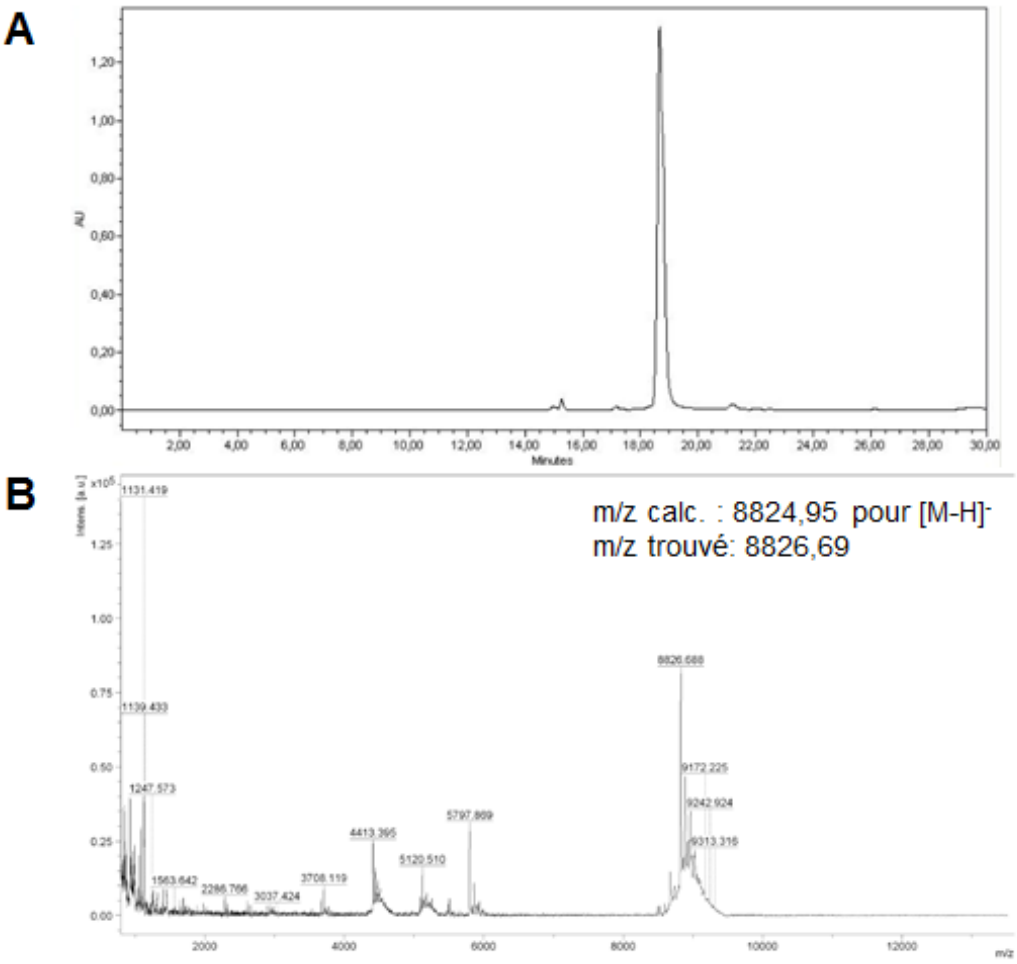


Figure 95 : (a) chromatogramme RP-HPLC et (b) analyse MALDI de l'édifice **Q1**

La seconde et dernière étape pour obtenir l'édifice **Q** est de réaliser la cycloaddition CuAAC sur l'intermédiaire **Q1**. Pour cela, des conditions usuelles de cycloaddtion CuAAC ont été utilisées. Toutefois, avant l'ajout des différents réactifs, le composé **Q1** est solubilisé dans le mélange tampon en présence de 0,1M de NaCl et est chauffé à 80°C pendant 5 minutes pour ensuite être laissé à température ambiante durant 2 heures. A noter également que cette cycloaddition s'effectue en milieu dilué afin de favoriser la réaction intramoléculaire. Ces aménagements permettent d'assurer la formation du produit antiparallèle **Q**[164]. En fin de réaction, le brut réactionnel est purifié par RP-HPLC. Le composé **Q** est obtenu avec un rendement de 60% et est caractérisé par RP-HPLC et spectrométrie de masse (**Figure 96**).

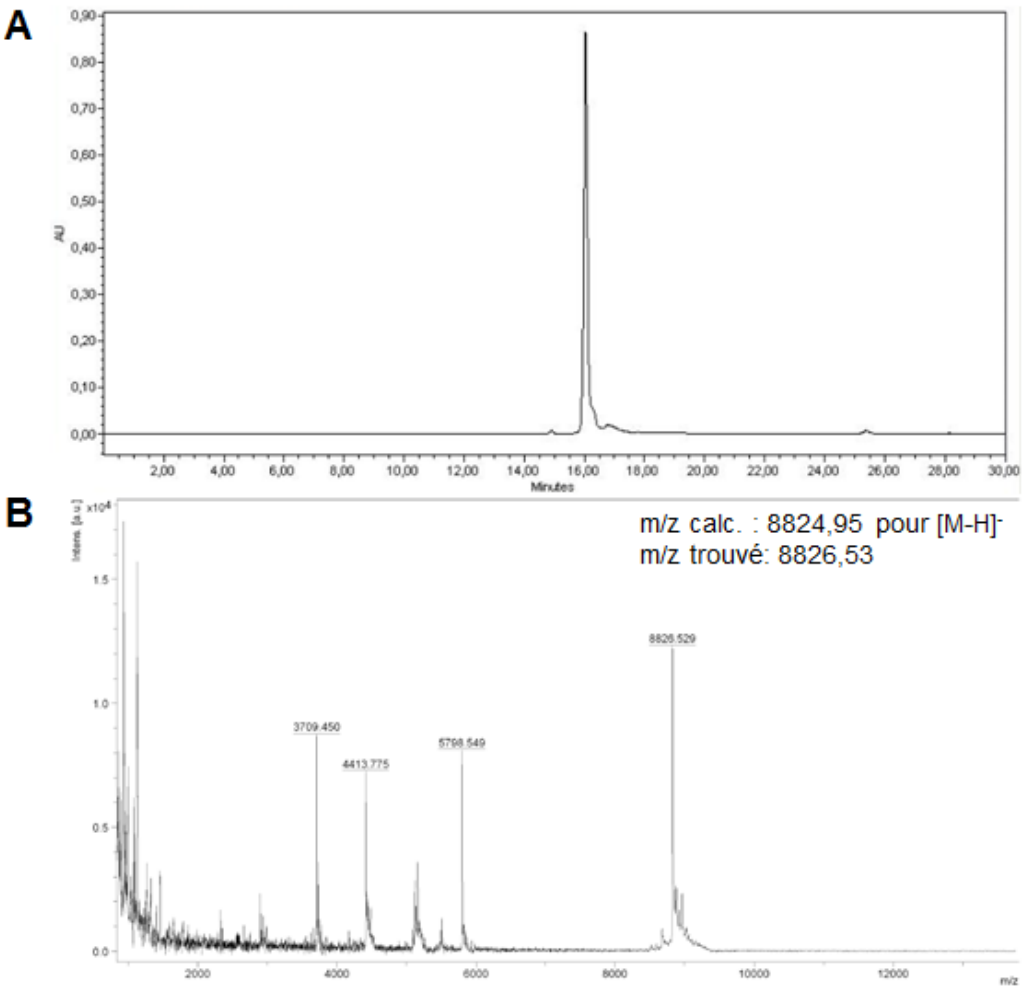


Figure 96 : (a) chromatogramme RP-HPLC et (b) analyse MALDI de l'édifice **Q**

Afin de s'assurer que l'édifice **Q** présente bien la topologie antiparallèle, une analyse par dichroïsme circulaire a été réalisée à température ambiante en présence de NaCl (**Figure 97a).** On observe les deux pics positifs caractéristiques à 242 nm et 294 nm et un pic négatif à 263 nm prouvant l'existence de la topologie antiparallèle. Une évaluation de la stabilité de la topologie a également été réalisée et une température de dénaturation de 72°C a été obtenue pour l'édifice **Q** (**Figure 97b**), valeur inférieure à celle répertoriée pour l'édifice **N** (81°C)[164].

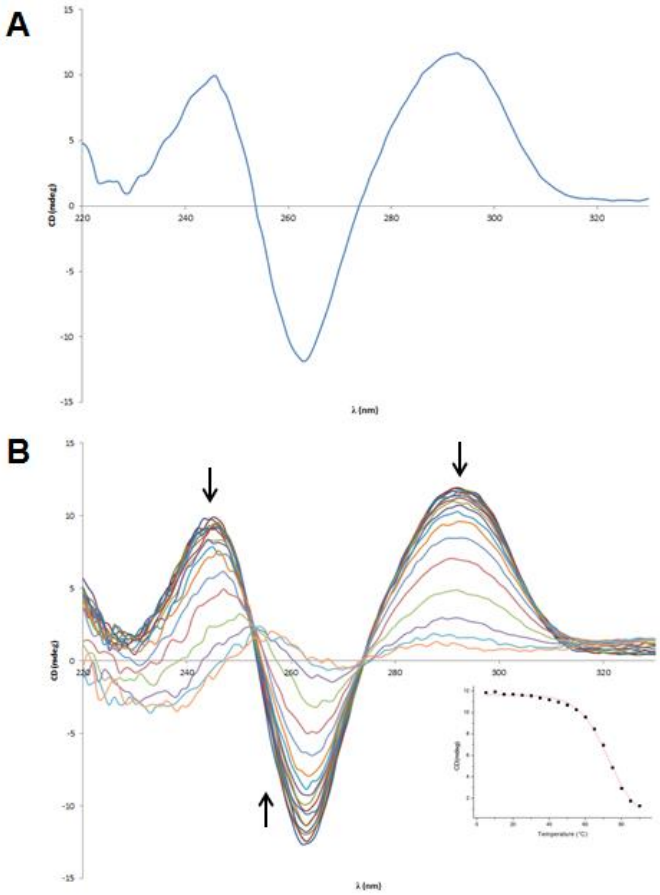
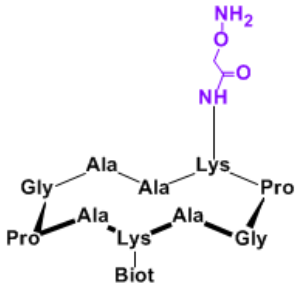


Figure 97 : (a) spectre CD de l'édifice **Q** à température ambiante (2µM de **Q** dans un tampon HEPES 10mM pH 7,4 avec 0,1M de NaCl) ; (b) spectre CD de dénaturation de l'édifice **Q** (2µM de **Q** dans un tampon HEPES 10mM pH 7,4 avec 0,1M de NaCl ; gamme de température : 5 à 90°C)

b) Synthèse de l'édifice U

i. Synthèse du cyclodécapeptide **25**

La synthèse du cyclodécapeptide **25**, portant une fonction aminooxy, (**Figure 98**) utilisé pour la synthèse de l'édifice **U** a été mise au point auparavant au sein de notre laboratoire[171] ainsi la description détaillée de ce composé ne sera pas mentionnée.



25

Figure 98 : représentation du cyclodécapeptide **25**

ii. <u>Synthèse de l'oligonucléotide **26**</u>

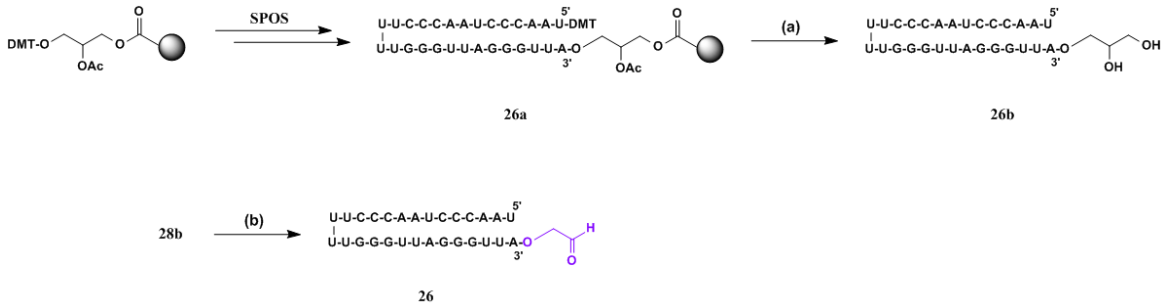


Schéma 38 : synthèse de l'oligonucléotide **26**
(a) 1) 1M DBU dans CH_3CN anhydre, 10min, t.a., 2) 28% NH_3 dans H_2O, 5h, t.a., 3) 80% AcOH dans H_2O, 20min, t.a. ; (b) $NaIO_4$, H_2O, 30min, t.a.

L'oligonucléotide **26** est obtenu au départ d'une résine 3'glyceryl CPG (**Schéma 38**). Après synthèse sur support solide, déprotection et élimination du support, l'oligonucléotide **26b** est purifié par RP-HPLC et est obtenu avec un rendement de 13%. Ce dernier est caractérisé par RP-HPLC et spectrométrie de masse (**Figure 99**). Comme précédemment, la fonction 3' aldéhyde de l'oligonucléotide **26** est obtenue par coupure oxydative en présence de périodate de sodium et cette réaction est considérée comme quantitative. A noter que la synthèse d'oligonucléotides d'ARN s'effectue également par SPOS, en utilisant la stratégie PivOM développée par l'équipe de F. Debart en 2008[266].

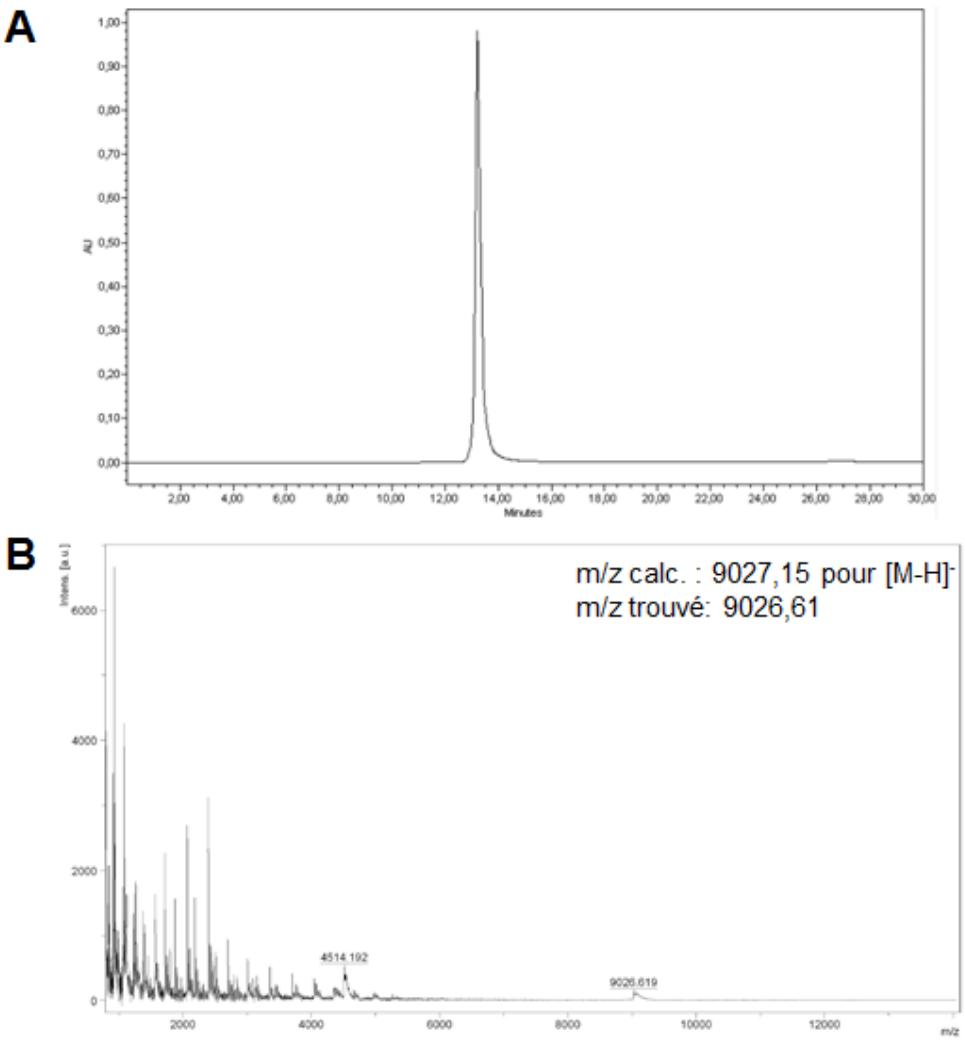


Figure 99 : (a) chromatogramme et (b) analyse MALDI de l'oligonucléotide **26b**

iii. Synthèse de l'édifice **U** par ligation oxime

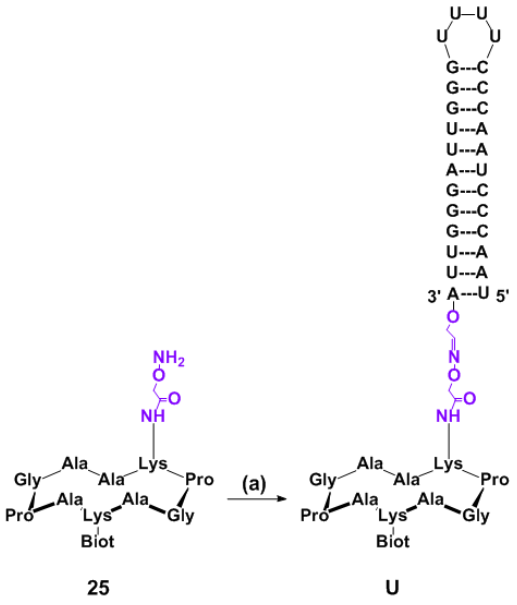


Schéma 39 : synthèse de l'édifice **U**
(a) oligonucléotide **26** (1eq), tampon NH_4OAc 0,4M pH 4,5, t.a., 13%

La synthèse de l'édifice **U** s'effectue entre le cyclodécapeptide **25** et l'oligonucléotide **26** dans un mélange tampon NH_4OAc 0,4M pH 4,5 à température ambiante (**Schéma 39**). La réaction est suivie par RP-HPLC et en fin de réaction, le brut est purifié par RP-HPLC. Le composé **U** est obtenu avec un rendement de 13% et est caractérisé par RP-HPLC et spectrométrie de masse (**Figure 100**).

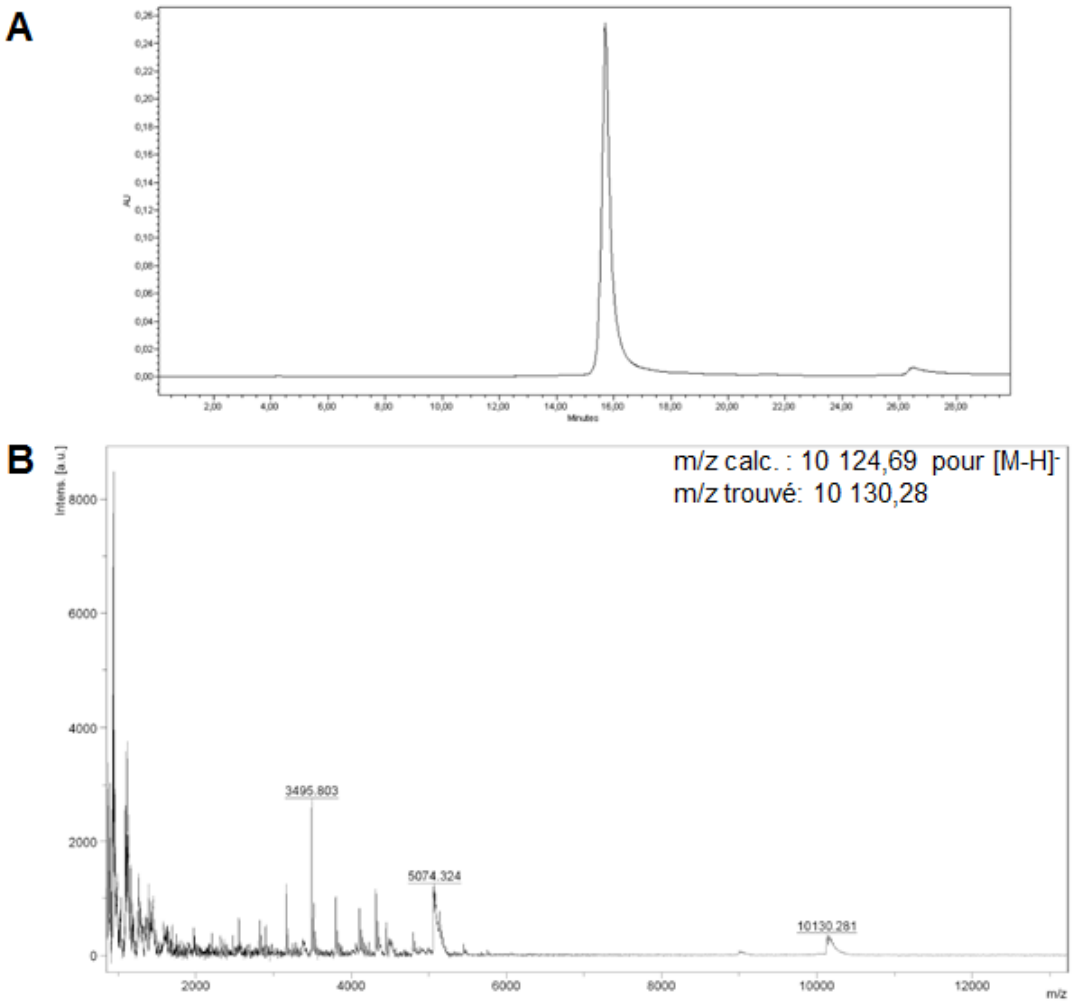


Figure 100 : (a) chromatogramme RP-HPLC et (b) analyse MALDI de l'édifice **U**

4. Etudes d'interactions G4/Protéines

L'objectif de cette étude est de pouvoir identifier des protéines en interactions avec des édifices G4 et non G4 (édifice contrôle) par le biais d'expériences de 'pêche à la protéine' (« pull-down » en anglais). Le principe de cette technique est illustré par le **Schéma 40**. Il est important de préciser que ces études ont été réalisées par A. Pipier, doctorante à l'IPBS et qu'il ne sera présenté que quelques résultats préliminaires illustrant ce travail.

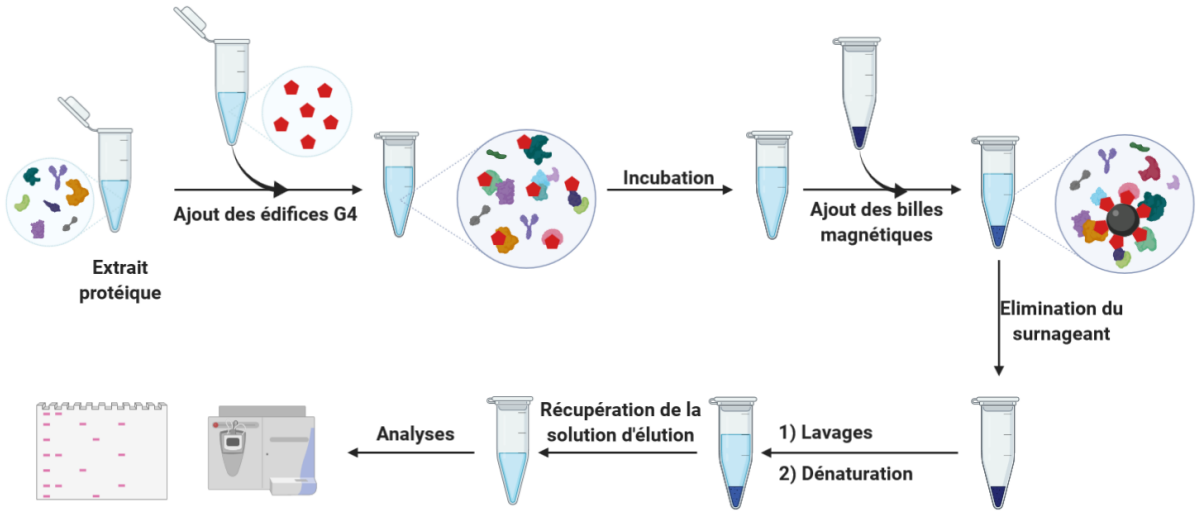


Schéma 40 : illustration de la méthode 'pêche à la protéine' mise en place

Au départ, un extrait protéique ainsi que le ou les édifice(s) G4 ou non G4 sont solubilisés dans une solution tampon HEPES 20mM pH 7,5. Après 1h30 d'incubation, des billes magnétiques recouvertes de streptavidine sont ensuite ajoutées au milieu afin de former un complexe billes magnétiques-édifices grâce à la forte interaction non-covalente entre les fonctions streptavidine et biotine (fonction présente sur la face inférieure de nos édifices). Le **surnageant** peut être ensuite éliminé et les complexes billes-édifices sont lavés plusieurs fois. Ensuite, une étape de dénaturation est réalisée dans le but de récupérer la solution d'**élution** qui contient les protéines qui se sont liées aux édifices. Cette étape s'effectue à l'aide d'un tampon Tris 60mM pH 8 à 95°C pendant 10 minutes. Enfin, les protéines présentes dans le **surnageant** et l'**élution** sont finalement identifiées et caractérisées par western blot et spectrométrie de masse.

Ces expériences de 'pêche à la protéine' ont été réalisés avec les édifices G4 parallèles (**L** et **M**), antiparallèles (**N** et **O**) et le contrôle **P**. De ces études, il en ressort différents points. Tout d'abord, un grand nombre de protéines ont été identifiées par spectrométrie de masse, un peu plus de 600 protéines. Il est d'ailleurs difficile de déterminer si les interactions observées pour ces protéines correspondent à de réelles interactions de type G4-protéines ou plutôt à des interactions *via* d'autres protéines. Pour limiter la présence d'artéfacts, plusieurs cycles de lavages ont été réalisés au cours de ces études. De plus, il est également possible qu'un phénomène de compétition puisse s'effectuer entre les différentes protéines présentes à différentes concentrations dans le milieu. Suite à une répétition des expériences et un affinage

des données, environ 350 protéines ont été identifiées et il a été observé que certaines d'entre elles se fixaient 'sélectivement' selon la configuration de l'édifice G4. En effet, certaines protéines interagissaient préférentiellement avec la topologie parallèle, d'autres avec la configuration antiparallèle et certaines avec les deux. La **Figure 101a** illustre ces résultats préliminaires. Par exemple, les protéines hnRNPA1 et CNBP interagissent préférentiellement avec les formes parallèles alors que les protéines eIF4G, WRN et Mre11 se fixent plutôt avec la configuration antiparallèle. La protéine DHX36 a une préférence pour l'édifice parallèle ARN et la protéine Ku70 quant à elle ne fixe que l'édifice contrôle.

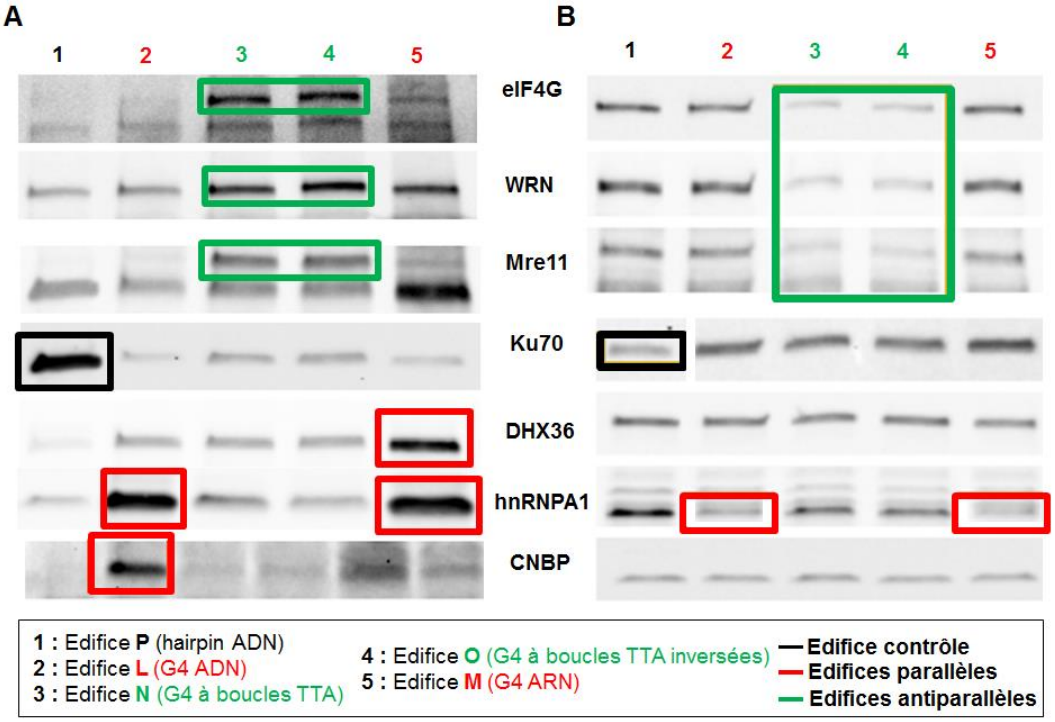


Figure 101 : analyses par western blot des protéines récupérées avec soit la solution d'**élution** (a), soit avec **le surnageant** (b)

La méthode utilisée semble être satisfaisante et ce pour plusieurs raisons. Tout d'abord, la mise en œuvre de ces expériences est relativement facile d'utilisation puisqu'elle permet de 'piéger' des protéines selon les édifices analysés. De plus, la quantité obtenue des protéines 'piégées' est suffisante pour réaliser des caractérisations par western blot et spectrométrie de masse. Les premières données récoltées sont rassurantes et encourageantes étant donné que les protéines identifiées sont connues dans la littérature pour interagir avec des structures G4. En effet, la protéine hnRNPA1 (pour « heterogeneous nuclear ribonucleoprotein A1 ») est connue pour se lier et déplier des structures G4 et être également impliquée dans la transcription[267,268]. Les protéines WRN (« Werner syndrome helicase ») et DHX36 (pour « DEAH box protein

36 ») font partie de la famille des hélicases. La première appartient au groupe d'hélicases RecQ et est impliquée notamment dans la préservation de l'intégrité de l'ADN pendant la réplication[262]. La protéine DHX36 fait partie de la famille des hélicases RHAU pour « RNA Helicase associated with AU-rich element ». Des études ont montrées qu'elle possède une grande affinité pour les structures G4[269] et est impliquée notamment dans le processus de traduction[262]. La protéine CNBP pour « Cellular Nucleic acid Binding Protein » est une protéine nucléaire-cytoplasmique qui se lie préférentiellement à des structures G4 ARN ou des séquences ADN simple brin et est impliquée dans les processus de traduction et transcription[270]. La protéine Mre11 interagit et s'associe avec d'autres protéines formant un complexe (le complexe MRX/N pour « Mre11, Rad50 et Xrs2/Nbs1 ») et ce dernier est impliqué dans différents évènements biologiques tels que la maintenance de la longueur des télomères[261]. La protéine eIF4G pour « eukaryotic translation Initiation Factor 4 G » quant à elle est impliquée dans l'initiation du processus de traduction eucaryote[271] et enfin la protéine Ku est un complexe de protéines impliqué dans la réparation d'ADN et se lie préférentiellement sur l'ADN double brin[272].

Le **surnageant** a également été caractérisé (**Figure 101b**) afin d'effectuer une vérification sur les résultats obtenus pour la solution d'**élution**. Il semble y avoir une concordance entre les deux expériences, c'est-à-dire que les protéines fixées par les différents édifices dans la solution d'**élution** se retrouvent en moindre mesure dans la **surnageant**, ce qui, encore une fois, est rassurant et encourageant sur la méthode utilisée.

Toutefois suite à ces données préliminaires, plusieurs interrogations ont été soulevées. La première concerne l'influence de la boucle TTA présente sur les édifices antiparallèles. Il a été observé que le sens de la boucle n'a pas réel importance puisque les résultats obtenus pour les édifices **N** et **O** sont similaires. Cependant, la question que l'on peut se poser est : « est-ce-que la nature de la boucle TTA a une influence ? ». Afin de confirmer ou infirmer cette hypothèse, il a été envisagé de synthétiser un édifice bimoléculaire antiparallèle présentant deux boucles TCT (composé **Q**) Autres interrogations, peut-il avoir une reconnaissance de ces protéines si une seule boucle est présente et s'il n'y pas de structure G-quadruplexe. L'utilisation des édifices **R** et **S** a été envisagée. Et enfin, une dernière question s'est posée concernant l'édifice parallèle ADN : « la présence d'une extension TTA en position 5' peut-elle avoir une influence sur l'interaction des protéines ? ». L'utilisation de l'édifice **T** au sein de ces études permettra de répondre à cette interrogation. A noter également qu'un édifice

double brin de type ARN (édifice **U**) a été synthétisé et pourra servir de contrôle ARN dans de futures études d'interactions.

A travers ces premiers travaux préliminaires, un autre point intéressant est à signaler. En effet, il a été observé que l'édifice parallèle **L** et antiparallèle **N** interagissent préférentiellement avec un composé multiprotéique appelée le complexe NELF pour « Negative ELongation Factor » (**Figure 102a**). Ce dernier est composé de quatre unités (NELF-A, B, C/D et E) et est connu pour être impliqué dans le processus de transcription et plus précisément dans ce que l'on appelle le « transcriptional pausing », c'est-à-dire la mise en pause de la transcription suite à l'interaction entre NELF et l'ARN polymérase II[273,274]. Lorsque le processus de transcription est en pause, il peut être observé la formation de « R-loop » que l'on peut définir comme étant une structure d'acides nucléiques à trois brins qui est composée d'un duplex hybride ADN/ARN et d'un simple brin ADN dit « déplacé »[253,275] (**Figure 102b**). La dénomination R-loop a été donnée en rapport à la structure D-loop que l'on peut apercevoir dans les télomères (voir introduction, page 17) et l'utilisation de la lettre « R » est utilisée pour désigner l'implication d'un brin ARN. La formation de ce genre de structure n'est pas exceptionnelle, elle peut être observée dans les régions promotrices mais également lors de différents évènements cellulaires[275,276]. Au sein de la R-loop si le simple brin déplacé d'ADN est riche en guanine, ce dernier peut potentiellement former des structures G-quadruplexes et sous l'action d'endonucléases, cela peut mener *in fine* à des coupures au sein du brin et provoquer une instabilité génomique[253]. Une récente étude a montré qu'au sein de cellules cancéreuse la stabilisation de structures G-quadruplexes par un ligand favorise la formation de R-loop et provoque des cassures double brin causant une instabilité du génome ou une mort cellulaire[277] (**Figure 102c**). Ainsi le fait d'observer une interaction G4-complexe NELF pourrait laisser supposer que ce complexe favorise/stabilise les structures G4 contribuant à l'augmentation de cassures double brins. Cela est encore purement hypothétique et est actuellement sous investigation.

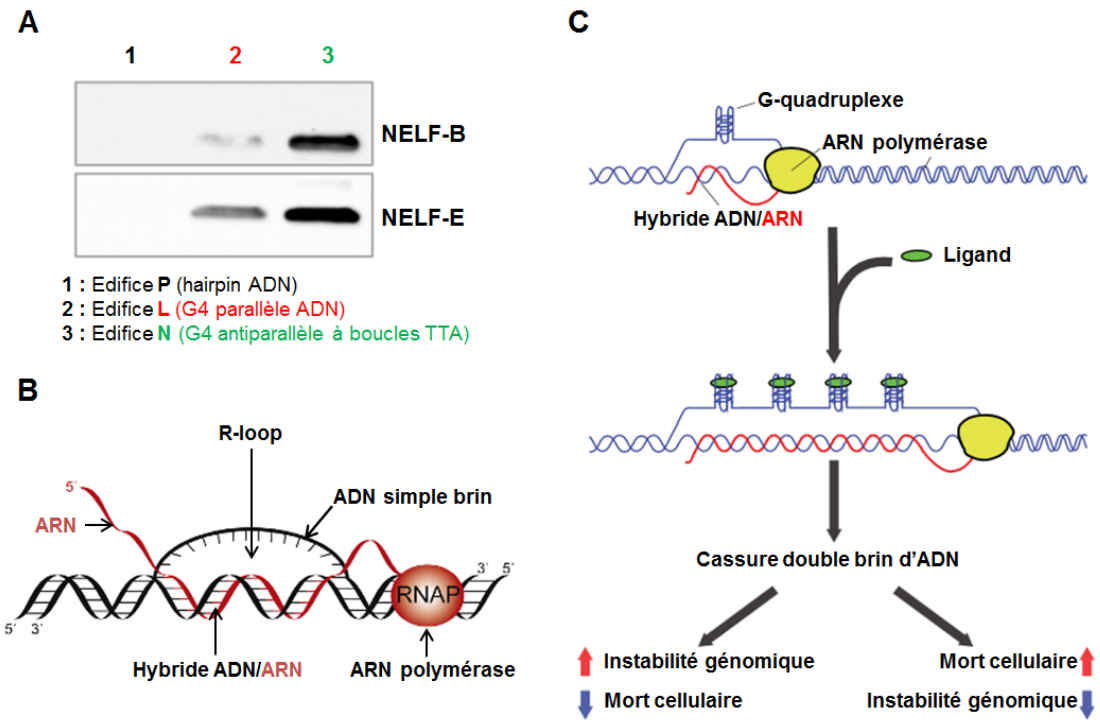


Figure 102 : (a) analyse par western blot des protéines NELFE-B et NELFE-E suite à l'expérience de 'pêche à la protéine' avec les édifices **P**, **L** et **N** ; (b) représentation schématique d'une structure R-loop. Figure issue et modifiée de Hamperl *et al. DNA Repair* **2014**, 19, 84–94 [253] ; (c) représentation schématique de l'impact de la stabilisation de structures G-quadruplexes par un ligand au sein de cellules cancéreuses. Figure issue et modifiée de De Magis *et al. Proc. Natl. Acad. Sci.* **2019**, 116 (3), 816–825 [277]

5. Bilan et perspectives

Dans ce troisième chapitre, les travaux se sont portés sur l'élaboration d'édifices G4 et non G4 de type ADN et ARN afin qu'ils soient utilisés dans le cadre d'études d'interactions avec des protéines cellulaires. Ces expériences d'identification et de caractérisation de protéines ont été initiées à l'aide d'une collaboration avec le laboratoire de D. Gomez à l'Institut de Pharmacologie et de Biologie Structurale (IPBS) à Toulouse et ont été réalisées par Angélique Pipier, doctorante de ce même institut.

Les premiers résultats qui ressortent de ces expériences de 'pêche à la protéine' se sont montrés satisfaisants. En effet, différentes protéines ont été identifiées pour avoir des interactions avec des édifices G4 parallèles ou antiparallèles. De plus, les protéines identifiées sont connues dans la littérature pour avoir un lien avec les G4 et sont impliquées dans différents processus biologiques. Les résultats préliminaires ont suscités différentes interrogations comme l'influence que peut avoir la nature et la présence de boucles sur l'édifice antiparallèle ou encore la présence d'une extension de bases TTA sur l'édifice tétramoléculaire parallèle. Ainsi

plusieurs composés ont été synthétisés et permettront de répondre à ces différentes questions. Dans le futur, d'autres édifices G4 parallèles, antiparallèles, voir hybrides peuvent être envisagés afin de compléter le panel de G4 déjà mis en place et permettront de voir quelles sont les caractéristiques qui influencent l'interaction G4-protéine. De plus, il a été observé que les unités B et C du complexe multiprotéique NELF interagissent avec les composés G4 parallèle et antiparallèle. Ces résultats sont intéressants puisqu'ils mettent en avant un possible lien direct entre G-quadruplexe, complexe NELF et cassure double brins bien que cela est encore sous investigation. En résumé, le système mis en place est relativement simple d'utilisation, il permet d'obtenir des quantités suffisantes de protéines pour réaliser l'identification et la caractérisation des protéines et enfin il peut être vu comme un outil permettant d'identifier si une protéine possède une affinité particulière pour une topologie bien spécifique et quelles sont les caractéristiques de cette topologie qui influence l'interaction G4-protéine.

Chapitre VI

Conclusions et perspectives

Depuis plusieurs années, les structures G-quadruplexes (G4) et i-motifs provenant par exemple de la séquence télomérique ou de séquences promotrices de gènes ont suscité un grand intérêt de par leurs implications dans de nombreux processus biologiques. L'un des domaines d'applications de ces structures est le ciblage spécifique par de petites molécules provoquant une modification de certains processus cellulaires. Toutefois en ce qui concerne les séquences d'acides nucléiques riches en guanines, celles-ci peuvent adopter une multitude de topologies en solution. De plus, la plupart de ces ligands ne montrent pas une spécificité suffisante pour une topologie bien particulière de G4. Ainsi dans le but de pallier à ces difficultés, notre laboratoire a développé ces dernières années un système permettant d'obtenir des édifices tétraplexes d'acides nucléiques avec une topologie bien définie. Le concept TASQ consiste à utiliser un support peptidique rigide permettant de greffer successivement les séquences d'acides nucléiques afin d'obtenir une conformation bien particulière. Le greffage de ces séquences sur le support peptidique s'effectue à l'aide de ligations chimiosélectives couramment utilisées dans notre laboratoire. L'utilisation de plusieurs ligations orthogonales permet de diversifier mais également d'envisager l'obtention d'édifices tétraplexes plus complexes. Dans ce contexte, l'un des objectifs de cette thèse a été de développer de nouveaux mimes d'ADN tétraplexes contraints et d'implémenter une quatrième ligation chimiosélective, orthogonale aux trois autres déjà mises en place au sein de notre laboratoire.

Suite à une introduction et une présentation du projet de thèse dans les deux premiers chapitres de ce manuscrit, le troisième chapitre décrit la synthèse d'un mime i-motif bimoléculaire antiparallèle dérivé de la séquence télomérique, l'édifice **F** (**Figure 103a**). La synthèse de celui-ci implique l'utilisation de trois ligations chimiosélectives orthogonales et où l'une d'entre elles (la cycloaddition CuAAC) a été réutilisée une seconde fois. La première séquence d'ADN bifonctionnalisée a été greffée en réalisant respectivement le couplage thiol-chloroacétamide et ensuite la première cycloaddition CuAAC. Afin de pouvoir réutiliser cette dernière, une modification sur le châssis peptidique a été réalisée. Effectivement, une fonction azoture a été générée à la surface supérieure du support à partir d'une fonction amine par le biais d'une réaction de diazo-transfert. A partir de là, il a été observé qu'une seule stratégie de synthèse permettait l'obtention du mime i-motif **F**. En effet, la voie faisant appel à la seconde cycloaddition CuAAC dans un premier temps et à la ligation oxime dans un second temps s'est montrée concluante. Lors de cette dernière réaction, un changement de temps de rétention et de masse a permis de confirmer la formation du composé final. De plus, des études de dichroïsme circulaire ont été effectuées afin de caractériser la topologie du mime i-motif **F**. Celles-ci ont

permis de confirmer que ce dernier adopte une structuration i-motif et de plus, cela sur une gamme de pH plus importante que la séquence télomérique naturelle. Des températures de dénaturation de 39,5°C et 31°C ont été obtenues respectivement, à pH 6 et 7. Et enfin, il a été confirmé que la topologie i-motif est de type intramoléculaire et non intermoléculaire. En ce qui concerne les perspectives concernant ce mime, il aurait été intéressant de caractériser l'édifice **F** par 1H-RMN afin de fournir une nouvelle preuve de l'existence de la structure i-motif de ce composé. De plus, ce dernier représente une cible pertinente et attrayante dans le cadre d'études d'interactions avec des ligands ou des protéines cellulaires.

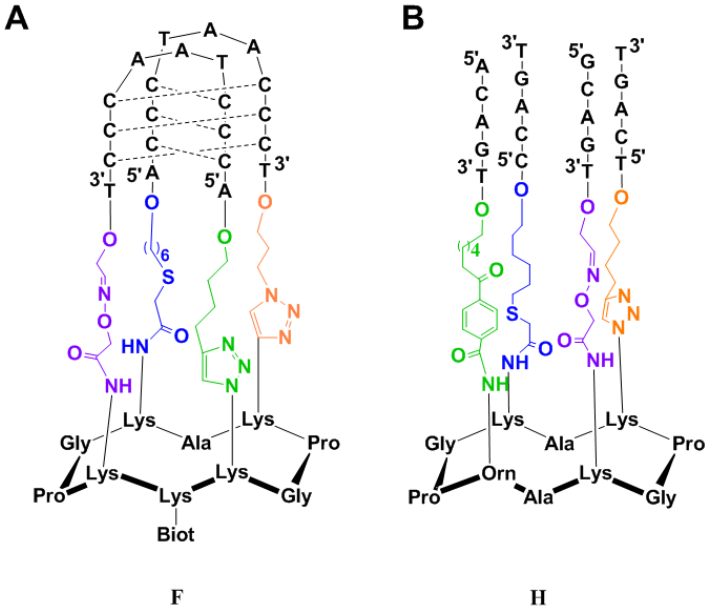


Figure 103 : représentation du mime i-motif **F** (a) et de l'édifice **H** (b)

Le quatrième chapitre décrit l'incorporation et l'utilisation d'une nouvelle ligation chimiosélective dîtes KAT dans la chimie des oligonucléotides. Au cours de cette partie, il a été décrit la méthodologie de cette nouvelle ligation pour les oligonucléotides mais également l'implémentation de celle-ci avec les trois autres chimies couramment utilisées dans notre laboratoire. La transposition de cette ligation à la synthèse de conjugués d'oligonucléotides est, à notre connaissance, le premier exemple de composé faisant intervenir cette ligation entre un partenaire oligonucléotidique et peptidique. L'incorporation de la ligation KAT aux trois autres chimies couramment utilisées dans notre laboratoire a présenté des difficultés. En effet, il a été observé une interférence entre les fonctions oxyamine et O-diéthylcarbamoyle. Toutefois, un traitement de post-déprotection adapté a été mis au point et a permis de conserver de manière intacte ces deux fonctions pour qu'elles soient utilisées dans leurs ligations respectives. Un édifice 'preuve de concept' (composé **H**) faisant appel à quatre ligations différentes a pu être

synthétisé (**Figure 103b**). Différentes voies de synthèse ont été investiguées mais une seule s'est montrée concluante. Au cours de cette investigation, différentes problématiques ont été rencontrées comme la réduction du groupement azoture dû à la présence de TCEP dans le milieu réactionnel, la fragilité du groupement O-diéthylcarbamoyle mais également des temps de rétention proches des différents composés rendant les purifications difficiles. Concernant les perspectives liées à l'utilisation de quatre ligations, réaliser certaines ligations en « one-pot » pourrait être envisagé comme par exemple effectuer la ligation oxime dans un premier temps et ensuite faire la ligation KAT étant donné que ces deux chimies peuvent se dérouler dans le même milieu réactionnel.

D'autre part, il serait également intéressant d'entreprendre la synthèse d'édifices tétraplexes avec ces quatre ligations orthogonales. En effet, cela permettrait d'obtenir un panel plus diversifié d'édifices tétraplexes possédant des topologies bien définies comme par exemple des conformations hybrides de G4 télomérique. Ces édifices pourraient ainsi être utilisés dans le cadre d'études d'interactions (études d'affinité et sélectivité de ligands de G4 ou encore des études d'interaction entre G4 et protéines). Par exemple, la synthèse de l'édifice **K**, mentionné au sein du chapitre IV, pourrait être ainsi envisagée. La **Figure 104** illustre un plan de synthèse envisageable pour l'obtention de ce dernier. La stratégie de synthèse proposée a été établie selon les informations obtenues au cours du chapitre IV. En effet, la première ligation à réaliser serait la ligation oxime. Ensuite, en raison des différents problèmes rencontrés il serait plus judicieux d'effectuer le couplage thiol-chloroacétamide en seconde position. Ensuite, le troisième couplage à réaliser serait la ligation KAT et ce, compte tenu de la fragilité du groupement O-diéthylcarbamoyle. Enfin, la dernière étape serait de réaliser la fermeture de la boucle par la cycloaddition CuAAC. Bien entendu, la voie de synthèse proposée n'est qu'un plan théorique, différentes difficultés peuvent être rencontrées en cours de synthèse provoquant un changement en ce qui concerne l'ordre des ligations et/ou un changement des fonctions chimiques présentes sur les partenaires oligonucléotidiques. De plus, il est important de signaler que selon le plan de stratégie proposé, la dernière ligation est la cycloaddition CuAAC. Ainsi, aucun changement de masse ou de charge ne pourra être observé lors de cette étape, il faut donc espérer qu'un changement de temps de rétention ait lieu lors de la formation du produit final.

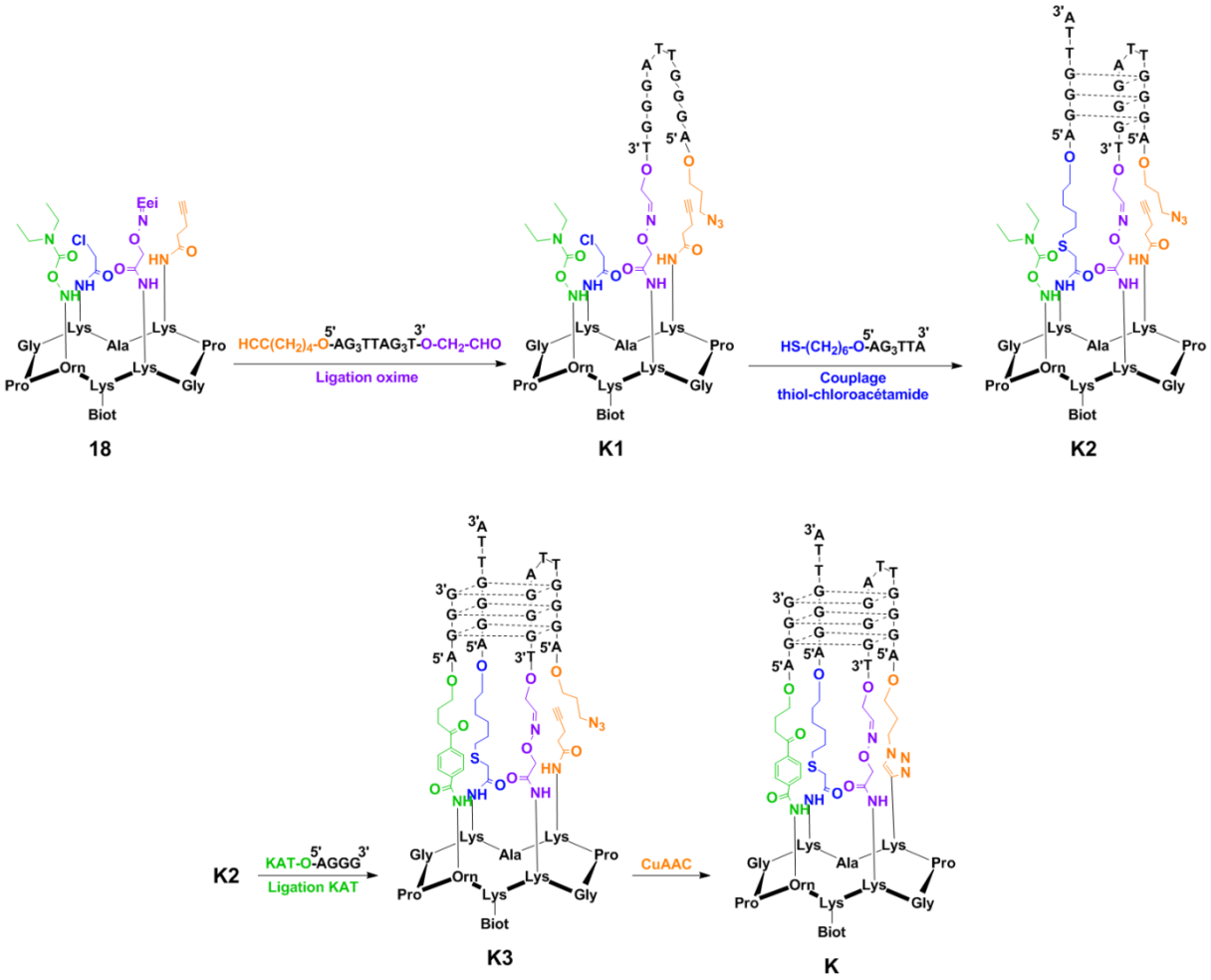


Figure 104 : Proposition d'une voie de synthèse pour l'édifice **K**

Le cinquième chapitre expose les résultats préliminaires d'études d'interactions entre des mimes G-quadruplexes et des protéines cellulaires. Ces résultats proviennent d'une collaboration établie avec le Dr. D. Gomez de l'université de Toulouse (Institut de Pharmacologie et de Biologie Structurale, IPBS). Ces expériences se sont montrées intéressantes puisque de multiples protéines ont été identifiées pour interagir avec des édifices G4 parallèles ou antiparallèles. De plus, on retrouve dans la littérature que ces protéines possèdent un lien avec les G4 et sont impliquées dans différents processus biologiques. Pareillement à ces résultats, il a été observé que des G4 parallèle et antiparallèle interagissent avec les unités B et C du complexe NELF, mettant en avant un possible lien direct entre G-quadruplexe, complexe NELF et cassure double brins. Les résultats préliminaires obtenus au cours de ces études ont soulevés différentes interrogations comme par exemple l'influence que peut avoir la nature et la présence de boucles sur l'édifice antiparallèle. A l'avenir, il peut être envisagé la synthèse d'autres édifices G4 parallèles, antiparallèles et hybrides afin de

complémenter le panel de G4 déjà mis en place ce qui permettrait de voir quelles sont les caractéristiques qui influencent l'interaction entre G4 et protéine.

Chapitre VII

Partie expérimentale

1. Materials and equipment

Protected amino acids were obtained from Activotec (Cambridge, UK), Bachem Biochimie SARL (Voisins-les-Bretonneux, France), France Biochem SA (Meudon, France), Merck Eurolab (Fontenay-sous-Bois, France) or Calbiochem-Novabiochem (Merck Biosciences - VWR, Limonest, France). PyBOP® was purchased from Calbiochem-Novabiochem. 2-cyanoethyl phosphoramidites and CPG resins were purchased from BioSolve BV (Valkenswaard, Netherlands), Eurogentec (Angers, France), Link Technologies (Glasgow, Scotland) or Chem Genes Corporation (Wilmington, USA). Trizma base (tris) and all the reagents for acrylamide gel preparation were from Sigma Aldrich (St Quentin Fallavier, France.

Other reagents were obtained from Aldrich (Saint-Quentin Fallavier, France), Acros (Noisy-le-Grand, France) and Eurogentec (Angers, France). All solvents and reagents used were of highest purity commercially available.

ESI mass spectra were performed on an Esquire 3000 spectrometer from Bruker or on an Acquity UPLC-MS system from Waters equipped with a SQ Detector 2. NMR spectra were obtained with an Avance III Bruker spectrometer. MALDI-Tof mass spectra were performed on an Autoflex Bruker using hydropiccolinic acid (HPA, 45 mg; ammonium citrate 4 mg in 500 µL H_2O/CH_3CN) as matrix. Peptides were analysed in positive mode and oligonucleotides and conjugates in negative mode. Water was of Milli-Q quality and was obtained after filtration of distilled water through a Milli-Q® cartridge system. CH_3CN and TFA were HPLC quality.

2. Peptides synthesis

The synthesis was performed on a Syro II synthesizer using Fmoc/tBu strategy on a 2-Chlorotrityl resin. The course of reactions were monitored by using UPLC-MS system Waters, it includes reverse phase chromatography using Nucleosil C18 column (130 Å, 2.1 x 50 mm, 1.7 µm) and detection by UV at 214 nm and 250 nm and by electron spray ionization mass spectrometry. A 0.6 mL/min flow linear gradient from 95% solvent A (0.1% citric acid in water) and 5% solvent B (0.1% citric acid in acetonitrile) to 100% B for 4 minutes was applied. RP-HPLC purifications were performed on a Gilson or Waters system with Nucleosil C18 column (100 Å, 250 x 21 mm, 7 µm) with UV monitoring at 214 nm and 250 nm. A 22 mL/min flow linear gradient was applied from 95% solvent A (0.1% citric acid in water) and 5% solvent B (0.1% citric acid in acetonitrile) to 100% B1 for 15 minutes.

2.1. General procedure for linear protected peptides synthesis by SPPS

All peptide were synthesis using commercially available 2-Chlorotrityl resin (loading of 0.78 mmol/g). Fmoc-Gly-OH (3 eq) was coupled on the resin in anhydrous DCM in a glass reaction vessel fitted with a sintered glass. pH 8 was adjusted using DIEA. The mixture was stirred for 1h at room temperature. The resin was washed using a DCM/MeOH/DIEA solution (17/2/1, v/v/v). Fmoc protecting group was removed using three washes with 20% piperidine in DMF (40 mL). The resin loading was monitored by quantification of free dibenzofulvene using UV absorbance at 299 nm (loading of 0.35 mmol/g). The elongation was performed on a Syro II using Fmoc/tBu strategy on the above-prepared resin. Fmoc-Ala-OH, Fmoc-Gly-OH, Fmoc-Lys(Boc)-OH, Fmoc-Lys(Dde)-OH, Fmoc-Pro-OH, and Fmoc-Lys(Alloc)-OH were commercially available. Fmoc-Lys(Biotine)-OH, Fmoc-azidononorleucine, Fmoc-Lys(-CO-CH2-O-N=Eei)-OH, Fmoc-Lys(pentynoyl)-OH and O-carbamoylhydroxylamine ornithine derivative [(S)-5-[N-(tert-butoxycarbonyl)-N-(diethylcarbamoyloxy)amino]-2-[N-(9-fluorenylmethoxy-carbonyl)amino]pentanoic acid] were obtained using reported protocols[172–174,249]. The synthesizer uses amino acid at 4 eq, DIEA and HBTU as activator. Supported peptide was cleaved from the resin using a DCM/TFE/AcOH solution (70/20/10, v/v/v). The solution was evaporated under vacuum and the peptides were precipitated in ether as a white powder.

2.2.General procedure for cyclisation reaction

Linear peptides were dissolved in DMF to reach a 10^{-3} M concentration and 1.2 eq of PyBOP was added. The pH was adjusted to 8-9 using DIEA and the solution was stirred at room temperature until the complete peptide cyclisation (UPLC monitoring). The solvent was evaporated under vacuum then the crude peptides were precipitated in ether. The crude products were purified on RP-HPLC and freeze-dried to obtain a white powder.

2.3. General procedure for N-ε Boc or oxyamine deprotection

Peptides were dissolved in TFA/TIS/H_2O (80/16/4, v/v/v) solution and the mixture was stirred for 1h at room temperature. The solvent was then evaporated under vacuum and the crude peptides were precipitated in ether as a white powder.

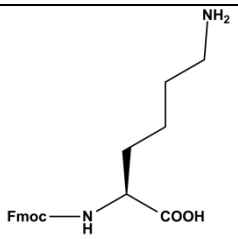
2.4. General procedure for N- ε Dde lysine deprotection

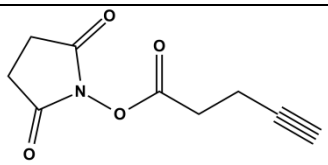
Peptides were dissolved in a 2% hydrazine solution in DMF. The solution was stirred at room temperature until the complete deprotection (UPLC monitoring). The solvent was evaporated under vacuum then the crude peptides were precipitated in ether as a white powder.

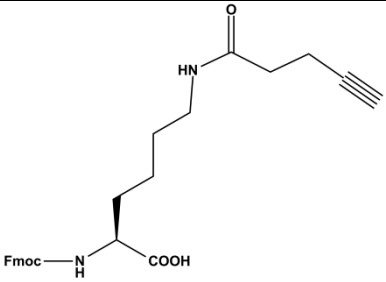
2.5. General procedure for coupling chloroacetamide linker on peptide

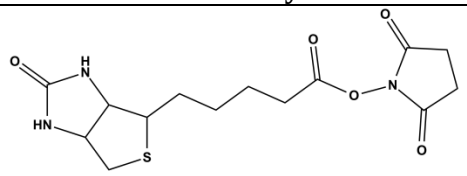
Peptides were dissolved in DMF to reach a 3.10^{-3} M concentration. The pH was adjusted to 8-9 using DIEA and chloroacetic anhydride (2 eq) was added. The reaction was stirred for 1h30 at room temperature until the complete coupling (UPLC monitoring). The solvent was evaporated under vacuum then the crude peptides were precipitated in ether. The crude products were purified on RP-HPLC and freeze-dried to obtain a white powder.

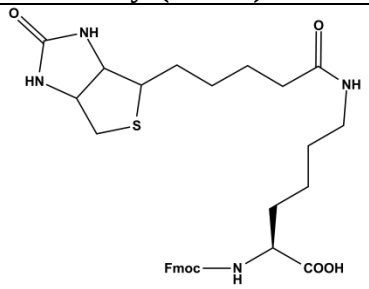
2.6.Experimental procedure

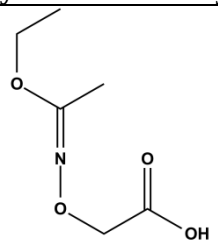
Fmoc-Lys-OH 2
 2
Fmoc-Lys(Boc)-OH (1 eq, 10.67 mmol, 5 g) was dissolved in HCl (37%)/Dioxane solution (1/2, v/v). The reaction was stirred at room temperature for 1h. Afterwards the solvent was evaporated under vacuum to obtain yellow oil. Then the product was dissolved in a H_2O/CH_3CN solution (6/4, v/v) and freeze-dried to obtain a white powder (9.81mmol, 3.61 g, yield: 92%). **ESI-MS ($^+$):** *m/z* calcd for $C_{21}H_{22}N_4O_4Na$: 368.43, *m/z* found: 368.54 $[M+H]^+$.

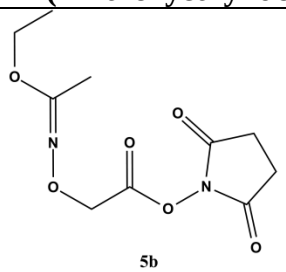
4-Pent-AcA-OSu 3a
 3a
3a was synthesized according to the published procedure described by Galibert *et al.*[172]

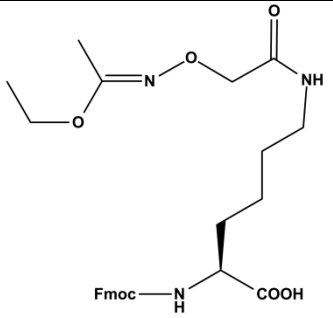
Fmoc-Lys(pentynoyl)-OH 3
 3
3 was synthesized according to the published procedure described by Galibert *et al.*[172]

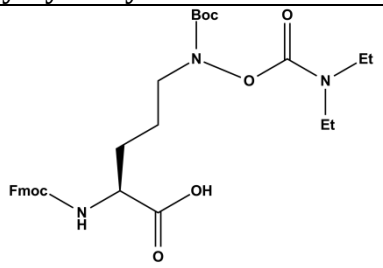
Biotin succinimidyl ester 4a
 4a
4a was synthesized according to the published procedure described by Youn *et al.*[173]

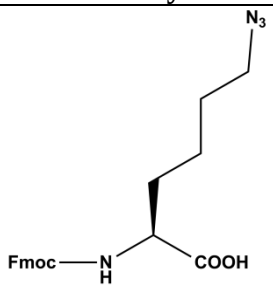
Fmoc-Lys(Biotin)-OH 4
 4
4 was synthesized according to the published procedure described by Youn *et al.*[173]

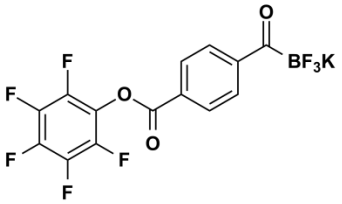
2-(1-Ethoxyethylideneaminooxy)acetic acid 5a
 5a
5a was synthesized according to the published procedure described by Foillard *et al.*[174]

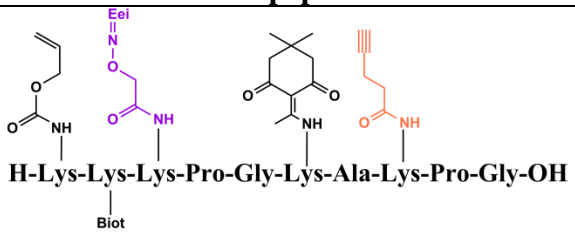
N-hydroxysuccunimidyl 2-(1-Ethoxyethylideneaminooxy)acetate 5b
 5b
5b was synthesized according to the published procedure described by Foillard *et al.*[174]

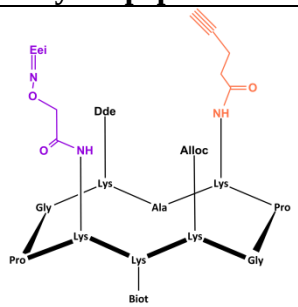
Fmoc-Lys(Eei-Aoa)-OH 5
 **5**
5 was synthesized according to the published procedure described by Foillard *et al.*[174]

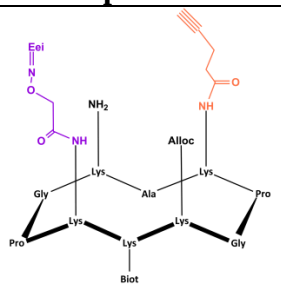
O-carbamoylhydroxylamine ornithine derivative 9
 **9**
9 was synthesized according to the published procedure described by Noda *et al.*[249] **^{1}H NMR** (400 MHz, CD_3OD): 7.79 (d, J = 7.6 Hz, 2H), 7.67 (dd, J = 7.2, 7.2 Hz, 2H), 7.38 (dd, J = 7.6, 7.6 Hz, 2H), 7.30 (dd, J = 7.2, 7.6 Hz, 2H), 4.38-4.31 (m, 2H), 4.23-4.17 (m, 2H), 3.64 (brs, 1H), 3.37-3.32 (m, 4H), 2.02-1.91 (m, 1H), 1.77-1.67 (m, 3H), 1.46 (s, 9H), 1.16 (brs, 6H) **^{13}C NMR** (100 MHz, CD_3OD): 175.6, 158.7, 156.6, 155.9, 145.2, 142.6, 128.8, 128.2, 126.3, 120.9, 83.4, 68, 54.9, 50.8, 48.4, 44.2, 42.9, 29.8, 28.5, 24.8, 14.5, 13.6 **ESI-MS:** *m/z* calcd for $C_{30}H_{39}N_3O_8$: 592.26, *m/z* found: 592.26 $[M+Na]^+$

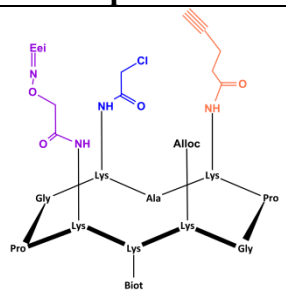
Fmoc-azidolysine 19
 **19**
19 was synthesized according to the published procedure described by Byrne *et al.*[251]

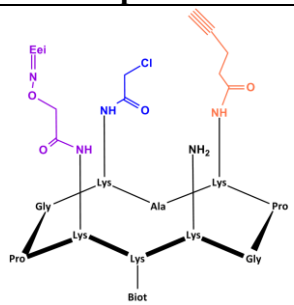
Potassium 4-(perfluorophénylcarboxy)benzoyltrifluoroborate 12
 **12**
12 was synthesized according to the published procedure described by Liu *et al.*[250] **^{1}H NMR** (400 MHz, acetone–d_6): δ 8.25 (d, J = 2.3 Hz, 4H) **^{13}C NMR** (100 MHz, acetone–d_6): δ 235.0 (br), 163.2, 146.6, 143.1 ({^{19}F}), 141.4 ({^{19}F}), 139.7 ({^{19}F}), 138.1 ({^{19}F}), 131.2, 129.5, 129.1 **^{19}F NMR** (376 MHz, acetone–d_6):δ −145.3, −155.0, −160.3, −164.8 **^{11}B NMR** (128.4 MHz, acetone–d_6): δ −0.90 **ESI-MS** : *m/z* calcd for $C_{14}H_4BF_8O_3$: 383.01, *m/z* found: 383.01 $[M-K]^-$

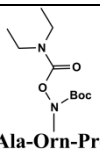
Linear peptide 1c
 **1c**
The linear peptide **1c** was synthesizer using general procedure **2.1** and modified amino acid **3**,**4** and **5**. The crude product was used without purification. Rt = 1.51 min **ESI-MS (+):** *m/z* calcd for $C_{82}H_{130}N_{18}O_{21}S$: 1734.93, *m/z* found: 1734.99 $[M+H]^+$

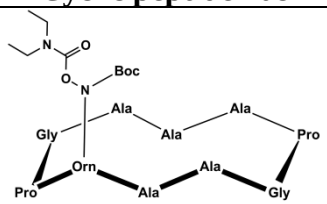
Cyclic peptide 1d
 **1d**
The linear peptide **1d** (0.129 mmol, 224 mg) was cyclized using general procedure **2.2**. The crude product was purified on RP-HPLC and freeze-dried to obtain a white powder (0.069 mmol, 117 mg, yield: 53% for two steps). Rt = 1.8 min **ESI-MS (+):** m/z calcd for $C_{82}H_{128}N_{18}O_{20}S$: 1716.92, m/z found: 1716.97 $[M+H]^+$.

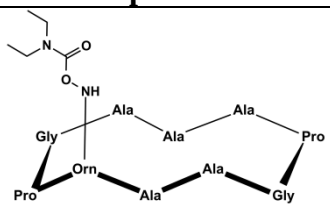
Peptide 1e
 **1e**
The peptide **1e** (0.069 mmol, 117 mg) was dissolved in DMF (10 mL). Allylic alcohol (5eq, 0.345 mmol) and 2%vol hydrazine were added. The solution was stirred at room temperature until the complete deprotection (UPLC monitoring). The solvent was evaporated under vacuum then the crude peptide was precipitated in ether as a white powder. The crude product was used without any further purification.

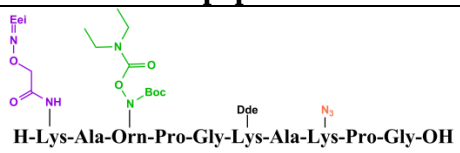
Peptide 1f
 **1f**
The peptide **1f** was obtained using general procedure **2.5**. and peptide **1e**. The crude product was purified on RP-HPLC and freeze-dried to obtain a white powder (0.034 mmol, 55 mg, yield: 49% for two steps). Rt = 1.71 min **ESI-MS (+):** m/z calcd for $C_{74}H_{118}ClN_{18}O_{19}S$: 1629.34, m/z found: 1629.95 $[M+H]^+$.

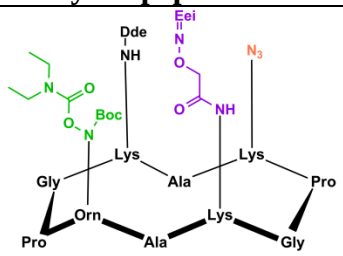
Peptide 1
 **1**
The peptide **1f** (25 mg, 15.3 µmol, 1 eq) was dissolved in anhydrous CH_2Cl_2/ anhydrous DMF (3/1, v/v). $Pd[P(Ph)_3]_4$ (0.2 eq, 3 µmol, 3.47 mg), $PhSiH_3$ (100 eq, 1.53 mmol, 188 µL) are added. The reaction was stirred at room temperature and quenching with 500 µL of MeOH. The solution is evaporated under vacuum and finally purified on RP-HPLC (7.1 µmol, 11 mg, yield: 44 %). Rt = 1.34 min **ESI-MS (+):** m/z calcd for $C_{70}H_{115}ClN_{18}O_{17}S$: 1546.27, m/z found: 1546.93 $[M+H]^+$.

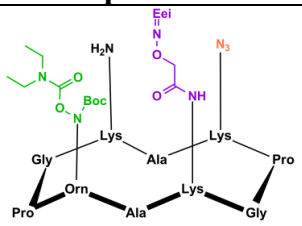
Linear peptide 10b
 **H-Ala-Ala-Orn-Pro-Gly-Ala-Ala-Ala-Pro-Gly-OH** **10b**
The linear peptide **10b** was synthesized using general procedure **2.1** and modified amino acid **9**. The crude product was used without purification.

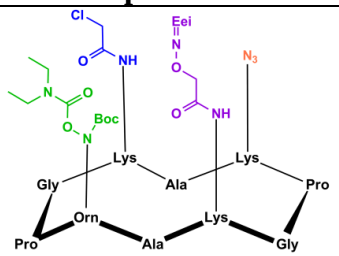
Cyclic peptide 10c
 **10c**
The linear peptide **10b** (1 eq, 0.198 mmol, 200 mg) was cyclized using general procedure **2.2**. The crude product was purified on RP-HPLC and freeze-dried to obtain a white powder (0.181 mmol, 140 mg, yield: 71% for two steps). Rt = 1.62 min **ESI-MS ($^+$):** *m/z* calcd for $C_{44}H_{73}N_{12}O_{14}$: 993.53, m/z found: 993.52 $[M+H]^+$.

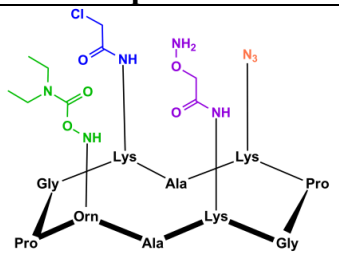
Peptide 10
 **10**
Peptide **10b** (2.31 µmol, 2.3 mg) was dissolved in a TFA/H_2O/TIS (70/20/10, v/v/v) solution (3 mL) and the reaction was stirred 2h at room temperature. The solvent was evaporated under vacuum then the crude peptide was precipitated in ether as a white powder. The crude product was used without any further purification. The yield was considered as quantitative. Rt = 1.29 min. **ESI-MS** ($^+$): m/z calcd for $C_{39}H_{65}N_{12}O_{12}$: 893.48, m/z found: 893.46 $[M+H]^+$.

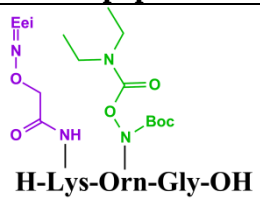
Linear peptide 17b
 17b
The linear peptide was synthesizer using general procedure **2.1** and modified amino acid **5**, **9** and **19**. The crude product was used without purification.

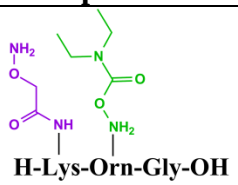
Cyclic peptide 17c
 17c
The linear peptide **17b** (1 eq, 0.094 mmol, 143 mg) was cyclized using general procedure **2.2**. The crude product was purified on RP-HPLC and freeze-dried to obtain a white powder (0.054 mmol, 81 mg, yield: 57% for two steps). Rt = 2.06 min **ESI-MS (+):** m/z calcd for $C_{69}H_{112}N_{18}O_{19}$: 1497.8, m/z found: 1497.74 $[M+H]^+$.

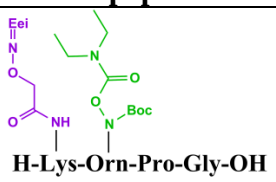
Peptide 17d
 17d
The peptide **17c** (0.054 mmol, 81 mg) was deprotected using general procedure **2.4**. The crude product was used without any further purification.

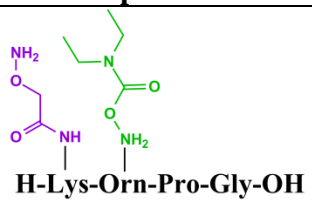
Peptide 17e
 **17e**
The peptide **17e** was obtained using general procedure **2.5**. and peptide **17d** (0.054 mmol, 81 mg). The crude product was purified on RP-HPLC and freeze-dried to obtain a white powder (0.0375 mmol, 53 mg, yield: 69% for two steps). Rt = 1.93 min **ESI-MS (+):** m/z calcd for $C_{61}H_{102}ClN_{18}O_{18}$: 1409.72, m/z found: 1409.65 $[M+H]^+$.

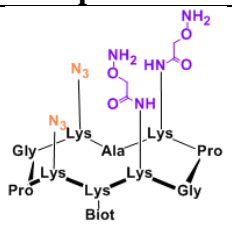
Peptide 17
 **17**
The peptide **17e** (5 mg) was dissolved in a TFA/H_2O/TIS (80/16/4, v/v/v) solution (300 µL) and the reaction was stirred 2h at room temperature. Then a solution of NaOH at 4M was added until reached a pH of 9 (up to 10) was reached. Afterwards, the crude product was purified using a Glen-Pak™ C-18 cartridge. After loading the crude peptide on the Glen-Pak™ column, 2 x 1 mL of H_2O was used in order to remove the salts. The peptide was recovered using 3 x 1 mL of H_2O/CH_3CN (1/1, v/v). The deprotection was considered quantitative and the solution was used without further treatment. Rt = 1.43 min. **UPLC-MS** (+): m/z calcd for $C_{52}H_{90}ClN_{18}O_{15}$: 1241.64, m/z found: 1241.24 $[M+H]^+$.

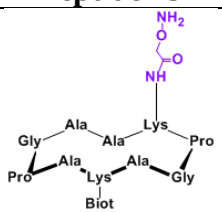
Linear peptide 21b
 **21b**
The linear peptide **21b** was synthesizer using general procedure **2.1** and modified amino acid **5** and **9**. The crude product was used without purification.

Peptide 21
 **21**
Peptide **21** was dissolved in a TFA/H_2O/TIS (95/2.5/2.5, v/v/v) solution (300 µL) and the reaction was stirred 2h at room temperature. Then a solution of NaOH at 4M was added until reached a pH of 9 (up to 10) was reached. Afterwards, the crude product was purified using a Glen-Pak™ C-18 cartridge. After loading the crude peptide on the Glen-Pak™ column, 2 x 1 mL of H_2O was used in order to remove the salts. The peptide was recovered using 3 x 1 mL of H_2O/CH_3CN (1/1, v/v). The deprotection was considered quantitative and the solution was used without further treatment. Rt = 1.41 min **UPLC-MS (+):** *m/z* calcd for $C_{20}H_{40}N_7O_8$: 506.29, *m/z* found: 506.35 $[M+H]^+$

Linear peptide 22b
 **22b**
The linear peptide **22b** was synthesizer using general procedure **2.1** and modified amino acid **5** and **9**. The crude product was used without purification. Rt = 1.49 min **UPLC-MS (+):** *m/z* calcd for $C_{34}H_{61}N_8O_{12}$: 773.44, *m/z* found: 773.41 $[M+H]^+$

Peptide 22
 **22**
Peptide **22b** (5 mg) was dissolved in a TFA/H_2O/TIS (95/2.5/2.5, v/v/v) solution (300 µL) and the reaction was stirred 2h at room temperature. Then a solution of NaOH at 4M was added until reached a pH of 9 (up to 10) was reached. Afterwards, the crude product was purified using a Glen-Pak™ C-18 cartridge. After loading the crude peptide on the Glen-Pak™ column, 2 x 1 mL of H_2O was used in order to remove the salts. The peptide was recovered using 3 x 1 mL of H_2O/CH_3CN (1/1, v/v). The deprotection was considered quantitative and the solution was used without further treatment. Rt = 0.96 min **UPLC-MS (+):** m/z calcd for $C_{25}H_{47}N_8O_9$: 603.34, m/z found: 603.4 $[M+H]^+$.

Peptide 23
 23
23 was synthesized according to the published procedure described by Dr. L. Bonnat[171]

Peptide 25
 25
25 was synthesized according to the published procedure described by Dr. L. Bonnat[171]

3. Oligonucleotides synthesis

RNA were prepared using PivOM chemistry on an Expedite synthesizer at 1 µmol scale. DNA were prepared using β-cyanoethylphosphoramidite chemistry on a 3400 DNA synthesizer at 1 µmol scale.

RP-HPLC analysis were performed on a Waters HPLC system using C18 Nucleosil column (Macherey-Nagel, 100 Å, 250 x 4.6 mm, 5 µm) with UV-monitoring at 260 nm and 280 nm. A 1 mL/min flow linear gradient was applied. Solvent A (50 mM triethylammonium acetate buffer with 5% acetonitrile) and solvent B (acetonitrile with 5% water) were used. Several gradients have been used: gradient I (0-30% B in 30 minutes), gradient II (0-50% B in 30 minutes), gradient III (0-70% B in 30 minutes) and gradient IV (0-100% B in 30 minutes).

The RP-HPLC purifications of oligonucleotides were performed on a Gilson system with Nucleosil C-18 colunm (Macherey-Nagel, 100 Å, 250 x 10 mm, 7 µm) using 4 mL/min with UV-monitoring at 260 nm and 280 nm. Solvent A (50 mM triethylammonium acetate buffer with 5% acetonitrile) and solvent B (acetonitrile with 5% water) were used.

Desalting of oligonucleotide was performed by SEC on NAP 25 cartridge using manufacturer's protocol.

Quantification of oligonucleotides was performed at 260 nm using Nanodrop (Molar extinction ε_{260nm} were estimated according to the nearest neighbour model).

3.1. General procedure for RNA synthesis

After synthesis, cyanoethyl protecting groups were removed using a solution of 1M DBU in acetonitrile. Cleavage from the resin and deprotection were performed in 28% aqueous NH_3 solution for 4h at room temperature. The product was purified on Dionex system with a Dionex column (DNAPac PA-100, 9x250 mm) and a 5 $mLmin^{-1}$ flow linear gradient with solvent A2 (Tris buffer 0.4M pH 7 with 5% acetonitrile) and B2 (Tris buffer 0.4M pH 7, lithium perchlorate trihydrate buffer 0.4M with 5% acetonitrile) at 75°C with UV monitoring at 260 and 280 nm. Then the product was desalted and freeze-dried.

3.2.General procedure for DNA synthesis

After DMTr-ON synthesis, cyanoethyl protecting groups were removed using a solution of 1M DBU in acetonitrile. Cleavage from the resin and deprotection were performed in 28% aqueous NH_3/methylamine (1/1) solution for 30 min at 60°C.

The product was purified on RP-HPLC then desalted and freeze-dried. When it's present the DMT protecting groups was removed using in 80% aqueous acetic acid solution for 20 min at room temperature.

3.3.General procedure for 3'diol DNA and RNA oxidation

30 eq of sodium periodate were added to a solution of 3'diol oligonucleotide in water to obtain a final concentration of 10^{-3} M. The mixture was stirred for 30 min at room temperature in dark conditions. Excess of $NaIO_4$ was then removed by SEC. The oxidation was considered quantitative and the crude oligonucleotide was used for oxime ligation without further purification.

3.4.General procedure for 3'or 5' amino DNA diazotransfer

The oligonucleotide (1 eq) was suspended in $NaHCO_3$ buffer (50 mM, pH 8.5)/MeOH (3/1, v/v) in a closed screw-cap plastic vial. $CuSO_4$ (3 eq) and imidazole-1-sulfonyl azide hydrochloride (ISAHC, 30 eq) were added. Afterwards, the solution was heated at 60°C. The reaction was followed by RP-HPLC and/ or UPLC-MS. Finally, the solution was cooled at room temperature and the solution was desalted by SEC.

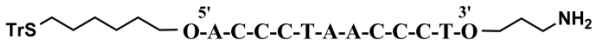
3.5.General procedure for 5' STrityl DNA deprotection

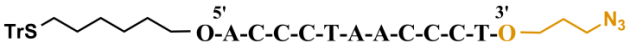
The oligonucleotide was suspended in triethylammonium acetate (TEAA) buffer (0.1M, pH 6.5). 0.15 volume of 1M aqueous silver nitrate solution and 0.5 volume of acetic acid were added. The mixture was mix and leave to react for 45 minutes at room temperature. Afterwards, 0.2 volume of 1M aqueous DTT solution was added. The mixture was mix and leave to react for 15 minutes at room temperature. After, the solution was centrifuge in order to remove the silver-DTT complex and the supernatant was recovered. The precipitate was washed with 3 volumes of TEAA buffer. Finally, the solution was desalted by SEC.

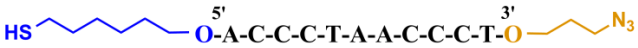
3.6. General procedure for the addition of the acyl trifluoroborate group

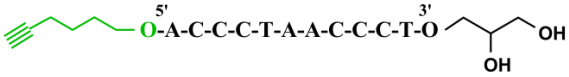
Potassium 4-(perfluorophenylcarboxy)benzoyltrifluoroborate **12** (2 eq) in 0.5 volume of DMF (synthesized according to published procedure) was added to the oligonucleotide (1 eq) in 0.5 volume of water. The pH was adjusted to 8-9 using DIEA. The reaction was monitored by RP-HPLC. After completion of the reaction, the solution was desalted by SEC and finally the product was purified on RP-HPLC.

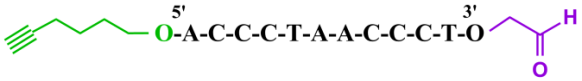
3.7. Experimental procedure

5'-C_6-STrityl 3'-C_3-Amino DNA 6a
 6a
Oligonucleotide **6a** was obtained from automated synthesis using commercially available 5'-thiolmodifier C6 phosphoramidite and 3'-PT-amino-modifier C3 CPG resin at 1 µmol scale, using general procedure **3.1**. The crude product was purified on RP-HPLC (430 nmol, yield: 43 %, ε_{260nm} = 98 100 $M^{-1}.cm^{-1}$). Rt = 15 min (RP-HPLC, Gradient IV). **ESI-MS** (-): m/z calcd for $C_{132}H_{171}N_{38}O_{69}P_{12}S$: 3797.73, m/z found: 3797.11 $[M-H]^-$.

5'-C_6-STrityl 3'-C_3-Azide DNA 6b
 6b
Oligonucleotide **6b** was obtained using the general procedure **3.4**. The crude product was used without any further purification. The yield was considered as quantitative (ε_{260nm} = 98 100 $M^{-1}.cm^{-1}$). Rt = 15 min (RP-HPLC, Gradient IV). **ESI-MS** (-): m/z calcd for $C_{132}H_{169}N_{40}O_{69}P_{12}S$: 3821.7, m/z found: 3821.5 $[M-H]^-$.

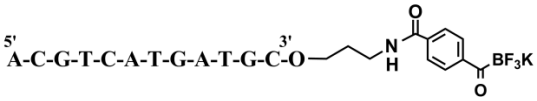
5'-C_6-SH 3'-C_3-Azide DNA 6
 6
Oligonucleotide **6** (500 nmol) was deprotected using general procedure **3.5**. The crude product was used without any further purification (330 nmol, yield: 66 %, ε_{260nm} = 98 100 $M^{-1}.cm^{-1}$). Rt = 13.3 min (RP-HPLC, Gradient III). **ESI-MS** (-): m/z calcd for $C_{113}H_{155}N_{40}O_{69}P_{12}S$: 3579.63, m/z found: 3579.58 $[M-H]^-$.

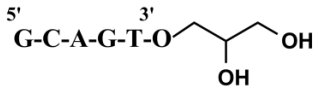
5' alkyne-3' diol DNA 7a
 7a
Oligonucleotide **7a** was obtained from automated synthesis using commercially available 5'-hexynyl (β-cyanoethyl) phosphoramidite and 3'-glyceryl CPG resin at 1 µmol scale, using general procedure **3.2**. The crude product was purified on RP-HPLC and freeze-dried (415 nmol, yield: 41%, ε_{260nm} = 98 100 $M^{-1}.cm^{-1}$). Rt = 15.6 min (RP-HPLC, Gradient I). **ESI-MS** (-): m/z calcd for $C_{113}H_{152}N_{37}O_{71}P_{12}$: 3534.65, m/z found: 3534.54 $[M-H]^-$.

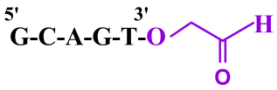
5’ alkyne-3’ aldehyde DNA 7
 **7**
Oligonucleotide **7a** was oxidized using general procedure **3.3**. The crude product was used without any further purification. The yield was considered as quantitative (ε_{260nm} = 98 100 $M^{-1}.cm^{-1}$). Rt = 15.8 min (RP-HPLC, Gradient I). **ESI-MS** ($^-$): m/z calcd for $C_{112}H_{148}N_{37}O_{70}P_{12}$: 3502.60, m/z found: 3502.68 $[M-H]^-$.

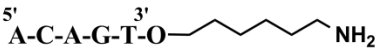
hTeloC DNA 8
5' C-C-C-T-A-A-C-C-C-T-A-A-C-C-C-T-A-A-C-C-C-T 3' **8**
Oligonucleotide **8** was obtained from automated synthesis using commercially available T SynBase™ CPG resin at 1 µmol scale, using general procedure **3.2**. The crude product was purified on RP-HPLC and freeze-dried (300 nmol, yield: 30%, ε_{260nm} = 193 700 $M^{-1}.cm^{-1}$). Rt = 14.2 min (RP-HPLC, Gradient II). **ESI-MS** ($^-$): m/z calcd for $C_{208}H_{269}N_{74}O_{128}P_{21}$: 6501.13, m/z found: 6501.56 $[M-H]^-$.

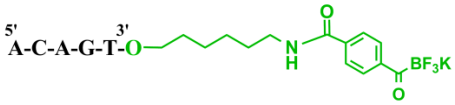
3’-C_3-Amino DNA 11a
11a
Oligonucleotide **11a** was obtained from automated synthesis on a 3’-PT-amino-modifier C3 CPG resin at 1 µmol scale using general procedure **3.1**. The crude product was purified on RP-HPLC (440 nmol, yield: 44%, ε_{260nm} = 115 700 $M^{-1}.cm^{-1}$). Rt = 14 min (RP-HPLC, Gradient I). **ESI-MS** ($^-$) m/z calcd for $C_{120}H_{156}N_{46}O_{73}P_{12}$: 3782.47, m/z found: 3782.6 $[M-H]^-$.

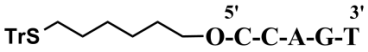
3’-C_3-NH-KAT DNA 11
 **11**
Oligonucleotide **11** (50 nmol) was obtained using general procedure **3.6**. The crude product was purified on RP-HPLC (28 nmol, yield: 56 %, ε_{260nm} = 115 700 $M^{-1}.cm^{-1}$). Rt = 15.7 min (RP-HPLC, Gradient I). **ESI-MS** ($^-$) m/z calcd $C_{128}H_{159}N_{46}O_{75}P_{12}BF_3$: 3979.69 m/z found: 3979.9 $[M-H]^-$.

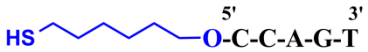
3’ diol DNA 13a
 **13a**
Oligonucleotide **13a** was obtained from automated synthesis on a 3’-glyceryl CPG resin at 1 µmol scale using general procedure **3.1**. The crude product was purified on RP-HPLC (480 nmol, yield: 48%, ε_{260nm} = 49 500 $M^{-1}.cm^{-1}$). Rt = 12 min (RP-HPLC, Gradient II). **ESI-MS** ($^-$): m/z calcd for $C_{52}H_{69}N_{20}O_{33}P_5$: 1656.3, m/z found: 1656.4 $[M-H]^-$.

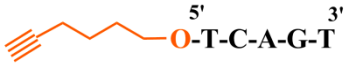
3’ aldehyde DNA 13
 **13**
Oligonucleotide **13a** (1.375 µmol) was oxidized using general procedure **3.3**. The crude product was used without any further purification. The yield was considered as quantitative (ε_{260nm} = 49 500 $M^{-1}.cm^{-1}$). Rt = 12 min (RP-HPLC, Gradient II). **ESI-MS** ($^-$): m/z calcd for $C_{52}H_{69}N_{20}O_{33}P_5$: 1624.2, m/z found: 1624.4 $[M-H]^-$.

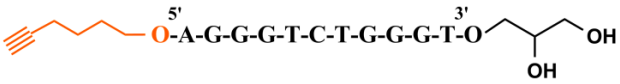
3’-C_6-amino DNA 14a
 **14a**
Oligonucleotide **14a** was obtained from automated synthesis on a 3’-PT-amino-modifier C6 CPG resin at 1 µmol scale using general procedure **3.1**. The crude product was purified on RP-HPLC (450 nmol, yield: 45%, ε_{260nm} = 53 100 $M^{-1}.cm^{-1}$). Rt = 14 min (RP-HPLC, Gradient II). **ESI-MS** ($^-$): m/z calcd for $C_{55}H_{75}N_{21}O_{30}P_5$: 1665.3, m/z found: 1665.3 $[M-H]^-$.

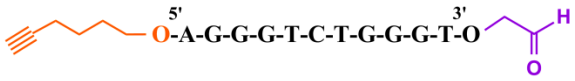
3’-C_6-NH-KAT DNA 14
 **14**
Oligonucleotide **14** (800 nmol) was obtained using general procedure **3.6**. The crude product was purified on RP-HPLC (440 nmol, yield: 55 %, ε_{260nm} = 53 100 $M^{-1}.cm^{-1}$). Rt = 16 min (RP-HPLC, Gradient II). **ESI-MS** ($^-$): m/z calcd for $C_{63}H_{80}N_{21}O_{32}P_5BF_3$: 1865.4, m/z found: 1865.2 $[M-H]^-$

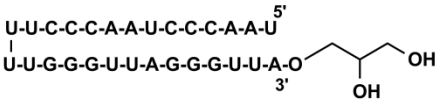
5’-C_6-STrityl DNA 15a
 **15a**
Oligonucleotide **15a** was obtained from automated synthesis using commercially available 5’-thiolmodifier C6 phosphoramidite at 1 µmol scale, using general procedure **3.1**. The crude product was purified on RP-HPLC (430 nmol, yield: 43 %, ε_{260nm} = 46 500 $M^{-1}.cm^{-1}$). Rt = 23.5 min (RP-HPLC, Gradient II).

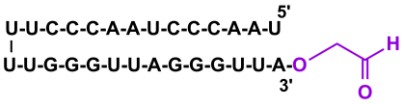
5’-C_6-SH DNA 15
 **15**
Oligonucleotide **15** (1 µmol) was deprotected using general procedure **3.5**. The crude product was used without any further purification (760 nmol, yield: 76 %, ε_{260nm} = 46500 $M^{-1}.cm^{-1}$). Rt = 17.7 min (RP-HPLC, Gradient II). **ESI-MS** ($^-$): m/z calcd for $C_{54}H_{74}N_{18}O_{31}P_5S$: 1658.3, m/z found: 1658.4 $[M-H]^-$.

5’-C_6-alkyne DNA 16
 **16**
Oligonucleotide **16** was obtained from automated synthesis using commercially available 5’-hexynyl (β-cyanoethyl) phosphoramidite at 1 µmol scale, using general procedure **3.1**. The crude product was purified on RP-HPLC (410 nmol, yield: 41 %, ε_{260nm} = 48 100 $M^{-1}.cm^{-1}$). Rt = 15 min (RP-HPLC, Gradient II). **ESI-MS** ($^-$): m/z calcd for $C_{55}H_{72}N_{17}O_{32}P_5$: 1637.3 m/z found: 1637.4 $[M-H]^-$.

5’ alkyne-3’ diol DNA 24a
 **24a**
Oligonucleotide **24a** was obtained from automated synthesis using commercially available 5’-hexynyl (β-cyanoethyl) phosphoramidite and 3’-glyceryl CPG resin at 1 µmol scale, using general procedure **3.2**. The crude product was purified on RP-HPLC and freeze-dried (410 nmol, yield: 41%, ε_{260nm} = 108 000 $M^{-1}.cm^{-1}$). Rt = 17 min (RP-HPLC, Gradient II). **ESI-MS** ($^-$): m/z calcd for $C_{118}H_{152}N_{44}O_{74}P_{12}$: 3742.4, m/z found: 3742.73 $[M-H]^-$.

5’ alkyne-3’ aldehyde DNA 24
 **24**
Oligonucleotide 5’alkyne-3’diol **24a** was oxidized using general procedure **3.3**. The crude product was used without any further purification. The yield was considered as quantitative (ε_{260nm} = 108 000 $M^{-1}.cm^{-1}$). Rt = 17 min (RP-HPLC, Gradient II).

3’ diol RNA 26b
 **26b**
Oligonucleotide 3’ diol **26a** was obtained from automated synthesis on a 3’-glyceryl CPG resin at 1 µmol scale using general procedure **3.1**. The crude product was purified on Dionex and freeze-dried (130 nmol, yield: 13%, ε_{260nm} = 267 400 $M^{-1}.cm^{-1}$). Rt = 13.8 min (RP-HPLC, Gradient II). **MALDI** ($^-$): m/z calcd for $C_{267}H_{334}N_{98}O_{203}P_{28}$: 9027.15, m/z found: 9026.61 $[M-H]^-$.

3’ aldehyde RNA 26
 **26**
Oligonucleotide 3’ diol **26a** (119 nmol) was oxidized using general procedure **3.3**. The crude product was used without any further purification. The yield was considered as quantitative (ε_{260nm} = 267 400 $M^{-1}.cm^{-1}$).

4. Conjugate synthesis

1. General details

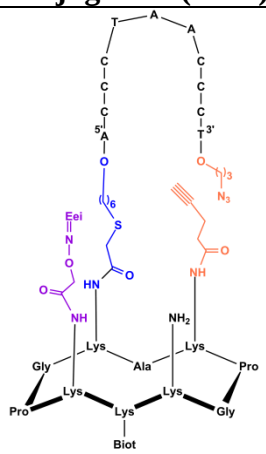
RP-HPLC analyses were performed on a Waters HPLC system using C18 Nucleosil column (Macherey-Nagel, 100 Å, 250 x 4.6 mm, 5 µm) with UV-monitoring at 260 nm and 280 nm. A 1 mL/min flow linear gradient was applied. Solvent A (50 mM triethylammonium acetate buffer with 5% acetonitrile) and solvent B (acetonitrile with 5% water) were used. Two gradients have been used: gradient I (0-30% B in 30 minutes) and gradient II (0-50% B in 30 minutes).

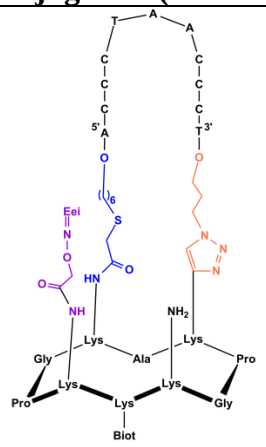
The course of reactions were monitored by using UPLC-MS system Waters, it includes reverse phase chromatography using Nucleosil C18 column (130 Å, 2.1 x 50 mm, 1.7 µm) and detection by UV at 260 nm and 280 nm and by electron spray ionization mass spectrometry. A 0.3 mL/min flow linear gradient with solvent C (triethylamine (15mM) and hexafluoro-2-propanol (50mM) in water) and solvent D (triethylamine (15mM) and hexafluoro-2-propanol (50mM) in methanol) is used. Several gradients have been used: gradient I (0-50% D in 12.5 minutes), gradient II (0-60% D in 12.5 minutes) and gradient III (0-80% D in 12.5 minutes).

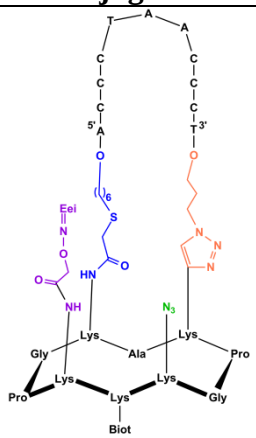
The RP-HPLC purifications of conjugates were performed on a Gilson system with Nucleosil C-18 colunm (Macherey-Nagel, 100 Å, 250 x 10 mm, 7 µm) with UV-monitoring at 260 nm and 280 nm using 4 mL/min flow linear gradient. Solvent A (50 mM triethylammonium acetate buffer with 5% acetonitrile) and solvent B (acetonitrile with 5% water) were used.

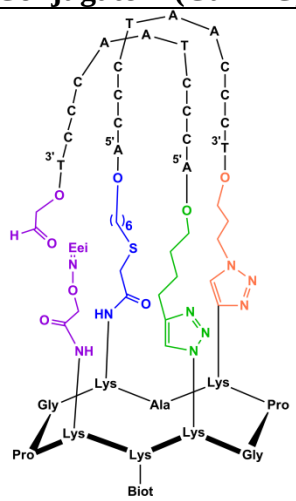
Quantification of oligonucleotides was performed at 260 nm using Nanodrop apparatus (molar extinction ε_{260nm} was estimated according to the nearest neighbor model).

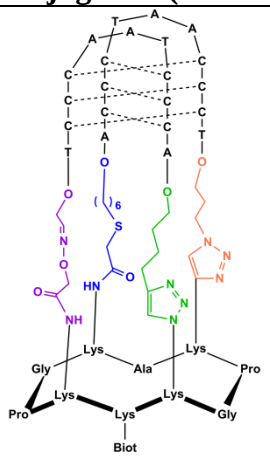
2. Experimental procedure

Conjugate A (CTC)
 A
To a solution of thiol oligonucleotide **6** (220 nmol, 1 eq) and peptide **1** (440 nmol, 2 eq) in a solution of H_2O were added TCEP (1 eq, 220 nmol), DIEA (60 eq, 13.2 µmol), KI (60 eq, 13.2 µmol). The reaction was stirred at 37°C for 90 minutes. The crude product is desalted by SEC and finally purified on RP-HPLC and freeze-dried (116 nmol, yield: 53%, ε_{260nm} = 98 100 $M^{-1}.cm^{-1}$). Rt = 5.43 min (UPLC, Gradient III). **ESI-MS** ($^-$) m/z calcd $C_{183}H_{268}N_{58}O_{86}P_{12}S_2$: 5089.46, m/z found: 5089.68 $[M-H]^-$.

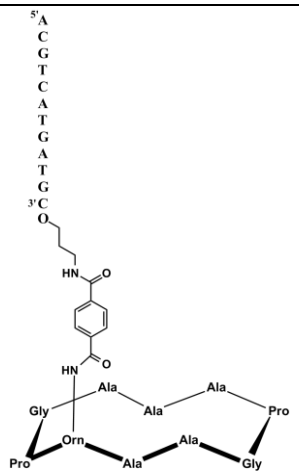
Conjugate B (CuAAC)
 B
To a solution of **conjugate A** (1 eq, 150 nmol) in 50 mM PBS buffer (pH 7.4) were added $CuSO_4$ (3 eq, 450 nmol), THPTA (30 eq, 4.5 µmol), sodium ascorbate (30 eq, 4.5 µmol), 500mM KCl and DMF (10%v). The reaction was stirred at 37°C for 2h and quenching with 0.5 M EDTA solution (50 eq, 7.5 µmol). The product was desalted by SEC then purified on RP-HPLC and freeze-dried (60 nmol, yield: 40%, ε_{260nm} = 98 100 $M^{-1}.cm^{-1}$). Rt = 4.80 min (UPLC-MS, Gradient III). **ESI-MS** ($^-$) m/z calcd $C_{183}H_{268}N_{58}O_{86}P_{12}S_2$: 5089.46, m/z found: 5089.62 $[M-H]^-$.

Conjugate C
 C
The **conjugate B** (1 eq, 90 nmol) was suspended in $NaHCO_3$ buffer (50 mM, pH 8.5)/MeOH (3/, v/v) in a closed screw-cap plastic vial. $CuSO_4$ (3 eq, 270 nmol) and imidazole-1-sulfonyl azide hydrochloride (ISAHC, 30 eq, 2.7 µmol) were added. Afterwards, the solution was heated at 60°C. The reaction was followed by UPLC-MS. Finally, the solution was cooled at room temperature and the solution was desalted by SEC. The yield was considered as quantitative. Rt = 5.18 min (UPLC, Gradient II). **ESI-MS** ($^-$) m/z calcd $C_{183}H_{267}N_{60}O_{86}P_{12}S_2$: 5116.45, m/z found: 5116.21 $[M-H]^-$.

Conjugate E (CuAAC)
 E
To a solution of oligonucleotide **7** (1.2 eq, 84 nmol) and **conjugate C** (1 eq, 70 nmol) in 50 mM PBS buffer (pH 7.4) were added $CuSO_4$ (3 eq, 210 nmol), THPTA (30 eq, 2.1 µmol), sodium ascorbate (30 eq, 2.1 µmol), 500mM KCl and DMF (10%v). The reaction was stirred at 60°C for 1h and quenching with 0.5 M EDTA solution (50 eq, 3.5 µmol). The product was desalted by SEC then purified on RP-HPLC and freeze-dried (42 nmol, yield: 48%, ε_{260nm} = 196 200 $M^{-1}.cm^{-1}$). Rt = 5.02 min (UPLC, Gradient II). **MALDI-Tof** ($^-$) m/z calcd $C_{295}H_{413}N_{97}O_{156}P_{24}S_2$: 8617.05, m/z found: 8613.61 $[M-H]^-$.

Conjugate F (Oxime)
 F
The **conjugate E** (1 eq, 9 nmol) was suspended in water (350 µL) in a closed screw-cap glass vial. TFA 2%vol (350 µL) is added in order to have a final solution of TFA 1%vol. The reaction was stirred at room temperature for 1h. The product was desalted by SEC then freeze-dried. The yield was considered as quantitative. Rt = 5.28 min (UPLC, Gradient I). **MALDI-Tof** ($^-$) m/z calcd $C_{291}H_{405}N_{97}O_{154}P_{24}S_2$: 8528.99, m/z found: 8526.68 $[M-H]^-$.

Conjugate G (KAT)



G

Condition A

Oligonucleotide **11** (5 nmol, 1 eq) was dissolved in tBuOH/H2O (1/1, v/v) solution and peptide **10** (10 nmol, 2 eq) was added. The solution was stirred at room temperature. The reaction is monitored by RP-HPLC. The crude was analyzed by ESI and MALDI.
Rt = 16.2 min (RP-HPLC, Gradient I).
MALDI-Tof ($^-$) m/z calcd $C_{162}H_{213}N_{57}O_{85}P_{12}$: 4690.4, m/z found: 4689.55 $[M-H]^-$.

Condition B

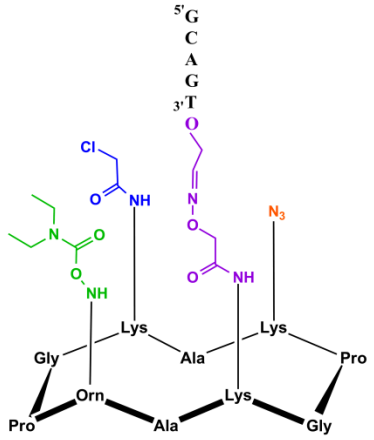
Oligonucleotide **11** (5 nmol, 1 eq) was dissolved in oxalic acid (100mM)/ potassium oxalate (10mM) solution and peptide **10** (10 nmol, 2 eq) was added. The solution was stirred at room temperature. The reaction is followed by RP-HPLC. The crude was analyzed by MALDI.
Rt = 15.9 min (RP-HPLC, Gradient I).
MALDI-Tof ($^-$) m/z calcd $C_{162}H_{213}N_{57}O_{85}P_{12}$: 4690.4, m/z found: 4687.68 $[M-H]^-$.

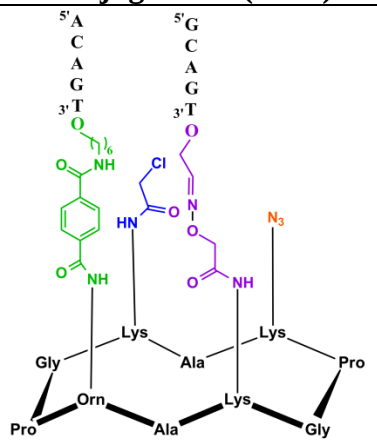
Condition C

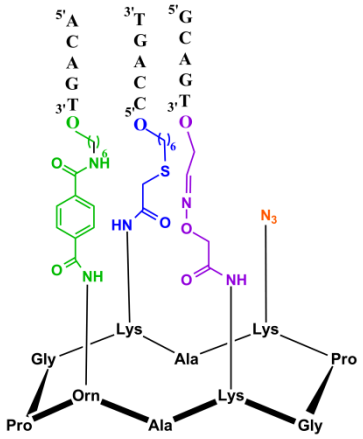
Oligonucleotide **11** (5 nmol, 1 eq) was dissolved in 0.4 M ammonium acetate buffer (pH 4.5) and peptide **10** (10 nmol, 2 eq.) was added. The solution was stirred at room temperature. The reaction is followed by RP-HPLC. The crude was analyzed by ESI and MALDI.
Rt = 16 min (RP-HPLC, Gradient I).
MALDI-Tof ($^-$) m/z calcd $C_{162}H_{213}N_{57}O_{85}P_{12}$: 4690.4, m/z found: 4688.23 $[M-H]^-$.

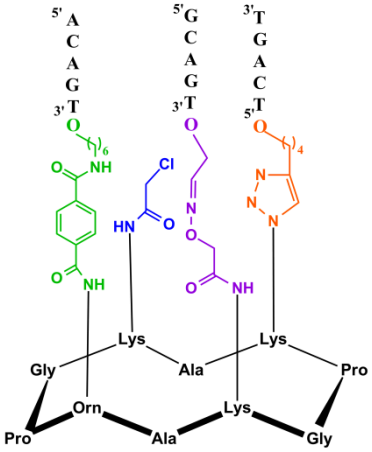
Condition D

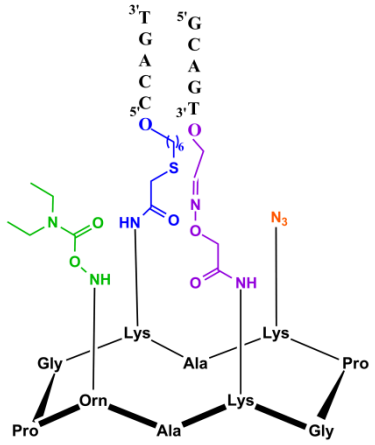
Oligonucleotide **11** (5 nmol, 1 eq) was dissolved in H_2O/DIEA (pH 8-9) solution and peptide **10** (10 nmol, 2 eq.) was added. The solution was stirred at room temperature. The reaction is followed by RP-HPLC. The crude was analyzed by ESI and MALDI.
Rt = 16.2 min (RP-HPLC, Gradient I).
MALDI-Tof ($^-$) m/z calcd $C_{162}H_{213}N_{57}O_{85}P_{12}$: 4690.4, m/z found: 4688.09 $[M-H]^-$.

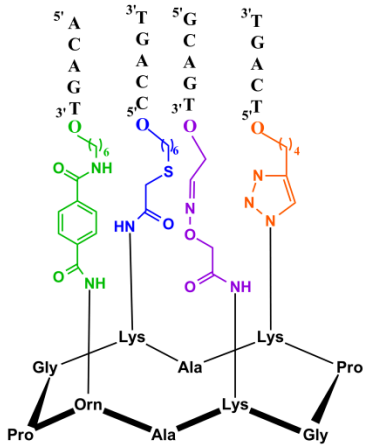
Conjugate H1 (Oxime)
 H1
Oligonucleotide **13** (350 nmol) was dissolved in 0.4 M ammonium acetate buffer (pH 4.5) and peptide **17** (in excess) was added. The solution was stirred at room temperature. The reaction is followed by UPLC-MS. The product was purified on RP-HPLC and freeze-dried (217 nmol, yield: 62%, ε_{260nm} = 49 500 $M^{-1}.cm^{-1}$). Rt = 7.28 min (UPLC, Gradient I). **ESI-MS** ($^-$) m/z calcd $C_{103}H_{150}N_{38}O_{46}P_5Cl$: 2844.89, m/z found: 2844.92 $[M-H]^-$.

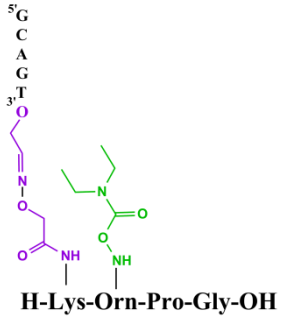
Conjugate H2 (KAT)
 H2
Conjugate **H1** (80 nmol, 1 eq) was dissolved in 0.4 M ammonium acetate buffer (pH 4.5) and oligonucleotide **14** (96 nmol, 1.2 eq) was added. The solution was stirred at room temperature. The reaction is followed by UPLC-MS. The product was purified on RP-HPLC and freeze-dried (45 nmol, yield: 56%, ε_{260nm} = 102 600 $M^{-1}.cm^{-1}$). Rt = 5.45 min (UPLC, Gradient I). **ESI-MS** ($^-$) m/z calcd $C_{161}H_{217}N_{58}O_{76}P_{10}Cl$: 4523.19, m/z found: 4523.67 $[M-H]^-$.

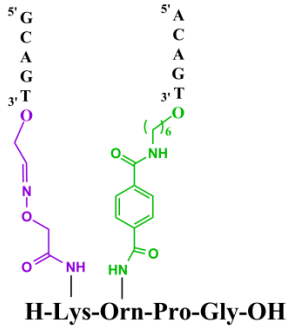
Conjugate H3 (CTC)
 H3
Conjugate **H5** (35 nmol, 1 eq) was dissolved in 0.4 M ammonium acetate buffer (pH 4.5) and oligonucleotide **14** (42 nmol, 1.2 eq) was added. The solution was stirred at room temperature. The reaction is followed by UPLC-MS. The product was purified on RP-HPLC and freeze-dried (20 nmol, yield: 57%, ε_{260nm} = 149 100 $M^{-1}.cm^{-1}$). Rt = 5.30 min (UPLC, Gradient I). **ESI-MS** ($^-$) m/z calcd $C_{215}H_{293}N_{76}O_{107}P_{15}S$: 6147.55, m/z found: 6147.57 $[M-H]^-$.

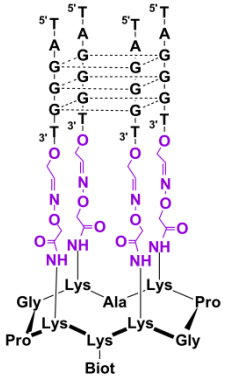
Conjugate H4 (CuAAC)
 H4
To a solution of alkyne oligonucleotide **16** (2 eq, 70 nmol) and conjugate **H2** (1 eq, 35 nmol) in 50 mM PBS buffer (pH 7.4) were added $CuSO_4$ (3 eq, 105 nmol), THPTA (30 eq, 1.05 µmol), sodium ascorbate (30 eq, 1.05 µmol) and DMF (10%v). The reaction was stirred at 37°C for 1h and quenching with 0.5 M EDTA solution (50 eq, 1.75 µmol). The product was desalted by SEC then purified on RP-HPLC and freeze-dried (24 nmol, yield: 69%, ε_{260nm} = 197 200 $M^{-1}.cm^{-1}$). Rt = 5.28 min (UPLC, Gradient I). **MALDI** ($^-$) m/z calcd $C_{216}H_{289}N_{75}O_{108}P_{15}ClNa$: 6183.5, m/z found: 6183.88 $[M-H]^-$.

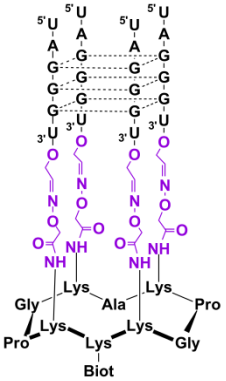
Conjugate H5 (CTC)
 H5
To a solution of thiol oligonucleotide **15** (60 nmol, 2 eq) and conjugate **H1** (30 nmol, 1 eq) in a solution of H_2O were added TCEP (2 eq, 60 nmol), 400 mM KCl, DIEA (60 eq, 1.8 µmol), KI (60 eq, 1.8 µmol). The reaction was stirred at 40°C for 90 minutes. The crude product was directly purified on RP-HPLC and freeze-dried (14 nmol, yield: 47%, ε_{260nm} = 96 000 $M^{-1}.cm^{-1}$). Rt = 6.30 min (UPLC, Gradient I). **ESI-MS** ($^-$) m/z calcd $C_{157}H_{224}N_{56}O_{77}P_{10}S$: 4467.24, m/z found: 4467.45 $[M-H]^-$.

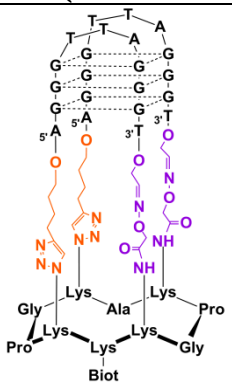
Conjugate H (CuAAC)
 H
To a solution of alkyne oligonucleotide **16** (1.2 eq, 42 nmol) and conjugate **H3** (1 eq, 35 nmol) in 50 mM PBS buffer (pH 7.4) were added $CuSO_4$ (10 eq, 350 nmol), THPTA (30 eq, 1.05 µmol), sodium ascorbate (30 eq, 1.05 µmol) and DMF (10%v). The reaction was stirred at 60°C for 1h and quenching with 0.5 M EDTA solution (50 eq, 1.75 µmol). The product was desalted by SEC then purified on RP-HPLC and freeze-dried (17 nmol, yield: 48%, ε_{260nm} = 197 200 $M^{-1}.cm^{-1}$). Rt = 5.28 min (UPLC, Gradient I). **ESI-MS** ($^-$) m/z calcd $C_{270}H_{363}N_{93}O_{139}P_{20}S$: 7786.8, m/z found: 7786.18 $[M-H]^-$.

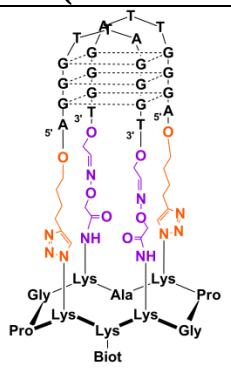
Conjugate I (Oxime)
 **I**
Condition A Oligonucleotide **13** (300 nmol, 1 eq) was dissolved in 0.4 M ammonium acetate buffer (pH 4.5) and peptide **22** (in excess) was added. The solution was stirred at room temperature. The reaction is followed by UPLC-MS. Rt = 5.4 min (UPLC, Gradient I). **UPLC-MS** ($^-$) m/z calcd $C_{76}H_{109}N_{28}O_{40}P_5$: 2209, m/z found: 2208.8 $[M-H]^-$. **Condition B** Oligonucleotide **13** (25 nmol, 1 eq) was dissolved in 0.3 M sodium phosphate buffer (pH 7.3), aniline (100 eq, 2.5 µmol) and peptide **22** (in excess) was added. The solution was stirred at room temperature. The reaction is followed by UPLC-MS. Finally, the crude is desalted by SEC and analyzed by UPLC-MS. Rt = 3.53 min (UPLC, Gradient III). **UPLC-MS** ($^-$) m/z calcd $C_{76}H_{109}N_{28}O_{40}P_5$: 2209, m/z found: 2208.8 $[M-H]^-$.

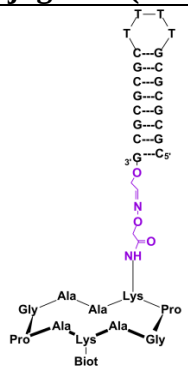
Conjugate J (KAT)
 **J**
Conjugate **I** (30 nmol) was dissolved in 0.4 M ammonium acetate buffer (pH 4.5) and oligonucleotide **14** (60 nmol, 2 eq) was added. The solution was stirred at room temperature. The reaction is followed by UPLC-MS. Rt = 4.25 min (UPLC, Gradient I). **UPLC-MS** ($^-$) m/z calcd $C_{134}H_{177}N_{48}O_{70}P_{10}$: 3887.91, m/z found: 3887.5 $[M-H]^-$.

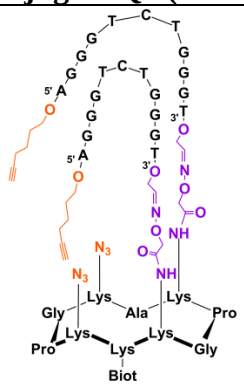
Conjugate L (Oxime)

L was synthesized according to the published procedure described by Bonnat *et al.*[165]

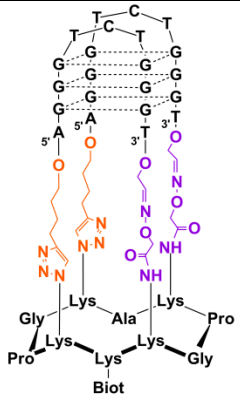
Conjugate M (Oxime)

M was synthesized according to the published procedure described by Bonnat *et al.*[165]

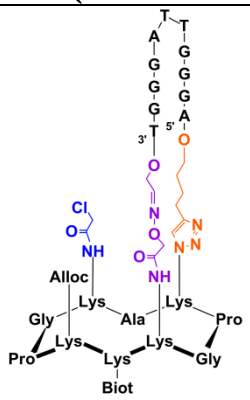
Conjugate N (Oxime and CuAAC)

N was synthesized according to the published procedure described by Bonnet *et al.*[164]

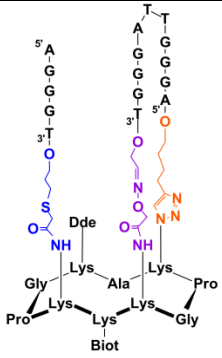
Conjugate O (Oxime and CuAAC)
 O
O was synthesized according to the published procedure described by Bonnet *et al.*[164]

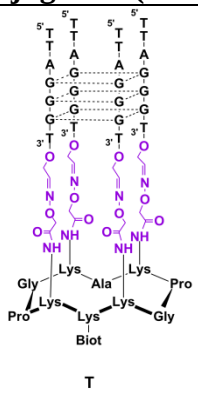
Conjugate P (Oxime)
 P
L was synthesized according to the published procedure described by Murat *et al.*[162]

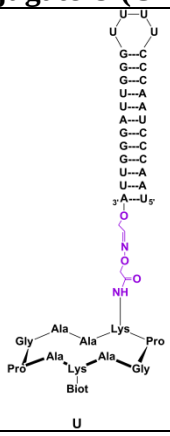
Conjugate Q1 (Oxime)
 Q1
Oligonucleotide **24** (400 nmol, 2.4 eq) was dissolved in 0.4 M ammonium acetate buffer (pH 4.5) and peptide **23** (1 eq, 167 nmol) was added. The solution was stirred at room temperature. The reaction is followed by HPLC. The product was purified on RP-HPLC and freeze-dried (130 nmol, yield: 32%, ε_{260nm} = 216 000 $M^{-1}.cm^{-1}$). Rt = 18.3 min (RP-HPLC, Gradient I). **MALDI** ($^-$) m/z calcd $C_{295}H_{394}N_{111}O_{160}P_{24}S$: 8824.95, m/z found: 8826.69 $[M-H]^-$.

Conjugate Q (CuAAC)
 Q
To a solution of conjugate **Q1** (1 eq, 150 nmol) in 50 mM PBS buffer (pH 7.4) were added $CuSO_4$ (6 eq, 900 nmol), THPTA (30 eq, 4.5 µmol), sodium ascorbate (30 eq, 4.5 µmol) and DMF (10%v). The reaction was stirred at 37°C for 90 minutes and quenching with 0.5 M EDTA solution (50 eq, 7.5 µmol). The product was desalted by SEC then purified on RP-HPLC and freeze-dried (90 nmol, yield: 60%, ε_{260nm} = 216 000 $M^{-1}.cm^{-1}$). Rt = 16.1 min (RP-HPLC, Gradient I). **MALDI** ($^-$) m/z calcd $C_{295}H_{394}N_{111}O_{160}P_{24}S$: 8824.95, m/z found: 8826.53 $[M-H]^-$.

Conjugate R (Oxime and CuAAC)
 R
L was synthesized according to the published procedure described by Dr. Laureen Bonnat[171]

Conjugate S (Oxime, CuAAC and CTC)

L was synthesized according to the published procedure described by Dr. Laureen Bonnat[171]

Conjugate T (Oxime)

L was synthesized according to the published procedure described by Murat *et al.*[162]

Conjugate U (Oxime)

Oligonucleotide **26** (45 nmol, 1 eq) was dissolved in 0.4 M ammonium acetate buffer (pH 4.5) and peptide **25** (2 eq, 90 nmol) was added. The solution was stirred at 55°C. The reaction is followed by HPLC. The product was purified on RP-HPLC and freeze-dried (6 nmol, yield: 13%, ε_{260nm} = 267 400 $M^{-1}.cm^{-1}$). Rt = 16.2 min (RP-HPLC, Gradient I). **MALDI** (-) m/z calcd $C_{316}H_{409}N_{113}O_{215}P_{28}S$: 10 124.69, m/z found: 10 130.28 $[M-H]^-$.

Références bibliographiques

(1) Dahm, R. Friedrich Miescher and the Discovery of DNA. *Developmental Biology* **2005**, *278* (2), 274–288. https://doi.org/10.1016/j.ydbio.2004.11.028.
(2) Avery, O. T.; MacLeod, C. M.; McCarty, M. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. *Journal of Experimental Medicine* **1944**, *79* (2), 137–158. https://doi.org/10.1084/jem.79.2.137.
(3) Paweletz, N. Walther Flemming: Pioneer of Mitosis Research. *Nature Reviews Molecular Cell Biology* **2001**, *2* (1), 72–75. https://doi.org/10.1038/35048077.
(4) Levene, P. A.; Mori, T. Ribodesose and Xylodesose and Their Bearing on the Structure of Thyminose. *Journal of Biological Chemistry* **1929**, *83* (3), 803–816.
(5) Levene, P. A.; Mikeska, L. A.; Mori, T. On the Carbohydrate of Thymonucleic Acid. *Journal of Biological Chemistry* **1930**, *85* (3), 785–787.
(6) Chargaff, E. Chemical Specificity of Nucleic Acids and Mechanism of Their Enzymatic Degradation. *Experientia* **1950**, *6* (6), 201–209. https://doi.org/10.1007/BF02173653.
(7) Zamenhof, S.; Brawerman, G.; Chargaff, E. On the Desoxypentose Nucleic Acids from Several Microorganisms. *Biochimica et Biophysica Acta* **1952**, *9*, 402–405. https://doi.org/10.1016/0006-3002(52)90184-4.
(8) Wilkins, M. H. F.; Stokes, A. R.; Wilson, H. R. Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature* **1953**, *171* (4356), 738–740. https://doi.org/10.1038/171738a0.
(9) Franklin, R. E.; Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **1953**, *171* (4356), 740–741. https://doi.org/10.1038/171740a0.
(10) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737–738. https://doi.org/10.1038/171737a0.
(11) Evich, M.; Spring, A.; Germann M. Impact of modified ribose sugars on nucleic acid conformation and function. *Heterocyclic Communications* **2017**, *23*, 155-165. https://doi.org/10.1515/hc-2017-0056.
(12) Herbert, A.; Rich, A. The Biology of Left-Handed Z-DNA. *Journal Biology Chemistry* **1996**, *271* (20), 11595–11598. https://doi.org/10.1074/jbc.271.20.11595.
(13) Wood, B. R. The Importance of Hydration and DNA Conformation in Interpreting Infrared Spectra of Cells and Tissues. *Chemistry Society Reviews* **2016**, *45* (7), 1980–1998. https://doi.org/10.1039/C5CS00511F.
(14) Batey, R. T.; Rambo, R. P.; Doudna, J. A. Tertiary Motifs in RNA Structure and Folding. *Angewandte Chemie International Edition* **1999**, *38* (16), 2326–2343.
(15) Jain, A.; Wang, G.; Vasquez, K. M. DNA Triple Helices: Biological Consequences and Therapeutic Potential. *Biochimie* **2008**, *90* (8), 1117–1130. https://doi.org/10.1016/j.biochi.2008.02.011.
(16) Duckett, D. R.; Murchie, A. I. H.; Clegg, R. M.; Bassi, G. S.; Giraud-Panis, M.-J. E.; Lilley, D. M. J. Nucleic Acid Structure and Recognition. *Biophysical Chemistry* **1997**, *68* (1–3), 53–62. https://doi.org/10.1016/S0301-4622(97)00007-0.
(17) Lilley, D. M. J. Structures of Helical Junctions in Nucleic Acids. *Quarterly Reviews of Biophysics* **2000**, *33* (2), 109–159. https://doi.org/10.1017/S0033583500003590.
(18) Hardin, C. C.; Watson, T.; Corregan, M.; Bailey, C. Cation-Dependent Transition between the Quadruplex and Watson-Crick Hairpin Forms of d(CGCG3GCG). *Biochemistry* **1992**, *31* (3), 833–841. https://doi.org/10.1021/bi00118a028.
(19) Dolinnaya, N. G.; Ogloblina, A. M.; Yakubovskaya, M. G. Structure, Properties, and Biological Relevance of the DNA and RNA G-Quadruplexes: Overview 50 Years after

Their Discovery. *Biochemistry Moscow* **2016**, *81* (13), 1602–1649. https://doi.org/10.1134/S0006297916130034.

(20) Podbevšek, P.; Šket, P.; Plavec, J. Stacking and Not Solely Topology of T_3 Loops Controls Rigidity and Ammonium Ion Movement within $d(G_4T_3G_4)_2$ G-Quadruplex. *Journal of the American Chemical Society* **2008**, *130* (43), 14287–14293. https://doi.org/10.1021/ja8048282.

(21) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Research* **2006**, *34* (19), 5402–5415. https://doi.org/10.1093/nar/gkl655.

(22) Dai, J.; Carver, M.; Yang, D. Polymorphism of Human Telomeric Quadruplex Structures. *Biochimie* **2008**, *90* (8), 1172–1183. https://doi.org/10.1016/j.biochi.2008.02.026.

(23) Limongelli, V.; De Tito, S.; Cerofolini, L.; Fragai, M.; Pagano, B.; Trotta, R.; Cosconati, S.; Marinelli, L.; Novellino, E.; Bertini, I.; Randazzo, A.; Luchinat, C.; Parrinello, M. The G-Triplex DNA. *Angewandte Chemie International Edition* **2013**, *52* (8), 2269–2273. https://doi.org/10.1002/anie.201206522.

(24) Rajendran, A.; Endo, M.; Hidaka, K.; Sugiyama, H. Direct and Single-Molecule Visualization of the Solution-State Structures of G-Hairpin and G-Triplex Intermediates. *Angewandte Chemie International Edition* **2014**, *53* (16), 4107–4112. https://doi.org/10.1002/anie.201308903.

(25) Gray, R. D.; Trent, J. O.; Chaires, J. B. Folding and Unfolding Pathways of the Human Telomeric G-Quadruplex. *Journal of Molecular Biology* **2014**, *426* (8), 1629–1650. https://doi.org/10.1016/j.jmb.2014.01.009.

(26) Hou, X.-M.; Fu, Y.-B.; Wu, W.-Q.; Wang, L.; Teng, F.-Y.; Xie, P.; Wang, P.-Y.; Xi, X.-G. Involvement of G-Triplex and G-Hairpin in the Multi-Pathway Folding of Human Telomeric G-Quadruplex. *Nucleic Acids Research* **2017**, *45* (19), 11401–11412. https://doi.org/10.1093/nar/gkx766.

(27) Jiang, H.-X.; Cui, Y.; Zhao, T.; Fu, H.-W.; Koirala, D.; Punnoose, J. A.; Kong, D.-M.; Mao, H. Divalent Cations and Molecular Crowding Buffers Stabilize G-Triplex at Physiologically Relevant Temperatures. *Scientific Reports* **2015**, *5* (1), 9255. https://doi.org/10.1038/srep09255.

(28) Zhang, D.-H.; Fujimoto, T.; Saxena, S.; Yu, H.-Q.; Miyoshi, D.; Sugimoto, N. Monomorphic RNA G-Quadruplex and Polymorphic DNA G-Quadruplex Structures Responding to Cellular Environmental Factors. *Biochemistry* **2010**, *49* (21), 4554–4563. https://doi.org/10.1021/bi1002822.

(29) Martadinata, H.; Phan, A. T. Structure of Propeller-Type Parallel-Stranded RNA G-Quadruplexes, Formed by Human Telomeric RNA Sequences in K^+ Solution. *Journal of the American Chemical Society* **2009**, *131* (7), 2570–2578. https://doi.org/10.1021/ja806592z.

(30) Tang, C.-F.; Shafer, R. H. Engineering the Quadruplex Fold: Nucleoside Conformation Determines Both Folding Topology and Molecularity in Guanine Quadruplexes. *Journal of the American Chemical Society* **2006**, *128* (17), 5966–5973. https://doi.org/10.1021/ja0603958.

(31) Joachimi, A.; Benz, A.; Hartig, J. S. A Comparison of DNA and RNA Quadruplex Structures and Stabilities. *Bioorganic & Medicinal Chemistry* **2009**, *17* (19), 6811–6815. https://doi.org/10.1016/j.bmc.2009.08.043.

(32) Huang, H.; Suslov, N. B.; Li, N.-S.; Shelke, S. A.; Evans, M. E.; Koldobskaya, Y.; Rice, P. A.; Piccirilli, J. A. A G-Quadruplex–Containing RNA Activates Fluorescence in a GFP-like Fluorophore. *Nature Chemical Biology* **2014**, *10* (8), 686–691. https://doi.org/10.1038/nchembio.1561.

(33) Xiao, C.-D.; Shibata, T.; Yamamoto, Y.; Xu, Y. An Intramolecular Antiparallel G-Quadruplex Formed by Human Telomere RNA. *Chemistry Communication* **2018**, *54* (32), 3944–3946. https://doi.org/10.1039/C8CC01427B.
(34) Zhang, A. Y. Q.; Bugaut, A.; Balasubramanian, S. A Sequence-Independent Analysis of the Loop Length Dependence of Intramolecular RNA G-Quadruplex Stability and Topology. *Biochemistry* **2011**, *50* (33), 7251–7258. https://doi.org/10.1021/bi200805j.
(35) Pandey, S.; Agarwala, P.; Maiti, S. Effect of Loops and G-Quartets on the Stability of RNA G-Quadruplexes. *Journal of Physical Chemistry B* **2013**, *117* (23), 6896–6905. https://doi.org/10.1021/jp401739m.
(36) Guiset Miserachs, H.; Donghi, D.; Börner, R.; Johannsen, S.; Sigel, R. K. O. Distinct Differences in Metal Ion Specificity of RNA and DNA G-Quadruplexes. *Journal of Biological Inorganic Chemistry* **2016**, *21* (8), 975–986. https://doi.org/10.1007/s00775-016-1393-4.
(37) Huppert, J. L. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Research* **2005**, *33* (9), 2908–2916. https://doi.org/10.1093/nar/gki609.
(38) Kikin, O.; D'Antonio, L.; Bagga, P. S. QGRS Mapper: A Web-Based Server for Predicting G-Quadruplexes in Nucleotide Sequences. *Nucleic Acids Research* **2006**, *34* (Web Server), W676–W682. https://doi.org/10.1093/nar/gkl253.
(39) Eddy, J.; Maizels, N. Gene Function Correlates with Potential for G4 DNA Formation in the Human Genome. *Nucleic Acids Research* **2006**, *34* (14), 3887–3896. https://doi.org/10.1093/nar/gkl529.
(40) Yadav, V. K.; Abraham, J. K.; Mani, P.; Kulshrestha, R.; Chowdhury, S. QuadBase: Genome-Wide Database of G4 DNA Occurrence and Conservation in Human, Chimpanzee, Mouse and Rat Promoters and 146 Microbes. *Nucleic Acids Research* **2007**, *36* (Database), D381–D385. https://doi.org/10.1093/nar/gkm781.
(41) Bedrat, A.; Lacroix, L.; Mergny, J.-L. Re-Evaluation of G-Quadruplex Propensity with G4Hunter. *Nucleic Acids Res* **2016**, *44* (4), 1746–1759. https://doi.org/10.1093/nar/gkw006.
(42) Rhodes, D.; Lipps, H. J. G-Quadruplexes and Their Regulatory Roles in Biology. *Nucleic Acids Research* **2015**, *43* (18), 8627–8637. https://doi.org/10.1093/nar/gkv862.
(43) Jansen, A.; Verstrepen, K. J. Nucleosome Positioning in Saccharomyces Cerevisiae. *Microbiology and Molecular Biology Reviews* **2011**, *75* (2), 301–320. https://doi.org/10.1128/MMBR.00046-10.
(44) Sandin, S.; Rhodes, D. Telomerase Structure. *Current Opinion in Structural Biology* **2014**, *25*, 104–110. https://doi.org/10.1016/j.sbi.2014.02.003.
(45) Giardini, M. A.; Segatto, M.; da Silva, M. S.; Nunes, V. S.; Cano, M. I. N. Telomere and Telomerase Biology. In *Progress in Molecular Biology and Translational Science*; Elsevier, 2014; Vol. 125, pp 1–40. https://doi.org/10.1016/B978-0-12-397898-1.00001-3.
(46) Griffith, J. D.; Comeau, L.; Rosenfield, S.; Stansel, R. M.; Bianchi, A.; Moss, H.; de Lange, T. Mammalian Telomeres End in a Large Duplex Loop. *Cell* **1999**, *97* (4), 503–514. https://doi.org/10.1016/S0092-8674(00)80760-6.
(47) Greider, C. W.; Blackburn, E. H. Identification of a Specific Telomere Terminal Transferase Activity in Tetrahymena Extracts. *Cell* **1985**, *43* (2), 405–413. https://doi.org/10.1016/0092-8674(85)90170-9.
(48) Sundquist, W. I.; Klug, A. Telomeric DNA Dimerizes by Formation of Guanine Tetrads between Hairpin Loops. *Nature* **1989**, *342* (6251), 825–829. https://doi.org/10.1038/342825a0.
(49) Wang, Y.; Patel, D. J. Guanine Residues in d(T2AG3) and d(T2G4) Form Parallel-Stranded Potassium Cation Stabilized G-Quadruplexes with Anti Glycosidic Torsion

Angles in Solution. *Biochemistry* **1992**, *31* (35), 8112–8119. https://doi.org/10.1021/bi00150a002.

(50) Neidle, S.; Parkinson, G. N. The Structure of Telomeric DNA. *Current Opinion in Structural Biology* **2003**, *13* (3), 275–283. https://doi.org/10.1016/S0959-440X(03)00072-1.

(51) Arora, R.; Brun, C. M. C.; Azzalin, C. M. TERRA: Long Noncoding RNA at Eukaryotic Telomeres. In *Long Non-Coding RNAs*; Ugarkovic, D., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; Vol. 51, pp 65–94. https://doi.org/10.1007/978-3-642-16502-3_4.

(52) Xu, Y.; Ishizuka, T.; Yang, J.; Ito, K.; Katada, H.; Komiyama, M.; Hayashi, T. Oligonucleotide Models of Telomeric DNA and RNA Form a Hybrid G-Quadruplex Structure as a Potential Component of Telomeres. *Journal of Biological Chemistry* **2012**, *287* (50), 41787–41796. https://doi.org/10.1074/jbc.M112.342030.

(53) Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. Inhibition of Telomerase by G-Quartet DMA Structures. *Nature* **1991**, *350* (6320), 718–720. https://doi.org/10.1038/350718a0.

(54) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proceedings of the National Academy of Sciences* **2002**, *99* (18), 11593–11598. https://doi.org/10.1073/pnas.182256799.

(55) Phan, A. T.; Modi, Y. S.; Patel, D. J. Propeller-Type Parallel-Stranded G-Quadruplexes in the Human *c-Myc* Promoter. *Journal of the American Chemical Society* **2004**, *126* (28), 8710–8716. https://doi.org/10.1021/ja048805k.

(56) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. Solution Structure of the Biologically Relevant G-Quadruplex Element in the Human c-MYC Promoter. Implications for G-Quadruplex Stabilization. *Biochemistry* **2005**, *44* (6), 2048–2058. https://doi.org/10.1021/bi048242p.

(57) Agrawal, P.; Hatzakis, E.; Guo, K.; Carver, M.; Yang, D. Solution Structure of the Major G-Quadruplex Formed in the Human VEGF Promoter in K+: Insights into Loop Interactions of the Parallel G-Quadruplexes. *Nucleic Acids Research* **2013**, *41* (22), 10584–10592. https://doi.org/10.1093/nar/gkt784.

(58) Dai, J.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D. An Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel G-Strands Formed in the Human BCL-2 Promoter Region in Solution. *Journal of the American Chemical Society* **2006**, *128* (4), 1096–1098. https://doi.org/10.1021/ja055636a.

(59) Dai, J.; Chen, D.; Jones, R. A.; Hurley, L. H.; Yang, D. NMR Solution Structure of the Major G-Quadruplex Structure Formed in the Human BCL2 Promoter Region. *Nucleic Acids Research* **2006**, *34* (18), 5133–5144. https://doi.org/10.1093/nar/gkl610.

(60) De Armond, R.; Wood, S.; Sun, D.; Hurley, L. H.; Ebbinghaus, S. W. Evidence for the Presence of a Guanine Quadruplex Forming Region within a Polypurine Tract of the Hypoxia Inducible Factor 1α Promoter. *Biochemistry* **2005**, *44* (49), 16341–16350. https://doi.org/10.1021/bi051618u.

(61) Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. Small-Molecule Interaction with a Five-Guanine-Tract G-Quadruplex Structure from the Human MYC Promoter. *Nature Chemical Biology* **2005**, *1* (3), 167–173. https://doi.org/10.1038/nchembio723.

(62) Gunaratnam, M.; Swank, S.; Haider, S. M.; Galesa, K.; Reszka, A. P.; Beltran, M.; Cuenca, F.; Fletcher, J. A.; Neidle, S. Targeting Human Gastrointestinal Stromal Tumor Cells with a Quadruplex-Binding Small Molecule. *Journal of Medicinal Chemistry* **2009**, *52* (12), 3774–3783. https://doi.org/10.1021/jm900424a.

(63) Jana, J.; Mondal, S.; Bhattacharjee, P.; Sengupta, P.; Roychowdhury, T.; Saha, P.; Kundu, P.; Chatterjee, S. Chelerythrine down Regulates Expression of VEGFA, BCL2 and KRAS by Arresting G-Quadruplex Structures at Their Promoter Regions. *Scientific Reports* **2017**, *7* (1), 40706. https://doi.org/10.1038/srep40706.

(64) Li, Y.; Zhang, X.; Gao, Y.; Shi, J.; Tang, L.; Sui, G. G-Quadruplexes in the BAP1 Promoter Positively Regulate Its Expression. *Experimental Cell Research* **2018**, *369* (1), 147–157. https://doi.org/10.1016/j.yexcr.2018.05.016.

(65) Zhao, Y.; Uhler, J. P. Identification of a G-Quadruplex Forming Sequence in the Promoter of UCP1. *Acta Biochimica et Biophysica Sinica* **2018**, *50* (7), 718–722. https://doi.org/10.1093/abbs/gmy059.

(66) Armas, P.; David, A.; Calcaterra, N. B. Transcriptional Control by G-Quadruplexes: In Vivo Roles and Perspectives for Specific Intervention. *Transcription* **2017**, *8* (1), 21–25. https://doi.org/10.1080/21541264.2016.1243505.

(67) Gray, L. T.; Vallur, A. C.; Eddy, J.; Maizels, N. G Quadruplexes Are Genomewide Targets of Transcriptional Helicases XPB and XPD. *Nature Chemical Biology* **2014**, *10* (4), 313–318. https://doi.org/10.1038/nchembio.1475.

(68) Fernando, H.; Sewitz, S.; Darot, J.; Tavaré, S.; Huppert, J. L.; Balasubramanian, S. Genome-Wide Analysis of a G-Quadruplex-Specific Single-Chain Antibody That Regulates Gene Expression. *Nucleic Acids Research* **2009**, *37* (20), 6716–6722. https://doi.org/10.1093/nar/gkp740.

(69) Xu, Y.; Kaminaga, K.; Komiyama, M. Human Telomeric RNA in G-Quadruplex Structure. *Nucleic Acids Symposium Series* **2008**, *52* (1), 175–176. https://doi.org/10.1093/nass/nrn089.

(70) Xu, Y.; Suzuki, Y.; Ito, K.; Komiyama, M. Telomeric Repeat-Containing RNA Structure in Living Cells. *Proceedings of the National Academy of Sciences* **2010**, *107* (33), 14579–14584. https://doi.org/10.1073/pnas.1001177107.

(71) Martadinata, H.; Phan, A. T. Structure of Human Telomeric RNA (TERRA): Stacking of Two G-Quadruplex Blocks in K^+ Solution. *Biochemistry* **2013**, *52* (13), 2176–2183. https://doi.org/10.1021/bi301606u.

(72) Gros, J.; Guédin, A.; Mergny, J.-L.; Lacroix, L. G-Quadruplex Formation Interferes with P1 Helix Formation in the RNA Component of Telomerase HTERC. *ChemBioChem* **2008**, *9* (13), 2075–2079. https://doi.org/10.1002/cbic.200800300.

(73) Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S. An RNA G-Quadruplex in the 5′ UTR of the NRAS Proto-Oncogene Modulates Translation. *Nature Chemical Biology* **2007**, *3* (4), 218–221. https://doi.org/10.1038/nchembio864.

(74) Beaudoin, J.-D.; Perreault, J.-P. 5’-UTR G-Quadruplex Structures Acting as Translational Repressors. *Nucleic Acids Research* **2010**, *38* (20), 7022–7036. https://doi.org/10.1093/nar/gkq557.

(75) Bugaut, A.; Balasubramanian, S. 5’-UTR RNA G-Quadruplexes: Translation Regulation and Targeting. *Nucleic Acids Research* **2012**, *40* (11), 4727–4741. https://doi.org/10.1093/nar/gks068.

(76) Paeschke, K.; McDonald, K. R.; Zakian, V. A. Telomeres: Structures in Need of Unwinding. *FEBS Letters* **2010**, *584* (17), 3760–3772. https://doi.org/10.1016/j.febslet.2010.07.007.

(77) Lin, W.; Sampathi, S.; Dai, H.; Liu, C.; Zhou, M.; Hu, J.; Huang, Q.; Campbell, J.; Shin-Ya, K.; Zheng, L.; Chai, W.; Shen, B. Mammalian DNA2 Helicase/Nuclease Cleaves G-Quadruplex DNA and Is Required for Telomere Integrity. *EMBO J* **2013**, *32* (10), 1425–1439. https://doi.org/10.1038/emboj.2013.88.

(78) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nature Chem* **2013**, *5* (3), 182–186. https://doi.org/10.1038/nchem.1548.
(79) Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. Visualization and Selective Chemical Targeting of RNA G-Quadruplex Structures in the Cytoplasm of Human Cells. *Nature Chem* **2014**, *6* (1), 75–80. https://doi.org/10.1038/nchem.1805.
(80) Krafcikova, M.; Hänsel-Hertsch, R.; Trantirek, L.; Foldynova-Trantirkova, S. In Cell NMR Spectroscopy: Investigation of G-Quadruplex Structures Inside Living Xenopus Laevis Oocytes. In *G-Quadruplex Nucleic Acids*; Yang, D., Lin, C., Eds.; Springer New York: New York, NY, 2019; Vol. 2035, pp 397–405. https://doi.org/10.1007/978-1-4939-9666-7_25.
(81) Arola, A.; Vilar, R. Stabilisation of G-Quadruplex DNA by Small Molecules. *CTMC* **2008**, *8* (15), 1405–1415. https://doi.org/10.2174/156802608786141106.
(82) Monchaud, D.; Teulade-Fichou, M.-P. A Hitchhiker's Guide to G-Quadruplex Ligands. *Organic & Biomolecular Chemistry* **2008**, *6* (4), 627–636. https://doi.org/10.1039/B714772B.
(83) Yan, Y.; Tan, J.; Ou, T.; Huang, Z.; Gu, L. DNA G-Quadruplex Binders: A Patent Review. *Expert Opinion on Therapeutic Patents* **2013**, *23* (11), 1495–1509. https://doi.org/10.1517/13543776.2013.833187.
(84) Gowan, S. M.; Harrison, J. R.; Patterson, L.; Valenti, M.; Read, M. A.; Neidle, S.; Kelland, L. R. A G-Quadruplex-Interactive Potent Small-Molecule Inhibitor of Telomerase Exhibiting in Vitro and in Vivo Antitumor Activity. *Molecular Pharmacology* **2002**, *61* (5), 1154–1162. https://doi.org/10.1124/mol.61.5.1154.
(85) Burger, A. M.; Dai, F.; Schultes, C. M.; Reszka, A. P.; Moore, M. J.; Double, J. A.; Neidle, S. The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Research* **2005**, *65* (4), 1489–1496. https://doi.org/10.1158/0008-5472.CAN-04-2910.
(86) Rodriguez, R.; Müller, S.; Yeoman, J. A.; Trentesaux, C.; Riou, J.-F.; Balasubramanian, S. A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres. *Journal of the American Chemical Society* **2008**, *130* (47), 15758–15759. https://doi.org/10.1021/ja805615w.
(87) Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. Small-Molecule–Induced DNA Damage Identifies Alternative DNA Structures in Human Genes. *Nature Chemical Biology* **2012**, *8* (3), 301–310. https://doi.org/10.1038/nchembio.780.
(88) Fedoroff, O. Yu.; Rangan, A.; Chemeris, V. V.; Hurley, L. H. Cationic Porphyrins Promote the Formation of I-Motif DNA and Bind Peripherally by a Nonintercalative Mechanism †. *Biochemistry* **2000**, *39* (49), 15083–15090. https://doi.org/10.1021/bi001528j.
(89) Grand, C. L.; Han, H.; Muñoz, R. M.; Weitman, S.; Von Hoff, D. D.; Hurley, L. H.; Bearss, D. J. The Cationic Porphyrin TMPyP4 Down-Regulates c-MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth in Vivo. *Molecular Cancer Therapeutics* **2002**, *1* (8), 565.
(90) Nakajima, A.; Tauchi, T.; Sashida, G.; Sumi, M.; Abe, K.; Yamamoto, K.; Ohyashiki, J. H.; Ohyashiki, K. Telomerase Inhibition Enhances Apoptosis in Human Acute Leukemia Cells: Possibility of Antitelomerase Therapy. *Leukemia* **2003**, *17* (3), 560–567. https://doi.org/10.1038/sj.leu.2402825.
(91) De Cian, A.; Cristofari, G.; Reichenbach, P.; De Lemos, E.; Monchaud, D.; Teulade-Fichou, M.-P.; Shin-ya, K.; Lacroix, L.; Lingner, J.; Mergny, J.-L. Reevaluation of

Telomerase Inhibition by Quadruplex Ligands and Their Mechanisms of Action. *Proceedings of the National Academy of Sciences* **2007**, *104* (44), 17347–17352. https://doi.org/10.1073/pnas.0707365104.

(92) Ahmed, S.; Henderson, E. Formation of Novel Hairpin Structures by Telemeric C-Strand Oligonucleotides. *Nucleic Acids Research* **1992**, *20* (3), 507–511. https://doi.org/10.1093/nar/20.3.507.

(93) Gehring, K.; Leroy, J.-L.; Guéron, M. A Tetrameric DNA Structure with Protonated Cytosine-Cytosine Base Pairs. *Nature* **1993**, *363* (6429), 561–565. https://doi.org/10.1038/363561a0.

(94) Kang, C. H.; Berger, I.; Lockshin, C.; Ratliff, R.; Moyzis, R.; Rich, A. Crystal Structure of Intercalated Four-Stranded d(C3T) at 1.4 A Resolution. *Proceedings of the National Academy of Sciences* **1994**, *91* (24), 11636–11640. https://doi.org/10.1073/pnas.91.24.11636.

(95) Berger, I.; Egli, M.; Rich, A. Inter-Strand C-H...O Hydrogen Bonds Stabilizing Four-Stranded Intercalated Molecules: Stereoelectronic Effects of O4' in Cytosine-Rich DNA. *Proceedings of the National Academy of Sciences USA* **1996**, *93* (22), 12116. https://doi.org/10.1073/pnas.93.22.12116.

(96) Guéron, M.; Leroy, J.-L. The I-Motif in Nucleic Acids. *Current Opinion in Structural Biology* **2000**, *10* (3), 326–331. https://doi.org/10.1016/S0959-440X(00)00091-9.

(97) Kanaori, K.; Maeda, A.; Kanehara, H.; Tajima, K.; Makino, K. 1H Nuclear Magnetic Resonance Study on Equilibrium between Two Four-Stranded Solution Conformations of Short d(CnT). *Biochemistry* **1998**, *37* (37), 12979–12986. https://doi.org/10.1021/bi980492g.

(98) Phan, A. T.; Guéron, M.; Leroy, J.-L. The Solution Structure and Internal Motions of a Fragment of the Cytidine-Rich Strand of the Human Telomere 1 1Edited by I. Tinoco. *Journal of Molecular Biology* **2000**, *299* (1), 123–144. https://doi.org/10.1006/jmbi.2000.3613.

(99) Lieblein, A. L.; Buck, J.; Schlepckow, K.; Fürtig, B.; Schwalbe, H. Time-Resolved NMR Spectroscopic Studies of DNA i-Motif Folding Reveal Kinetic Partitioning. *Angewandte Chemie* **2012**, *51* (1), 250–253. https://doi.org/10.1002/anie.201104938.

(100) Benabou, S.; Aviñó, A.; Eritja, R.; González, C.; Gargallo, R. Fundamental Aspects of the Nucleic Acid I-Motif Structures. *RSC Advances* **2014**, *4* (51), 26956–26980. https://doi.org/10.1039/C4RA02129K.

(101) Lacroix, L.; Mergny, J.-L. Chemical Modification of Pyrimidine TFOs: Effect on i-Motif and Triple Helix Formation. *Archives of Biochemistry and Biophysics* **2000**, *381* (1), 153–163. https://doi.org/10.1006/abbi.2000.1934.

(102) Bhavsar-Jog, Y. P.; Van Dornshuld, E.; Brooks, T. A.; Tschumper, G. S.; Wadkins, R. M. Epigenetic Modification, Dehydration, and Molecular Crowding Effects on the Thermodynamics of i-Motif Structure Formation from C-Rich DNA. *Biochemistry* **2014**, *53* (10), 1586–1594. https://doi.org/10.1021/bi401523b.

(103) Xu, B.; Devi, G.; Shao, F. Regulation of Telomeric I-Motif Stability by 5-Methylcytosine and 5-Hydroxymethylcytosine Modification. *Organic & Biomolecular Chemistry* **2015**, *13* (20), 5646–5651. https://doi.org/10.1039/C4OB02646B.

(104) Wright, E. P.; Lamparska, K.; Smith, S. S.; Waller, Z. A. E. Substitution of Cytosine with Guanylurea Decreases the Stability of I-Motif DNA. *Biochemistry* **2017**, *56* (36), 4879–4883. https://doi.org/10.1021/acs.biochem.7b00628.

(105) Mir, B.; Solés, X.; González, C.; Escaja, N. The Effect of the Neutral Cytidine Protonated Analogue Pseudoisocytidine on the Stability of I-Motif Structures. *Scientific Reports* **2017**, *7* (1), 2772. https://doi.org/10.1038/s41598-017-02723-y.

(106) Tsvetkov, V. B.; Zatsepin, T. S.; Belyaev, E. S.; Kostyukevich, Y. I.; Shpakovski, G. V.; Podgorsky, V. V.; Pozmogova, G. E.; Varizhuk, A. M.; Aralov, A. V. I-Clamp Phenoxazine for the Fine Tuning of DNA i-Motif Stability. *Nucleic Acids Research* **2018**, *46* (6), 2751–2764. https://doi.org/10.1093/nar/gky121.

(107) Wright, E. P.; Huppert, J. L.; Waller, Z. A. E. Identification of Multiple Genomic DNA Sequences Which Form I-Motif Structures at Neutral PH. *Nucleic Acids Research* **2017**, *45* (6), 2951–2959. https://doi.org/10.1093/nar/gkx090.

(108) Fleming, A. M.; Ding, Y.; Rogers, R. A.; Zhu, J.; Zhu, J.; Burton, A. D.; Carlisle, C. B.; Burrows, C. J. 4n–1 Is a "Sweet Spot" in DNA i-Motif Folding of 2′-Deoxycytidine Homopolymers. *Journal of the American Chemical Society* **2017**, *139* (13), 4682–4689. https://doi.org/10.1021/jacs.6b10117.

(109) Brazier, J. A.; Shah, A.; Brown, G. D. I-Motif Formation in Gene Promoters: Unusually Stable Formation in Sequences Complementary to Known G-Quadruplexes. *Chemical Communication* **2012**, *48* (87), 10739–10741. https://doi.org/10.1039/c2cc30863k.

(110) Mir, B.; Serrano, I.; Buitrago, D.; Orozco, M.; Escaja, N.; González, C. Prevalent Sequences in the Human Genome Can Form Mini I-Motif Structures at Physiological PH. *Journal of the American Chemical Society* **2017**, *139* (40), 13985–13988. https://doi.org/10.1021/jacs.7b07383.

(111) Brooks, T. A.; Kendrick, S.; Hurley, L. Making Sense of G-Quadruplex and i-Motif Functions in Oncogene Promoters: G-Quadruplex and i-Motif in Oncogene Promoters. *FEBS Journal* **2010**, *277* (17), 3459–3469. https://doi.org/10.1111/j.1742-4658.2010.07759.x.

(112) Yang, Y.; Sun, Y.; Yang, Y.; Xing, Y.; Zhang, T.; Wang, Z.; Yang, Z.; Liu, D. Influence of Tetra(Ethylene Glycol) (EG_4) Substitution at the Loop Region on the Intramolecular DNA *i* -Motif. *Macromolecules* **2012**, *45* (6), 2643–2647. https://doi.org/10.1021/ma300230q.

(113) Malliavin, T. E.; Gau, J.; Snoussi, K.; Leroy, J.-L. Stability of the I-Motif Structure Is Related to the Interactions between Phosphodiester Backbones. *Biophysical Journal* **2003**, *84* (6), 3838–3847. https://doi.org/10.1016/S0006-3495(03)75111-X.

(114) Kanehara, H.; Mizuguchi, M.; Tajima, K.; Kanaori, K.; Makino, K. Spectroscopic Evidence for the Formation of Four-Stranded Solution Structure of Oligodeoxycytidine Phosphorothioate. *Biochemistry* **1997**, *36* (7), 1790–1797. https://doi.org/10.1021/bi961528c.

(115) Mergny, J. Kinetics and Thermodynamics of I-DNA Formation: Phosphodiester versus Modified Oligodeoxynucleotides. *Nucleic Acids Research* **1998**, *26* (21), 4797–4803. https://doi.org/10.1093/nar/26.21.4797.

(116) Lacroix, L.; Mergny, J.-L.; Leroy, J.-L.; Hélène, C. Inability of RNA To Form the I-Motif: Implications for Triplex Formation. *Biochemistry* **1996**, *35* (26), 8715–8722. https://doi.org/10.1021/bi960107s.

(117) Collin, D.; Gehring, K. Stability of Chimeric DNA/RNA Cytosine Tetrads: Implications for i-Motif Formation by RNA. *Journal of the American Chemical Society* **1998**, *120* (17), 4069–4072. https://doi.org/10.1021/ja973346r.

(118) Robidoux, S.; Damha, M. J. D-2-Deoxyribose and D-Arabinose, but Not D-Ribose, Stabilize the Cytosine Tetrad (i-DNA) Structure. *Journal of Biomolecular Structure and Dynamics* **1997**, *15* (3), 529–535. https://doi.org/10.1080/07391102.1997.10508963.

(119) Mergny, J.-L.; Lacroix, L.; Han, X.; Leroy, J.-L.; Helene, C. Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an I-DNA Motif. *Journal of the American Chemical Society* **1995**, *117* (35), 8887–8898. https://doi.org/10.1021/ja00140a001.

(120) Benabou, S.; Ferreira, R.; Aviñó, A.; González, C.; Lyonnais, S.; Solà, M.; Eritja, R.; Jaumot, J.; Gargallo, R. Solution Equilibria of Cytosine- and Guanine-Rich Sequences

near the Promoter Region of the n-Myc Gene That Contain Stable Hairpins within Lateral Loops. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2014**, *1840* (1), 41–52. https://doi.org/10.1016/j.bbagen.2013.08.028.

(121) Miyoshi, D.; Matsumura, S.; Nakano, S.; Sugimoto, N. Duplex Dissociation of Telomere DNAs Induced by Molecular Crowding. *Journal of the American Chemical Society* **2004**, *126* (1), 165–169. https://doi.org/10.1021/ja036721q.

(122) Cui, J.; Waltman, P.; Le, V.; Lewis, E. The Effect of Molecular Crowding on the Stability of Human C-MYC Promoter Sequence I-Motif at Neutral PH. *Molecules* **2013**, *18* (10), 12751–12767. https://doi.org/10.3390/molecules181012751.

(123) Day, H. A.; Huguin, C.; Waller, Z. A. E. Silver Cations Fold I-Motif at Neutral PH. *Chemical Communication* **2013**, *49* (70), 7696–7698. https://doi.org/10.1039/c3cc43495h.

(124) Abdelhamid, M. A.; Fábián, L.; MacDonald, C. J.; Cheesman, M. R.; Gates, A. J.; Waller, Z. A. Redox-Dependent Control of i-Motif DNA Structure Using Copper Cations. *Nucleic Acids Research* **2018**, *46* (12), 5886–5893. https://doi.org/10.1093/nar/gky390.

(125) Saxena, S.; Joshi, S.; Shankaraswamy, J.; Tyagi, S.; Kukreti, S. Magnesium and Molecular Crowding of the Cosolutes Stabilize the I-Motif Structure at Physiological PH. *Biopolymers* **2017**, *107* (7), e23018. https://doi.org/10.1002/bip.23018.

(126) Phan, A. T. Human Telomeric DNA: G-Quadruplex, i-Motif and Watson-Crick Double Helix. *Nucleic Acids Research* **2002**, *30* (21), 4618–4625. https://doi.org/10.1093/nar/gkf597.

(127) Marsich, E.; Xodo, L. E.; Manzini, G. Widespread Presence in Mammals and High Binding Specificity of a Nuclear Protein That Recognises the Single-Stranded Telomeric Motif (CCCTAA)n. *European Journal of Biochemistry* **1998**, *258* (1), 93–99. https://doi.org/10.1046/j.1432-1327.1998.2580093.x.

(128) Lacroix, L. Identification of Two Human Nuclear Proteins That Recognise the Cytosine-Rich Strand of Human Telomeres in Vitro. *Nucleic Acids Research* **2000**, *28* (7), 1564–1575. https://doi.org/10.1093/nar/28.7.1564.

(129) Simonsson, T.; Pribylova, M.; Vorlickova, M. A Nuclease Hypersensitive Element in the Human C-Myc Promoter Adopts Several Distinct i-Tetraplex Structures. *Biochemical and Biophysical Research Communications* **2000**, *278* (1), 158–166. https://doi.org/10.1006/bbrc.2000.3783.

(130) González, V.; Hurley, L. H. The C-MYC NHE III1: Function and Regulation. *Annual Review of Pharmacology and Toxicology* **2010**, *50* (1), 111–129. https://doi.org/10.1146/annurev.pharmtox.48.113006.094649.

(131) Brooks, T. A.; Hurley, L. H. Targeting MYC Expression through G-Quadruplexes. *Genes & Cancer* **2010**, *1* (6), 641–649. https://doi.org/10.1177/1947601910377493.

(132) Kendrick, S.; Akiyama, Y.; Hecht, S. M.; Hurley, L. H. The I-Motif in the Bcl-2 P1 Promoter Forms an Unexpectedly Stable Structure with a Unique 8:5:7 Loop Folding Pattern. *Journal of the American Chemical Society* **2009**, *131* (48), 17667–17676. https://doi.org/10.1021/ja9076292.

(133) Kendrick, S.; Kang, H.-J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D.; Hecht, S. M.; Hurley, L. H. The Dynamic Character of the BCL2 Promoter I-Motif Provides a Mechanism for Modulation of Gene Expression by Compounds That Bind Selectively to the Alternative DNA Hairpin Structure. *Journal of the American Chemical Society* **2014**, *136* (11), 4161–4171. https://doi.org/10.1021/ja410934b.

(134) Nevins, J. R. The Rb/E2F Pathway and Cancer. *Human Molecular Genetics* **2001**, *10* (7), 699–703. https://doi.org/10.1093/hmg/10.7.699.

(135) Xu, Y. Formation of the G-Quadruplex and i-Motif Structures in Retinoblastoma Susceptibility Genes (Rb). *Nucleic Acids Research* **2006**, *34* (3), 949–954. https://doi.org/10.1093/nar/gkj485.
(136) Guo, K.; Pourpak, A.; Beetz-Rogers, K.; Gokhale, V.; Sun, D.; Hurley, L. H. Formation of Pseudosymmetrical G-Quadruplex and i-Motif Structures in the Proximal Promoter Region of the *RET* Oncogene. *Journal of the American Chemical Society* **2007**, *129* (33), 10220–10228. https://doi.org/10.1021/ja072185g.
(137) Guo, K.; Gokhale, V.; Hurley, L. H.; Sun, D. Intramolecularly Folded G-Quadruplex and i-Motif Structures in the Proximal Promoter of the Vascular Endothelial Growth Factor Gene. *Nucleic Acids Research* **2008**, *36* (14), 4598–4608. https://doi.org/10.1093/nar/gkn380.
(138) Bucek, P.; Jaumot, J.; AviÃ±Ã^3, A.; Eritja, R.; Gargallo, R. PH-Modulated Watson–Crick Duplex–Quadruplex Equilibria of Guanine-Rich and Cytosine-Rich DNA Sequences 140 Base Pairs Upstream of the *c-Kit* Transcription Initiation Site. *Chemistry: A European Journal* **2009**, *15* (46), 12663–12671. https://doi.org/10.1002/chem.200901631.
(139) Bucek, P.; Gargallo, R.; Kudrev, A. Spectrometric Study of the Folding Process of I-Motif-Forming DNA Sequences Upstream of the c-Kit Transcription Initiation Site. *Analytica Chimica Acta* **2010**, *683* (1), 69–77. https://doi.org/10.1016/j.aca.2010.10.008.
(140) Dzatko, S.; Krafcikova, M.; Hänsel-Hertsch, R.; Fessl, T.; Fiala, R.; Loja, T.; Krafcik, D.; Mergny, J.-L.; Foldynova-Trantirkova, S.; Trantirek, L. Evaluation of the Stability of DNA I-Motifs in the Nuclei of Living Mammalian Cells. *Angewandte Chemie.* **2018**, *57* (8), 2165–2169. https://doi.org/10.1002/anie.201712284.
(141) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. I-Motif DNA Structures Are Formed in the Nuclei of Human Cells. *Nature Chemistry* **2018**, *10* (6), 631–637. https://doi.org/10.1038/s41557-018-0046-3.
(142) Qin, T.; Liu, K.; Song, D.; Yang, C.; Su, H. Porphyrin Bound to I-Motifs: Intercalation versus External Groove Binding. *Chemistry Asian Journal* **2017**, *12* (13), 1578–1586. https://doi.org/10.1002/asia.201700398.
(143) Wang, L.; Wu, Y.; Chen, T.; Wei, C. The Interactions of Phenanthroline Compounds with DNAs: Preferential Binding to Telomeric Quadruplex over Duplex. *International Journal of Biological Macromolecules* **2013**, *52*, 1–8. https://doi.org/10.1016/j.ijbiomac.2012.08.015.
(144) Alberti, P.; Ren, J.; Teulade-Fichou, M. P.; Guittat, L.; Riou, J.-F.; Chaires, J. B.; Hélène, C.; Vigneron, J.-P.; Lehn, J.-M.; Mergny, J.-L. Interaction of an Acridine Dimer with DNA Quadruplex Structures. *Journal of Biomolecular Structure and Dynamics* **2001**, *19* (3), 505–513. https://doi.org/10.1080/07391102.2001.10506758.
(145) Xu, L.; Hong, S.; Sun, N.; Wang, K.; Zhou, L.; Ji, L.; Pei, R. Berberine as a Novel Light-up i-Motif Fluorescence Ligand and Its Application in Designing Molecular Logic Systems. *Chemical Communication* **2016**, *52* (1), 179–182. https://doi.org/10.1039/C5CC08242K.
(146) Sheng, Q.; Neaverson, J. C.; Mahmoud, T.; Stevenson, C. E. M.; Matthews, S. E.; Waller, Z. A. E. Identification of New DNA I-Motif Binding Ligands through a Fluorescent Intercalator Displacement Assay. *Organic & Biomolecular Chemistry* **2017**, *15* (27), 5669–5673. https://doi.org/10.1039/C7OB00710H.
(147) Pagano, A.; Iaccarino, N.; Abdelhamid, M. A. S.; Brancaccio, D.; Garzarella, E. U.; Di Porzio, A.; Novellino, E.; Waller, Z. A. E.; Pagano, B.; Amato, J.; Randazzo, A. Common G-Quadruplex Binding Agents Found to Interact With i-Motif-Forming

DNA: Unexpected Multi-Target-Directed Compounds. *Frontiers in Chemistry* **2018**, *6*, 281. https://doi.org/10.3389/fchem.2018.00281.

(148) Chen, Y.; Qu, K.; Zhao, C.; Wu, L.; Ren, J.; Wang, J.; Qu, X. Insights into the Biomedical Effects of Carboxylated Single-Wall Carbon Nanotubes on Telomerase and Telomeres. *Nature Communication* **2012**, *3* (1), 1074. https://doi.org/10.1038/ncomms2091.

(149) Brown, R. V.; Wang, T.; Chappeta, V. R.; Wu, G.; Onel, B.; Chawla, R.; Quijada, H.; Camp, S. M.; Chiang, E. T.; Lassiter, Q. R.; Lee, C.; Phanse, S.; Turnidge, M. A.; Zhao, P.; Garcia, J. G. N.; Gokhale, V.; Yang, D.; Hurley, L. H. The Consequences of Overlapping G-Quadruplexes and i-Motifs in the Platelet-Derived Growth Factor Receptor β Core Promoter Nuclease Hypersensitive Element Can Explain the Unexpected Effects of Mutations and Provide Opportunities for Selective Targeting of Both Structures by Small Molecules To Downregulate Gene Expression. *Journal of the American Chemical Society* **2017**, *139* (22), 7456–7475. https://doi.org/10.1021/jacs.6b10028.

(150) Shu, B.; Cao, J.; Kuang, G.; Qiu, J.; Zhang, M.; Zhang, Y.; Wang, M.; Li, X.; Kang, S.; Ou, T.-M.; Tan, J.-H.; Huang, Z.-S.; Li, D. Syntheses and Evaluation of New Acridone Derivatives for Selective Binding of Oncogene C- *Myc* Promoter i-Motifs in Gene Transcriptional Regulation. *Chemical Communication* **2018**, *54* (16), 2036–2039. https://doi.org/10.1039/C8CC00328A.

(151) Debnath, M.; Ghosh, S.; Chauhan, A.; Paul, R.; Bhattacharyya, K.; Dash, J. Preferential Targeting of I-Motifs and G-Quadruplexes by Small Molecules. *Chemical Science* **2017**, *8* (11), 7448–7456. https://doi.org/10.1039/C7SC02693E.

(152) Mergny, J.-L. Thermal Difference Spectra: A Specific Signature for Nucleic Acid Structures. *Nucleic Acids Research* **2005**, *33* (16), e138–e138. https://doi.org/10.1093/nar/gni134.

(153) Carvalho, J.; Queiroz, J. A.; Cruz, C. Circular Dichroism of G-Quadruplex: A Laboratory Experiment for the Study of Topology and Ligand Binding. *Journal of Chemical Education* **2017**, *94* (10), 1547–1551. https://doi.org/10.1021/acs.jchemed.7b00160.

(154) Adrian, M.; Heddi, B.; Phan, A. T. NMR Spectroscopy of G-Quadruplexes. *Methods* **2012**, *57* (1), 11–24. https://doi.org/10.1016/j.ymeth.2012.05.003.

(155) Phan, A. T.; Patel, D. J. Two-Repeat Human Telomeric d(TAGGGTTAGGGT) Sequence Forms Interconverting Parallel and Antiparallel G-Quadruplexes in Solution: Distinct Topologies, Thermodynamic Properties, and Folding/Unfolding Kinetics. *Journal of the American Chemical Society* **2003**, *125* (49), 15021–15027. https://doi.org/10.1021/ja037616j.

(156) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA. *Nature* **2002**, *417* (6891), 876–880. https://doi.org/10.1038/nature755.

(157) Campbell, N. H.; Parkinson, G. N. Crystallographic Studies of Quadruplex Nucleic Acids. *Methods* **2007**, *43* (4), 252–263. https://doi.org/10.1016/j.ymeth.2007.08.005.

(158) Liu, C.; Geng, Y.; Miao, H.; Shi, X.; You, Y.; Xu, N.; Zhou, B.; Zhu, G. G-Quadruplex Structures Formed by Human Telomeric DNA and C9orf72 Hexanucleotide Repeats. *Biophysics Reviews* **2019**, *11* (3), 389–393. https://doi.org/10.1007/s12551-019-00545-y.

(159) Palacký, J.; Vorlíčková, M.; Kejnovská, I.; Mojzeš, P. Polymorphism of Human Telomeric Quadruplex Structure Controlled by DNA Concentration: A Raman Study. *Nucleic Acids Research* **2012**, *41* (2), 1005–1016. https://doi.org/10.1093/nar/gks1135.

(160) Tuchscherer, G.; Dömer, B.; Sila, U.; Kamber, B.; Mutter, M. The TASP Concept : Mimetics of Peptide Ligands, Protein Surfaces and Folding Units. *Tetrahedron* **1993**, *49* (17), 3559–3575. https://doi.org/10.1016/S0040-4020(01)90215-7.
(161) Dumy, P.; Eggleston, I. M.; Esposito, G.; Nicula, S.; Mutter, M. Solution Structure of Regioselectively Addressable Functionalized Templates: An NMR and Restrained Molecular Dynamics Investigation. *Biopolymers* **1996**, *39* (3), 297–308. https://doi.org/10.1002/(SICI)1097-0282(199609)39:3<297::AID-BIP3>3.0.CO;2-J.
(162) Murat, P.; Bonnet, R.; Van der Heyden, A.; Spinelli, N.; Labbé, P.; Monchaud, D.; Teulade-Fichou, M.-P.; Dumy, P.; Defrancq, E. Template-Assembled Synthetic G-Quadruplex (TASQ): A Useful System for Investigating the Interactions of Ligands with Constrained Quadruplex Topologies. *Chemistry - A European Journal* **2010**, *16* (20), 6106–6114. https://doi.org/10.1002/chem.200903456.
(163) Murat, P.; Gennaro, B.; Garcia, J.; Spinelli, N.; Dumy, P.; Defrancq, E. The Use of a Peptidic Scaffold for the Formation of Stable Guanine Tetrads: Control of a H-Bonded Pattern in Water. *Chemistry: A European Journal* **2011**, *17* (21), 5791–5795. https://doi.org/10.1002/chem.201003556.
(164) Bonnet, R.; Lavergne, T.; Gennaro, B.; Spinelli, N.; Defrancq, E. Construction of Anti-Parallel G-Quadruplexes through Sequential Templated Click. *Chemical Communication* **2015**, *51* (23), 4850–4853. https://doi.org/10.1039/C4CC09744K.
(165) Bonnat, L.; Dejeu, J.; Bonnet, H.; Génnaro, B.; Jarjayes, O.; Thomas, F.; Lavergne, T.; Defrancq, E. Templated Formation of Discrete RNA and DNA:RNA Hybrid G-Quadruplexes and Their Interactions with Targeting Ligands. *Chemistry: A European Journal* **2016**, *22* (9), 3139–3147. https://doi.org/10.1002/chem.201504351.
(166) Bonnat, L.; Bar, L.; Génnaro, B.; Bonnet, H.; Jarjayes, O.; Thomas, F.; Dejeu, J.; Defrancq, E.; Lavergne, T. Template-Mediated Stabilization of a DNA G-Quadruplex Formed in the HIV-1 Promoter and Comparative Binding Studies. *Chemistry: A European Journal* **2017**, *23* (23), 5602–5613. https://doi.org/10.1002/chem.201700417.
(167) Bonnat, L.; Dautriche, M.; Saidi, T.; Revol-Cavalier, J.; Dejeu, J.; Defrancq, E.; Lavergne, T. Scaffold Stabilization of a G-Triplex and Study of Its Interactions with G-Quadruplex Targeting Ligands. *Organic & Biomolecular Chemistry* **2019**, *17* (38), 8726–8736. https://doi.org/10.1039/C9OB01537J.
(168) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **1963**, *85* (14), 2149–2154. https://doi.org/10.1021/ja00897a025.
(169) Matteucci, M. D.; Caruthers, M. H. Synthesis of Deoxyoligonucleotides on a Polymer Support. *Journal of the American Chemical Society* **1981**, *103* (11), 3185–3191. https://doi.org/10.1021/ja00401a041.
(170) Bonnet, R.; Murat, P.; Spinelli, N.; Defrancq, E. Click–Click Chemistry on a Peptidic Scaffold for Easy Access to Tetrameric DNA Structures. *Chemical Communication* **2012**, *48* (48), 5992–5994. https://doi.org/10.1039/C2CC32010J.
(171) Bonnat, L. Synthèse et Étude d'ADN et d'ARN G-Quadruplexes à Topologies Contrôlées. Applications Pour La Caractérisation et La Sélection de Ligands. PhD Thesis, 2017.
(172) Galibert, M.; Dumy, P.; Boturyn, D. One-Pot Approach to Well-Defined Biomolecular Assemblies by Orthogonal Chemoselective Ligations. *Angewandte Chemie* **2009**, *48* (14), 2576–2579. https://doi.org/10.1002/anie.200806223.
(173) Youn, Y. S.; Chae, S. Y.; Lee, S.; Kwon, M. J.; Shin, H. J.; Lee, K. C. Improved Peroral Delivery of Glucagon-like Peptide-1 by Site-Specific Biotin Modification: Design,

Preparation, and Biological Evaluation. *European Journal of Pharmaceutics and Biopharmaceutics* **2008**, *68* (3), 667–675. https://doi.org/10.1016/j.ejpb.2007.07.009.
(174) Foillard, S.; Rasmussen, M. O.; Razkin, J.; Boturyn, D.; Dumy, P. 1-Ethoxyethylidene, a New Group for the Stepwise SPPS of Aminooxyacetic Acid Containing Peptides. *The Journal of Organic Chemistry* **2008**, *73* (3), 983–991. https://doi.org/10.1021/jo701628k.
(175) Meyer, A.; Spinelli, N.; Dumy, P.; Vasseur, J.-J.; Morvan, F.; Defrancq, E. Oligonucleotide Sequential Bis-Conjugation via Click−Oxime and Click−Huisgen Procedures. *The Journal of Organic Chemistry* **2010**, *75* (11), 3927–3930. https://doi.org/10.1021/jo100599m.
(176) Lartia, R.; Murat, P.; Dumy, P.; Defrancq, E. Versatile Introduction of Azido Moiety into Oligonucleotides through Diazo Transfer Reaction. *Organic Letters* **2011**, *13* (20), 5672–5675. https://doi.org/10.1021/ol202397e.
(177) Kölmel, D. K.; Kool, E. T. Oximes and Hydrazones in Bioconjugation: Mechanism and Catalysis. *Chemical Reviews* **2017**, *117* (15), 10358–10376. https://doi.org/10.1021/acs.chemrev.7b00090.
(178) Dirksen, A.; Hackeng, T. M.; Dawson, P. E. Nucleophilic Catalysis of Oxime Ligation. *Angewandte Chemie* **2006**, *45* (45), 7581–7584. https://doi.org/10.1002/anie.200602877.
(179) Wang, S.; Gurav, D.; Oommen, O. P.; Varghese, O. P. Insights into the Mechanism and Catalysis of Oxime Coupling Chemistry at Physiological PH. *Chemistry: A European Journal* **2015**, *21* (15), 5980–5985. https://doi.org/10.1002/chem.201406458.
(180) Crisalli, P.; Kool, E. T. Water-Soluble Organocatalysts for Hydrazone and Oxime Formation. *The Journal of Organic Chemistry* **2013**, *78* (3), 1184–1189. https://doi.org/10.1021/jo302746p.
(181) Ollivier, N.; Olivier, C.; Gouyette, C.; Huynh-Dinh, T.; Gras-Masse, H.; Melnyk, O. Synthesis of Oligonucleotide–Peptide Conjugates Using Hydrazone Chemical Ligation. *Tetrahedron Letters* **2002**, *43* (6), 997–999. https://doi.org/10.1016/S0040-4039(01)02315-2.
(182) Achilles, K.; Kiedrowski, G. v. Kinetic Model Studies on the Chemical Ligation of Oligonucleotides via Hydrazone Formation. *Bioorganic & Medicinal Chemistry Letters* **2005**, *15* (4), 1229–1233. https://doi.org/10.1016/j.bmcl.2004.11.068.
(183) Lartia, R. Synthesis of 5′-Aldehyde Oligonucleotide. *Current Protocols in Nucleic Acid Chemistry* **2016**, *64* (1), 4.70.1-4.70.12. https://doi.org/10.1002/0471142700.nc0470s64.
(184) Meyer, A.; Vasseur, J.-J.; Dumy, P.; Morvan, F. Phthalimide-Oxy Derivatives for 3′- or 5′-Conjugation of Oligonucleotides by Oxime Ligation and Circularization of DNA by "Bis- or Tris-Click" Oxime Ligation: Phthalimide-Oxy Derivatives for 3′- or 5′-Conjugation of Oligonucleotides by Oxime Ligation and Circularization of DNA by "Bis- or Tris-Click" Oxime L. *The Journal of Organic Chemistry* **2017**, *2017* (46), 6931–6941. https://doi.org/10.1002/ejoc.201701317.
(185) Noël, M.; Clément-Blanc, C.; Meyer, A.; Vasseur, J.-J.; Morvan, F. Solid Supports for the Synthesis of 3′-Aminooxy Deoxy- or Ribo-Oligonucleotides and Their 3′-Conjugation by Oxime Ligation. *The Journal of Organic Chemistry* **2019**, *84* (22), 14854–14860. https://doi.org/10.1021/acs.joc.9b00848.
(186) Karskela, M.; Helkearo, M.; Virta, P.; Lönnberg, H. Synthesis of Oligonucleotide Glycoconjugates Using Sequential Click and Oximation Ligations. *Bioconjugate Chemistry* **2010**, *21* (4), 748–755. https://doi.org/10.1021/bc900529g.
(187) Estalayo-Adriàn, S.; Lartia, R.; Meyer, A.; Vasseur, J.-J.; Morvan, F.; Defrancq, E. Assessment of the Full Compatibility of Copper(I)-Catalyzed Alkyne-Azide

Cycloaddition and Oxime Click Reactions for Bis-Labelling of Oligonucleotides. *ChemistryOpen* **2015**, *4* (2), 169–173. https://doi.org/10.1002/open.201402099.
(188) Saxon, E.; Bertozzi, C. R. Cell Surface Engineering by a Modified Staudinger Reaction. *Science* **2000**, *287* (5460), 2007–2010. https://doi.org/10.1126/science.287.5460.2007.
(189) Lin, F. L.; Hoyt, H. M.; van Halbeek, H.; Bergman, R. G.; Bertozzi, C. R. Mechanistic Investigation of the Staudinger Ligation. *Journal of the American Chemical Society* **2005**, *127* (8), 2686–2695. https://doi.org/10.1021/ja044461m.
(190) Tornøe, C. W.; Christensen, C.; Meldal, M. Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *The Journal of Organic Chemistry* **2002**, *67* (9), 3057–3064. https://doi.org/10.1021/jo011148j.
(191) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angewandte Chemie International Edition* **2002**, *41* (14), 2596–2599.
(192) Meldal, M.; Tornøe, C. W. Cu-Catalyzed Azide−Alkyne Cycloaddition. *Chemical Reviews* **2008**, *108* (8), 2952–3015. https://doi.org/10.1021/cr0783479.
(193) Sun, X.-L.; Stabler, C. L.; Cazalis, C. S.; Chaikof, E. L. Carbohydrate and Protein Immobilization onto Solid Surfaces by Sequential Diels−Alder and Azide−Alkyne Cycloadditions. *Bioconjugate Chemistry* **2006**, *17* (1), 52–57. https://doi.org/10.1021/bc0502311.
(194) Gramlich, P. M. E.; Wirges, C. T.; Manetto, A.; Carell, T. Postsynthetic DNA Modification through the Copper-Catalyzed Azide-Alkyne Cycloaddition Reaction. *Angewandte Chemie* **2008**, *47* (44), 8350–8358. https://doi.org/10.1002/anie.200802077.
(195) El-Sagheer, A. H.; Brown, T. Click Chemistry with DNA. *Chemical Society Reviews* **2010**, *39* (4), 1388–1405. https://doi.org/10.1039/b901971p.
(196) Lartia, R.; Bonnat, L.; Murat, P.; Defrancq, E. Preparation of Azido Containing Oligonucleotides Through Diazo Transfer Reaction. *Current Protocols in Nucleic Acid Chemistry* **2012**, *50* (1). https://doi.org/10.1002/0471142700.nc0450s50.
(197) Brewer, G. J. Risks of Copper and Iron Toxicity during Aging in Humans. *Chemical Research in Toxicology* **2010**, *23* (2), 319–326. https://doi.org/10.1021/tx900338d.
(198) Yang, M.; Jalloh, A. S.; Wei, W.; Zhao, J.; Wu, P.; Chen, P. R. Biocompatible Click Chemistry Enabled Compartment-Specific PH Measurement inside E. Coli. *Nature Communication* **2014**, *5* (1), 4981. https://doi.org/10.1038/ncomms5981.
(199) Hong, V.; Presolski, S.; Ma, C.; Finn, M. Analysis and Optimization of Copper-Catalyzed Azide–Alkyne Cycloaddition for Bioconjugation. *Angewandte Chemie International Edition* **2009**, *48* (52), 9879–9883.
(200) Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. G. Labeling Live Cells by Copper-Catalyzed Alkyne−Azide Click Chemistry. *Bioconjugate Chemistry* **2010**, *21* (10), 1912–1916. https://doi.org/10.1021/bc100272z.
(201) Blomquist, A. T.; Liu, L. H. Many-Membered Carbon Rings. VII. Cycloöctyne. *Journal of the American Chemical Society* **1953**, *75* (9), 2153–2154. https://doi.org/10.1021/ja01105a039.
(202) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A Strain-Promoted [3 + 2] Azide−Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *Journal of the American Chemical Society* **2004**, *126* (46), 15046–15047. https://doi.org/10.1021/ja044996f.

(203) Agard, N. J.; Baskin, J. M.; Prescher, J. A.; Lo, A.; Bertozzi, C. R. A Comparative Study of Bioorthogonal Reactions with Azides. *ACS Chemical Biology* **2006**, *1* (10), 644–648. https://doi.org/10.1021/cb6003228.
(204) Jewett, J. C.; Bertozzi, C. R. Cu-Free Click Cycloaddition Reactions in Chemical Biology. *Chemical Society Reviews* **2010**, *39* (4), 1272–1279.
(205) Debets, M. F.; van Berkel, S. S.; Dommerholt, J.; Dirks, A. (Ton) J.; Rutjes, F. P. J. T.; van Delft, F. L. Bioconjugation with Strained Alkenes and Alkynes. *Accounts of Chemical Research* **2011**, *44* (9), 805–815. https://doi.org/10.1021/ar200059z.
(206) Dommerholt, J.; Schmidt, S.; Temming, R.; Hendriks, L. J. A.; Rutjes, F. P. J. T.; van Hest, J. C. M.; Lefeber, D. J.; Friedl, P.; van Delft, F. L. Readily Accessible Bicyclononynes for Bioorthogonal Labeling and Three-Dimensional Imaging of Living Cells. *Angewandte Chemie International Edition* **2010**, *49* (49), 9422–9425. https://doi.org/10.1002/anie.201003761.
(207) Takayama, Y.; Kusamori, K.; Nishikawa, M. Click Chemistry as a Tool for Cell Engineering and Drug Delivery. *Molecules* **2019**, *24* (1), 172. https://doi.org/10.3390/molecules24010172.
(208) Jawalekar, A.; Malik, S.; Verkade, J.; Gibson, B.; Barta, N.; Hodges, J.; Rowan, A.; van Delft, F. Oligonucleotide Tagging for Copper-Free Click Conjugation. *Molecules* **2013**, *18* (7), 7346–7363. https://doi.org/10.3390/molecules18077346.
(209) Haque, M. M.; Peng, X. DNA-Associated Click Chemistry. *Science China Chemistry* **2014**, *57* (2), 215–231. https://doi.org/10.1007/s11426-013-5035-1.
(210) Shangguan, N.; Katukojvala, S.; Greenberg, R.; Williams, L. J. The Reaction of Thio Acids with Azides: A New Mechanism and New Synthetic Applications. *Journal of the American Chemical Society* **2003**, *125* (26), 7754–7755. https://doi.org/10.1021/ja0294919.
(211) Mhidia, R.; Bézière, N.; Blanpain, A.; Pommery, N.; Melnyk, O. Assembly/Disassembly of Drug Conjugates Using Imide Ligation. *Organic Letters* **2010**, *12* (18), 3982–3985. https://doi.org/10.1021/ol101049g.
(212) Xie, S.; Fukumoto, R.; Ramström, O.; Yan, M. Anilide Formation from Thioacids and Perfluoroaryl Azides. *The Journal of Organic Chemistry* **2015**, *80* (9), 4392–4397. https://doi.org/10.1021/acs.joc.5b00240.
(213) Rijkers, D. T. S.; Merkx, R.; Yim, C.-B.; Brouwer, A. J.; Liskamp, R. M. J. 'Sulfo-Click' for Ligation as Well as for Site-Specific Conjugation with Peptides, Fluorophores, and Metal Chelators. *Journal of Peptide Science* **2010**, *16* (1), 1–5. https://doi.org/10.1002/psc.1197.
(214) Clavé, G.; Dursun, E.; Vasseur, J.-J.; Smietana, M. An Entry of the Chemoselective Sulfo-Click Reaction into the Sphere of Nucleic Acids. *Organic Letters* **2020**, *22* (5), 1914–1918. https://doi.org/10.1021/acs.orglett.0c00265.
(215) Hermanson, G. T. The Reactions of Bioconjugation. In *Bioconjugate Techniques*; Elsevier, 2013; pp 229–258. https://doi.org/10.1016/B978-0-12-382239-0.00003-0.
(216) Lindgren, J.; Eriksson Karlström, A. Intramolecular Thioether Crosslinking of Therapeutic Proteins to Increase Proteolytic Stability. *ChemBioChem* **2014**, *15* (14), 2132–2138. https://doi.org/10.1002/cbic.201400002.
(217) Joshi, T.; Patra, M.; Spiccia, L.; Gasser, G. Di-Heterometalation of Thiol-Functionalized Peptide Nucleic Acids. *Artificial DNA: PNA & XNA* **2013**, *4* (1), 11–18. https://doi.org/10.4161/adna.24019.
(218) Olszewska, A.; Pohl, R.; Brázdová, M.; Fojta, M.; Hocek, M. Chloroacetamide-Linked Nucleotides and DNA for Cross-Linking with Peptides and Proteins. *Bioconjugate Chemistry* **2016**, *27* (9), 2089–2094. https://doi.org/10.1021/acs.bioconjchem.6b00342.

(219) Viault, G.; Dautrey, S.; Maindron, N.; Hardouin, J.; Renard, P.-Y.; Romieu, A. The First "Ready-to-Use" Benzene-Based Heterotrifunctional Cross-Linker for Multiple Bioconjugation. *Organic & Biomolecular Chemistry* **2013**, *11* (16), 2693–2705.
(220) Paris, C.; Brun, O.; Pedroso, E.; Grandas, A. Exploiting Protected Maleimides to Modify Oligonucleotides, Peptides and Peptide Nucleic Acids. *Molecules* **2015**, *20* (4), 6389–6408. https://doi.org/10.3390/molecules20046389.
(221) Ning, X.; Temming, R. P.; Dommerholt, J.; Guo, J.; Ania, D. B.; Debets, M. F.; Wolfert, M. A.; Boons, G.-J.; van Delft, F. L. Protein Modification by Strain-Promoted Alkyne-Nitrone Cycloaddition. *Angewandte Chemie International Edition* **2010**, *49* (17), 3065–3068. https://doi.org/10.1002/anie.201000408.
(222) Jawalekar, A. M.; Reubsaet, E.; Rutjes, F. P. J. T.; van Delft, F. L. Synthesis of Isoxazoles by Hypervalent Iodine-Induced Cycloaddition of Nitrile Oxides to Alkynes. *Chemical Communication* **2011**, *47* (11), 3198–3200.
(223) Pasinszki, T.; Hajgató, B.; Havasi, B.; Westwood, N. P. C. Dimerisation of Nitrile Oxides: A Quantum-Chemical Study. *Physical Chemistry Chemical Physics* **2009**, *11* (26), 5263–5272. https://doi.org/10.1039/b823406j.
(224) Singh, I.; Vyle, J. S.; Heaney, F. Fast, Copper-Free Click Chemistry: A Convenient Solid-Phase Approach to Oligonucleotide Conjugation. *Chemical Communication* **2009**, No. 22, 3276–3278. https://doi.org/10.1039/b904185k.
(225) Gutsmiedl, K.; Fazio, D.; Carell, T. High-Density DNA Functionalization by a Combination of Cu-Catalyzed and Cu-Free Click Chemistry. *Chemistry: A European Journal* **2010**, *16* (23), 6877–6883. https://doi.org/10.1002/chem.201000363.
(226) Gutsmiedl, K.; Wirges, C. T.; Ehmke, V.; Carell, T. Copper-Free "Click" Modification of DNA via Nitrile Oxide−Norbornene 1,3-Dipolar Cycloaddition. *Organic Letters* **2009**, *11* (11), 2405–2408. https://doi.org/10.1021/ol9005322.
(227) Hansen, T. V.; Wu, P.; Fokin, V. V. One-Pot Copper(I)-Catalyzed Synthesis of 3,5-Disubstituted Isoxazoles. *The Journal of Organic Chemistry* **2005**, *70* (19), 7761–7764. https://doi.org/10.1021/jo050163b.
(228) Kankala, S.; Vadde, R.; Vasam, C. S. N-Heterocyclic Carbene-Catalyzed 1,3-Dipolar Cycloaddition Reactions: A Facile Synthesis of 3,5-Di- and 3,4,5-Trisubstituted Isoxazoles. *Organic & Biomolecular Chemistry* **2011**, *9* (22), 7869–7876. https://doi.org/10.1039/c1ob06072d.
(229) Grecian, S.; Fokin, V. V. Ruthenium-Catalyzed Cycloaddition of Nitrile Oxides and Alkynes: Practical Synthesis of Isoxazoles. *Angewandte Chemie International Edition* **2008**, *47* (43), 8285–8287. https://doi.org/10.1002/anie.200801920.
(230) Singh, I.; Heaney, F. Solid Phase Strain Promoted "Click" Modification of DNAvia[3+2]-Nitrile Oxide–Cyclooctyne Cycloadditions. *Chemical Communication* **2011**, *47* (9), 2706–2708. https://doi.org/10.1039/C0CC03985C.
(231) McKay, C. S.; Moran, J.; Pezacki, J. P. Nitrones as Dipoles for Rapid Strain-Promoted 1,3-Dipolar Cycloadditions with Cyclooctynes. *Chemical Communication* **2010**, *46* (6), 931–933. https://doi.org/10.1039/B921630H.
(232) Oliveira, B. L.; Guo, Z.; Bernardes, G. J. L. Inverse Electron Demand Diels–Alder Reactions in Chemical Biology. *Chemical Society Reviews* **2017**, *46* (16), 4895–4950. https://doi.org/10.1039/C7CS00184C.
(233) Devaraj, N. K.; Weissleder, R. Biomedical Applications of Tetrazine Cycloadditions. *Accounts of Chemical Research* **2011**, *44* (9), 816–827. https://doi.org/10.1021/ar200037t.
(234) Schoch, J.; Wiessler, M.; Jäschke, A. Post-Synthetic Modification of DNA by Inverse-Electron-Demand Diels−Alder Reaction. *Journal of the American Chemical Society* **2010**, *132* (26), 8846–8847. https://doi.org/10.1021/ja102871p.

(235) Song, W.; Wang, Y.; Qu, J.; Madden, M. M.; Lin, Q. A Photoinducible 1,3-Dipolar Cycloaddition Reaction for Rapid, Selective Modification of Tetrazole-Containing Proteins. *Angew. Chem. Int. Ed.* **2008**, *47* (15), 2832–2835. https://doi.org/10.1002/anie.200705805.

(236) Yu, Z.; Pan, Y.; Wang, Z.; Wang, J.; Lin, Q. Genetically Encoded Cyclopropene Directs Rapid, Photoclick-Chemistry-Mediated Protein Labeling in Mammalian Cells. *Angewandte Chemie International Edition* **2012**, *51* (42), 10600–10604. https://doi.org/10.1002/anie.201205352.

(237) Kaya, E.; Vrabel, M.; Deiml, C.; Prill, S.; Fluxa, V. S.; Carell, T. A Genetically Encoded Norbornene Amino Acid for the Mild and Selective Modification of Proteins in a Copper-Free Click Reaction. *Angewandte Chemie International Edition* **2012**, *51* (18), 4466–4469. https://doi.org/10.1002/anie.201109252.

(238) Wang, Y.; Song, W.; Hu, W. J.; Lin, Q. Fast Alkene Functionalization In Vivo by Photoclick Chemistry: HOMO Lifting of Nitrile Imine Dipoles. *Angewandte Chemie International Edition* **2009**, *48* (29), 5330–5333. https://doi.org/10.1002/anie.200901220.

(239) Favalli, N.; Bassi, G.; Zanetti, T.; Scheuermann, J.; Neri, D. Screening of Three Transition Metal-Mediated Reactions Compatible with DNA-Encoded Chemical Libraries. *HCA* **2019**, *102* (4), e1900033. https://doi.org/10.1002/hlca.201900033.

(240) Kölmel, D. K.; Barandun, L. J.; Kool, E. T. Efficient Synthesis of Fluorescent Alkynyl C-Nucleosides via Sonogashira Coupling for the Preparation of DNA-Based Polyfluorophores. *Organic & Biomolecular Chemistry* **2016**, *14* (27), 6407–6412. https://doi.org/10.1039/C6OB01199C.

(241) de Pedro Beato, E.; Priego, J.; Gironda-Martínez, A.; González, F.; Benavides, J.; Blas, J.; Martín-Ortega, M. D.; Toledo, M. Á.; Ezquerra, J.; Torrado, A. Mild and Efficient Palladium-Mediated C–N Cross-Coupling Reaction between DNA-Conjugated Aryl Bromides and Aromatic Amines. *ACS Combinatorial Science* **2019**, *21* (2), 69–74. https://doi.org/10.1021/acscombsci.8b00142.

(242) Dong, J.; Krasnova, L.; Finn, M. G.; Sharpless, K. B. Sulfur(VI) Fluoride Exchange (SuFEx): Another Good Reaction for Click Chemistry. *Angewandte Chemie International Edition* **2014**, *53* (36), 9430–9448. https://doi.org/10.1002/anie.201309399.

(243) Xu, H.; Ma, F.; Wang, N.; Hou, W.; Xiong, H.; Lu, F.; Li, J.; Wang, S.; Ma, P.; Yang, G.; Lerner, R. A. DNA-Encoded Libraries: Aryl Fluorosulfonates as Versatile Electrophiles Enabling Facile On-DNA Suzuki, Sonogashira, and Buchwald Reactions. *Advanced Science* **2019**, *6* (23), 1901551. https://doi.org/10.1002/advs.201901551.

(244) Wurm, F. R.; Klok, H.-A. Be Squared: Expanding the Horizon of Squaric Acid-Mediated Conjugations. *Chemical Society Reviews* **2013**, *42* (21), 8220–8236. https://doi.org/10.1039/c3cs60153f.

(245) Fonvielle, M.; Sakkas, N.; Iannazzo, L.; Le Fournis, C.; Patin, D.; Mengin-Lecreulx, D.; El-Sagheer, A.; Braud, E.; Cardon, S.; Brown, T.; Arthur, M.; Etheve-Quelquejeu, M. Electrophilic RNA for Peptidyl-RNA Synthesis and Site-Specific Cross-Linking with TRNA-Binding Enzymes. *Angewandte Chemie International Edition* **2016**, *55* (43), 13553–13557. https://doi.org/10.1002/anie.201606843.

(246) Schreiber, C. L.; Smith, B. D. Molecular Conjugation Using Non-Covalent Click Chemistry. *Nature Reviews Chemistry* **2019**, *3* (6), 393–400. https://doi.org/10.1038/s41570-019-0095-1.

(247) Bode, J. W.; Fox, R. M.; Baucom, K. D. Chemoselective Amide Ligations by Decarboxylative Condensations of N-Alkylhydroxylamines and α-Ketoacids.

Angewandte Chemie International Edition **2006**, *45* (8), 1248–1252. https://doi.org/10.1002/anie.200503991.
(248) Dumas, A. M.; Molander, G. A.; Bode, J. W. Amide-Forming Ligation of Acyltrifluoroborates and Hydroxylamines in Water. *Angewandte Chemie International Edition* **2012**, *51* (23), 5683–5686. https://doi.org/10.1002/anie.201201077.
(249) Noda, H.; Erős, G.; Bode, J. W. Rapid Ligations with Equimolar Reactants in Water with the Potassium Acyltrifluoroborate (KAT) Amide Formation. *Journal of the American Chemical Society* **2014**, *136* (15), 5611–5614. https://doi.org/10.1021/ja5018442.
(250) Liu, S. M.; Mazunin, D.; Pattabiraman, V. R.; Bode, J. W. Synthesis of Bifunctional Potassium Acyltrifluoroborates. *Organic Letters* **2016**, *18* (20), 5336–5339. https://doi.org/10.1021/acs.orglett.6b02652.
(251) Byrne, C.; McEwan, P. A.; Emsley, J.; Fischer, P. M.; Chan, W. C. End-Stapled Homo and Hetero Collagen Triple Helices: A Click Chemistry Approach. *Chemical Communication* **2011**, *47* (9), 2589–2591. https://doi.org/10.1039/C0CC04795C.
(252) Valton, A.-L.; Prioleau, M.-N. G-Quadruplexes in DNA Replication: A Problem or a Necessity? *Trends in Genetics* **2016**, *32* (11), 697–706. https://doi.org/10.1016/j.tig.2016.09.004.
(253) Hamperl, S.; Cimprich, K. A. The Contribution of Co-Transcriptional RNA:DNA Hybrid Structures to DNA Damage and Genome Instability. *DNA Repair* **2014**, *19*, 84–94. https://doi.org/10.1016/j.dnarep.2014.03.023.
(254) Tornaletti, S. Transcriptional Processing of G4 DNA. *Molecular Carcinogenesis* **2009**, *48* (4), 326–335. https://doi.org/10.1002/mc.20513.
(255) Leppek, K.; Das, R.; Barna, M. Functional 5′ UTR MRNA Structures in Eukaryotic Translation Regulation and How to Find Them. *Nature Reviews Molecular Cell Biology* **2018**, *19* (3), 158–174. https://doi.org/10.1038/nrm.2017.103.
(256) Grinstein, E.; Du, Y.; Santourlidis, S.; Christ, J.; Uhrberg, M.; Wernet, P. Nucleolin Regulates Gene Expression in CD34-Positive Hematopoietic Cells. *Journal of Biological Chemistry* **2007**, *282* (17), 12439–12449. https://doi.org/10.1074/jbc.M608068200.
(257) González, V.; Guo, K.; Hurley, L.; Sun, D. Identification and Characterization of Nucleolin as a C- *Myc* G-Quadruplex-Binding Protein. *Journal of Biological Chemistry* **2009**, *284* (35), 23622–23635. https://doi.org/10.1074/jbc.M109.018028.
(258) Arimondo, P. B. Interaction of Human DNA Topoisomerase I with G-Quartet Structures. *Nucleic Acids Research* **2000**, *28* (24), 4832–4838. https://doi.org/10.1093/nar/28.24.4832.
(259) Sun, H.; Yabuki, A.; Maizels, N. A Human Nuclease Specific for G4 DNA. *Proceedings of the National Academy of Sciences* **2001**, *98* (22), 12444–12449. https://doi.org/10.1073/pnas.231479198.
(260) Assenmacher, N.; Hopfner, K.-P. MRE11/RAD50/NBS1: Complex Activities. *Chromosoma* **2004**, *113* (4), 157–166. https://doi.org/10.1007/s00412-004-0306-4.
(261) Ghosal, G.; Muniyappa, K. Saccharomyces Cerevisiae Mre11 Is a High-Affinity G4 DNA-Binding Protein and a G-Rich DNA-Specific Endonuclease: Implications for Replication of Telomeric DNA. *Nucleic Acids Research* **2005**, *33* (15), 4692–4703. https://doi.org/10.1093/nar/gki777.
(262) Mendoza, O.; Bourdoncle, A.; Boulé, J.-B.; Brosh, R. M., Jr; Mergny, J.-L. G-Quadruplexes and Helicases. *Nucleic Acids Research* **2016**, *44* (5), 1989–2006. https://doi.org/10.1093/nar/gkw079.

(263) Sauer, M.; Paeschke, K. G-Quadruplex Unwinding Helicases and Their Function in Vivo. *Biochemical Society Transactions* **2017**, *45* (5), 1173–1182. https://doi.org/10.1042/BST20170097.
(264) Bernstein, K. A.; Gangloff, S.; Rothstein, R. The RecQ DNA Helicases in DNA Repair. *Annual Review of Genetics* **2010**, *44* (1), 393–417. https://doi.org/10.1146/annurev-genet-102209-163602.
(265) Brosh, R. M.; Bohr, V. A. Human Premature Aging, DNA Repair and RecQ Helicases. *Nucleic Acids Research* **2007**, *35* (22), 7527–7544. https://doi.org/10.1093/nar/gkm1008.
(266) Lavergne, T.; Bertrand, J.-R.; Vasseur, J.-J.; Debart, F. A Base-Labile Group for 2′-OH Protection of Ribonucleosides: A Major Challenge for RNA Synthesis. *Chemistry – A European Journal* **2008**, *14* (30), 9135–9138. https://doi.org/10.1002/chem.200801392.
(267) Krüger, A. C.; Raarup, M. K.; Nielsen, M. M.; Kristensen, M.; Besenbacher, F.; Kjems, J.; Birkedal, V. Interaction of HnRNP A1 with Telomere DNA G-Quadruplex Structures Studied at the Single Molecule Level. *European Biophysics Journal* **2010**, *39* (9), 1343–1350. https://doi.org/10.1007/s00249-010-0587-x.
(268) Paramasivam, M.; Membrino, A.; Cogoi, S.; Fukuda, H.; Nakagama, H.; Xodo, L. E. Protein HnRNP A1 and Its Derivative Up1 Unfold Quadruplex DNA in the Human KRAS Promoter: Implications for Transcription. *Nucleic Acids Research* **2009**, *37* (9), 2841–2853. https://doi.org/10.1093/nar/gkp138.
(269) Creacy, S. D.; Routh, E. D.; Iwamoto, F.; Nagamine, Y.; Akman, S. A.; Vaughn, J. P. G4 Resolvase 1 Binds Both DNA and RNA Tetramolecular Quadruplex with High Affinity and Is the Major Source of Tetramolecular Quadruplex G4-DNA and G4-RNA Resolving Activity in HeLa Cell Lysates. *Journal of Biological Chemistry* **2008**, *283* (50), 34626–34634. https://doi.org/10.1074/jbc.M806277200.
(270) David, A. P.; Pipier, A.; Pascutti, F.; Binolfi, A.; Weiner, A. M. J.; Challier, E.; Heckel, S.; Calsou, P.; Gomez, D.; Calcaterra, N. B.; Armas, P. CNBP Controls Transcription by Unfolding DNA G-Quadruplex Structures. *Nucleic Acids Research* **2019**, *47* (15), 7901–7913. https://doi.org/10.1093/nar/gkz527.
(271) Song, J.; Perreault, J.-P.; Topisirovic, I.; Richard, S. RNA G-Quadruplexes and Their Potential Regulatory Roles in Translation. *Translation* **2016**, *4* (2), e1244031. https://doi.org/10.1080/21690731.2016.1244031.
(272) Anisenko, A. N.; Knyazhanskaya, E. S.; Zatsepin, T. S.; Gottikh, M. B. Human Ku70 Protein Binds Hairpin RNA and Double Stranded DNA through Two Different Sites. *Biochimie* **2017**, *132*, 85–93. https://doi.org/10.1016/j.biochi.2016.11.001.
(273) Liu, X.; Kraus, W. L.; Bai, X. Ready, Pause, Go: Regulation of RNA Polymerase II Pausing and Release by Cellular Signaling Pathways. *Trends in Biochemical Sciences* **2015**, *40* (9), 516–525. https://doi.org/10.1016/j.tibs.2015.07.003.
(274) Chen, F. X.; Smith, E. R.; Shilatifard, A. Born to Run: Control of Transcription Elongation by RNA Polymerase II. *Nature Reviews Molecular Cell Biology* **2018**, *19* (7), 464–478. https://doi.org/10.1038/s41580-018-0010-5.
(275) Sollier, J.; Cimprich, K. A. Breaking Bad: R-Loops and Genome Integrity. *Trends in Cell Biology* **2015**, *25* (9), 514–522. https://doi.org/10.1016/j.tcb.2015.05.003.
(276) Niehrs, C.; Luke, B. Regulatory R-Loops as Facilitators of Gene Expression and Genome Stability. *Nature Reviews Molecular Cell Biology* **2020**, *21* (3), 167–178. https://doi.org/10.1038/s41580-019-0206-3.
(277) De Magis, A.; Manzo, S. G.; Russo, M.; Marinello, J.; Morigi, R.; Sordet, O.; Capranico, G. DNA Damage and Genome Instability by G-Quadruplex Ligands Are

Mediated by R Loops in Human Cancer Cells. *Proceedings of the National Academy of Sciences* **2019**, *116* (3), 816–825. https://doi.org/10.1073/pnas.1810409116.

Conception et application d'ADN tétraplexe à topologie contrôlée

Les acides nucléiques riches en cytosines ou en guanines peuvent se replier pour former des structures i-motifs ou G-quadruplexes (G4). G4 et i-motifs sont localisés au niveau des télomères et dans les régions promotrices de nombreux gènes, y compris des oncogènes. Il a été démontré que ces derniers contribuent à la régulation des cellules normales et pathologiques et sont désormais considérés comme des cibles thérapeutiques potentielles pour le traitement de nombreuses maladies et anomalies. Une caractéristique majeure des G4 est leur susceptibilité à adopter, en particulier in vitro, différentes topologies. D'autre part, les i-motifs nécessitent des conditions acides pour se former in vitro. Dans ce contexte, notre laboratoire a développé le concept de TASQ (pour « Template Assembled Synthetic G-Quadruplex ») pour assembler des mimes stables de G4 et d'i-motif.

La première partie de notre étude a consisté à synthétiser une structure stable i-motif provenant de la séquence télomérique en utilisant une plateforme cyclopeptidique fonctionnelle et des ligations chimiosélectives distinctes. Puis dans un second temps, une nouvelle ligation a été introduite aux trois autres chimies déjà en place dans notre laboratoire afin de synthétiser de nouvelles structures tétraplexes. Enfin, un panel de topologies G4 a été synthétisé afin de réaliser des études d'interactions visant à identifier des protéines qui interagissent avec des G4 adoptant des topologies bien définies.

Mots clés : G-quadruplexe, i-motif, ligation chimiosélective, gabarit cyclodécapeptidique, études d'interactions avec des protéines

Design and application of DNA tetraplex with controlled topology

Cytosines or guanines rich nucleic acids can fold into i-motifs or G-quadruplex (G4) structures. G4s and i-motifs are found at telomeres and in promoter regions of numerous genes including oncogenes. Both have been shown to contribute to the regulation of normal and pathological cells and are now viewed as potential therapeutic targets for the treatment of numerous diseases and defects. One major feature of G4s is their susceptibility to adopt, especially in vitro, different topologies. On the other hand i-motifs require acidic conditions to fold in vitro. In this context, our laboratory has developed the concept of TASQ (for « Template Assembled Synthetic G-Quadruplex ») to assemble stable mimics of G4 and i-motif structures.

The first part of our study consisted in synthesizing a stable i-motif structure from the telomeric sequence using an addressable cyclopeptide platform and distinct chemoselective ligations. Then in a second part of the study, a novel ligation was introduced along with three other chemistries already optimized in our laboratory in order to synthetize novel tetraplex structures. Finally, a panel of G4 mimics adopting various topologies were synthesized and used in pull-down assays to identify the proteins that interact with G4 of specific topologies.

Keywords : G-quadruplex, i-motif, chemoselective ligation, cyclodecapeptidic scaffold, pull-down assays